NEUROSCIENCE

RESEARCH ARTICLE

S. D. McDougle / Neuroscience 486 (2022) 37–45

Post-error Slowing During Instrumental Learning is Shaped by Working Memory-based Choice Strategies

Samuel D. McDougle

Department of Psychology, Yale University, United States

Abstract—Post-error slowing (PES) – a relative increase in response time for a decision on trial *t* given an error on trial $t-1$ – is a well-known effect in studies of human decision-making. Post-error processing is reflected in neural signatures such as reduced activity in sensorimotor regions and increased activity in medial prefrontal cortex. PES is thought to reflect the deployment of executive resources to get task performance back on track. This provides a general account of PES that cuts across perceptual decision-making, memory, and learning tasks. With respect to PES and learning, things are complicated by the fact that learning often reflects multiple qualitatively different processes with distinct neural correlates. It is unclear if multiple processes shape PES during learning, or if PES reflects a policy for reacting to errors generated by one particular process (e.g., cortico-striatal reinforcement learning). Here we provide behavioral and computational evidence that PES is influenced by the operation of multiple distinct processes. Human subjects learned a simple visuomotor skill (arbitrary visuomotor association learning) under low load conditions more amenable to simple working memory-based strategies, and high load conditions that were putatively more reliant on trial-by-trial reinforcement learning. PES decreased with load, even when the progress of learning (i.e., reinforcement history) was accounted for. This result suggested that PES during learning is influenced by the recruitment of working memory. Indeed, observed PES effects were approximated by a computational model with parallel working memory and reinforcement learning systems that are differentially recruited according to cognitive load.

This article is part of a Special Issue entitled: SI: Error Processing. **© 2021 IBRO. Published by Elsevier Ltd. All rights reserved.**

Key words: post-error slowing, reinforcement learning, working memory, cognitive control.

INTRODUCTION

Why do people typically slow down after making errors? This intuitive behavior resonates in folk psychology, for example, when a coach instructs a pupil to “slow down and focus” after they err. The large body of research on post-error slowing (PES; Rabbitt, 1966) typically uses laboratory tasks requiring rapid decision-making, such as perceptual judgment (e.g., dot-motion discrimination; Purcell and Kiani, 2016) and response conflict tasks (e.g., the Stroop task; Botvinick et al., 2001), where each trial involves an independent choice. In such tasks, PES is often thought of as a compensatory mechanism to improve future goal-directed behavior by recruiting cognitive control (Dutilh et al., 2012a). This cognitive control account has been linked to activities in the medial prefrontal cortex (Botvinick et al., 2001; Gehring and Fencsik, 2001; Kerns, 2004; Narayanan and Laubach, 2008; Cavanagh et al., 2010; Danielmeier et al., 2011), a key correlate of executive functioning.

What about PES during *learning*? It is reasonable to assume that PES would also be present in learning tasks, where performance should be monitored on an ongoing basis. Indeed, one influential study using a standard reinforcement learning task showed that PES does indeed occur during instrumental learning, is linearly related to a computationally derived reward prediction error, and covaries with frontal theta power (Cavanagh et al., 2010). In addition to demonstrating PES during instrumental learning, these findings linked PES to a theoretical construct of error derived from internally represented stimulus or action values (Schultz et al., 1997).

Work over the past several years has shown that in many human instrumental learning tasks, multiple processes contribute to the learning curve. For example, a body of research by Collins and colleagues shows that during the learning of simple visuomotor skills (i.e., the learning of arbitrary visuomotor associations), both working memory strategies and incremental reinforcement learning simultaneously contribute to peoples’ choices (Collins and Frank, 2012; Collins et al., 2014; Collins et al., 2017; Collins and Frank, 2018; Collins, 2018; Master et al., 2020; McDougle and

Abbreviations: ERN, event-related negativity; PES, post-error slowing; RL, reinforcement learning; RT, reaction time; WM, working memory.
E-mail address: samuel.mcdougle@yale.edu

https://doi.org/10.1016/j.neuroscience.2021.10.016

Collins, 2021). Similarly, simple sequential instrumental learning tasks have revealed (at least) two dissociable learning systems – ‘‘model-based” and ‘‘model-free” reinforcement learning – with the former involving the representation of state transitions and the latter involving the caching of a running average of action outcomes (Daw et al., 2011; Doll et al., 2012; Otto et al., 2015). The model-based process has been linked to cognitive control and working memory systems (Otto et al., 2015), while the model-free process has been linked to canonical cortico-striatal learning systems (Gläscher et al., 2010).

These findings highlight a lacuna in accounts of PES during instrumental learning – is the relationship between PES and learning mediated by one particular learning system? Here we propose that cognitive decision-making processes, which operate alongside reinforcement learning processes, shape PES effects. Consequently, PES during learning should be affected by cognitive load in a manner predicted by a multi-system account of learning. This view would accord with neurophysiological findings in both humans and model organisms that link PES with processing in the prefrontal cortex (Gehring and Fencsik, 2001; Narayanan and Laubach, 2008). Moreover, this prediction follows from results in the category learning literature demonstrating that PES is more closely related to adoption of ‘‘rule-based” learning strategies linked to prefrontal function than incremental ‘‘information integration” learning strategies linked to striatal function (Ashby and Maddox, 2005; Tam et al., 2013). If confirmed, this result would complicate the assumption that PES during learning is diagnostic of error monitoring within one particular learning system.

We analyzed a large data set (total $N = 119$) from two previously-published studies (Collins and Frank, 2012; Collins and Frank, 2018), where human subjects learned arbitrary visuomotor mappings under varying cognitive loads. Load was operationalized as the ‘‘set size” (i.e., the number of unique visuomotor associations to be learned) in a given task block. Nearly a decade of research using this method (Collins and Frank, 2012; McDougle and Collins, 2021) has confirmed that both working memory strategies (i.e., active memory traces of correct stimulus–response associations) and conventional reinforcement learning processes (i.e., trial-by-trial integration of stimulus–response values) can operate in parallel. Such data provide a natural testbed for characterizing how PES might relate to different learning strategies.

