The effect of delayed rewards on human goal-directed actions

Omar D. Perez[1,2,3] & Gonzalo P. Urcelay[4]

[1] Department of Industrial Engineering, University of Chile, Santiago, Chile.

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

[3] Instituto Sistemas Complejos de Ingeniería (ISCI), Chile.

[4] School of Psychology, University of Nottingham, UK.

Author Note

Correspondence regarding this article should be sent to Omar D. Perez (omar.perez.r@uchile.cl) or Gonzalo P. Urcelay (gonzalo.urcelay@nottingham.ac.uk).

Abstract

Goal-directed actions are defined by their sensitivity to the causal association between actions and outcomes, as well as the subjective value ascribed to those outcomes. When this sensitivity diminishes, actions may transition into habitual behavior. Based on recent findings from animal studies, we hypothesized that delaying outcomes relative to actions would weaken sensitivity to outcome revaluation and reduce action rates. In three experiments ($N = 290$), participants made fictitious investments in companies within different contexts providing immediate or delayed feedback. After training, participants were informed of a market crash that affected the performance of both companies, favoring one over the other. Across all experiments, action rates were lower in the delayed-feedback condition, and outcome revaluation had a more pronounced effect in the immediate-feedback condition. Self-reported action-outcome knowledge revealed that participants were less aware of the contingencies in the delayed condition. These findings suggest that delays in reinforcement weaken the action-outcome association critical for goal-directed control. We discuss the potential mechanisms underlying this phenomenon and its implications for understanding real-life decisions.

Keywords: habits, goal-directed, delay

The effect of delayed rewards on human goal-directed actions

Goal-directed actions are defined as those sensitive to the causal attribution of their consequences and the subjective value that subjects ascribe to those consequences. This latter feature is typically examined through outcome-revaluation tests, in which the value of the reward associated with an action is modified. Adams and Dickinson (1981) were the first to demonstrate how animals' behavior could be under goal-directed control with the use of this technique. They trained hungry rats to lever press for a rewarding outcome (or reinforcer) while an alternative reinforcer was delivered non-contingently to responding. To decrease the relative value of one of the outcomes with respect to the other, Adams and Dickinson established a flavor aversion to one of them by pairing its consumption with gastric malaise until the animals no longer ate the outcome when freely presented (i.e., the value of the outcome was reduced by the devaluation manipulation). During a test phase, they gave animals the opportunity to press the same lever as in training but suspended outcome delivery (i.e., on extinction). Critically, they found that animals whose devalued outcome was the one contingent to responding decreased responding compared to animals that had the non-contingent outcome devalued.

The Adams and Dickinson's (1981) study clearly demonstrated the capacity of animals to perform a lever response in anticipation of the outcome while encoding the relevance of the outcome to their current motivational state (i.e., subjective value). In humans, the goal-directed status of an action was later confirmed experimentally in an fMRI study in by Valentin, Dickinson, and O'Doherty (2007), who trained thirsty participants to perform two different actions yielding different liquid rewards. The revaluation procedure consisted of devaluing one of the outcomes by satiation (i.e., allowing participants to freely consume the outcome) before conducting a final extinction test with the two previously trained actions. In accord with the animal results, participants responded less to the action associated with the devalued outcome compared to the non-devalued outcome. Activity in the ventromedial prefrontal cortex correlated with the value of the reward and participants' behavior.

There is also evidence suggesting that numerous variables can render responses insensitive to revaluation of the outcome. In this so-called habit mode, and in contrast to the flexibility of goal-directed actions, subjects do not change their behavior after an outcome has been revalued; they keep doing what was good in the past. Among such factors, using interval as opposed to ratio reward schedules of reinforcement (Dickinson *et al.*, 1983; Gremel and Costa, 2013), single rather than multiple responses and outcomes (Colwill *et al.*, 1985; Holland, 2004; Kosaki and Dickinson, 2010), pre-exposing the outcome non-contingent to responding (Adams, 1982), and extended periods of training (Adams, 1982; Pool *et al.*, 2022; Tricomi *et al.*, 2009; but see de Wit et al. (2018)) can all promote habits. More recently, Pool et al. (2022) found important individual differences in the extent to which people deploy habitual control, and found that these differences could be linked to differences in psychological traits and states such as stress and anxiety (see also Schwabe and Wolf, 2009) . These results have spurred a gamut of theories where training conditions are critical in modulating the extent to which behavior is controlled by goal-directed or habitual systems (Daw *et al.*, 2005; Dezfouli and Balleine, 2012; Keramati *et al.*, 2011; Lee *et al.*, 2014; Miller *et al.*, 2019; Perez and Dickinson, 2020).

Another factor, which has received comparably less attention, is the delay of reinforcement. Under these procedures, the outcome of an action is delivered after a period of time has elapsed from the performance of the action. Delays of reinforcement have been shown to affect response rates and weaken causal beliefs, both of which are critical for goal-directed control. For example, Dickinson et al. (1992) demonstrated that rats trained with delayed reinforcement exhibited lower response rates than those trained with immediate reinforcement. In humans, Okouchi (2009) found similar results. In his study, participants were required to perform a specific sequence of responses and showed decreased response rates and increased post-reinforcement pauses as the delay between the action and the outcome increased, further emphasizing the importance of contiguity for acquiring instrumental actions. The importance of delay of reinforcement in attributing causal beliefs was demonstrated by Shanks (1989), who found that humans causally rate action-outcome relationships less favorably when reinforcement is delayed. Across three experiments, participants performed key presses to produce outcomes on a computer screen under varying delays and rated the extent to which their actions caused the outcomes. Causality ratings progressively decreased as delays increased, with participants in delayed conditions attributing less causality to their actions compared to those in immediate conditions. These results demonstrate the detrimental impact of delays on forming strong causal action-outcome links.

