npj | science of learning

Article

Published in partnership with The University of Queensland

https://doi.org/10.1038/s41539-025-00325-2

Delayed rewards weaken human goal directed actions

Check for updates

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

Goal-directed actions are sensitive to the causal association between actions and outcomes, as well as the value of those outcomes. Such sensitivity diminishes when actions become habitual. Based on recent findings in animals, we tested if 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 contexts that provided either immediate or delayed feedback. After training, participants were informed of a change in markets which affected both companies (one improved and the other worsened). Across all experiments, action rates were lower in the delayed-feedback condition, and outcome revaluation was stronger in the immediate-feedback condition. In addition, self-reported action-outcome knowledge was weaker in the delayed-feedback 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 in light of a contemporary theory of goal-directed behavior.

Goal-directed actions are 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[1] 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 more than animals that had the non-contingent outcome devalued.

The Adams and Dickinson's[1] study clearly demonstrated the capacity of animals to integrate action-outcome knowledge with outcome value to guide instrumental behavior. The goal-directed status of an action was later confirmed experimentally in an fMRI study in by Valentin, Dickinson, and O'Doherty[2], who trained thirsty human 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 rewarding in the past. Among such factors, using interval as opposed to ratio reward schedules of reinforcement[3,4] single rather than multiple responses and outcomes[5–7], pre-exposing the outcome non-contingent to responding[8], and extended periods of training[8–10] (but see ref. 11,12) can all promote habits. More recently, Pool et al.[10] found important individual differences in the extent to which human participants 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[13]), both of which tend to increase the speed of habit formation. 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[14–19].

[1]Department of Industrial Engineering, University of Chile, Santiago, Chile. [2]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. [3]Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile. [4]School of Psychology, University of Nottingham, Nottingham, UK.
e-mail: omar.perez.r@uchile.cl; gonzalo.urcelay@nottingham.ac.uk

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.[20] demonstrated that rats trained with delayed reinforcement exhibited lower response rates than those trained with immediate reinforcement. In humans, Okouchi[21] found similar results. In this study, participants were required to perform a specific sequence of responses and showed decreased response rates as the delay between the action and the outcome increased, further emphasizing the importance of action-outcome contiguity for acquiring instrumental actions.

The importance of delay of reinforcement in attributing causal beliefs to actions was demonstrated by Shanks[22], who found that humans causally rated action-outcome relationships less favorably when reinforcement was 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 the formation of strong causal action-outcome links.

Despite these well-established effects of reinforcement delay 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[23]. In one of their experiments, rats were trained to press two levers 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 more effective in reducing responding for the devalued action in the immediate context than in the delayed context, suggesting that delayed reinforcement weakens responding sensitivity to changes in outcome value, affecting goal-directed control.

Perez and Dickinson[18,24] have recently postulated a theory that is able to capture these results by assuming that agents deploy a goal-directed system that computes the correlation between response and outcome rates continuously by keeping responses and outcome in different contiguous working-memory slots. When rewards are delayed, they fall in slots that are further ahead in time, degrading the experienced rate correlation between actions and outcomes and the strength of goal-directed behavior. Following this theory, the effect of changes in outcome value is directly related to the experienced action-outcome rate correlation established by the reward schedule in effect, and a weaker revaluation effect is expected with delay of reinforcement.

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, as anticipated by Perez and Dickinson's (2020) theory. 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.

Results

Participants completed an action-outcome or instrumental task in which they played the role of stockbrokers, making investments in stocks from two fictitious companies ("Initech" and "Globex") in two different cities or contexts (Paris and London). The key manipulation involved the timing of feedback: in one context, feedback about stock purchases was provided immediately, while in the other, feedback was delayed by 5 s. The assignment of contexts to immediate or delayed feedback, the response keys used to make investments, and the company affected by the revaluation manipulation were counterbalanced across participants. Each participant completed four blocks of training, two blocks in the immediate and two in the delayed contexts. The order of contexts (immediate first or delay first) was randomly assigned for each participant at the beginning of the experiment and then repeated in the same order for the second block.

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]

Following training, participants were informed that one company's stock had crashed while the other was doing better, serving as the revaluation manipulation. In the final phase, participants continued to make investments, but no feedback was provided. This design allowed us to examine whether delayed feedback influenced response rates during training and whether it impacted sensitivity to revaluation during the test phase. Based on prior findings in animal learning and reinforcement delay effects, we expected that delayed reinforcement would weaken action-outcome associations, leading to lower response rates and reduced sensitivity to outcome revaluation.