EXPERIMENTAL PROCEDURES

Arbitrary visuomotor association learning task

Detailed methods for the behavioral task can be found in the source studies (Collins and Frank, 2012; Collins and Frank, 2018), though we offer a summary here. The protocol for all behavioral tasks was approved by the institutional review board at Brown University and all subjects gave informed consent. A combined sample size of $N = 119$ was included in our analysis, and consisted of neurologically healthy, right-handed, young-adult participants with normal or corrected-to-normal vision. The design of the task is depicted in Fig. 1. Subjects were seated in front of a computer monitor where they responded to stimuli using a USB computer keyboard. Subjects were tasked with learning which of three responses (one of three button presses, using the ‘‘J,” ‘‘K,” or ‘‘L” keys) was associated with each presented image in order to maximize reward feedback. Key presses (e.g. J, K, or L) were produced with the index, middle, or ring finger, respectively. On correct trials, positive feedback (‘‘+1” points) was displayed centrally in green font, and on incorrect trials, negative feedback (‘‘0” points) was displayed centrally in red font. Subjects had to respond within 1.4 seconds to receive feedback. We excluded trials where responses were too slow or exceedingly rapid (<200 ms).

Each experiment consisted of several distinct blocks of trials. Unique sets of arbitrary, discriminable visual images were used in each block (e.g., shape line-drawings, colored blobs, vegetables, vehicles, scenes, etc.). Each block was associated with a particular load (or ‘‘set size”), defined as the number of individual stimulus–response associations the subject was required to learn during that block. Block ordering was designed to approximate an even distribution of high and low load blocks across the two halves of the experiment, and to avoid repeated blocks of the same load (>2 in a row). Each experiment was completed in a single session without breaks.

On each trial, one image was displayed on screen at a time over a black background (stimulus visual angle, ~8°). Each stimulus was presented for a minimum of 9 iterations over the block, with a maximum of 15 iterations. Blocks were complete after either 15 iterations of each stimulus were seen, or when subjects selected the correct action for three of the four last iterations for all stimuli. The specific sequence of stimuli within a block was pseudorandomized. In the first data set (from Collins and Frank, 2012) 18 blocks were completed (loads 1–6; load = 1 was not analyzed here as it was not implemented in both data sets; average number of trials = 670; experiment time: ~45 minutes), and in the second data set (from Collins and Frank, 2018) 22 blocks were completed (loads 2–6; average number of trials = 750; experiment time: ~50 minutes). Across all subjects, the mean number of stimulus iterations experienced for loads 2–6 was, respectively, 9.7, 9.9, 10.1, 11.0, and 11.7, and the mode was 9 iterations across all loads. Thus, while subjects completed n*iterations more trials per block as load increased by n, the number of iterations performed per stimulus were comparable across loads.

Behavioral analysis

Learning curves were computed by taking the mean percent correct (i.e., percent trials where subjects performed the correct action in response to a stimulus) relative to the stimulus iteration. This allows us to analyze learning with respect to each stimulus, rather than with respect to the raw number of trials (which is linearly confounded with load). Reaction time (RT) was

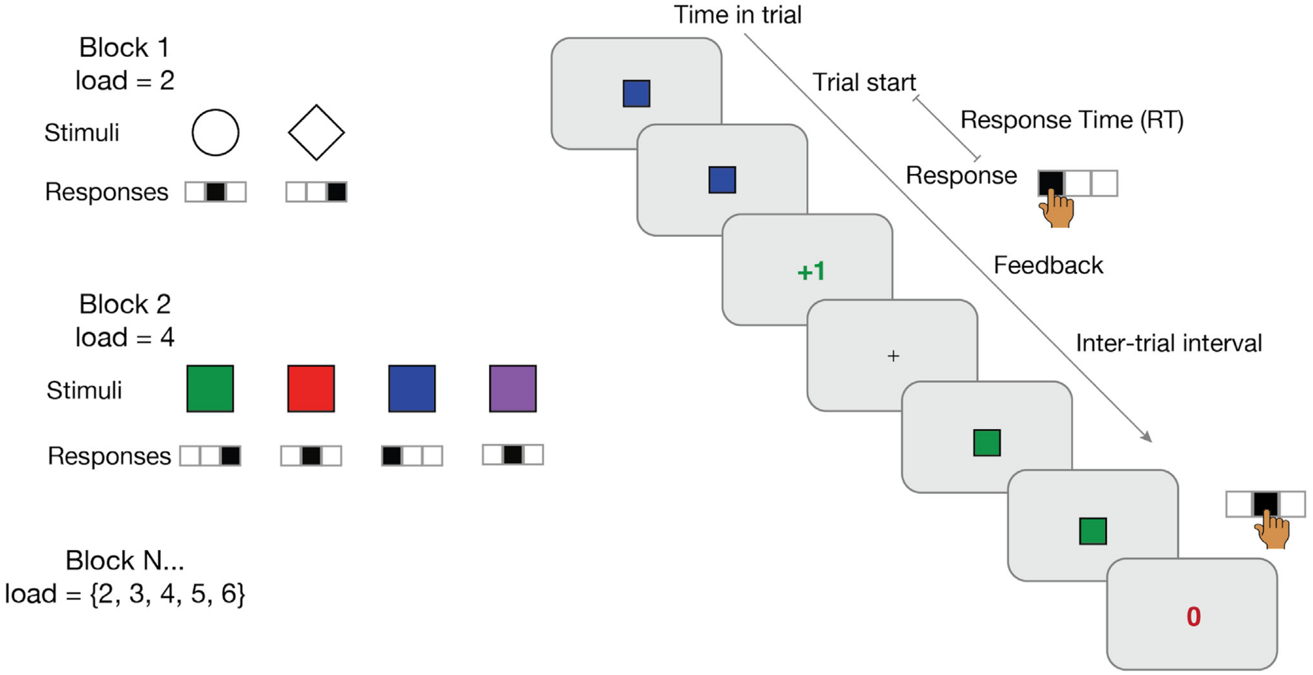

Fig. 1. Task. In this arbitrary visuomotor association learning task, subjects learn stimulus–response associations over independent trial blocks. Example block designs are shown on the left, demonstrating how the cognitive load manipulation is implemented. Within each block, stimuli were presented in a pseudorandomized sequence (right), and each stimulus was seen 9–15 times within a block.

computed as the amount of time (in seconds) between the appearance of the stimulus and the moment of the key press.