Despite these well-established effects of reinforcement delays on response rates and causal beliefs, its influence on devaluation sensitivity—the cardinal marker to infer goal-directed control—remains underexplored. The only evidence comes from a single study in rodents by Urcelay and Jonkman (2019). In one of their experiments, rats were trained in two different contexts, each associated with different outcomes and reinforcement delays. In one context, lever presses produced an immediate sucrose pellet reward, whereas in the other, lever presses led to a delayed chocolate-flavored pellet (counterbalanced across subjects). Following training, the rats underwent a satiety-specific devaluation procedure, in that they were pre fed with the different reinforcers before lever-press tests on extinction. The results revealed a significant interaction between context and devaluation: outcome devaluation was effective in reducing responding in the immediate-reward context but not in the delayed-reward context. These findings suggest that delayed reinforcement attenuates sensitivity to changes in outcome value, weakening goal-directed control.

In summary, there is evidence that delays in reinforcement weaken responding and causal action-outcome attribution in humans, suggesting that it may be affecting goal-directed control, but no evidence has been provided that they would also affect sensitivity to revaluation. This is the goal of the present study. Here, we test the hypothesis that delayed rewards modulates outcome revaluation sensitivity and response rates in human participants. We expect delays of reinforcement to reduce the impact of revaluation and lead to lower response rates, highlighting the importance of temporal contiguity in maintaining human goal-directed control.

General methods

In the three experiments reported in this paper, participants played the role of a stockbroker, making investments in stocks from two companies (with the fictitious names "Initech" and "Globex") in two different cities (contexts; Paris and London). In one context, immediate feedback on stock purchases was provided, while in the other context feedback was provided after a 5-sec delay . The delayed context, the instrumental response to make the investments and the company that was revalued were counterbalanced across subjects.

Participants

A total of 290 participants took part across the experiments reported in this paper. Experiment 1 (conducted in the laboratory with a sample composed of undergraduate students: N=53; 6 males and 47 females, ages ranging from 18 to 31 (M = 19.67 years, SD = 1.93.)), Experiment 2 (run online with a sample composed of undergraduate students: N=39, 14 males and 24 females, ages ranging from 18 to 43 (M = 20.64 years, SD = 3.88)).; and Experiment 3 (preregistered experiment run online with a sample recruited in Prolific: N=198; 95 males and 97 females, ages ranging from 18 to 66 (M = 37.10 years, SD = 11.38)). Participants were recruited through university subject pools (Experiments 1 and 2) and the Prolific online platform (Experiment 3). This latter experiment was preregistered as a replication of the first two. On the basis of Experiments 1 and 2, we performed a power analysis (using ANOVAexact) which yielded that 198 participants were required to to achieve .90 power to detect the interaction effect between Delay (immediate/delay context) and Value (valued/devalued outcome) during Test. All participants provided informed consent prior to participation. Participants in Experiment 3 (the online sample) were required to be fluent in English, with no other specific inclusion criteria applied prior to participating in the experiment (participants in Experiments 1 and 2 were undergraduate students at UK institutions). In Experiments 1 and 2, participants took part in exchange for course credit. In Experiment 3, participants were compensated with 8 pounds per hour for their participation. Participants in Experiments 2 and 3 were also asked to report knowledge about the action-outcome contingencies in each context.

Apparatus and Materials

For task presentation and data collection in-lab, the experiments were programmed using the PsychoPy (version 1.82) software (Peirce, 2007). JavaScript was used for the online samples. Participants performed the tasks on personal computers using their keyboards. The main task involved key presses to simulate stock purchases, with immediate (continuous reinforcement (CRF) or fixed ratio 1 (FR1)) or delayed feedback (CRF + delay of 5 seconds) provided in different contexts (cities). A debounce time of 1 second was imposed in the program, so that only one response every 1 second was effective in producing the an investment.

Figure 1 (bottom) shows the general design of the experiments. The instructions provided to participants at the beginning of the experiment stated their role in the task and what they should be trying to achieve (purchase shares for two different companies [R1 and R2], in two different cities [S1 and S2]). During training, participants experienced 2 blocks each of 2 minutes, in the two different contexts, S1 and S2, and the same two responses were possible in each context (R1 and R2). To press the left button (and buy the share indicated below) participants had to press the "a" key on the keyboard whereas to press the right button the "l" key was required to be pressed. Participants were instructed that key presses would earn them 3 shares, but that there was a cost of 1 share associated with each investment (key press).

Training	Devaluation	Test
S1: R1 → O1 R2 → O2 (immediate)	O1 ↓ O2 ↑	S1: R1 vs R2
S2: R1 →...O2 R2 →...O1 (delayed)		S2: R1 vs R2

Figure 1. **Design of the task.** During training, participants were presented with two different buttons (R1, R2) to purchase stocks from two different companies (with the fictitious names "Globex" and "Initech"). In one context (S1: Paris, in the figure), the feedback about stock purchase was given immediately, whereas in another context (S2: London in the figure) the feedback was delayed for 5 seconds. During the revaluation manipulation, participants were informed that one of the companies had crashed while the other was doing better (signified by the arrows point down or up, respectively). During the test phase, participants were presented in the same contexts (S1 and S2) as before, but this time no feedback or information about which company was associated with each button was provided. The difference between responses to the valued and devalued companies in each context indicates the degree to which participants were sensitive to revaluation of one of the outcomes (companies) in each context.