We first analyzed the effect of delay on responses rates. The final mean response rates and the 95% CIs attained by the participants are shown in Table 1; the acquisition curves are shown in the top panel of Fig. 1. Following previous findings in animals and humans[20,21] we hypothesized that response rates would be lower in the delay condition. 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, [$t(52) = 9.25$, $p < 0.001$, 95%CI[1.65, 2.57]]. Similarly, in Experiment 2, response rates were higher in the immediate condition compared to the delayed condition, [$t(38) = 3.41$, $p = 0.002$, 95%CI[2.64, 10.36]]. This pattern was also observed in Experiment 3, with significantly higher response rates in the immediate condition [$t(197) = 4.71$, $p < 0.001$, 95%CI[1.28, 3.13]]. These results indicate that reinforcement delays reduce response rates, consistent with the theoretical idea that temporal delay has a role in modulating goal-directed strength.

We next analyzed the effect of delay on revaluation sensitivity. The results of the final tests between the revalued (*val*) and devalued (*deval*) companies are illustrated in the bottom panel of Fig. 1. 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 3, we preregistered our statistical analysis in line with the previous study on delay of reinforcement and revaluation sensitivity by Urcelay and Jonkman[23]. 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.

The test data were analysed with 2 (Delay: immediate vs. delayed) x 2 (Val: valued vs. devalued) within-subjects ANOVAs. For Experiment 1, the main effect of *Delay* was not significant, [$F(1, 52) = 2.49$, $p = 0.121$, $\eta^2 = 0.01$]. The main effect of *Val* was significant, [$F(1, 52) = 5.77$, $p = 0.020$, $\eta^2 = 0.06$], indicating that responses were influenced by valuation. The

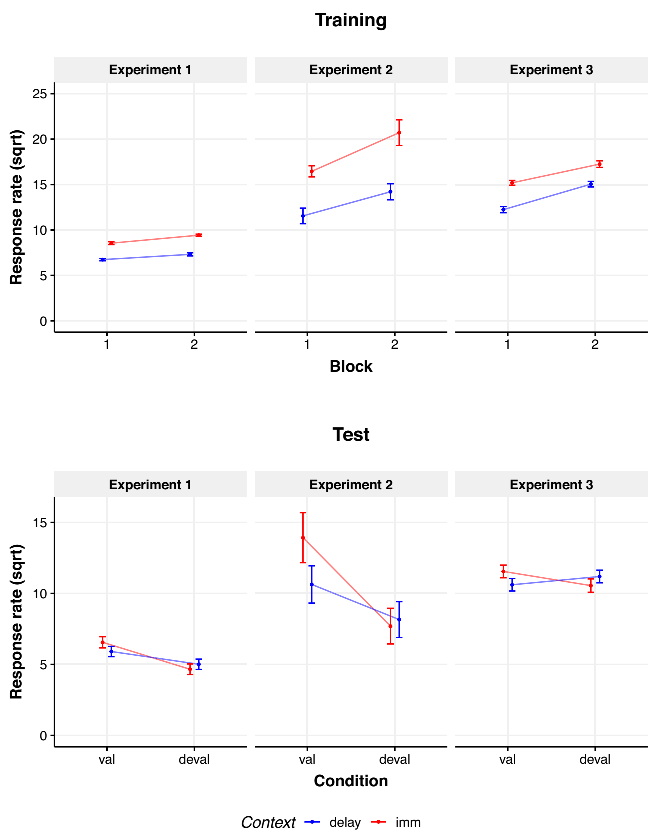

Fig. 1 | Average response rates per block during the training and test phases in each experiment. Top panel. Response rates (sqrt transformed) during the two blocks of training in each context. Bottom panel. Response rates during the test phase under extinction. Error bars represent within-subject standard errors of the mean[39].

interaction between *Delay* and *Val* was not significant [$F\ (1, 52) = 3.63$, $p = 0.062$, $\eta^2 = 0.01$]. For Experiment 2, the main effect of *Delay* was not significant, [$F\ (1, 38) = 2.53$, $p = 0.120$, $\eta^2 = 0.01$] and the main effect of *Val* was significant, [$F\ (1, 38) = 5.31$, $p = 0.027$, $\eta^2 = 0.05$], demonstrating that valuation affected response rates. The interaction between *Delay* and *Val* was not significant, [$F\ (1, 38) = 2.15$, $p = 0.151$, $\eta^2 = 0.01$]. For Experiment 3, the main effects of *Delay* and *Val* were both not significant, [$F\ (1, 197) = 0.58$, $p = 0.448$, $\eta^2 = 0.00$], and [$F\ (1, 197) = 0.13$, $p = 0.724$, $\eta^2 < 0.00$], respectively. The interaction between *Delay* and *Val* was not significant, [$F\ (1, 197) = 2.76$, $p = 0.098$, $\eta^2 = 0.00$].