A key recent study revealed that conventional operationalizations of PES – simply analyzing response times on all trials that follow errors versus correct responses – are susceptible to significant temporal confounds (Dutilh et al., 2012b). As demonstrated by Dutilh et al. (2012b), PES can be masked (or erroneously amplified) by global changes in accuracy, speed, attentional vigilance, and meta-learning across an experiment. To address this issue those authors developed a simple solution: PES should be computed on pairwise comparisons of RTs within strings of trials, wherein the error of interest is flanked by (at least) two correct trials preceding it and another correct trial succeeding it. PES is then quantified as the difference between RTs on the trial after the error versus the trial just before the error.

This so-called ‘‘robust” method (Dutilh et al., 2012b) can mostly account for the aforementioned global confounds and was thus used throughout our analyses. Critically, these global confounds are especially present in learning tasks, where accuracy and speed are not evenly distributed over trials. Moreover, to control for the well-known effect of trial repetition on RT (Hale, 1969), effects which are avoided in typical decision-making tasks with independent trials, we did not analyze trials where the same choice stimulus was shown successively. We note that implementing this rather restrictive measure of PES was afforded by our large sample size ($N = 119$). Overall, an average of 21.45 robust PES trials were analyzed per subject, with a total of 2552 trials in the analysis. (We note that our main findings were replicated using traditional PES metrics.)

We performed additional control analyses on the PES data. In our ‘‘reinforcement history” analysis, we separately visualized PES effects for each load based on the number of times the stimulus associated with the error response had previously been correctly responded to (i.e., rewarded). The goal of this analysis was to confirm that an effect of load on PES was not confounded by the number of rewards accrued to the stimulus leading up to the error trial preceding PES. If so, the effect could reflect mainly the distribution of ‘‘surprising” trials sampled across loads, with more surprising trials (i.e., errors for well-learned stimuli) being over-represented in the lower loads. (We note that this over-representation is related to better learning in the lower load conditions, which is itself taken as evidence of a role for working memory in this task; see *Results*.) We thus visualized PES as a function of the cumulative reinforcement history for the current stimulus at the moment of the error, focusing on 0, 1, 2, 3, 4, and > 4 cumulative rewards (> 80% of valid robust PES trials occurred with 4 or fewer cumulative rewards for the stimulus on that error trial).

To quantitatively control for effects of reinforcement history and additional factors potentially influencing PES, we performed a multiple linear regression analysis. For each subject, we designated three predictor variables: the current Load condition for each PES trial, the current Delay number for each PES trial (i.e., the number of intervening trials between the error trial and the last time the same stimulus was observed), and the current cumulative reinforcement history for the presented stimulus (see above). Each individual variable was z-scored before being entered into a linear regression. All interaction terms were included as well, and the regression was carried out using the *glmfit* function in MATLAB (version 2020a; Natick, Massachusetts). Because of occasional low trial numbers due to our restrictive PES operationalization, we excluded subjects from the regression analysis if they had fewer than 10 valid PES trials or did not have robust PES trials in more than one load condition (nine subjects total). The regression analysis allowed us to simultaneously control for multiple variables that could shape response times in this task.

Averages are depicted using the mean, and error bars are computed using the standard error of the mean, with the exception of the regression results which are presented using box plots (median marked plus non-outlier range). T-tests are used for pairwise comparisons; ANOVAs are used to measure the effects of Load, as a repeated measure, on learning and PES. All reported statistical tests (ANOVAs and *t*-tests) are two-tailed, with alpha set at 0.05. Analyses and statistics were performed using MATLAB (version 2020a; Natick, Massachusetts) and R (R Core Team).

Computational modeling analysis

We performed a computational modeling analysis to further test the idea that PES is shaped by multiple interacting learning processes. Collins and Frank (2012) formalized how working memory (WM) and reinforcement learning (RL) work in parallel during instrumental learning (the "RL + WM" model). In the RL + WM model, two dissociable modules learn stimulus–response associations (i.e., state-action values). Learning is modeled using standard RL equations, where the action (a) value in a given state (s) – Q(s,a) – is updated on each trial, t, using the delta rule:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\delta_t \quad (1)$$

$$\delta_t = r - Q_t(s,a) \quad (2)$$

where α is the learning rate, δ is the reward prediction error, and *r* is the reward received (1 or 0).

Values are transformed into probabilities via the softmax function,

$$p(a|s) = \frac{e^{Q(s,a)\beta}}{\Sigma_i e^{Q(s,a_i)\beta}} \quad (3)$$

where β constitutes the inverse temperature parameter, and the sum in the denominator is taken over the three possible actions, a_i. The RL module of the RL + WM model is defined in Eqs. (1) and (2). The WM module simultaneously learns stimulus–response associations (V), but with a fixed learning rate of 1, capturing immediate commitment to memory rather than proper incremental learning:

$$W_{t+1}(s,a) = W_t(s,a) + (r - W_t(s,a)) = r \quad (4)$$

Working memory, being vulnerable to short-term forgetting, undergoes trial-by-trial decay of W,

$$W_t(s_j,a_i) = W_t(s_j,a_i) + \phi(W_0 - W_t(s_j,a_i)) \quad (5)$$

where ϕ draws W (over all stimuli *j* and actions *i*) toward their initial values (W_0) of $\frac{1}{3}$.