Procedure

The procedure was as follows:

1. **Training Phase:** Participants were trained to press keys ("a" for the left button and "l" for the right button) to purchase stocks from the two companies. The feedback was presented immediately after the investment in one city, and after a 5-sec delay in the other. The feedback consisted of a text with the phrase: *"You have purchased a stock from [name of the company]"*.

2. **Revaluation Phase:** Participants were informed that one company's stock value had

crashed while the other company's stock value had improved. The text presented to subject was as follows: *"The economy is booming, but not for all companies. The company below has crashed: [name of the devalued company]. However, this company is doing better: [name of the non-devalued company]"*. The company that crashed was counterbalanced across subjects.

3. **Test Phase:** Participants were asked to make stock purchases without receiving feedback or information about the companies associated with each button. The instructions for this phase were as follows: *"You shall continue trading in London and Paris. However, due to a malfunction with the trading equipment you will not receive any feedback. Press 'c' to continue."*. This cover story ensured that participants would not see the outcome of their investments, so that no new learning was allowed during the tests.

At issue were 1) The effect of delay on the response rates performed during training and 2) the effect of revaluation of the outcome on the number of stocks purchased for the revalued and devalued companies in each context. If delayed rewards have an impact on goal-directed behavior, we should expect response rates to be lower in the delay condition and, in addition, to observe a difference in stock purchasing for valued and devalued companies that is larger in the immediate condition than in the delayed condition.

Data Analysis

The dependent measure in these experiments was the amount of presses in each block of training, and during the two blocks of tests following revaluation. Because with count (i.e., keypresses) data the variance increases with the mean, all data were transformed (for data analyses and presentation) by calculating the square root of presses during each block during training and test.

We preregistered our statistical analysis in line with the previous study on delay of reinforcement and revaluation sensitivity by Urcelay and Jonkman (2019). For each experiment, we ran an ANOVA with Delay (immediate/delay) and Valued (valued/devalued) as within-subject factors. The details of this preregistered analysis can be found in https://aspredicted.org/tbjy-6mtg.pdf. For response rate comparison, Welch t-tests were performed on the final response rates (Block 2) of each experiment. Our analyses were all performed using the `R` programming language under the RStudio IDE (RStudio Team, 2020).

Results

Preregistered analysis

Effect of Delay on Response rates

The final mean response rates and 95% CIs attained by participants are shown in Table 1; the acquiscition curves are shown in the top panel of Figure 2. Following previous findings in animal and human studies (Dickinson *et al.*, 1992; Okouchi, 2009), we hypothesized that response rates would be affected by delays in reinforcement. This hypothesis was supported by comparing the final response rates in the immediate and delayed feedback conditions in each experiment. For Experiment 1, participants demonstrated significantly higher response rates in the immediate condition compared to the delayed condition, [$t(69.26) = 7.66$, $p = 0.00$, 95% CI [1.56, 2.66]]. Similarly, in Experiment 2, response rates were higher in the immediate condition compared to the delayed condition, [$t(70.73) = 2.57$, $p = .01$, 95% CI [1.46, 11.55]]. This pattern persisted in Experiment 3, with significantly higher response rates in the immediate condition [$t(391.61) = 2.16$, $p = .03$, 95% CI [0.20, 4.22]]. These results indicate that delaying reinforcement systematically reduces response rates, consistent with the theoretical importance of temporal delay in modulating goal-directed strength.

