Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Reviews Errors lead to transient impairments in memory formation

Alexandra Decker*, Amy Finn¹, Katherine Duncan¹

University of Toronto, Canada

ARTICLE INFO

Keywords: Errors Memory formation Arousal Norepinephrine Visual engagement

ABSTRACT

Making an error triggers a host of cognitive and behavioral adjustments theorized to boost task engagement and facilitate learning. Yet how errors influence memory formation – a cognitive process foundational to learning – remains unknown. Adaptive cognitive accounts of error processing propose that errors increase arousal, taskengagement, and attention, and should therefore enhance subsequent memory formation. Conversely, nonadaptive accounts of error processing and related research in arousal-mediated memory selectivity predict that errors could impair subsequent memory formation. We tested these divergent predictions in two experiments. In experiment 1, participants categorized trial-unique images as 'living' or 'nonliving', and following a short delay, performed a surprise recognition memory task. In contrast to what adaptive cognitive accounts of error processing would predict, people formed memories more poorly after errors, even when performance after errors was accurate. In experiment 2, we asked whether poorer memory formation after errors correlated with arousal or visual engagement after errors. Participants performed a modified Simon task in which they categorized trialunique images as 'natural' or 'man-made', while we recorded pupil dilation and visual fixations. Recognition memory was subsequently tested. We found that people who encoded memories more poorly after errors had larger pupillary responses to errors and spent less time fixating on stimuli after errors relative to before. Our results support non-adaptive theories of error processing by showing that errors transiently impair memory formation, possibly by increasing arousal and capturing attention.

1. Introduction

Realizing that we have made an error, even during a mundane task, is a powerful experience. Errors evoke marked changes in task accuracy, engender slower and more cautious responding (Houtman & Notebaert, 2013; Jentzsch & Dudschig, 2009; Notebaert et al., 2009), and increase autonomic arousal, as evidenced by increased blood pressure, ventilation, plasma norepinephrine and pupil size (a marker of norepinephrine mediated arousal) (Aarts, De Houwer, & Pourtois, 2013; Critchley, Tang, Glaser, Butterworth, & Dolan, 2005; Hoshikawa & Yamamoto, 1997; Inzlicht, Bartholow, & Hirsh, 2015; Maier, Ernst, & Steinhauser, 2019; Saunders, Lin, Milyavskaya, & Inzlicht, 2017; van Steenbergen & Band, 2013). These post-error adjustments have long been theorized to boost attention and task engagement, and - critically - to facilitate learning (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Holroyd & Coles, 2002; Yeung, Botvinick, & Cohen, 2004). Yet despite the theorized link between errors and subsequent learning, how mistakes shape episodic memory formation - a critical component of learning - remains unstudied. This gap is made all the more surprising by the ubiquity of errors in pedagogical and training contexts, where learning is the primary goal.

Though uninvestigated, theories of error processing make clear predictions about how errors should influence subsequent memory formation. Adaptive neurocognitive theories propose that errors generate cognitive conflict which energizes conflict resolution processes, leading to increases in arousal and task engagement (Botvinick et al., 2001; Kerns et al., 2004; Yeung et al., 2004). This adaptive theory of error processing proposes that the anterior cingulate cortex (ACC), a brain region involved in performance monitoring (Yeung et al., 2004), detects errors and signals to other brain regions to increase attention (Botvinick et al., 2001; Kerns et al., 2004; Yeung et al., 2004) and arousal (Critchley et al., 2005) to prevent future errors. Adaptive error signalling is thought to slow subsequent responding and improve accuracy (Botvinick et al., 2001; Kerns et al., 2004; Yeung et al., 2004). Consistent with this adaptive theory, errors on non-memory tasks have been shown to be followed by slower responses and improved accuracy, which is attributed to a more conservative response threshold and adaptive control adjustments (Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; Maier, Yeung, & Steinhauser, 2011; Marco-Pallarés, Camara, Münte, & Rodríguez-Fornells, 2008). Post-

* Corresponding author at: Department of Psychology, University of Toronto, 100 St. George street, M5S 3G3 Toronto, Canada.

E-mail address: alexandraleerdecker@gmail.com (A. Decker).

¹ Denotes equal contributions.

https://doi.org/10.1016/j.cognition.2020.104338

Received 2 July 2019; Received in revised form 22 April 2020; Accepted 25 May 2020 0010-0277/ © 2020 Elsevier B.V. All rights reserved.







error behavioral adjustments correlate with increased ACC activity (Kerns et al., 2004) and norepinephrine release (a measure of autonomic arousal), that is thought to mediate performance improvements (Critchley et al., 2005; Hajcak, Mcdonald, & Simons, 2003; Notebaert et al., 2009). According to the adaptive cognitive account of error processing, heightened arousal and task-engagement after errors should *improve* subsequent memory formation, as both arousal and task engagement benefit memory (deBettencourt, Norman, & Turk-Browne, 2018; Krebs, Boehler, De Belder, & Egner, 2015; Mather, Clewett, Sakaki, & Harley, 2016). This adaptive account also predicts that memory should be worse for content that precedes errors, owing to the cognitive control failures that led to an error.

However, more recent cognitive accounts of error processing would predict a different pattern. The non-adaptive account of error processing argues that errors are highly salient events that capture attention and divert attentional resources from the task at hand, impairing subsequent performance (Houtman & Notebaert, 2013; Notebaert et al., 2009). In line with this perspective, some studies show that people detect targets more poorly after errors, particularly when the interval between the error and the to-be-detected target is short (i.e., < 500 ms intertrial intervals (ITI); (Buzzell, Beatty, Paquette, Roberts, & McDonald, 2017; Jentzsch & Dudschig, 2009; Notebaert et al., 2009; Rabbitt & Rodgers, 1977)). Accordingly, the non-adaptive account of error processing predicts that errors should impair subsequent memory formation, owing to fewer attentional resources devoted to task relevant processing. Moreover, this account proposes that the potential benefits of errors on memory formation endorsed by the adaptive accounts of error processing should only be observed after a sufficient delay, when attentional resources have been re-directed towards task goals (Buzzell et al., 2017; Jentzsch & Dudschig, 2009; Murphy, van Moort, & Nieuwenhuis, 2016).

Strikingly, the predictions of this non-adaptive account better reflect the complex relationship between arousal and memory formation which has been characterized in a separate literature. Memory for emotionally arousing content has long been known to be superior than memory for neutral content (Doerksen & Shimamura, 2001; Hurlemann et al., 2005; Kensinger, 2007; Kensinger & Corkin, 2003), consistent with the intuition that arousal benefits new learning. However, arousalrelated enhancements in memory has been shown to come at a cost to memory for concurrently or subsequently presented neutral content (Clewett, Sakaki, Nielsen, Petzinger, & Mather, 2017; Kensinger, 2007; Loftus, Loftus, & Messo, 1987; Mather et al., 2016; Mather & Sutherland, 2011; Waring & Kensinger, 2011). Of note, the presentation of an emotionally arousing stimulus can enhance memory for related neutral content, when the neutral content is motivationally salient (Sakaki, Fryer, & Mather, 2014). This pattern of trade-offs in memory formation for arousing and neutral content are thought to be mediated by the actions of norepinephrine-induced arousal; once evoked by an arousing stimulus, norepinephrine is thought to prioritize the processing of salient content but diminish the processing of less important neutral content (Aston-Jones & Cohen, 2005; Clewett et al., 2017; Clewett, Huang, Velasco, Lee, & Mather, 2018; Hurlemann et al., 2005; Mather & Sutherland, 2011). Given that errors evoke arousal (Inzlicht et al., 2015; Maier et al., 2019; Saunders et al., 2017), this framework would predict a bi-directional effect of errors on memory, such that memory is enhanced for stimuli that trigger arousal and was processed at its peak (i.e., the stimulus which elicited the error) but diminished for subsequent neutral content. Further, these effects should correlate with the degree to which errors elicit arousal responses.