RL and WM choice policies (π_{RL}/π_{WM}) are separately computed using a softmax function with a fixed β of 100 (Eq. (3); see Collins and Frank, 2018 and McDougle et al., 2021) and are then combined into a final policy as a weighted sum,

$$\pi = weight * \pi_{WM} + (1 - weight) * \pi_{RL} \quad (6)$$

where the "weight" is a proxy for the degree to which WM is currently recruited. This variable is determined by two free parameters, the working memory capacity (i.e., resource limit) *K*, and the initial WM weighting ρ ,

$$weight = \rho * min(1, \frac{K}{load(b)}) \quad (7)$$

This equation simply determines that WM recruitment for a block of trials (b) is reduced as the load exceeds *K*. An additional parameter (ϵ) models undirected noise in the final policy, which reflects "slips" of action,

$$\pi = (1 - \epsilon) * \pi + \epsilon * Unif \quad (7)$$

where "Unif" represents the uniform action policy (*p* = $\frac{1}{3}$ for each action).

Finally, the model also captures the neglect of negative feedback often observed in this task by reducing the learning rate (multiplicatively) on error trials:

$$\alpha = \gamma\alpha \quad (8)$$

where γ controls the degree of perseveration. (Perseveration occurs for both modules; for WM the fixed learning rate following negative feedback trials is simply γ).

For completeness, we compared the fit of the RL + WM model to two RL-only alternative models tested in previous studies (Collins and Frank, 2012; Collins and Frank, 2018) – the "Multi-α" model, where a separate RL learning rate (Eq. (1)) is fit to each load condition, and the "Basic RL" model, where only a single learning rate is used across all load conditions. Both alternative models also include the noise and perseveration parameters. We note that the Basic RL model cannot capture any effects of load, and thus serves as a baseline. According to previous work, the Multi-α model can capture load effects, but does not fit the data as well as the RL + WM model, both quantitatively and qualitatively (Collins and Frank, 2012; Collins and Frank, 2018; see *Results* and Supplemental Fig. 1 for replication of these computational modeling results).

Models were fit to choice data using maximum likelihood estimation, minimizing the negative log likelihood using the MATLAB function *fmincon*. Initial parameter values were randomized for each fitting iteration, with 50 iterations per fitting run to avoid local minima. Parameter constraints were: α = [0,1]; γ = [0,1]; ϕ = [0,1]; ρ = [0,1]; ϵ = [0,1]; *K* = [2,6]. Models were compared using the Akaike Information Criterion (Akaike, 1974). We used simulations to validate the model and try to measure the relationship between modeled choice behavior and PES: We simulated each model using the best fit parameters from the fitting procedure over each subject's actual observed stimulus sequences, simulating each subject 100 times and averaging the results.

RESULTS

Learning of arbitrary visuomotor associations under varying cognitive load

We analyzed data from an instrumental learning task (Fig. 1) that requires human subjects to learn an arbitrary visuomotor mapping between visual stimuli (e.g., shapes, color patches, etc.) and discrete actions (e.g., button presses) under various cognitive loads (Collins and Frank, 2012; Collins and Frank, 2018). One of the more straightforward pieces of evidence that working memory-based strategies may influence behavior in this task are the learning curves it produces (Fig. 2A). Critically, when performance (e.g., percent correct) is plotted as a function of the number of iterations a choice stimulus has been observed, people's learning curves show striking load effects, with more incremental learning seen as the load increases (repeated measures ANOVA capturing the effect of load on percent correct: *F*(1,118) = 207.20, $p < 0.001$). In theory, if learning were fully driven by a high-capacity reinforcement learning system this effect should not occur, implicating the recruitment of a short term memory-based strategy (Collins and Frank, 2012).

PES during instrumental learning is modulated by cognitive load

Is PES during learning affected by cognitive load? Fig. 2B depicts average response times in the PES trial sequences we analyzed in each load condition. PES, defined as RTs on trial C + 1 minus RTs on trial C-1 (i.e., the correct trials flanking the error trial), was significant in all load conditions (all Bonferroni-corrected *p's* < 0.05; Fig. 2C). This was observed using both the robust method of quantifying PES and the typical (confounded) method, where average post-correct RTs are subtracted from average post-error RTs (all Bonferroni-corrected *p's* < 0.005). We note here that the overall magnitude of PES was not significantly different between the two data sets analyzed (*t*(117) = 0.66, $p = 0.51$)

As shown in Fig. 2C, PES was attenuated as a function of load: As load increased, the magnitude of PES significantly decreased. We quantified this effect by performing a one-way repeated measures ANOVA on the average PES effect over each load condition. We found a significant effect of load on the magnitude of PES (*F*(1,118) = 6.77, $p = 0.010$). In support of our hypothesis, these results suggest that PES during