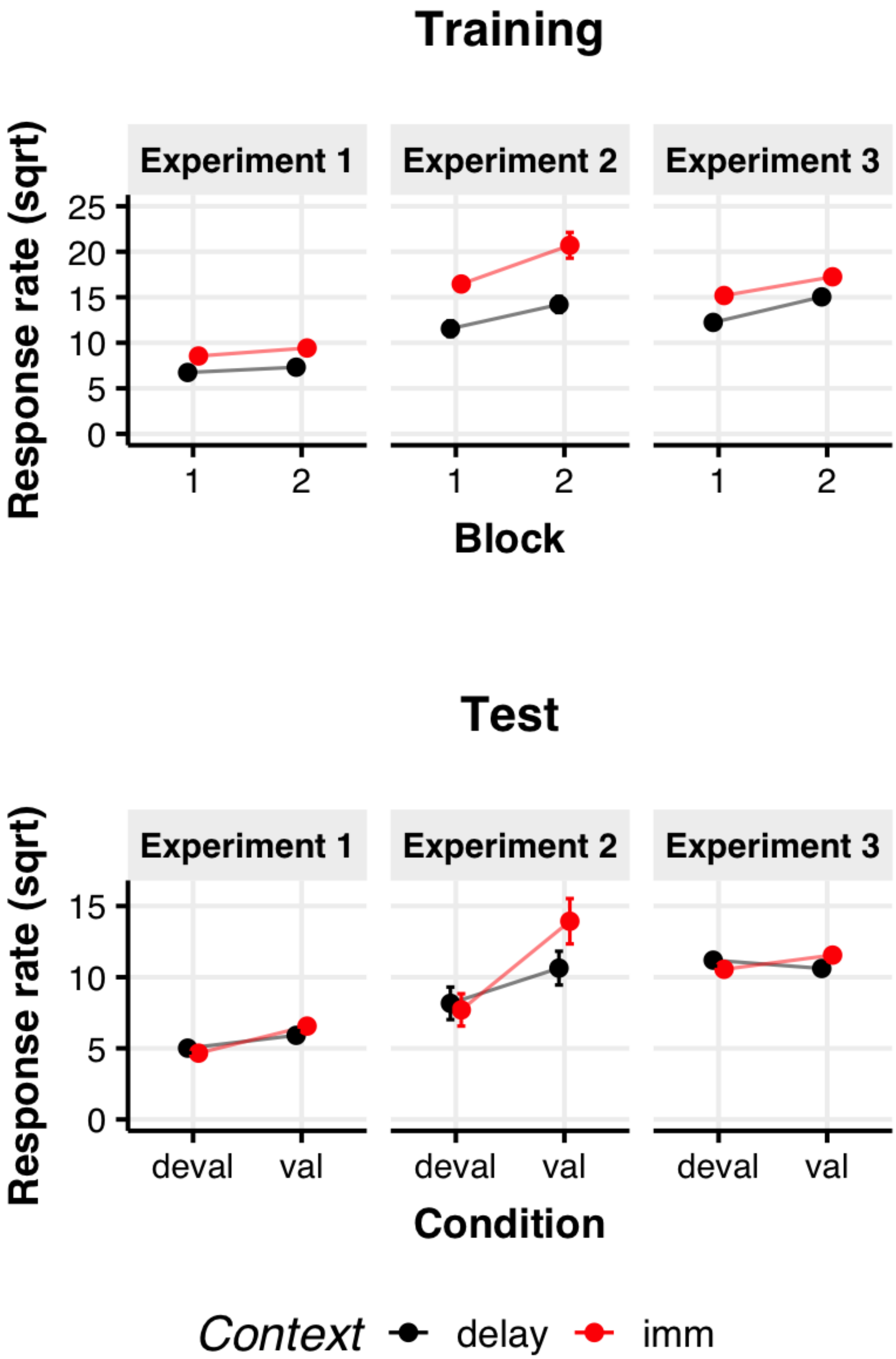

Figure 2. **Average response rates per block during the training and test phases in each experiment**. **a)** Response rates during the two blocks of training in each context. **b)** Response rates during the test phase under extinction. Error bars represent within-subject standard errors of the mean (Morey, 2008).

Table 1

Mean final response rates (sqrt transformed) by context and experiment.

Experiment	Context	Mean	95% CI
Experiment 1 (N=53)	Immediate	9.43	[9.17, 9.68]
	Delayed	7.32	[7.07, 7.57]
Experiment 2 (N=39)	Immediate	20.71	[18.59, 22.82]
	Delayed	14.20	[12.09, 16.32]
Experiment 3 (N=198)	Immediate	17.25	[16.74, 17.75]
	Delayed	15.04	[14.54, 15.55]

Effect of Delay on Revaluation

The results of the final tests between the revalued (*val*) and devalued (*deval*) companies are illustrated in the bottom panel of Figure 2. Visual inspection suggests that the difference between response rates between the revalued and devalued companies was more pronounced in the immediate (*imm*) than in the delay (*delay*) condition.

For Experiment 1, the main effect of *Delay* was not significant, $[F(1,52) = 2.49, p = .12, \eta_g^2 = .01]$. The main effect of *Val* was significant, $[F(1,52) = 5.77, p = .02, \eta_g^2 = .06]$, indicating that responses were influenced by valuation. The interaction between *Delay* and *Val* approached significance $[F(1,52) = 3.63, p = .06, \eta_g^2 = .01]$. For Experiment 2, the main effect of *Delay* was not significant, $[F(1,38) = 2.53, p = .12, \eta_g^2 = .01]$. The main effect of *Val* was significant, $[F(1,38) = 5.31, p = .03, \eta_g^2 = .05]$, demonstrating that valuation affected response rates. The interaction between *Delay* and *Val* was not significant, $[F(1,38) = 2.15, p = .15, \eta_g^2 = .01]$. For Experiment 3, the main effects of *Delay* and *Val* were both not significant, $[F(1,197) = 0.58, p = .45, \eta_g^2 = .00]$, and $[F(1,197) = 0.13, p = .72, \eta_g^2 < .00]$, respectively. The interaction between *Delay* and *Val* approached significance, $[F(1,197) = 2.76, p = .10, \eta_g^2 = .00]$.

Full sample analysis

Given the consistent direction of the effect and the fact that the task was the same in all experiments, we performed the same preregistered analysis on the full dataset, whose response rates for training and testing are shown in Figure 3. When collapsing data across all experiments, the repeated-measures ANOVA revealed a marginal main effect of *Delay*, $[F(1, 289) = 3.11, p = .08, \eta_g^2 = .01]$, and a significant main effect of *Val*, $[F(1, 289) = 4.01, p = .05, \eta_g^2 = .01]$. Importantly, the interaction between *Delay* and *Val* was significant, $[F(1, 289) = 5.69, p = .02, \eta_g^2 = .02]$, showing that the influence of devaluation on response rate was moderated by the delay condition: participants were less sensitive to devaluation in the delay context compared to the immediate context. Consistent with the interaction reported above, planned comparisons revealed an effect of revaluation in the Immediate context ($p = .00$) but not in the Delay context ($p = .86$)

Exploratory analysis

The preregistered analysis showed that the interaction effect between Delay and Val was evident only when the full dataset was analyzed. This result highlights the value of combining data across experiments to enhance statistical power, particularly when the tasks and hypotheses are identical. Moreover, the analysis of the full sample provides evidence that delayed rewards reduce revaluation sensitivity and strengthens the idea that temporal delays play a critical role in goal-directed control.