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, and the response rates for training and testing are shown in Fig. 2. When collapsing data across all experiments, the repeated-measures ANOVA revealed a non-significant main effect of *Delay*, [$F\ (1, 289) = 3.11$, $p = 0.079$, $\eta^2 = 0.01$], and a significant main effect of *Val*, [$F\ (1, 289) = 4.01$, $p = 0.046$, $\eta^2 = 0.01$]. Importantly, the interaction between *Delay* and *Val* was significant, [$F\ (1, 289) = 5.69$, $p = 0.018$, $\eta^2 = 0.02$], 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, we found a significant difference in responding between valued and devalued actions in the Immediate context [$t(289) = 2.91$, $p = 0.004$, $95\%CI[0.60, 3.13]$] but not in the Delay context [$t(289) = 0.17$, $p = 0.866$, $95\%CI[-1.06, 1.26]$].

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.

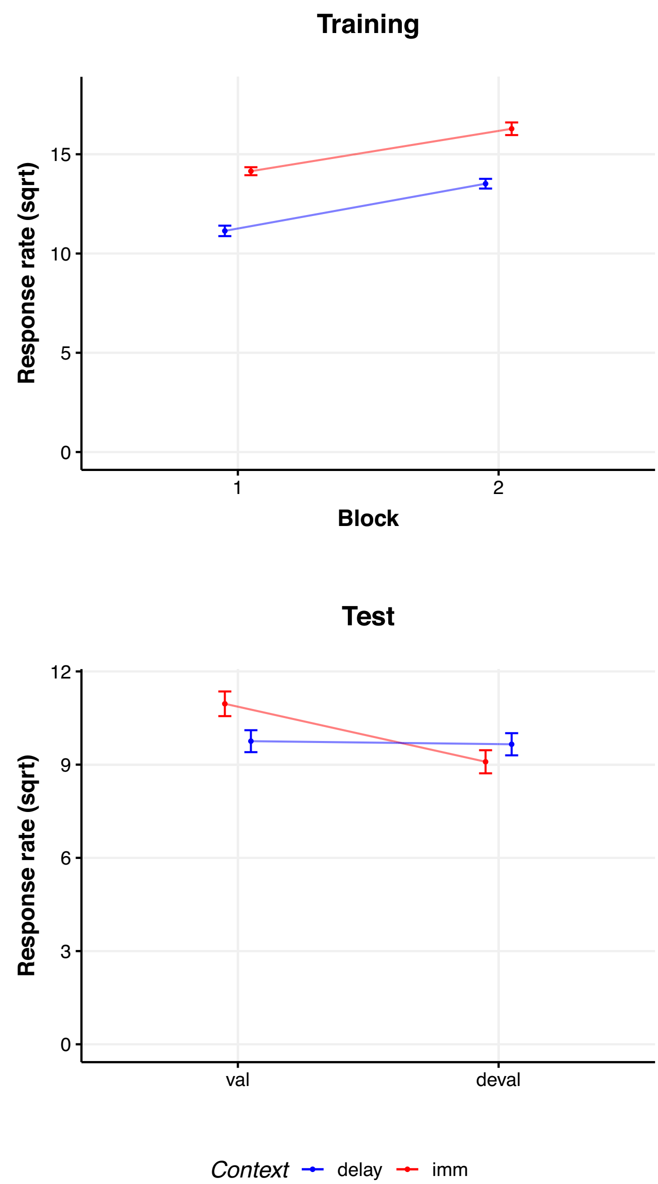

Fig. 2 | Average response rates per block during the training and test phases in all three experiments collapsed. Top panel. Response rates during the two blocks of training in each context. Bottom panel. Response rates during the test phase under extinction. Error bars represent within-subject standard errors of the mean[39].