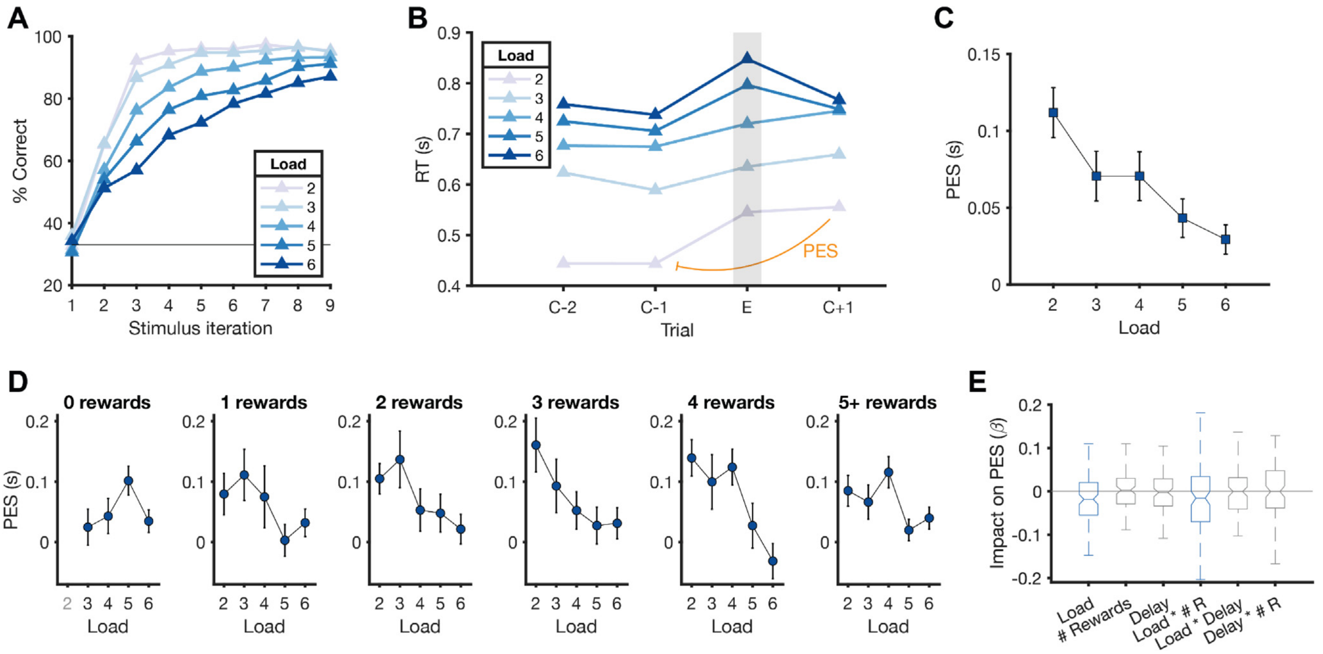

Fig. 2. PES results. **(A)** Learning curves (means), plotted as a function of stimulus iteration, separated by load. **(B)** Post-error slowing (PES) of response times (RTs) was computed using the "robust" method of Dutilh et al. (2012b). This method controls for global confounds on PES by targeting strings of trials where errors are flanked by a subsequent correct trial and a preceding correct trial (that is itself a post-correct trial). **(C)** Average PES as a function of Load. **(D)** PES as a function of load given different reinforcement histories (number of cumulative rewards earned for the presented stimulus at time of error). (Note that in the 0 rewards case, no valid trials were extracted in the Load = 2 condition.) **(E)** Box plot of beta coefficients from a multiple regression analysis on post-error slowing (PES) that accounted for the effects of Load, Delay (the number of intervening trials between the analyzed error trial and the last trial that presented the same choice stimulus), and reinforcement history (same as D), as well as their interactions. Distributions outlined in blue connote significance at $p < 0.05$. Error bars in panels C-*D* = 1 s.e.m.

learning may be shaped by processes sensitive to cognitive load, such as working memory.

REINFORCEMENT HISTORY DOES NOT EXPLAIN THE ROLE OF LOAD ON RESPONSE SLOWING

There are alternatives to our interpretation of these PES effects. First and most importantly, differences in reward history could determine how load interacts with PES. That is, in higher load conditions, where learning is slower (Fig. 2A), prediction errors from the sampled PES trials should be smaller on average (i.e., errors should be less surprising), logically leading to attenuated PES. One straightforward way to test this interpretation is to visualize PES while holding reinforcement history constant. Fig. 2D depicts PES effects as a function of both load (abscissas) and the number of cumulative rewards earned for the stimulus seen on the analyzed error trials (panels from left to right). As predicted, reinforcement history did not appear to fully explain the observed effects: Higher load conditions tended to show reduced PES in spite of reinforcement history. Statistically quantifying the effects plotted in Fig. 2D was not optimal given the 30 + factors for each mean PES value (moreover, our restrictive PES criteria led to abundant missing values across the sample). We thus opted for a multiple regression approach that used the full set of PES trials (further details of this analysis are given in the *Experimental Procedures* section.)

The results of the regression analysis (Fig. 2E) confirmed our observation that Load attenuates PES, while simultaneously accounting for other variables of interest (effect of Load: $t(109) = 2.44$, $p = 0.016$). Delay did not have a significant effect on PES ($t(109) = 0.16$, $p = 0.87$). Reinforcement history (# Rewards) had a numeric but nonsignificant on PES ($t(109) = 1.59$, $p = 0.11$). Lastly, reinforcement history displayed a robust negative interaction with Load ($t(109) = 3.10$, $p = 0.003$); in combination with Fig. 2D, this interaction appears to suggest that in lower load conditions, PES effects were more sensitive to reinforcement history. None of the other two-way interactions were significant (all p*'s* > 0.05). Together, these results are inconsistent with a simple account of PES effects during instrumental learning. Instead, they suggest that PES during instrumental learning may be contingent on the particular cognitive processes currently contributing to behavior (Tam et al., 2013).

We note that in some cases, PES has been interpreted as reflecting a generic re-orienting of attention after unexpected events rather than an error-based adjustment *per se* (Notebaert et al., 2009). To test this in the context of our task, we re-ran our analysis but flipped the trials of interest to post-correct trials (i.e., ones flanked by error trials, again using the robust method). We did not see significant post-correct slowing in 4 out of the 5 load conditions (Bonferroni-corrected $p < 0.05$ for load = 5; all p*'s* > 0.05 for all other conditions; Supplemental Fig. 2). Thus, in this context, PES effects appeared to be restricted to errors rather than to an unsigned ‘‘surprise” signal.

Working memory contributions to learning shape PES effects

According to previous studies, a computational model that involves a mixture of working memory maintenance and reinforcement learning (the ‘‘RL + WM” model) best explains human data in this task (both choices and response times; Collins and Frank, 2012; Collins et al., 2014; Collins et al., 2017; Collins, 2018; Collins and Frank, 2018; Master et al., 2020; McDougle and Collins, 2021). Moreover, independent measures of working memory capacity correlate with the relevant parameters of the working memory component of the model (Collins et al., 2014), and physiological measures (EEG, fMRI, genetic correlates of executive control and reinforcement learning neural circuits) further corroborate the hybrid model framework (Collins and Frank, 2012; Collins et al., 2017; Collins and Frank, 2018). We asked how this computational framework may relate to PES.