Motivated by evidence from prior studies in humans (Pool *et al.*, 2022) which suggest that individual sensitivity to outcome devaluation varies substantially across subjects, we performed an exploratory analysis on our data. This variability is particularly noticeable in free-operant, non-signaled tasks like the one employed in our experiments, where participants differ widely in their ability to integrate reinforcement contingencies over time and respond to changes in outcome value. Following visual inspection of the data, it became clear that there was significant heterogeneity in participants' sensitivity to devaluation across contexts,

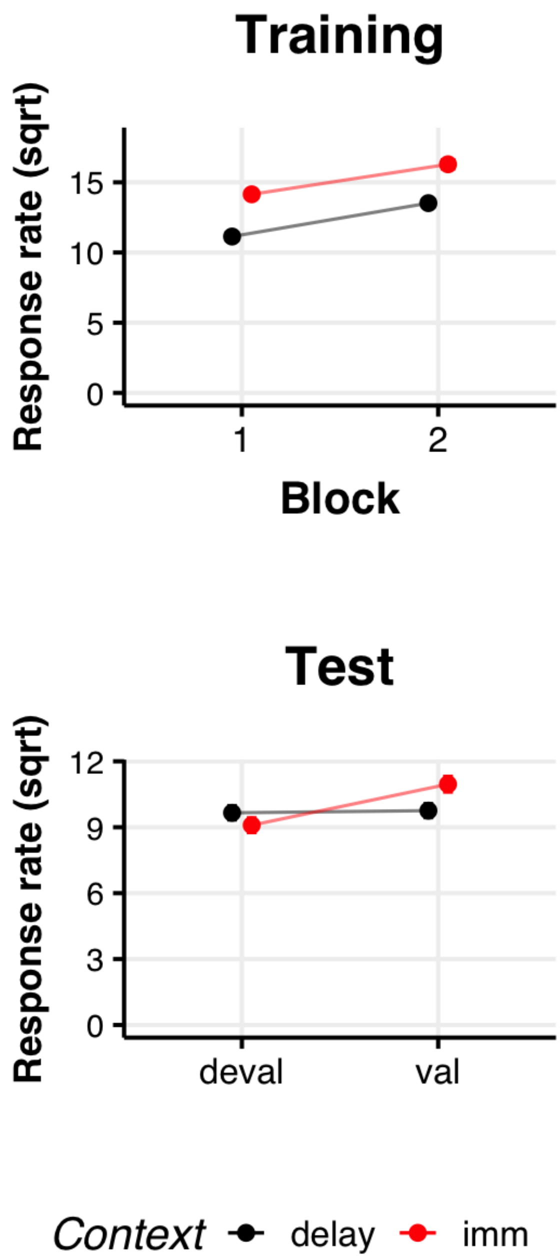

Figure 3. **Average response rates per block during the training and test phases in all three experiments collapsed**. **a)** Response rates during the two blocks of training in each context. **b)** Response rates during the test phase under extinction. Error bars represent within-subject standard errors of the mean (Morey, 2008).

reinforcing the need for a more nuanced and appropriate statistical approach for these data.

To formally account for these individual differences, we employed mixed-effects models with random intercepts and slopes. These models allow for participant-level variability in sensitivity to devaluation (random slopes for valuation) and delay (random slopes for delay) while estimating fixed effects at the group level. To justify the inclusion of random effects, we conducted a formal model comparison using likelihood-ratio tests and evaluated fit statistics with the Akaike Information Criterion (AIC).

The models were fitted using the `lme4` package in `R`, with response rates (square-root transformed, *sqNResp*) as the dependent variable, and *Delay* (immediate vs. delayed) and *Val* (valued vs. devalued) as within-subject predictors. Each model incorporated a varying degree of complexity in the random effects structure to determine the optimal balance between explanatory power and parsimony. Each model included fixed effects for the interaction between *Delay* and *Val*. Model comparisons were performed using Akaike Information Criterion (AIC), with lower AIC values indicating better model fit. Table 2 summarizes the fit statistics for each model. The model with the lowest AIC was **M2**, which included random intercepts and a random slope for *Val* (revaluation sensitivity). This result suggests that accounting for individual differences in how participants respond to outcome revaluation substantially improves the model fit compared to a random-intercept-only model (**M1**), which was the assumption in the preregistered analysis. In contrast, models including random slopes for *Delay* (**M3**) or both *Val* and *Delay* (**M4**) did not provide additional improvements in model fit.

The comparison indicated that the model with random intercepts and random slopes for devaluation provided the best fit to the data, consistent with our expectation that participants exhibit considerable variability in how they change responding after changes in outcome value (Pool *et al.*, 2022). The same analysis showed that there was no significant individual differences in how participants change their responding between immediate and delay condition.

Table 2

Model fit statistics for mixed-effects models.