Motivated by evidence from prior studies in humans[10] 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 can respond at any time and as many times as they want and therefore may differ widely in their integration of reinforcement contingencies over time and changes in outcome value. Following visual inspection of the data, we found significant heterogeneity in participants' sensitivity to devaluation across contexts, reinforcing the need for a more nuanced and appropriate statistical approach for these data.

To formally account for individual differences in sensitivity to devaluation and delay, we employed a mixed-effects model with random intercepts and slopes for both *Val* (valued vs. devalued) and *Delay* (immediate vs. delayed), as well as their interaction. This model captures

Table 2 | Results of Mixed-Effects Model predicting response rates in experiments 1, 2, 3, and all experiments combined

Coefficient	Exp. 1	Exp. 2	Exp. 3	All Exps.
Delay (immediate)	−0.650	−3.295	−0.939	−1.203
	(0.279) $p = 0.021$	(1.543) $p = 0.034$	(0.514) $p = 0.068$	(0.412) $p = 0.004$
Val (revalued)	−1.900	−6.231	−0.997	−1.866
	(0.615) $p = 0.002$	(2.181) $p = 0.005$	(0.690) $p = 0.149$	(0.571) $p = 0.001$
Delay × Val	1.003	3.756	1.579	1.767
	(0.393) $p = 0.012$	(2.146) $p = 0.082$	(0.727) $p = 0.030$	(0.582) $p = 0.002$
Num. Observations	212	156	792	1160

The table displays estimated betas with standard errors in parentheses and corresponding *p*-values. The hypothesis of delay affecting devaluation sensitivity is illustrated by the Delay × Val interaction.

participant-level variability while estimating fixed effects at the group level. The mixed-effects model was 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. Following recommendations for analyzing within-subject designs[25], we specified a maximal random effects structure, including random intercepts and slopes for both *Delay* and *Val*, but not for their interaction, (due to the limited number of observations per subject), to account for individual differences in sensitivity to delay and revaluation. We used the ‘immediate’ and ‘valued’ conditions as the reference levels for the Delay and Val factors, respectively.

The results of the mixed-effects model analysis are summarized in Table 2. The table reports the estimates for the influence of delay, valuation, and their interaction on response rates, while accounting for participant-level variability. As shown in the table, the analysis yielded significant interactions for *Delay* × *Val* in Experiments 1 and 3. Experiment 2 did not reach significance, possibly due to the smaller sample size. Additionally, the analysis of the full dataset confirmed the interaction effect, providing further evidence that delayed rewards weaken revaluation sensitivity.

Our main hypothesis was that reinforcement delays would affect goal-directed strength by weakening the causal connection of actions 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 (the company). Participants were assigned a total accuracy score by counting the number of correct action-outcome contingencies they reported. Therefore, for each context, the score ranged from 0 to 2. 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 ($t(236) = 3.03$, $p = 0.002$), with a mean difference of 0.29 (95% CI = [0.10, 0.47]). These findings show 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.

To examine whether response rates were affected by contingency knowledge during training, we conducted a mixed-effects model including Delay, Contingency Knowledge (accuracy score), and their interaction as fixed effects, with a random intercept for participants and a random slope for delay. The results indicated that contingency knowledge did not significantly predict response rates (coefficient = −1.86, $p = 0.216$), nor did it significantly interact with Delay (coefficient = 0.76, $p = 0.596$). The main effect of Delay was also non-significant (coefficient = 4.35, $p = 0.169$).

These findings suggest that differences in contingency knowledge do not directly account for the lower response rates observed in the delay condition. Instead, both lower contingency knowledge and lower response rates likely stem from a common underlying mechanism: a weakened action-outcome rate correlation.

Finally, we examined whether contingency knowledge influenced sensitivity to revaluation, including Delay, valuation (Val), contingency knowledge (accuracy score), and their interactions as fixed effects in a mixed-effects model with a random slope for both Delay and Val by participant. The results indicated that contingency knowledge did not significantly predict response rates during test (coefficient = −0.47, $p = 0.598$), nor did it significantly interact with Delay (coefficient = 0.32, $p = 0.633$), Val (coefficient = −0.83, $p = 0.393$), or both (three-way interaction, coefficient = −0.72, $p = 0.449$).