Here, in two experiments, we tested the divergent predictions of these theories. We tested the possibility that these theories make accurate predictions, but that their predictions emerge across different time courses (e.g., that errors at first impair, but then enhance memory formation).

In experiment 1, participants incidentally encoded images during a classification task, and then performed a surprise recognition memory

task. We examined how making an error influenced memory for images experienced on error trials and the ensuing trials. In experiment 2, a separate group of participants incidentally encoded images within a more cognitively demanding classification paradigm to assess how changes in cognitive control after errors related to changes in memory formation. Moreover, we further attempted to adjudicate between the adaptive and non-adaptive cognitive theories of error processing and the arousal framework. As a first step, we tested the opposing predictions of the adaptive cognitive control and arousal frameworks regarding how error-driven arousal influenced post-error memory formation (Joshi, Li, Kalwani, & Gold, 2016; McGinley, David, & McCormick, 2015: Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014: Reimer et al., 2014, 2016: Vinck, Batista-Brito, Knoblich, & Cardin, 2015). In particular, we assessed how increases in arousal after errors (indexed via pupillometry) influenced memory formation. We also used eye tracking to assess the opposing predictions of the adaptive and nonadaptive accounts of error processing regarding whether attention is pulled towards or pushed away from the task at hand

2. Experiment 1

In the first experiment, we examined how errors influenced subsequent memory formation. Participants completed a categorization task in which they incidentally encoded trial-unique images of animals and inanimate objects while categorizing each as living or nonliving (Fig. 1A). Critically, there was a high frequency of images from the nonliving category (~90% of trials), and thus, participants developed a prepotent response and, a high likelihood of making a categorization error when presented with an image from the living category (~10% of trials). Immediately after the categorization task, participants completed a surprise recognition task used to assess memory for each image that appeared during categorization (Fig. 1B). This allowed us to determine whether images presented after errors were remembered more or less often than images presented before errors, or at other points in the task.

2.1. Materials and methods

2.1.1. Participants

Prior to collecting the full sample, we collected a pilot sample of 20 young adult participants. This pilot sample was used to inform our confirmatory analyses. After piloting, we collected the full sample of 64 undergraduates at the University of Toronto, who all completed the study for course credit. The design and analyses were pre-registered on the Open Science Framework (https://osf.io/kbtze/) prior to data analysis. Following our pre-registered exclusion criteria, we removed 4 participants due to psychiatric illness (n = 3) and outlying memory scores that fell 3 SD below the sample mean (n = 1). The final analysis included 60 participants. A minimum sample size of 60 participants was chosen based on a power analysis indicating that this sample size would provide over 80% power to detect a medium to large effect for within subject analyses, similar to what we observed in our pilot data (d = 1.32). Participants had normal or corrected-to-normal vision and reported no history of head trauma, neurological or psychiatric illness. All experimental procedures were approved by the local ethics committee, and all participants provided written informed consent.

2.1.2. Stimuli

Stimuli consisted of 69 colored images of unique animals and 375 colored images of unique inanimate objects. Twenty of these images were used in the practice session (five of which were animals) and 424 were used in the experimental session of the study (64 of which were animals). Stimuli were presented using Psychopy software (Peirce, 2007) on a 13-inch MacBook.



Fig. 1. A schematic of the task structure in experiment 1. A) Participants encoded 332 images while categorizing each as living (~10%; infrequent category) or nonliving (~90%; frequent category). B) Immediately after, participants completed a surprise recognition task to test memory for each image encoded during the categorization task. Participants were asked to decide whether each image was 'old' (from the categorization task) or 'new'. After making each self-paced old/new memory decision, participants were asked to indicate their confidence on a 4-point scale.

2.1.3. Categorization task

Participants viewed 332 trial-unique images that appeared for 1500 ms each in the center of a white computer screen, Fig. 1A. Three hundred of these images were of inanimate objects (~90%; *frequent*, 'nonliving' category) and 32 were of animals (~10%; *infrequent*, 'living' category). Participants were asked to categorize each image as *living* or *nonliving* within a 1.5 second time window by pressing 'k' or 'j' on a keyboard. Response mappings were presented below each image and remained on the screen throughout the task. There was no ITI. Prior to starting the task, participants completed a practice block of 20 trials. The order of stimulus presentation was randomized across participants. The task took approximately 8.3 min to complete.

2.1.4. Surprise recognition task

Immediately after finishing the categorization task, participants completed a surprise recognition task. Participants were presented with an image and asked to indicate whether the image was *old* (from the categorization task; 'k') or *new* (images appearing *only* in the surprise recognition task; 'j', Fig. 1B). All of the categorized images were presented (n = 332) in an intermixed fashion with new images (n = 92; 32 living, 60 nonliving). Immediately after each memory decision, participants were asked to rate how confident they were in their decision on a four-point scale (just guessing = 1, not quite sure = 2, pretty sure = 3, 100% sure = 4). There was no time limit for making memory or confidence judgements. Image assignment to the old and new condition was counterbalanced across participants so that each image was similarly likely to be in the new condition across participants.

2.1.5. Statistical analyses

The pre-registered design and analyses can be found on the Open Science Framework (https://osf.io/6cmsa). Our pre-registration outlines analyses for two separate lines of questions, and we report the relevant analyses for the current questions involving errors in this manuscript. Statistical analyses were performed using R (R Core Team, 2013). For all analyses, estimates and statistics were obtained by fitting general linear and logistic mixed-effects regression models. We used the *glmer* and *lmer* function in the *lme4* packages (version 1.1-18) (Bates, Mächler, Bolker, & Walker, 2015) and the *lmerTest* package was used to obtain *p*-values for linear models using Satterthwaite's degrees of freedom method. Models included random intercepts but did not include random slopes for within subject variables included as fixed effects, unless otherwise indicated. We omitted random slopes from our primary analyses because some participants had only 1 trial contributing to a particular level of our variable of interest (e.g., 1 error trial, 1 post error trial). Statistically significant results from exploratory analyses were corrected for the family-wise error rate using false discovery rate (FDR) correction. Therefore, for all significant exploratory results, we report q-values (corrected *p* values) obtained from FDR correction. Data and scripts are available through the following link: https://osf.io/w7n3z/.

2.1.6. Data transformations and indices

Prior to running analyses, we removed linear drift in RT. Specifically, we regressed RT onto trial number (i.e., $RT \sim trial number$), separately for each participant, and extracted the residuals which were used in place of RT in all analyses. We operationalized *memory hits* as *old* responses to old images that were accompanied by high confidence (ratings of 3 or 4), and *memory misses as new* (forgotten) responses to old images. Low confidence old responses (i.e., with ratings of 1 or 2) were excluded because when deciding whether an image was old or new, participants had a 50% chance of being correct when guessing. Thus, removing low confidence old responses allowed us to eliminate guesses and thereby remove noise from the data. We also excluded trials from the recognition test which elicited responses that were faster than 300 ms seconds (see pre-registration). Prior to running statistical models, memory accuracy was dummy coded (i.e., 1 = memory hit, 0 = memory miss).

2.1.7. Confirmatory analyses

We first ran linear mixed effects models to test whether categorization RT, accuracy or memory (d prime) differed between images from the frequent (~90% of trials) and infrequent category (~10% of trials). The models examining RT and accuracy included trial type as both a fixed effect and random effect (e.g., $RT \sim$ trial type + (trial type|participant); see Supplementary Table 1, Models 1, 2 & 3 for model specifications).