Model	**logLik**	**AIC**	**BIC**
M1: Random intercept only	-3977.4	7966.9	7997.2
M2: Random intercept + slope for Val	-3921.0	**7858.0**	7898.5
M3: Random intercept + slope for Delay	-3977.4	7970.0	8011.3
M4: Random intercept + slopes for Val and Delay	-3920.6	7863.3	7918.9

The results of the mixed-effects model analysis are summarized in Table 3. The table reports the estimates for the influence of delay, revaluation, and their interaction on response rates, while accounting for individual differences in revaluation sensitivity across participants. As can be appreciated in the table, the analysis yielded significant interactions for Delay x Val in Experiments 1 and 3. Experiment 2 did not reach significance, possibly because the sample size was still too small to detect the effect. In addition, the whole-sample analysis confirmed our results from the preregistered analysis, in that the interaction was again significant, providing further evidence that delayed rewards have a detrimental effect on revaluation sensitivity.

Contingency knowledge

Our main hypothesis was that reinforcement delays would affect goal-directed strength by weakening the causal connection with the outcome. To test the extent to which knowledge of the causal link between actions and outcomes was encoded by subjects, at the end of Experiments 2 and 3 we asked participants to report their knowledge about the contingency between each action (the key press) and each outcome (shares of each company). Participants were assigned a total accuracy score by counting the number of correct action-outcome contingencies they reported correctly. Therefore, for each context, the score ranged from 0 to 2, and the total accuracy ranged from 0 to 4, indicating the total number of action-outcome contingencies they reported correctly. We expected that the delay context would affect the action-outcome strength, and therefore that this contingency knowledge score would be lower in this context than in the immediate context.

Consistent with our hypothesis, contingency knowledge scores were higher in the immediate condition ($M = 1.24$, $SD = 1.03$) compared to the delay condition ($M = 0.96$, $SD = 1.03$). A paired t-test revealed a significant difference between the two conditions ($t(236) = 3.03$, $p = 0.00$), with a mean difference of 0.29 (95% CI = $[0.10, 0.47]$). These findings suggest that delays in reinforcement negatively impact participants' ability to encode action-outcome contingencies, consistent with the hypothesis that temporal contiguity is crucial for forming strong action-outcome associations.

Table 3

Results of Mixed-Effects Model predicting response rates in experiments 1, 2, 3 and all experiments collapsed. The values are the estimated betas (and standard errors in parenthesis) for each factor. The hypothesis of delay affecting devaluation sensitivity is illustrated by the Delay x Val interaction.

Coefficient	**Exp. 1**	**Exp. 2**	**Exp. 3**	**All Exps.**
Delay (imm)	-0.35 (0.28)	-0.46 (1.56)	-0.64 (0.51)	-0.56 (0.41)
	$p = .20$	$p = .77$	$p = .21$	$p = .17$
Val (val)	0.90 (0.62)	2.48 (2.19)	-0.58 (0.69)	0.10 (0.57)
	$p = .15$	$p = .26$	$p = .40$	$p = .86$
Delay × Val	1.00 (0.40)	3.76 (2.21)	1.58 (0.73)	1.77 (0.58)
	$p = .01$ *	$p = .09$ +	$p = .03$ *	$p = .00$ **
Num. Observations	212	156	792	1160

Discussion

Theories of goal-directed actions define them by their sensitivity to revaluation of outcomes and changes in the causal association between actions and their outcomes (Dickinson, 1994; Perez and Dickinson, 2020). Environmental factors, such as the reinforcement schedule (Dickinson and Balleine, 1998; Perez and Dickinson, 2020), or psychological states, like stress and anxiety (Pool *et al.*, 2022), have been shown to modulate goal-directed behavior. The goal of this study was to assess whether outcome revaluation sensitivity is influenced by the temporal delay of action-outcome relationships in human participants.

We found that when outcomes were delayed by just 5 seconds relative to the actions producing them, participants attained lower action rates and showed reduced sensitivity to revaluation compared to when outcomes were delivered with no delay. Although the preregistered analyses did not detect significant interactions within each experiment, the overall effect of delay on revaluation sensitivity was consistent across experiments and evident when the data were analyzed as a whole. Furthermore, an analysis including individual differences in sensitivity to revaluation supported the idea that delayed rewards reduce revaluation sensitivity by showing the effect in Experiments 1 and 3 and in the whole dataset. These findings align with the results obtained by Urcelay and Jonkman (2019) in animal studies, where delayed reinforcement weakened outcome revaluation sensitivity.

In humans, delay of reinforcement has been predominantly studied in behavioral economics. These studies focus on how time affects the subjective value of rewards, estimating parameters like the discount rate to capture individual differences in valuing immediate versus delayed outcomes. Participants consistently devalue future outcomes relative to immediate ones, as modeled by hyperbolic discounting functions (Laibson, 1997; Mazur and Vaughan, 1987). A similar emphasis on outcome value weighting is found in Reinforcement Learning (RL) theories, where delays are assumed to down-weight the expected utility of outcomes as agents solve the *credit assignment problem* (Sutton and Barto, 2018) by ascribing value to actions in each state according to how far they are from the delivery of the outcome.

These theories, therefore, focus on value rather than causal associations. In contrast, our task explicitly tied feedback to participants' actions, with no monetary contingencies.

The observed effects of delay in our study are therefore unlikely to reflect changes in subjective outcome value. The most plausible explanation for our findings is that delays weakened participants' perceived causal link between their actions and outcomes, reducing both response rates and sensitivity to revaluation. This interpretation is consistent with studies showing that delay decreases causal ratings and response rates in humans (Okouchi, 2009; Shanks, 1989). We also observed that participants encoded action-outcome contingencies more strongly in the immediate condition, further supporting the idea that delays disrupt such action-outcome associations.