The Delay × Val interaction remained in the expected direction but was not significant in this model (coefficient = 3.02, $p = 0.189$). Taken together, these findings suggest that contingency knowledge does not directly predict revaluation sensitivity. Instead, our results support the hypothesis that the primary determinant of goal-directed behavior is the strength of the action-outcome association itself, which is weakened by reinforcement delay. This aligns with our interpretation that delay reduces the experienced correlation between action and outcome rates, which in turn affects both contingency awareness and sensitivity to revaluation.

Discussion

Goal-directed actions are characterized by their sensitivity to outcome revaluation and the causal association between actions and their consequences[18] Environmental factors, such as the reinforcement schedule[18,26], or psychological states, like stress and anxiety[10], 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 s relative to the actions producing them, participants attained lower response 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[23] in animal studies, where delayed reinforcement weakened outcome revaluation sensitivity in rodents.

In humans, delay of reinforcement has been predominantly studied in Behavioral Economics and Reinforcement Learning (RL). Economic theories focus on how time affects the subjective value of rewards, estimating parameters—most notably the discount rate—to capture individual differences in how people devalue future rewards relative to immediate ones[27,28] RL theories incorporate a similar assumption, where delays down-weight the expected utility of an outcome as agents solve the *credit assignment problem*[29]. In both frameworks, delay does not alter the probability that an action leads to an outcome, but rather reduces the value of the outcome itself, which is reflected in the expected value of the action that leads to the outcome ($EV_{action} = probability \times magnitude$). Given that the strength of the action-outcome link is given by the probability of an action leading to a reward, this is assumed to remain unaffected by delays.

In contrast, our data suggest that delay weakens the perceived causal relationship between actions and outcomes. Indeed, if delay merely influenced subjective reward value, we would expect participants to retain contingency knowledge but respond less due to lower expected utility. Instead, weaker contingency knowledge and response rate in the delay condition suggests that reinforcement delays disrupt causal associations, rather than affecting outcome value per se. This aligns with the view that goal-directed strength depends on the experienced correlation between action and outcome rates[18]. In this framework, the expected value of an action is determined not only by the magnitude of the outcome but also by the strength of the action-outcome link, given by this correlation ($EV_{action} = r \times magnitude$, where r is the rate correlation). When reinforcement is delayed, the expected value of the action is reduced, making actions less sensitive to changes in outcome value.

It could be argued that the differences in responding between immediate and delay conditions we observed may arise from differences in motivation. However, the within-subject nature of our study makes this possibility unlikely. Because each participant experienced both immediate and delayed conditions under the same motivational state, differences in revaluation sensitivity cannot be attributed to changes in motivation. Furthermore, if the effect was solely due to changes in outcome value, then devaluing a company should have equally affected responses in both conditions. Instead, the revaluation effect was mainly observed in the immediate condition, while responses in the delayed condition remained relatively insensitive to changes in outcome value.

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[30], see also ref. 31). 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[23] 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.

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[18,24] 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, this 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[3,32] causal action-outcome beliefs[33] and outcome revaluation sensitivity[3] 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.[34] 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.

Our main claim is that delays of reinforcement degrade the causal attribution of actions to outcomes, leading to reduced contingency awareness and revaluation sensitivity. This effect is mediated by a reduction in goal-directed strength, as postulated by a rate-correlation mechanism[18,24]. However, alternative explanations are possible. Under free-operant conditions, the 5-second delay between an action and its outcome may allow for the occurrence of other actions within that interval, which could lead to confusion among participants about which action is attributable to which outcome, thereby explaining the deleterious effect of delay on contingency awareness and revaluation sensitivity. Despite the merits of this explanation, a rate-correlation system still predicts reduced causal attribution in the delay condition, as it computes the rate-correlation separately for each action, rendering them independent of one another and providing a computationally-grounded explanation for our results.

The idea that sensitivity to devaluation is related to the strength of an action-outcome association has been demonstrated in a different scenario by Liljeholm and colleagues[35]. In an fMRI study, participants performed actions to fill fictitious empty beakers displayed on a screen while being scanned. In one condition, called the *high divergence condition*, participants saw a single empty beaker and a cue that signaled a sequence of actions required to fill it. Because multiple actions were needed, participants could not associate any single action with a specific sensory outcome, weakening all individual action-outcome associations. In contrast, in the *zero divergence condition*, participants consistently performed the same action across trials to fill specific beakers, strengthening the action-outcome association with respect to the sensory properties of the outcomes.