Since the primary goal of our investigation was to assess whether RT and memory formation changed after errors, we ran two linear mixed effects models comparing RT and memory hits *after* errors relative to baseline. In these two models, either RT or memory hits was the dependent variable, and a trial's position relative to an error was the independent variable (e.g., $RT \sim trial position + (1|participant)$; see Supplementary Table 1, Models 4 & 7 for model specifications). The factor levels of the *trial position* variable coded for a trial's position

relative to an error (*post error* 1 = trials that appeared one trial after an error; *post error* 2 = trials that appeared 2 trials after an error; *baseline* = other correctly categorized trials). Critically, because *infrequent* trials (10%, living category) were highly salient, we excluded correct responses to infrequent trials from this analysis. We also ran analogous models to those described above to examine how making errors on a frequent versus infrequent trials separately influenced subsequent RT and memory hits (see models 5–6, 8–9).

Finally, we tested whether the simple presentation of an infrequent trial affected subsequent RT and memory formation. In these two models, either memory hits or RT was the dependent variable and a trial's position relative to a correct response on an infrequent trial was the independent variable (see Supplementary Table 1, Models 20 & 21).

2.1.8. Exploratory within subject analyses

In addition to comparing post-error memory formation to baseline, we ran two analogous exploratory models to assess whether RT or memory formation differed *before* versus after errors. This analysis allowed us to disentangle the effects of errors from the poor cognitive states that preceded them. These models were similar to the confirmatory models described above, except that the independent variable included a factor level that labelled trials appearing 1–3 trials *before* an error (see Supplementary Table 1, Models 10 & 13). We also ran four analogous exploratory models to test how memory and RT changed after errors made on frequent and infrequent trials separately (see models 11–12, 14–15).

In addition to our analyses examining how *all* errors influenced memory formation, we also ran four exploratory models to test how commission and omission errors separately influenced RT and memory formation. In these models, *trial position* (relative to either commission or omission errors) was the independent variable, and RT and memory hits were the dependent variables in separate models. We did this separately for errors made on frequent and infrequent trials (see Supplementary Table 1, Models 16–19).

2.1.9. Exploratory individual differences analyses

We ran individual differences analyses to better understand why some people had greater changes to memory formation after errors. Specifically, using Pearson's correlations, we tested whether people with the *greatest* changes to memory after errors had (1) the *best* memory for error trials or (2) displayed the greatest post-error slowing. Of note, an adjusted *error trial memory* variable was calculated by subtracting overall *memory hits* from *memory on error trials*. *Post-error memory changes* were calculated by subtracting memory hits 1 trial after errors from mean memory hits 3 trials before an error. Additionally, *post-error slowing* was operationalized as RT 1 trial after an error minus mean RT 3 trials before an error. One participant was excluded from these analyses because their post-error memory changes fell 3 standard deviations from the sample mean. Excluding this participant did not change the pattern of results.

2.2. Results

2.2.1. Better categorization accuracy on frequent trials, but better memory for infrequent trials

As expected, participants were slower and less accurate categorizing infrequent trials (that elicited a high degree of response conflict) than frequent trials (*RT difference:* b = 0.09, SE = 0.006, t(59) = 15.13, p < .001, r = 0.89; accuracy difference: b = -2.85, SE = 0.15, z = -18.74, p < .001, r = 0.62; categorization accuracy for frequent trials = 99%, infrequent trials = 87%; see Supplementary Fig. 2A and B & Table 1 for RTs on infrequent and frequent trials). However, because infrequent trials only made up 10% of the task, there were a similar number of frequent and infrequent trial errors (mean errors for infrequent items = 4.3, mean for frequent items = 3.2). Among errors, approximately 77% were commission and 23% were omission errors.

Table 1

Respo	nse time	on frequent	and infrequent	trials by	accuracy.

Trial type	Correct RT	Error RT
Frequent	464 ms	604 ms
Infrequent	582 ms	360 ms

People also remembered the infrequent trials better than the frequent trials (b = 0.48, SE = 0.06, t(59) = 8.35, p < .001, r = 0.74, Supplementary Fig. 2C), which may reflect the greater response conflict or salience related to infrequent trials.

2.2.2. Making an error leads to slower response times

Participants displayed reliably slower responses after errors (see Table 2). Indeed, responses were slower 1–2 trials after an error relative to baseline and before errors (all ps < .001, see Fig. 2A). These findings suggest that participants slow down after errors, a phenomenon attributed to increased response caution (Botvinick et al., 2001), as well as attentional capture by errors (Houtman & Notebaert, 2013; Notebaert et al., 2009). See Supplementary Table 3 for results from exploratory analyses on post-error slowing after commission versus omission errors.

2.2.2.1. Errors on both infrequent and frequent trials lead to post-error slowing. Responses were slower after errors on infrequent trials relative to baseline and before errors (all ps < .001; see Table 2 & Supplementary Fig. 3A). Similarly, errors on frequent trials led to slower RTs relative to baseline (all ps < .001; see Supplementary Fig. 3B), though not slower RTs relative to before errors (see Table 2).

People also slowed down after responding correctly to an infrequent trial, all ps < .001, suggesting that infrequent trials lead to slowing even when people do not make an error (see Supplementary Table 5).

2.2.3. Making an error impairs subsequent memory formation

People were less likely to recognize images presented on post error 1 (p < .001) and post error 2 trials (p = .052) relative to baseline images; see Table 3 & Fig. 2B. Critical to our hypothesis, memory was worse on post error 1 trials relative to *before* errors (p < .001), suggesting that errors, and not poor attentional states that precede errors, impair memory formation. These effects appear to be short-lived, as memory on post error 2 trials was not worse than memory before errors, p = .20.

2.2.3.1. Errors on infrequent trials impair subsequent memory formation. Post-error memory decrements were observed even after restricting analyses to errors made on infrequent trials (see Table 3). Relative to baseline, errors on infrequent trials impaired memory formation on post error 1 and 2 trials (p < .01). Relative to before errors, memory formation was impaired on post error 1 trials (p < .001), but not post error 2 trials (p = .11), see Supplementary Fig. 4A). This pattern – that errors lead to memory dips one trial after, but not two trials after an error - further supports the idea that posterror memory decrements are short-lived. To further understand these post-error memory decrements, we assessed whether they were partially driven by the presentation of an infrequent trial - rather than an error itself. Therefore, we examined whether infrequent trials impaired subsequent memory formation even when people responded correctly. We found that the simple presentation of an infrequent trial to which participants responded correctly did not impair subsequent memory (all ps > .05; see Supplementary Fig. 6; Supplementary Table 5). These findings suggest that errors - and not the presentation of an infrequent trial - impair subsequent memory formation.

2.2.3.2. Errors on frequent trials impair subsequent memory formation. When restricting analyses to errors made on frequent

Results from analyses on post-error slowing from study 1.