There are a number of theories of human actions that are consistent with present findings by proposing that action-outcome relations or associations are weakened in the delay condition. Hommel and colleagues postulate that such associations are contingent upon the simultaneous activation of action codes and effect codes (outcomes; 2004, see also De Houwer *et al.*, 2018). They assert that the action-outcome contingency, defined as the likelihood of an outcome being produced by the action as opposed to other potential causes, establishes a robust action-outcome association, whereas reinforcement delays diminish its strength. In one of their experiments, they observed that delays exceeding one second resulted in decreased responsiveness and attenuated priming effects; specifically, presenting the outcome associated with an action in the delay condition led to a reduced probability of that action being executed upon presentation of the outcome.

In Urcelay and Jonkman's (2019) study, the authors showed that delaying reinforcement shifted causal attribution from actions to contextual cues, reducing sensitivity to revaluation. In one of their experiments, the researchers extinguished the contextual association in the delay condition and demonstrated that revaluation sensitivity was reinstated. This finding supports the idea that reinforcement delays interfere with goal-directed control by disrupting the action-outcome link. Specifically, when delays are long in a controlled and stable environment—as is often the case in rodent studies—the context itself becomes causally associated with the outcome. By extinguishing this contextual association, the causal link between the animal's actions and the outcomes is restored, thereby reinstating sensitivity to outcome revaluation during the devaluation test.

Perhaps the formal framework that best captures all the above findings, including the effect of extinguishing the context in restoring revaluation sensitivity is the goal-directed system proposed by Perez and Dickinson (2024; 2020). In this theory, goal-directed strength is determined by subjects' experienced correlation between action and outcome rates and is directly related to instrumental performance and sensitivity to outcome revaluation. Delays disrupt this experienced correlation, weakening goal-directed control. Importantly, the experienced correlation that subjects compute is given by a mnemonic system that includes the representation of time samples, some of which are empty when the reward is delayed, assigning to the context a causal relationship with the outcome. When the context is extinguished, such relationship is weakened and the rate correlation experienced becomes positive again. This framework integrates the observed effects of delay on revaluation sensitivity, response rates, and action-outcome contingency knowledge, providing a psychologically and computationally coherent account of our data. Furthermore, the theory explains why ratio training supports higher action rates (Dickinson *et al.*, 1983; Perez, 2021), causal action-outcome beliefs (Reed, 2001) and outcome revaluation sensitivity (Dickinson *et al.*, 1983) than interval training, and anticipates that other manipulations of a causal action-outcome association, such as degrading the contingency between actions and outcomes, should also have an impact on action rates and outcome revaluation.

Crimmins et al. (2022) have recently provided evidence for this latter hypothesis. They trained rats to perform two actions, each leading to different rewards, and degraded the action-outcome contingency for one action by equalizing the probability of the outcome occurring both in the absence and presence of the action. This manipulation was achieved using a bidirectional lever, ensuring that both actions were equally associated with their outcomes and neutralizing any Pavlovian motivational effects on responding after devaluation. Using this design, they found that action rates decreased in the degraded action, and that sensitivity to revaluation was stronger for the action that maintained a contingent relationship with the outcome.

Even when studies of delay of reinforcement have been mostly theoretical, using either animal subjects or humans playing fictitious tasks in the laboratory, the importance of this variable in affecting goal-directed behavior cannot be underestimated. The implications of these findings extend beyond laboratory settings and provide insight into real-world decision-making processes in domains such as finance, medicine, and retirement savings. Most real-life decisions involve rewards that are delayed, sometimes by significant periods. The present findings suggest that poor self-control or unnecessary risk-taking in these contexts may not solely arise from temporal discounting of future rewards, or the uncertainty of impending rewards, as traditionally postulated. Instead, they may reflect a weakened causal attribution of actions to delayed outcomes. For example, in the domain of retirement savings, individuals might undervalue consistent contributions due to the long delay in seeing tangible benefits, attributing less causal weight to their contributions. In finance, risky investment decisions could result from attributing success or failure to stochastic factors rather than the quality of their choices. Therefore, it is possible that the timing of rewards—often beyond the control of individuals—can weaken the perceived action-outcome link, highlighting the need for interventions that enhance it for long-term goal adherence.

In conclusion, our findings underscore the critical role of temporal contiguity in maintaining goal-directed control, revealing that reinforcement delays systematically weaken action-outcome associations, as reflected in diminished response rates, revaluation sensitivity, and contingency knowledge.These results build on previous findings, showing that delay weakens goal-directed strength and offering a computationally grounded explanation through the rate correlation approach (Perez and Dickinson, 2020). Whether other manipulations of causal action-outcome associations, such as extinction (where actions stop leading to outcomes) or omission training (where actions and outcomes are inversely correlated), influence goal-directed strength, remains to be tested.

Author note

Omar D. Perez is supported by Instituto Sistemas Complejos de Ingenieria (ISCI) ANID PIA/PUENTE AFB220003, ANID-SIA 85220023 and ANID-FONDECYT 1231027.

Transparency and Openness Statement

In this paper, we utilized AI-assisted text editing to refine grammar and enhance readability. Specifically, OpenAI's GPT-4 (ChatGPT) was employed to assist non-native English-speaking authors in improving the linguistic clarity of the manuscript. The AI was not involved in content creation, data analysis, or interpretation; its role was strictly confined to grammatical and stylistic enhancements.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology*: 37–41: 37–41.