Following this training phase, the authors conducted a devaluation procedure and found that devaluation only reduced the probability of actions in the *zero divergence condition*, where action-outcome associations had been stronger. Importantly, they also found that activation in the right supramarginal gyrus—a brain region previously implicated in sense of agency attributions—was directly related to devaluation sensitivity. These results align with the present findings, as delayed reinforcement also weakens the action-outcome association critical for goal-directed control, which in turn affects both contingency awareness and causal action-outcome ratings, as previously observed in human studies[22].

Even when studies of delay of reinforcement have been mostly theoretical, using either animal subjects or humans playing fictitious decision-making tasks in the laboratory, the importance of delay 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[36]. 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.

A few factors may limit the generality of our findings. First, the task was entirely fictitious, and while participants made financial decisions, the absence of real monetary consequences may have reduced engagement with the instrumental contingencies. Second, although our results demonstrate a clear effect of delay on revaluation sensitivity, we only tested a single 5 s delay, leaving open the question of how different delay durations modulate goal-directed control. Lastly, our findings were based on financial decision-making, raising the question of whether similar effects extend to primary reinforcers such as food or social rewards.

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[18]. Whether other manipulations of causal action-outcome associations, such as extinction (where actions stop leading to outcomes) non-contingent training (where action and outcomes are uncorrelated) or omission training (where actions and outcomes are inversely correlated), influence goal-directed control, remains to be tested.

Methods

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 the library ANOVAexact in R) which yielded that 198 participants were required 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.

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee at the University of Leicester (application number: 14345-sc629-neuroscience,psychologyandbehaviour) and the University of Nottingham (reference number: 893).

Apparatus and Materials

For task presentation and data collection in-lab, the experiments were programmed using the PsychoPy (version 1.82) library for Python[37]. 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 s) provided in different contexts (cities). A debounce time of 1 s was imposed in the program, so that only one response every 1 s was effective in producing an investment.

Figure 3 (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 min, 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). The share count was displayed in red, using Calibri font (size 12), with the numerical value presented below it in black. Feedback text was also in Calibri but in green, while the company name appeared in an opaque gray.

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 s delay in the other. The feedback

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

Fig. 3 | 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 s. 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.

consisted of a text with the phrase: *"You have purchased a stock from [name of the company]"*, together with a short beep sound.
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 number 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. Welch t-tests were performed on the final response rates (Block 2) of each experiment, and a preregistered within-subject ANOVA was run for the choice test (see Methods) Our analyses were all performed using the R programming language under the RStudio IDE[38].

Data availability

https://osf.io/rxtzm.

Code availability

Scripts for the analysis of this paper can be found at https://osf.io/rxtzm.