	Post-error slowing (all errors)	Post-error slowing (infrequent trial errors)	Post-error slowing (frequent trial errors)
Baseline vs. post error 1 Baseline vs. post error 2 Before vs. post error 1 Before vs. post error 2	$ b = 0.11, SE = 0.008, t(17428) = 13.88, \\ p < .001, r = 0.10 \\ b = 0.07, SE = 0.008, t(17428) = 8.29, \\ p < .001, r = 0.06 \\ b = 0.13, SE = 0.01, t(1472) = 10.13, p < .001, \\ q < 0.001, r = 0.26 \\ b = 0.09, SE = 0.01, t(1472) = 6.97, p < .001, \\ q < 0.001, r = 0.18 $	$ \begin{array}{l} b = 0.12, SE = 0.01, t(17547) = 11.67, \\ p < .001, r = 0.09 \\ b = 0.07, SE = 0.01, t(17547) = 7.02, p < .001, \\ r = 0.05 \\ b = 0.17, SE = 0.01, t(1004) = 13.27, p < .001, \\ q < 0.001, r = 0.39 \\ b = 0.11, SE = 0.01, t(1004) = 9.06, p < .001, \\ q < 0.001, r = 0.28 \\ \end{array} $	b = 0.08, SE = 0.01, t(17656) = 6.47, p < .001, r = 0.05 b = 0.07, SE = 0.01, t(17657) = 5.16, p < .001, r = 0.04 b = 0.04, SE = 0.02, t(544) = 1.71, p = .088, r = 0.07 b = 0.03, SE = 0.03, t(543) = 1.00, p = .319, r = 0.04

Bold denotes significant results.



Fig. 2. A) RTs during the categorization task was reliably slower on post error 1 and 2 trials relative to before errors, and relative to mean RT on other frequent trials, *'baseline RT'*, see horizontal line; grey shading reflects the standard error, all ps < .001, q < 0.001. Note that *'before'* reflects response time on frequent trials that appeared 1–3 trials before an error. B) Memory formation was reliably worse on trials that appeared after an error on the categorization task, *'post 1'*, relative to before errors and baseline memory, ps < .001, q < 0.001, see horizontal line; grey shading reflects standard error, ps > .05, but was marginally worse than baseline memory, ps < .053. There was no evidence of post-error memory decrements 3 trials after an error. C) There was no relationship between participants' post-error memory decrements and their memory for images presented on error trials (adjusted for overall memory hit rate), r(57) = 0.15, p = .25.

trials, we observed that relative to baseline, memory formation was worse on post error 1 (p < .05), but not post error 2 trials (p = .90). Relative to before errors, memory formation was also marginally worse on post error 1 trials (p = .06), but not worse on post error 2 trials (p = .77), see Table 3 & Supplementary Fig. 4B.

2.2.4. Error trials are not remembered differently than correct trials

People were not more likely to later recognize images presented on error trials than those on correct trials, all ps > .05. This was true for errors on infrequent trials (b = -0.18, SE = 0.17, z = -1.02, p = .305, r = -0.05) and frequent trials (*Baseline vs. frequent error*: b = -0.24, SE = 0.18, z = -1.39, p = .166, r = -0.07). People were also not more likely to later recognize images presented on error trials than those presented before errors (*Before vs. frequent error*: b = 0.24, SE = 0.21, z = 1.12, p = .262, r = 0.07).

Furthermore, exploratory analyses revealed that neither omission nor commission errors on infrequent trials were remembered differently than correct trials (all ps > .05). In contrast, omission errors on *frequent* trials were remembered *worse* than baseline and before errors (baseline vs. omission errors: b = -0.69, SE = 0.23, z = -3.01, p = .003, q = 0.007, r = -0.19; before vs. omission errors: b = -0.71, SE = 0.26, z = -2.77, p = .006, q = 0.02, r = -0.19). A different pattern was present for *commission* errors. On frequent trials, commission errors were remembered marginally *better* than baseline memory (b = 0.56, SE = 0.30, z = 1.88, p = .061, q = 0.08, r = 0.15), though no different than before errors (b = 0.53, SE = 0.32, z = 1.65, p = .100, r = 0.15). These findings suggest that memory formation is *worse* on omission error trials, but *better* on commission error trials relative to baseline. Thus, the act of committing a commission error may *enhance* memory formation, though further work is

Table	3
-------	---

Results from analyses on post-error memory decrements from study 1.

	Memory hits after errors (all errors)	Memory hits after errors (infrequent trial errors)	Memory hits after errors (frequent trial errors)
Baseline vs. post error 1 Baseline vs. post error 2 Before vs. post error 1 Before vs. post error 2	$ \begin{split} b &= -0.71, \text{SE} = 0.13, \text{z} = -5.52, \textbf{p} < .001, \\ r &= -0.19 \\ b &= -0.25, \text{SE} = 0.13, \textbf{z} = -1.95, \textbf{p} = .052, \\ r &= -0.07 \\ b &= -0.65, \text{SE} = 0.15, \textbf{z} = -4.35, \textbf{p} < .001, \\ \textbf{q} &< 0.001 \textbf{p} < .001, \textbf{r} = -0.18 \\ b &= -0.19, \text{SE} = 0.15, \textbf{z} = -1.26, \textbf{p} = .209, \\ r &= -0.05 \end{split} $	$ \begin{array}{l} b = -0.87, SE = 0.16, z = -5.33, p < .001, \\ r = -0.23 \\ b = -0.42, SE = 0.15, z = -2.75, p = .006, \\ r = -0.12 \\ b = -0.73, SE = 0.18, z = -3.94, p < .001, \\ q < 0.001, r = -0.20 \\ b = -0.28, SE = 0.18, z = -1.58, p = .113, \\ r = -0.08 \end{array} $	b = -0.43, SE = 0.20, z = -2.10, p = .035, r = -0.12 b = -0.03, SE = 0.21, z = -0.13, p = .898, r = -0.01 b = -0.46, SE = 0.24, z = -1.89, p = .058, r = -0.13 b = -0.07, SE = 0.24, z = -0.29, p = .768, r = -0.02

Bold denotes significant results.

needed to better disentangle how different errors types influence concurrent memory formation.

2.2.5. Assessing individual differences: post-error memory decrements correlate with post-error slowing, but not memory for error trials

People with larger post-error memory decrements displayed greater post-error slowing, r(57) = 0.29, p = .028, q = 0.056, Supplementary Fig. 7. These findings raise the possibility that the mechanisms involved in post-error slowing (error detection, response caution, internally ruminating on the error) interfere with memory formation. Interestingly, people with greater post-error memory decrements did not have the better memory for errors, r(57) = 0.15, p = .253, Fig. 2C. Thus, post-error memory decrements were not necessarily caused by ruminating on the *content* of the error trial or attentional capture by the image presented on error trials.

3. Experiment 2

In experiment 1, we found that errors transiently impaired subsequent memory formation. Memory formation was worse after errors relative to baseline memory and before errors. Because memory formation was impaired relative to before errors, it is unlikely that memory formation failures were due to poor cognitive states that preceded errors and then lingered into post-error trials.

These findings appear inconsistent with the adaptive account of error processing that predicts errors boost arousal and attention and in doing so, should improve post-error memory (Botvinick et al., 2001; Kerns et al., 2004; Yeung et al., 2004). Instead, they appear consistent with non-adaptive theories of error processing proposing that errors divert attention from task-goals and interfere with task performance (Houtman & Notebaert, 2013; Notebaert et al., 2009). It is possible, though, that post-error memory decrements reflect the costs of cognitive control; that is, the very same cognitive state that boosts task performance after errors may also impair memory formation. We could not determine whether this was the case in experiment 1, because it was not designed to measure how errors impact concurrent task performance: 90% of trials required the same response and therefore did not require cognitive control; indeed, most accurate responses could be produced by simply falling back on the prepotent response. Moreover, without an index of arousal or task engagement, it is unclear whether memory formation deficits after errors coincided with increases in arousal, as would be predicted by the emotional arousal framework (Mather et al., 2016; Mather & Sutherland, 2011; Waring & Kensinger, 2011), or strictly coincided with poorer attention and task engagement, as would be predicted by non-adaptive theories of error processing (Houtman & Notebaert, 2013; Notebaert et al., 2009).