Adams, C. D. and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B : Comparative and Physiological Psychology* **33**: 109–121. ISBN: 0272-4995: 109–121.

Colwill Ruth, M, Rescorla, R. A., and Colwill, R. M. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes* **11**: 520–536: 520–536.

Crimmins, B., Burton, T. J., McNulty, M., Laurent, V., Hart, G., and Balleine, B. W. (2022). Response-independent outcome presentations weaken the instrumental response-outcome association. *Journal of Experimental Psychology: Animal Learning and Cognition* **48**: 396. Publisher: American Psychological Association: 396.

Daw, N. D., Niv, Y., and Dayan, P. (Dec. 2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience* **8**: 1704–1711: 1704–1711.

De Houwer, J., Tanaka, A., Moors, A., and Tibboel, H. (2018). Kicking the habit: Why evidence for habits in humans might be overestimated. *Motivation Science* **4**: 50–59: 50–59.

Dezfouli, A. and Balleine, B. W. (Apr. 2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience* **35**: 1036–1051: 1036–1051.

Dickinson, A. (1994). Instrumental conditioning. *Animal Cognition and Learning*. Ed. by N. J. Mackintosh. London: Academic Press: 45–78.

Dickinson, A. and Balleine, B. W. (1998). Motivational Control of Instrumental Action. **4**: 162–167: 162–167.

Dickinson, A., Nicholas, D. J. J., 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. ISBN: 0272-4995 {%}[August 09, 2010: 35–51.

Dickinson, A., Watt, a., and Griffiths, W. J. H. (1992). Free-operant acquisition with delayed reinforcement. *The Quarterly Journal of Experimental Psychology Section B* **45**: 241–258: 241–258.

Eisner, B. and Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research* **68**: 138–154. ISBN: 0340-0727 (Print)$\$n0340-0727 (Linking): 138–154.

Gremel, C. M. and Costa, R. M. (2013). Orbifrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications* **4**: 1–12. Publisher: Nature Publishing Group: 1–12.

Holland, P. C. (Apr. 2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of experimental psychology. Animal behavior processes* **30**: 104–117. ISBN: 0097-7403: 104–117.

Keramati, M., Dezfouli, A, and Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology* **7**. ISBN: 1553-7358 (Electronic)$\$n1553-734X (Linking).

Kosaki, Y. and Dickinson, A. (July 2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *Journal of experimental psychology. Animal behavior processes* **36**: 334–342: 334–342.

Laibson, D (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*.

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

Mazur, J. E. and Vaughan, W (1987). Molar optimization versus delayed reinforcement as explanations of choice between fixed-ratio and progressive-ratio schedules. *Journal of the experimental analysis of behavior* **48**: 251–261: 251–261.

Miller, K. J., Shenhav, A., and Ludvig, E. A. (2019). Habits without values. *Psychological Review* **126**: 292–311. Publisher: American Psychological Association: 292–311.

Morey, R. D. and others (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology* **4**: 61–64: 61–64.

Okouchi, H. (May 2009). RESPONSE ACQUISITION BY HUMANS WITH DELAYED REINFORCEMENT. en. *Journal of the Experimental Analysis of Behavior* **91**: 377–390: 377–390.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods* **162**: 8–13. Publisher: Elsevier: 8–13.

Perez, O. D. (2021). Instrumental behavior in humans is sensitive to the correlation between response rate and reward rate. *Psychonomic Bulletin & Review* **28**: 649–656. Publisher: Springer: 649–656.

Perez, O. D. (2024). The Rate-Correlation Theory of Goal-Directed Behavior: An Update. en. *Habits: Their Definition, Neurobiology, and Role in Addiction*. Ed. by Y. Vandaele. Cham: Springer International Publishing: 69–86.

Perez, O. D. and Dickinson, A. (2020). A Theory of Actions and Habits: The Interaction of Rate Correlation and Contiguity Systems in Free-Operant Behavior. *Psychological Review*.

Pool, E. R., Gera, R., Fransen, A., Perez, O. D., Cremer, A., Aleksic, M., Tanwisuth, S., Quail, S., Ceceli, A. O., and Manfredi, D. A. (2022). Determining the effects of training duration on the behavioral expression of habitual control in humans: a multilaboratory investigation. *Learning & Memory* **29**: 16–28. Publisher: Cold Spring Harbor Lab: 16–28.

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: 187–195.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. Place: Boston, MA.

Schwabe, L. and Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience* **29**: 7191–7198. Publisher: Soc Neuroscience: 7191–7198.

Shanks, D. R. (1989). Temporal contiguity and the judgement of causality by human subjects. *The Quarterly Journal of ...*: 139–159: 139–159.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (June 2009). A specific role for posterior dorsolateral striatum in human habit learning. *The European journal of neuroscience* **29**: 2225–2232: 2225–2232.

Urcelay, G. P. and Jonkman, S. (2019). Delayed rewards facilitate habit formation. *Journal of Experimental Psychology: Animal Learning and Cognition* **45**: 413. Publisher: American Psychological Association: 413.

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (Apr. 2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**: 4019–4026: 4019–4026.

Wit, S. de, Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., and Gillan, C. M. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General* **147**: 1043. Publisher: American Psychological Association: 1043.