Received: 31 December 2024; Accepted: 7 May 2025;
Published online: 07 June 2025

References

1. Adams, C. D. & Dickinson, A. Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. B* **33**, 109–121 (1981).
2. Valentin, V. V., Dickinson, A. & O'Doherty, J. P. Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* **27**, 4019–4026 (2007).
3. Dickinson, A., Nicholas, D. J. J. & Adams, C. D. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Q. J. Exp. Psychol.* **35**, 35–51 (1983).
4. Gremel, C. M. & Costa, R. M. Orbifrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* **4**, 1–12 (2013).
5. Colwill, R. M. & Rescorla, R. A. Postconditioning devaluation of a reinforcer affects instrumental responding. *J. Exp. Psychol. Anim. Behav. Process.* **11**, 520–536 (1985).
6. Holland, P. C. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.* **30**, 104–117 (2004).
7. Kosaki, Y. & Dickinson, A. Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *J. Exp. Psychol. Anim. Behav. Process.* **36**, 334–342 (2010).
8. Adams, C. D. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol.* **34**, 77–98 (1982).
9. Tricomi, E., Balleine, B. W. & O'Doherty, J. P. A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* **29**, 2225–2232 (2009).
10. Pool, E. R. et al. Determining the effects of training duration on the behavioral expression of habitual control in humans: a multilaboratory investigation. *Learn. Mem.* **29**, 16–28 (2022).
11. De Wit, S. et al. Shifting the balance between goals and habits: five failures in experimental habit induction. *J. Exp. Psychol. Gen.* **147**, 1043–1065 (2018).
12. Garr, E., Padovan-Hernandez, Y., Janak, P. H. & Delamater, A. R. Maintained goal directed control with overtraining on ratio schedules. *Learning & Memory* **28**, 435–439 (2021).
13. Schwabe, L. & Wolf, O. T. Stress prompts habit behavior in humans. *J. Neurosci.* **29**, 7191–7198 (2009).
14. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
15. Lee, S. W., Shimojo, S. & O'Doherty, J. P. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* **81**, 687–699 (2014).
16. Keramati, M., Dezfouli, A. & Piray, P. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput. Biol.* **7**, e1002055 (2011).
17. Dezfouli, A. & Balleine, B. W. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* **35**, 1036–1051 (2012).
18. Perez, O. D. & Dickinson, A. A theory of actions and habits: the interaction of rate correlation and contiguity systems in free-operant behavior. *Psychol. Rev.* **127**, 487–517 (2020).
19. Miller, K. J., Shenhav, A. & Ludvig, E. A. Habits without values. *Psychol. Rev.* **126**, 292–311 (2019).
20. Dickinson, A., Watt, A. & Griffiths, W. J. H. Free-operant acquisition with delayed reinforcement. *Q. J. Exp. Psychol. B* **45**, 241–258 (1992).
21. Okouchi, H. Response acquisition by humans with delayed reinforcement. *J. Exp. Anal. Behav.* **91**, 377–390 (2009).
22. Shanks, D. R. Temporal contiguity and the judgment of causality by human subjects. *Q. J. Exp. Psychol.* **41**, 139–159 (1989).
23. Urcelay, G. P. & Jonkman, S. Delayed rewards facilitate habit formation. *J. Exp. Psychol. Anim. Learn. Cogn.* **45**, 413–425 (2019).
24. Perez, O. D. in Habits: Their Definition, Neurobiology, and Role in Addiction (ed. Vandaele, Y.) 69–86 (Springer, Cham, 2024).
25. Singmann, H. & Kellen, D. An introduction to mixed models for experimental psychology. *N. Methods Cogn. Psychol.* **4**, 4–31 (2019).
26. Dickinson, A. & Balleine, B. W. Motivational control of instrumental action. *Anim. Learn. Behav.* **4**, 162–167 (1998).
27. Laibson, D. Golden eggs and hyperbolic discounting. *Q. J. Econ.* **112**, 443–477 (1997).
28. Mazur, J. E. & Vaughan, W. M. Molar optimization versus delayed reinforcement as explanations of choice between fixed-ratio and progressive-ratio schedules. *J. Exp. Anal. Behav.* **48**, 251–261 (1987).
29. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* (MIT Press, 2018).
30. Eisner, B. & Hommel, B. Contiguity and contingency in action-effect learning. *Psychol. Res.* **68**, 138–154 (2004).
31. De Houwer, J., Tanaka, A., Moors, A. & Tibboel, H. Kicking the habit: why evidence for habits in humans might be overestimated. *Motiv. Sci.* **4**, 50–59 (2018).
32. Perez, O. D. Instrumental behavior in humans is sensitive to the correlation between response rate and reward rate. *Psychon. Bull. Rev.* **28**, 649–656 (2021).

33. Reed, P. Schedules of reinforcement as determinants of human causality judgments and response rates. *J. Exp. Psychol. Anim. Behav. Process.* **27**, 187–195 (2001).
34. Crimmins, B. et al. Response-independent outcome presentations weaken the instrumental response-outcome association. *J. Exp. Psychol. Anim. Learn. Cogn.* **48**, 396 (2022).
35. Liljeholm, M., Dunne, S. & O'Doherty, J. P. Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control. *Eur. J. Neurosci.* **41**, 1358–1371 (2015).
36. Wood, W. & Neal, D. T. The habitual consumer. *J. Consum. Psychol.* **19**, 579–592 (2009).
37. Peirce, J. W. PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
38. RStudio Team. *RStudio: Integrated Development Environment for R* (Boston, MA, 2020).
39. Morey, R. D. et al. Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* **4**, 61–64 (2008).

Acknowledgements

Omar David Perez is supported by Instituto Sistemas Complejos de Ingenieria (ISCI) ANID PIA/PUENTE AFB220003, ANID-SIA 85220023 and ANID-FONDECYT 1231027. The authors would like to thank Richard Kirkden for his help in deploying the task online for data collection.

Author contributions

O.D.P. programmed the task, analyzed the data, and wrote the manuscript. G.P.U. designed the task, funded the studies, collected and analyzed the data, and provided critical revisions and additions to the manuscript. Both authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Omar David Perez or Gonzalo P. Urcelay.

Reprints and permissions information is available at http://www.nature.com/reprints

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