To address these outstanding questions, in experiment 2, we investigated whether changes in memory formation after errors correlated with changes in arousal and/or stimulus engagement. Participants first incidentally encoded trial unique images during a classification task while pupillometry and eye fixation data were recorded. To increase the level of cognitive control required for accurate performance, we used a modified Simon task (Hommel, Proctor, & Vu, 2004). After the classification task, participants performed a surprise recognition task, in which memory for each encoded image was tested, along with a set of new images. We first examined whether classification accuracy decreased after error trials (i.e., whether people made more errors after errors), and also whether worse accuracy after errors correlated with worse memory formation after errors. We also investigated whether post-error memory decrements coincided with arousal responses (measured by pupillometry) and/or attentional engagement (measured by visual fixations).

3.1. Materials and methods

3.1.1. Participants

Prior to collecting the full sample, we collected a pilot sample of young adult participants (n = 13). Pilot data was used to ensure our experiment was working properly and inform our confirmatory hypotheses. After piloting, we collected a full sample of sixty-seven undergraduates from the University of Toronto who participated for course credit. Data from 7 participants were excluded due to psychiatric illness (n = 4), very low memory hit rate (n = 1), categorization accuracy that fell 3 SD below the sample mean (n = 1) and pressing the wrong button throughout the task (n = 1). Furthermore, for between subject analyses, we identified and excluded 1 participant whose posterror memory decrements fell 3 standard deviations and 3 absolute deviations from the sample mean. The final sample included 60 participants for within subject analyses and 59 participants for between subject analyses. We expected that post-error memory decrements would show a similar effect size to that observed in experiment 1. A power analysis indicated a sample of 60 participants would achieve at least 80% power for detecting a smaller effect size of d = 0.35 for within subject effects. Moreover, this sample size achieves 80% power to detect correlations as small as r = 0.24, at the standard alpha of 0.05. All experimental procedures were approved by the local ethics committee, and all participants provided written informed consent.

3.1.2. Stimuli

Stimuli consisted of 476 luminance-matched (Willenbockel et al., 2010) grey scale images of objects that were either man-made (n = 228; tools, furniture, etc.) or natural (n = 228; fruits, animals, plants, etc.). Twenty of these images were used in a practice phase and 456 were used in the experimental phase of the study. Stimuli were presented on a white background on a 21.5-inch iMac using SR Research Experiment Builder.

3.1.3. Simon categorization task

Participants performed the modified Simon task (Fig. 3A) (Hommel et al., 2004) while eye position and pupillometry data were recorded using the EyeLink 1000 Desktop Mount eye tracker. The task was performed in a dimly lit room and participants rested their heads on a chinrest positioned approximately 50.8 cm from the computer screen throughout.

Participants viewed 228 trial-unique images that appeared for 1000 ms each on either the left or right side of the computer screen. Half of these images were of man-made objects and half were of natural things. Participants were asked to categorize each image as natural or man-made within the 1000 ms stimulus presentation, by pressing the 'f' or 'j' key with their left and right index fingers. This was followed by a 500 ms inter-stimulus interval, in which a white screen with a black cross appeared. Critically, on half of the trials, the image's screen position (left vs. right) spatially matched the correct keyboard response option (congruent trial) and on half the trials it did not (incongruent trial). Image assignment to the congruent and incongruent category was counterbalanced across subjects. Response mappings were presented below each image and remained on the screen throughout the task. Stimulus presentation was structured such that no more than two incongruent trials appeared in a row but was otherwise randomized across participants. Prior to completing the task, participants completed 20 practice trials with feedback. The task took about 5.7 min to complete.

3.1.4. Surprise recognition task

Immediately after the Simon categorization task, participants completed a surprise recognition task to test memory for the images presented during the Simon task, Fig. 3B. Eye tracking and pupillometry data was recorded during this session but were not used in the preregistered analyses. During the surprise recognition task, participants



Fig. 3. A schematic of the task structure in experiment 2. A) Participants completed a Simon task in which they encoded 228 trial-unique grey scale images that were positioned either on the left or right side of the computer screen. Participants were asked to categorize each image as either 'man-made' (n = 114) or 'natural' (n = 114) by pressing 'f' and 'j' on a keyboard. On half of the trials, the image was presented on the same side as the correct keyboard response, *congruent trials*, whereas on half the trials, the image was positioned on the opposite side to that of the correct keyboard response, *incongruent trials*. B) Immediately after, participants completed a surprise recognition task designed to test memory for each image presented during the Simon categorization task. Participants were asked to decide whether each image was old (from the categorization task) or new. After making each self-paced memory decision, participants were probed to indicate their confidence on a 4-point scale.

viewed all of the images from the categorization task (n = 228), plus 228 new images that were not previously shown (114 new man-made). Participants were asked to indicate whether each image was old or new and were provided with as much time as needed to make a decision. As in experiment 1, participants indicated their confidence on a 4-point scale. Image assignment to the old and new condition was counterbalanced across participants.

3.1.5. Statistical analysis

The pre-registered design and analyses for experiment 2 can be found on the Open Science Framework: https://osf.io/4ndg6. For all models with repeated measures, general mixed-effects regression models (Bates et al., 2015) were fit using the *lme4* package (version 1.1–18) and the *lmerTest* package was used to obtain *p*-values for linear models using Satterthwaite's degrees of freedom method. All models testing within subject effects included random intercepts and slopes for within subject variables included as fixed effects, nested within participants, unless otherwise indicated. For across subject analyses, Pearson's correlations were used, unless otherwise noted. As in experiment 1, we corrected all statistically significant exploratory results for multiple comparisons using FDR correction. For these analyses, we report qvalues for all significant results. All raw data is available through the following link: https://osf.io/cm5bg/.

3.1.6. Data transformations and indices for behavioral data

Both RT and memory hits were treated in the same way as outlined in experiment 1; we removed linear drift in RT by residualizing out the effects of trial number, and excluded single trial data from the recognition test that was faster than 300 ms. Only old responses to old items accompanied by high confidence were considered *memory hits*.

3.1.7. Data transformations and indices for pupil and fixation data

The EyeLink system uses a 35 mm lens, 5-point gaze location calibration, and a monocular right-eye sampling rate of 1000 Hz. Blink artifacts that were detected using Eyelink's blink detection algorithm were removed from pupillometry data using linear interpolation from 200 ms prior to and 200 ms post-blink onset (Cavanagh & Frank, 2014). Time windows where pupil data was missing for \geq 500 ms were not interpolated, and instead treated as missing data because blinks typically do not last > 500 ms (Lin, Saunders, Hutcherson, & Inzlicht, 2018). Pupillometry data was down sampled to 100 Hz prior to analysis. Linear drift in pupil data due to time dependent effects and individual differences in mean pupil size were removed by regressing pupil size at each sample onto trial number (within participants) and extracting the residuals. After extracting residuals, we calculated mean pupil size for each trial by averaging pupil size in the 1-2 second time window after stimulus onset, see Fig. 6A. This time window was chosen because pupil dilations for images that are later remembered vs. forgotten diverge around 1 s after stimulus onset (Naber, Frässle, Rutishauser, & Einhäuser, 2013). Prior to analysis, mean pupil size in this time range (1-2 s post stimulus onset) was mean centered within participants.

Table 4

Response time on congruent and incongruent trials by accuracy.

Trial type	Correct RT	Error RT
Congruent	627 ms	673 ms
Incongruent	637 ms	657 ms

3.1.8. Confirmatory analyses

3.1.8.1. Within subject analyses. We first fit three mixed effects models to test whether (1) accuracy, (2) RT, or (3) memory hits differed between congruent and incongruent trials (see Supplementary Table 2, Models 1–3 for model specifications).