We fit three computational models to subjects' choice behavior in the task. The RL + WM model (see *Experimental Procedures* for model details), posits that a fast-learning, fast-forgetting working memory module and a slower procedural reinforcement learning system operate in parallel and both contribute to choice (Fig. 3A). Replicating previous work, we compared the RL + WM model to two RL-only models, both of which it outperformed (Fig. 3B; t-tests comparing RL + WM vs Multi-αAIC values: $t(118) = 6.04$, $p < 0.001$; RL + WM vs Basic RL: $t(118) = 16.67$, $p < 0.001$; Multi-αvs Basic RL: $t(118) = 15.40$, $p < 0.001$). As shown in Supplemental Fig. 1, simulations of both the RL + WM and Multi-α model captured load effects on learning, although the former model provided a better quantitative and qualitative fit to the data; the Basic RL model cannot capture any load effects on learning. Given that the RL + WM model better captures choice behavior in the task, it follows that it may also better explain other behavioral effects, such as PES.

PES is often thought to reflect the recruitment of cognitive control during performance monitoring (Botvinick et al., 2001). What determines the size of the PES effect? While our models do not explicitly simulate response times, we posited that one possible correlate of PES would be the subject's deviation from their choice policy on error trials (‘‘policy deviance”). That is, when a subject has a ‘‘confident” choice policy and errors are thus unlikely, they should show robust PES to get back on track; in contrast, when the policy is less definitive, there could be relatively less PES, perhaps reflecting a weaker policy error (Bennett et al., 2021). If true, the choice policy derived from fitting the superior RL + WM model should better match the pattern of observed PES load effects then the choice policy derived from the next best model, the RL-only Multi-α model.

To perform this analysis, we simulated each model using its best-fit parameters (see *Experimental Procedures* for simulation details). We then performed the robust PES analysis on the simulated data to

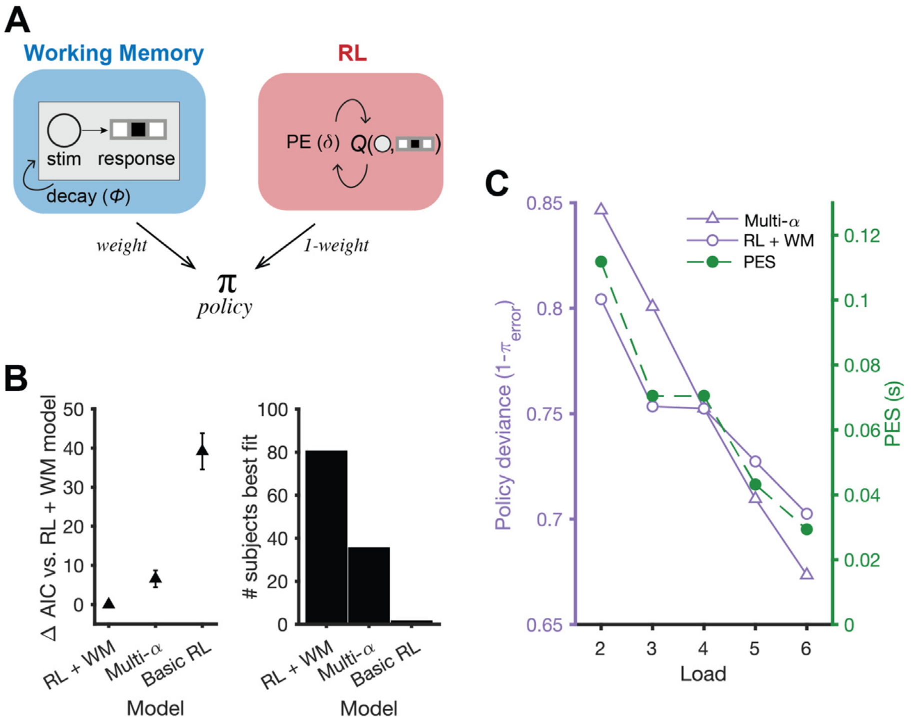

Fig. 3. Modeling results. **(A)** Schematic of RL + WM model (Collins and Frank, 2012). A working memory-based (WM) module stores stimulus–response associations in a short term store, but is susceptible to temporal decay. A reinforcement learning (RL) module performs trial-by-trial learning from reward prediction errors (PEs), incrementally updating the value (Q) of individual stimulus–response associations. Both modules are differentially weighted, based on load, to produce a combined action selection policy (π). **(B)** Comparison of the RL + WM model and both a load-sensitive RL model variant with different learning rates for each load condition (Multi-α) and a baseline standard RL model with no load-sensitive parameters (Basic RL). **(C)** Model simulation results. For error trials in the simulated agents, we computed the degree to which that error deviated from the current correct choice policy (“policy deviance”; purple functions/ordinate), operationalized as one minus the inferred probability of error on that trial. While both the RL + WM and Multi-α models demonstrate an effect of Load on policy deviance, the RL + WM model better captures the observed post-error slowing function (green function/ordinate). Error bars = 95% C.I.

specify trials that satisfied the criteria used to compute PES on the actual subject data (Fig. 2B). Using Eq. (3), we then quantified policy deviance on the error trials preceding the specified PES trials as 1-p(error) (we note that for the RL + WM model we used the weighted hybrid policy in Eq. (6)). Both models showed a negative effect of load on policy deviance (Fig. 3C), which was expected given that both models showed a negative effect of load on learning (Supplemental Fig. 1). Critically, however, the shape of the RL + WM model’s policy deviance function provided a better match to the observed PES function (Fig. 3C). This strong qualitative correspondence, in addition to the model fitting results (Fig. 3B), further suggests that PES during learning is shaped by an action selection policy that reflects the operation of multiple learning processes, rather than a single reinforcement learning system.

DISCUSSION

The phenomenon of post-error slowing (PES; Rabbitt, 1966) – an increase of choice response times after errors – is seen throughout human decision-making, from simple perceptual decisions (Purcell and Kiani, 2016), to recognition memory (Rae et al., 2014) and learning (Cavanagh et al., 2010; Verstynen et al., 2012). The most common account of PES argues that it represents the cost of exerting cognitive control to restore successful performance after errors (Botvinick et al., 2001; Gehring and Fencsik, 2001; Danielmeier and Ullsperger, 2011; Dutilh et al., 2012a; but see Notebaert et al., 2009). Neural correlates of PES appear to accord with the recruitment of control, such as increases in the event-related negativity (ERN) both in scalp EEG (Debener et al., 2005) and intracranially (Fu et al., 2019), and increased BOLD activity in medial prefrontal cortex (Gehring and Fencsik, 2001; Narayanan and Laubach, 2008), both of which are linked to cognitive control processes.