As in experiment 1, we also fit linear mixed effects models to investigate whether errors influenced RT and memory formation. In these two models either memory hits or RT was the dependent variable and *trial position* relative to an error was the independent variable. The trial position included separate levels for trials that occurred 1–3 trials before an error (labelled as *before*), error trials (labelled as *error*), post error trials that appeared 1–3 trials after an error (labelled as *before*) (see Supplementary Table 2, Models 4 & 5 for model specifications). Of note, as in Experiment 1, we also explored how commission versus omission errors influenced subsequent RT and memory formation. Model specifications and results from these analyses are detailed in the supplement (see Supplementary Table 2 & 6, Models 6–9).

Additionally, we performed analyses to test whether participants displayed post-error classification decrements. Therefore, for each participant, we calculated (1) the likelihood of making an error after an error and (2) the likelihood of making an error after a correct response and compared these values. Since these variables were not normally distributed, we used a Wilcoxon signed rank test with continuity correction which does not assume a normal distribution. Similarly, we used a Wilcoxon signed rank test to explore whether the likelihood of making an error differed on post error 1, 2, and 3 trials relative to (1) before errors and (2) at baseline.

Finally, to examine changes to pupil size after errors, we fit a model to test whether pupil size differed after errors (post 1, 2, 3) relative to before errors and at baseline. In this model, pupil size was the dependent variable, and a trial's position relative to an error was the independent variable (see Supplementary Table 2, Model 10).

3.1.8.2. Between subject analyses. We calculated post-error accuracy decrements for each participant by subtracting each participant's likelihood of making an error after a correct response from the likelihood of making an error after an error. Post-error memory decrements were calculated in the same way as in experiment 1 (mean memory hits before errors minus mean memory hits on post error 1 trials). Using these two variables, we tested whether people who had the worst memory decrements after errors had a higher likelihood of making an error after an error than after a correct response. We used Spearman's correlations for this analysis because the accuracy decrement variable was not normally distributed. Additionally, Pearson's correlations were used to assess whether people with the

greatest post-error memory decrements displayed the greatest changes in pupil size on error trials (mean pupil size on error trials minus pupil size 1–3 trials before errors) or on post error 1 trials (mean pupil size on post error 1 trials minus pupil size 1–3 trials before errors). Of note, while the former model investigating pupil dilation and memory decrements was exploratory, the latter model was pre-registered.

3.1.9. Exploratory analyses

Using linear mixed effects modelling, we ran two models to explore whether (1) individuals spent less time fixating on images after errors or (2) had fewer fixations on images after errors relative to before errors and at baseline (see Supplementary Table 2, Model 11 & 12 for model specifications). In these models, *either* time fixating on stimuli or the number of fixations on images were the dependent variables. Trial position was the independent variable in both models. We also fit a linear mixed effects model to test whether memory hits on any given trial was related to pupil size (see Supplementary Table 2, Model 13).

We also ran across subject analyses to test whether post-error memory decrements correlated with (1) individual differences in fixation time after errors (mean fixation time 1–3 trials before an error minus 1 trial after an error), (2) memory for error trials (mean memory hit rate on error trials minus overall hit rate) or (3) post-error slowing. To assess the relationship between fixations and memory decrements, we performed Spearman's correlations, because the fixation difference score was not normally distributed.

3.2. Results

3.2.1. Categorization accuracy and recognition memory

As expected, participants were faster and more accurate categorizing congruent (hit rate: 90%) than incongruent trials (hit rate: 88%; *RT difference*: b = 7.30, SE = 2.20, t(58) = 3.31, p = .002, r = 0.40; *accuracy difference*: b = -0.20, SE = 0.07, z = -3.06, p = .002, r = -0.06; see Table 4 for raw RTs), suggesting that participants experienced more response conflict on incongruent trials, Supplementary Fig. 8A & B. Among error trials, 68% were commission and 32% were omission errors. There were no differences in memory hits between congruent and incongruent trial types, b = 0.03, SE = 0.04, z = 0.82, p = .410, r = 0.01, Supplementary Fig. 8C, indicating that the additional conflict associated with incongruent trials did not impact memory formation.

3.2.2. Making an error slows subsequent response time

Participants responded slower on post error 1 and 2 trials relative to baseline and before errors (ps < .05; see Table 5 & Fig. 4A). In contrast, there was no significant evidence of post-error slowing on post error 3 trials relative to baseline or before errors (all ps > .05), demonstrating that post-error slowing recovers within 4.5 s (in our task, 3 trials) after an error. See Supplementary Table 6 for model estimates from analyses examining post-error slowing separately for omission and commission errors.

3.2.3. Errors lead to memory formation deficits that recover quickly Consistent with experiment 1, memory formation was worse

Table 5

Results from analyse	es on post-error	slowing and n	nemory decrements	from study 2

	Post-error slowing	Post-error memory decrements
Baseline vs. post error 1 Baseline vs. post error 2 Baseline vs. post error 3 Before vs. post error 1 Before vs. post error 2 Before vs. post error 3	$ b = 36.58, SE = 4.80, t(70) = 7.62, p < .001, r = 0.67 \\ b = 12.58, SE = 4.70, t(97) = 2.67, p = .009, r = 0.26 \\ b = 8.84, SE = 5.05, t(78) = 1.75, p = .084, r = 0.19 \\ b = 35.41, SE = 5.35, t(64) = 6.62, p < .001, r = 0.64 \\ b = 11.41, SE = 5.02, t(113) = 2.27, p = .025, r = 0.21 \\ b = 7.67, SE = 5.40, t(66) = 1.42, p = .160, r = 0.17 \\ \end{cases} $	$ \begin{split} b &= -0.33, \text{SE} = 0.09, z = -3.72, p < .001, r = -0.09 \\ b &= -0.17, \text{SE} = 0.09, z = -1.87, p = .061, r = -0.05 \\ b &= 0.10, \text{SE} = 0.10, z = 0.95, p = .342, r = 0.03 \\ b &= -0.16, \text{SE} = 0.10, z = -1.69, p = .092, r = -0.04 \\ b &= -0.005, \text{SE} = 0.10, z = -0.05, p = .957, r = 0 \\ b &= 0.26, \text{SE} = 0.11, z = 2.42, p = .016, r = 0.07 \end{split} $

Bold denotes significant results.



Fig. 4. Post-error slowing on the Simon categorization task and post-error memory decrements. Grey shading shows standard error. A) RT was reliably slower on trials that occurred 1–2 trials after categorization errors relative to before errors, ps < .05, and relative to baseline, see horizontal line, ps < .01. RT was not reliably slower 3 trials after an error, relative to before an error, or relative to baseline RT, all ps > .05. Note that 'before' reflects the average RT on frequent trials that appeared 3 trials before an error. Standard error bars are shown. B) Memory formation was marginally worse on trials that occurred immediately after a categorization error relative to baseline memory, p < .001. Memory formation three trials after an error was reliably better than before an error, p < .05. C) Participants who displayed better memory for error trials in experiment 2 had larger post-error memory decrements, p < .01, q = 0.032.

relative to baseline on post error 1 trials (p < .001), and marginally worse on post error 2 trials (p = .06), but not worse on post error 3 trials (see Table 5 for model estimates). Additionally, memory was marginally worse relative to before errors on post error 1 trials (p = .092; see Table 5 & Fig. 4B), but not worse on post error 2 trials (p = .96). By post error 3 trials, memory was *better* than memory before errors (p = .016; see Table 5 & Fig. 4B). The timing of these effects suggests that memory dips quickly after errors, but soon after recovers to baseline (by about 4.5 s in our task; see Fig. 4B).

3.2.4. Memory for error trials is worse than baseline memory

Memory formation was worse on error trials relative to baseline (baseline vs. error: b = -0.19, SE = 0.08, z = -2.51, p = .012, r = -0.05), but not worse relative to before errors (before vs. error: b = -0.03, SE = 0.09, z = -0.29, p = .769, r = -0.01). When restricting analyses to commission errors, there was no difference between memory for errors relative to baseline or before errors (baseline vs. errors: b = -0.07, SE = 0.09, z = -0.77, p = .441, r = -0.02; before vs. errors: b = 0.03, SE = 0.10, z = 0.34, p = .735, r = 0.01). In contrast, omission error trials were remembered more poorly relative to baseline (b = -0.35, SE = 0.12, z = -2.84, p = .005, r = -0.09) and but not worse than before errors (b = -0.16, SE = 0.14, z = -1.11, p = .268, r = -0.04). This pattern of results suggests that poor attentional states that lead to omission errors are also present on error trials.

3.2.5. Concurrent task accuracy is impaired after errors

In contrast to the predictions of adaptive cognitive control theories that propose errors should boost attention and enhance accuracy (Botvinick et al., 2001; Kerns et al., 2004; Yeung et al., 2004), errors led to worse classification accuracy. This is consistent with the non-adaptive account of error processing that proposes that errors orient attention away from a task (Houtman & Notebaert, 2013; Notebaert et al., 2009). Indeed, people were more likely to make an error after an error than after a correct response, V = 383, p < .001, Fig. 5A. We also observed that accuracy was worse on post error 1 trials relative to mean accuracy across the task (V = 1425, p < .001), and relative to before errors (V = 1335, p < .001; Fig. 5B). Decreased accuracy after errors recovered by post-error 2 trials, however, such that accuracy was better on post error 2 trials than before errors (V = 461, p = .01), and no different from accuracy across the task (V = 786, p = .34). By post error 3 trials, accuracy was no different than before errors (V = 682, p = .25), or mean task accuracy (V = 1003, p = .51), Fig. 5B. This combination of findings suggests that errors - and not poor attentional states that precede errors - transiently impair task performance.

3.2.6. Impaired concurrent task accuracy after errors is unrelated to impaired memory formation after errors

Of note, decreased accuracy on post error 1 trials did not correlate with post-error memory decrements, $r_s(57) = 0.04$, p = .78, Fig. 5C. Thus, separate mechanisms likely underlie post-error memory and accuracy decrements.

3.2.7. Pupil size increases after errors

Pupil size was larger on post error 1 and 2 trials relative to baseline and before errors (see Fig. 6B; *baseline* vs. *post error* 1: b = 60.81, SE = 9.37, t(53) = 6.49, p < .001, r = 0.67; *baseline* vs. *post-error* 2: b = 32.42, SE = 7.81, t(58) = 4.15, p < .001, r = 0.48; before vs. post *error* 1: b = 61.20, SE = 11.40, t(52) = 5.37, p < .001, r = 0.60; *before* vs. *post error* 2: b = 33.24, SE = 9.76, t(51) = 3.41, p = .001, r = 0.43). Since pupil dilation is thought to be a measure of arousal (Critchley et al., 2005), these findings suggest that errors are followed by increases in arousal. Pupil size on post error 3 trials did not differ from baseline or before errors (baseline vs. post 3: b = 8.39, SE = 6.23, t(96) = 1.35, p = .181, r = 0.14; before vs. post 3b = 9.78, SE = 7.85, t(60) = 1.25, p = .217, r = 0.16, Fig. 6B), suggesting that error-related pupil responses recover by 4.5 s after errors.

3.2.8. Post-error memory decrements coincide with larger pupils

Across individuals, post-error memory decrements (memory hits before errors minus memory hits on post error 1 trials) correlated with adjusted pupil size on error trials (pupil size on error trials minus pupil size before errors), r(57) = 0.31, p = .017, q = 0.04, Fig. 6C. This suggests that people who had the greatest post-error memory decrements had the greatest increases in arousal after errors. In contrast, post-error memory decrements were unrelated to adjusted pupil size on post error 1 trials (pupil size on post error 1 trials minus pupil size before errors), r(57) = 0.20, p = .13. Of note, these correlations did not differ significantly from each other, b = -0.0004, SE = 0.0004, t (115) = -0.95, p = .342, r = 0.09.

3.2.9. Exploratory results

3.2.9.1. Smaller pupils correlate with better memory formation. Larger pupil sizes correlated with worse memory formation across the task (b = -11.48, SE = 2.93, t(47) = -3.92, p < .001, r = 0.50), suggesting that on this particular task, when people are in an optimal state for forming memories, their pupil sizes were relatively smaller (see Supplementary Fig. 12).



Fig. 5. A) Participants were more likely to make an error after an error than after a correct response (p < .001). B) Accuracy was worse on post error 1 trials relative to before errors (p < .001) and mean task accuracy (p < .001). However, accuracy was better on post error 2 trials than before errors (p = .01) and no different than mean task accuracy (p = .34). On post error 3 trials, accuracy was no different than mean task accuracy (p = .25). C) There was no relationship between participants' post-error accuracy decrements and post-error memory decrements, $r_s(57) = 0.04$, p = .78.

3.2.9.2. Post-error memory decrements are related to less visual engagement. People did not spend less time fixating on stimuli after errors and did not have fewer fixations after errors (ps > .05; Fig. 7A & Supplementary Table 7). Interestingly, people did have more fixations 3 trials after an error relative to baseline and before errors, ps < .05, qs < 0.09. In addition, people who spent less time fixating on stimuli on post error 1 trials (relative to before errors) had the largest posterror memory decrements $r_s(57) = 0.26$, p = .04, q = 0.053, Fig. 7B. These findings indicate that, while people generally did not fixate less after errors, those who did, had the most difficulty forming memories during that time. Thus, post-error memory decrements may have been in part driven by less visual engagement after errors.

3.2.9.3. Post-error memory decrements correlate with memory for error trials. People with the largest post-error memory decrements had the best memory for error trials (adjusted for overall hit rate), r(57) = 0.34, p = .008, q = 0.032, Fig. 4C. These findings suggest that individuals who attended more to images that they mis-classified devoted less resources to processing the next trial. In contrast to our findings in experiment 1, people who slowed more after errors did not experience the greatest memory decrements, b = 0.00001, SE = 0.0005, t (57) = 0.02, p = .984, r = 0.003.

4. Discussion

Here, we uncover how making an error influences memory formation – a previously unexplored phenomenon of key theoretical importance because of the divergent consequences ascribed to errors from adaptive (Botvinick et al., 2001; Maier et al., 2011; Yeung et al., 2004) and non-adaptive theories of error processing (Notebaert et al., 2009), as well as emotional arousal perspectives (Inzlicht et al., 2015; Saunders et al., 2017). In contrast to the idea that errors enhance task performance and prepare people for learning opportunities, we consistently observed that errors impaired subsequent memory formation. Critically, memory formation was poorer immediately after errors relative to before errors, suggesting that post-error memory decrements were not simply due to pre-existing attentional failures that lingered into post error trials.

In support of our interpretation that error processing drives memory decrements, in experiment 2, we observed that people with the worst post-error memory decrements had the largest pupil dilation on error trials and spent less time fixating after errors relative to before errors. Because increased pupil size is a marker of phasic NE release (Clewett et al., 2018; Joshi et al., 2016; Murphy et al., 2014; Varazzani, San-Galli, Gilardeau, & Bouret, 2015) and arousal (Joshi et al., 2016; Reimer et al., 2014, 2016; Vinck et al., 2015), this finding raises the

possibility that heightened arousal interfered with memory formation. Notably, because those with the worst memory decrements also visually engaged with stimuli less after errors, it is also possible that transient decreases in task-engagement contributed to post-error memory decrements.