Computationally, studies of PES during instrumental learning have related the magnitude of PES effects to the reward prediction error experienced by the learner when an error is committed, as well as to frontal theta oscillations (Cavanagh et al., 2010). That is, on trials where the learner expected a reward but did not receive one, the PES effect (and frontal theta power) is larger than on trials where reward expectations were less certain (Cavanagh et al., 2010). Our results here show that a key factor influencing PES during instrumental learning is cognitive load (Fig. 2), which, to our knowledge, has not been manipulated in previous assays of PES during reward-based learning. These results appear to complicate a straightforward prediction error story, instead suggesting that PES may be primarily observed in contexts where top-down executive function guides decision-making (i.e., lower cognitive load), and would thus be especially efficacious in restoring performance after errors. Thus, executive function may mediate the relationship between prediction errors and PES. Indeed, according to our computational modeling analysis, deviation from a combined choice policy that reflects both reinforcement learning and working memory processes provided a compelling account of PES effects during learning (Fig. 3). These effects could not be captured by a model implementing reinforcement learning alone.

Similar results to those reported here have been observed in the domain of category learning: Tam et al. (2013) found that PES was most reliable early on in a category learning task, even though errors were made throughout. Crucially, PES was only reliably related to performance in a condition that required memory-based rule-following rather than incremental feature integration; that is, PES seemed to be related to a particular type of learning strategy, one which is closer to the kind of goal-directed behavior implied in standard studies of PES (Ashby and Maddox, 2005; Tam et al., 2013). While these findings are specific to category learning, they suggest that the relationship between PES and learning is not straightforward – PES may only appear when decision-making requires some form of explicit deliberation, memory retrieval, or online memory maintenance, such as in typical decision-making tasks, simple learning contexts (i.e., low cognitive load; explicit rule learning), and/or during the earliest phases of training.

Our study was inspired by discoveries showing that learning rarely reflects the workings of a single, monolithic learning system (Rmus et al., 2021). Indeed, multiple qualitatively distinct learning processes act in parallel (or opposition) across a variety of learning tasks in addition to category learning (Ashby and Maddox, 2005) and instrumental learning tasks (Collins and Frank, 2012), including sequential reinforcement learning tasks (Otto et al., 2015) and various motor learning tasks (Krakauer et al., 2019; McDougle and Taylor, 2019). In this study we examined PES in an instrumental learning task where people needed to learn to associate different stimuli with discrete actions. As mentioned earlier, even this relatively simple task has been shown to recruit (at least) two distinct learning strategies – top-down, working memory-based behavior that stores stimulus–response associations in short term memory, and incremental procedural reinforcement learning (Collins and Frank, 2012; Collins et al., 2014; Collins et al., 2017; Collins and Frank, 2018; Collins, 2018; Master et al., 2020; McDougle and Collins, 2021). The former strategy has been linked to executive processes in the prefrontal cortex, and the latter to the cortico-striatal dopamine system (Collins et al., 2017). More broadly, this study fits into a larger revision of many assumptions about instrumental learning – instead of reflecting isolated procedural learning systems, instrumental behavior often reflects complex interactions between executive functions (e.g., attention, working memory, rule learning) and lower-level learning circuits (Radulescu et al., 2019; McDougle et al., 2021; Rmus et al., 2021). This has implications for fields such as computational psychiatry, where researchers often attempt to computationally quantify behavioral effects for diagnostic purposes; our results suggest that PES, like action selection, is complex and cannot be easily linked to a single learning circuit.

One important caveat to any account of PES is that PES itself can, in practice, have multiple purposes (Purcell and Kiani, 2016). For example, it has been demonstrated that in certain conditions slowing can be seen when any rare event occurs, regardless of it being an error trial or a successful trial (Notebaert et al., 2009). We did not find robust evidence for that effect in these data, though this alternative ‘‘orienting” account of PES does not necessarily preclude a cognitive control account (Danielmeier and Ullsperger, 2011; Dutilh et al., 2012a). Future experiments could use probabilistic reward feedback to tightly control the frequency of error and correct trials to test alternative models of PES during learning.

Overall, our data suggest that in instrumental learning settings, PES may reflect a kind of normative accounting where the efficacy of cognitive control determines the utility of slowing after errors. Going forward, behavioral and neural techniques (such as EEG and fMRI) can be used to further elucidate how classic psychological effects like PES interact with a more holistic, multi-system understanding of human learning.

ACKNOWLEDGEMENTS

We thank Anne Collins and the CCN Lab (Berkeley) for their helpful guidance and comments. SDM was supported by National Institute of Mental Health fellowship F32 MH119797.

REFERENCES

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705.

Ashby FG, Maddox WT (2005) Human category learning. Ann Rev Psychol 56(1):149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217.

Bennett D, Niv Y, Langdon AJ (2021) Value-free reinforcement learning: policy optimization as a minimal model of operant behavior. Curr Opin Behav Sci 41:114–121. https://doi.org/10.1016/j.cobeha.2021.04.020.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108(3):624–652. https://doi.org/10.1037/0033-295X.108.3.624.

Cavanagh JF, Frank MJ, Klein TJ, Allen JJB (2010) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. NeuroImage 49(4):3198–3209. https://doi.org/10.1016/j.neuroimage.2009.11.080.

Collins AGE (2018) The tortoise and the hare: interactions between reinforcement learning and working memory. J Cogn Neurosci 30(10):1422–1432. https://doi.org/10.1162/jocn_a_01238.

Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 35(7):1024–1035.

Collins AGE, Frank MJ (2018) Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proc Natl Acad Sci 115(10):2502–2507. https://doi.org/10.1073/pnas.1720963115.

Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ (2014) Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci 34(41):13747–13756.

Collins AGE, Ciullo B, Frank MJ, Badre D (2017) Working memory load strengthens reward prediction errors. J Neurosci 37(16):4332–4342. https://doi.org/10.1523/JNEUROSCI.2700-16.2017.

Danielmeier C, Ullsperger M (2011) Post-error adjustments. Front Psychol 2. https://doi.org/10.3389/fpsyg.2011.00233.

Danielmeier C, Eichele T, Forstmann BU, Tittgemeyer M, Ullsperger M (2011) Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. J Neurosci 31(5):1780–1789. https://doi.org/10.1523/JNEUROSCI.4299-10.2011.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69(6):1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25(50):11730–11737. https://doi.org/10.1523/JNEUROSCI.3286-05.2005.

Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 22(6):1075–1081. https://doi.org/10.1016/j.conb.2012.08.003.

Dutilh G, van Ravenzwaaij D, Nieuwenhuis S, van der Maas HLJ, Forstmann BU, Wagenmakers E-J (2012a) How to measure post-error slowing: a confound and a simple solution. J Math Psychol 56(3):208–216. https://doi.org/10.1016/j.jmp.2012.04.001.

Dutilh G, Vandekerckhove J, Forstmann BU, Keuleers E, Brysbaert M, Wagenmakers E-J (2012b) Testing theories of post-error slowing. Attent Percept Psychophys 74(2):454–465. https://doi.org/10.3758/s13414-011-0243-2.

Fu Z, Wu D-A-J, Ross I, Chung JM, Mamelak AN, Adolphs R, Rutishauser U (2019) Single-neuron correlates of error monitoring and post-error adjustments in human medial frontal cortex. Neuron 101(1):165–177.e5. https://doi.org/10.1016/j.neuron.2018.11.016.

Gehring WJ, Fencsik DE (2001) Functions of the medial frontal cortex in the processing of conflict and errors. J Neurosci 21 (23):9430–9437. https://doi.org/10.1523/JNEUROSCI.21-23-09430.2001.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66 (4):585–595. https://doi.org/10.1016/j.neuron.2010.04.016.

Hale DJ (1969) Repetition and probability effects in a serial choice reaction task. Acta Psychol 29:163–171. https://doi.org/10.1016/0001-6918(69)90011-0.

Kerns JG (2004) Anterior cingulate conflict monitoring and adjustments in control. Science 303(5660):1023–1026. https://doi.org/10.1126/science.1089910.

Krakauer JW, Hadjiosif AM, Xu J, Wong AL, Haith AM (2019) Motor learning. Comprehensive Physiol 9(2):613–663. https://doi.org/10.1002/cphy.c170043.

Master SL, Eckstein MK, Gotlieb N, Dahl R, Wilbrecht L, Collins AGE (2020) Distentangling the systems contributing to changes in learning during adolescence. Dev Cogn Neurosci 41. https://doi.org/10.1016/j.dcn.2019.100732 100732.

McDougle SD, Taylor JA (2019) Dissociable cognitive strategies for sensorimotor learning. Nat Commun 10(1). https://doi.org/10.1038/s41467-018-07941-0.

McDougle SD, Ballard IC, Baribault B, Bishop SJ, Collins AGE (2021) Executive function assigns value to novel goal-congruent outcomes. Cereb Cortex bhab205. https://doi.org/10.1093/cercor/bhab205.

McDougle SD, Collins AGE (2021) Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. Psychonomic Bull Rev 28(1):20–39. https://doi.org/10.3758/s13423-020-01774-z.

Narayanan NS, Laubach M (2008) Neuronal correlates of post-error slowing in the rat dorsomedial prefrontal cortex. J Neurophysiol 100(1):520–525. https://doi.org/10.1152/jn.00035.2008.

Notebaert W, Houtman F, Opstal FV, Gevers W, Fias W, Verguts T (2009) Post-error slowing: an orienting account. Cognition 111 (2):275–279. https://doi.org/10.1016/j.cognition.2009.02.002.

Otto AR, Skatova A, Madlon-Kay S, Daw ND (2015) Cognitive control predicts use of model-based reinforcement learning. J Cogn Neurosci 27(2):319–333. https://doi.org/10.1162/jocn_a_00709.

Purcell BA, Kiani R (2016) Neural mechanisms of post-error adjustments of decision policy in parietal cortex. Neuron 89 (3):658–671. https://doi.org/10.1016/j.neuron.2015.12.027.

Rabbitt PM (1966) Errors and error correction in choice-response tasks. J Exp Psychol 71(2):264–272. https://doi.org/10.1037/h0022853.

Radulescu A, Niv Y, Ballard I (2019) Holistic Reinforcement learning: the role of structure and attention. Trends Cogn Sci 23 (4):278–292. https://doi.org/10.1016/j.tics.2019.01.010.

Rae B, Heathcote A, Donkin C, Averell L, Brown S (2014) The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. J Exp Psychol: Learn Memory Cogn 40 (5):1226–1243. https://doi.org/10.1037/a0036801.

Rmus M, McDougle SD, Collins AG (2021) The role of executive function in shaping reinforcement learning. Curr Opin Behav Sci 38:66–73. https://doi.org/10.1016/j.cobeha.2020.10.003.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275(5306):1593–1599.

Tam H, Maddox WT, Huang-Pollock CL (2013) Posterror slowing predicts rule-based but not information-integration category learning. Psychonomic Bull Rev 20(6):1343–1349. https://doi.org/10.3758/s13423-013-0441-0.

Verstynen T, Phillips J, Braun E, Workman B, Schunn C, Schneider W, Balasubramaniam R (2012) Dynamic sensorimotor planning during long-term sequence learning: the role of variability, response chunking and planning errors. PLOS ONE 7(10): e47336. https://doi.org/10.1371/journal.pone.0047336.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroscience.2021.10.016.

(Received 21 February 2021, Accepted 13 October 2021)
(Available online 23 October 2021)