The finding that increases in pupil size after errors correlated with post-error memory decrements aligns with the predictions of the emotional arousal framework. This framework postulates that norepinephrine-mediated arousal impairs memory for neutral events that occur after an arousing event (Clewett et al., 2017; Clewett et al., 2018; Hurlemann et al., 2005; Mather & Sutherland, 2011). Indeed, arousal has been argued to influence the selection of which information is later remembered (Clewett et al., 2017; Clewett et al., 2018; Mather et al., 2016; Mather & Sutherland, 2011), such that arousing stimuli are prioritized for processing over neutral content experienced close in time (Aston-Jones & Cohen, 2005; Mather et al., 2016; Mather & Sutherland, 2011). Adding to this theorized trade-off in memory, we observed a positive correlation between post-error memory decrements and memory for error trials as well as pupil dilation after errors. This result raises the possibility that greater NE release in response to errors may have enhanced memory for information processed on error trials, while impairing memory for subsequent information. Crucially, however, because we did not directly manipulate arousal, it is unclear whether arousal caused or simply correlated with post-error memory decrements. Indeed, it is possible that another individual difference that covaried with error-evoked arousal - for instance, the tendency to orient away from the task - drove the effects that we observed. It is our hope that our correlative findings inspire future work that directly manipulates arousal after errors to more directly assess if arousal causes post-error memory decrements. Furthermore, future work should examine how error awareness influences post-error memory decrements. Identifying a physiological marker of error-driven memory decrements, for example, could help identify people who need more time to recover after errors, and times when memory is most vulnerable to errors. This information is particularly important in pedagogical contexts, such as in classrooms, where errors are frequent and forming new memories is so crucial.

Though those with superior error trial memory were most susceptible to post-error memory decrements, on average, people did not have superior memory for images presented on error trials, and in some cases showed especially poor memory for these trials. This pattern contrasts with studies showing that arousal enhances concurrent memory formation (Cahill, Babinsky, Markowitsch, & McGaugh, 1995; Cahill & McGaugh, 1995, 1998; Kensinger, 2007; Kensinger & Corkin, 2003; McGaugh, 2004). Importantly, however, the error-driven arousal that we observed likely peaked *after* people made errors – and thus may



Fig. 6. A) Pupil size across time during the simon classification task for error trials and trials in close proximity to an error. Grey shaded box shows time window in which the stimulus is on the screen. Dashed box denotes the time window of interest that was used to compute mean pupil size for each trial that was used in subsequent analyses. B) Average pupil size on the Simon classification task on error trials and trials in close proximity to an error. Pupil size reflects the average pupil response in the time window 1–2 s after stimulus onset (panel A). Participants' pupil size was larger on error trials than before errors and relative to other trials which participants responded to correctly (not shown; all ps < .001). Participants' pupil size was also larger up to 2 trials after an error relative to before an error and relative to other trials on which participants responded correctly, all ps < .001. Horizontal grey line reflects mean pupil size before errors, and shaded grey area shows confidence interval. C) Participants who had the largest pupil size on error trials (relative to before errors) experienced the largest post-error memory decrements, p = .017, q = 0.04.

have only briefly coincided with the presentation of an image. The timing of the arousal response after errors may explain why we did not observe enhanced memory formation on error trials. It may also explain why we observed *impaired* memory for omission errors across the two experiments, and marginally *enhanced* memory for commission errors only in study 1; in particular, images in experiment 1 remained on

screen for an average of 1 s after commission errors; this timing may have allowed error-related increases in arousal to benefit memory formation. On the other hand, arousal after omission errors was likely only engaged after stimulus offset and thus, may not have overcome the negative effects of the attentional lapses that drove the error (deBettencourt et al., 2018). Another possible explanation for why we did not observe enhanced memory on error trials is that errors, and subsequent increases in arousal, are somewhat divorced from the neutral stimulus that elicited the error. Thus, the association between the neutral stimulus and the error may not have saturated the stimulus with enough salience for arousal to prioritize its representation in memory (Clewett et al., 2017: Clewett et al., 2018: Hurlemann et al., 2005: Mather & Sutherland, 2011). This may be particularly true if the adaptive target of attention after errors is the internal decision process that led to the error. Indeed, if people orient inwards after making an error, it is unlikely that increases in arousal would enhance memory for images presented on error trials. While orienting inwards may impair memory in the moment, it could be potentially adaptive over the long term, allowing one to better manage cognitive control in future contexts. Another factor that may explain why error memory was not enhanced is that prior research suggests that arousing stimuli are better remembered because they are perceptually and conceptually more distinct (Sommer, Gläscher, Moritz, & Büchel, 2008; Talmi, 2013) - a factor that the stimuli in our study would not benefit from. Thus, multiple factors could explain why we did not observe enhanced memory for images presented on error trials.

Adaptive cognitive control theories propose that detecting an error (and response conflict more broadly) is precisely the cue needed to overcome lapses in task engagement and energize goal-relevant processing (Botvinick et al., 2001; Maier et al., 2011; Marco-Pallarés et al., 2008; Yeung et al., 2004). Consistent with this theory, we see robust response slowing after errors - a phenomenon often interpreted as adaptive upregulation of controlled processing and response caution (Danielmeier & Ullsperger, 2011; Verguts, Notebaert, Kunde, & Wühr, 2011). However, despite what on the surface may seem like greater controlled processing and increased response caution, participants were less likely to later remember these stimuli. This performance cost is consistent with the non-functional orienting account of error processing, which predicts that attentional capture by errors divert cognitive resources (Jentzsch & Dudschig, 2009; Notebaert et al., 2009). Indeed, these post-error memory decrements may have occurred because attention was oriented inwards during these periods. Experiment 2 provided further support for the non-functional account by showing that errors also impaired performance on the task at hand, and that participants whose memory was more impaired by errors visually engaged with stimuli less following an error.

Our results help to reconcile adaptive (Botvinick et al., 2001; Yeung et al., 2004) and non-functional (Notebaert et al., 2009) accounts of error-regulated control. Specifically, we show that errors lead to impaired accuracy and memory formation before returning to pre-error levels. In experiment 2, we found that people formed better memories 3 trials after an error relative to before an error, demonstrating that memory recovered to baseline levels. Though we did not observe this effect in experiment 1, the findings from experiment 2 are broadly consistent with recent cognitive accounts arguing that errors enhance task performance relative to pre-error levels, but only after an initial dip (Jentzsch & Dudschig, 2009; Murphy et al., 2016). From this perspective, experiments showing post-error performance decrements (Carp & Compton, 2009; Jentzsch & Dudschig, 2009; Notebaert et al., 2009; Núñez Castellar, Kühn, Fias, & Notebaert, 2010; Verguts et al., 2011) and enhancements (Danielmeier et al., 2011; Maier et al., 2011; Marco-Pallarés et al., 2008) may simply reflect different points in the error resolution process. Based on these findings, errors could impair or enhance memory depending on when participants are asked to process information after errors. Studies with a long interval between trials may observe enhanced memory formation after errors, whereas studies with



Fig. 7. A) Time fixating on stimuli on error trials and trials in close proximity to an error during the Simon categorization task. There were no differences in the amount of time that participants fixated on stimuli before and after errors, and no differences relative to other correct trials, ps > .05. B) Participants who displayed the largest post-error memory decrements also spent less time fixating on stimuli after categorization errors relative before errors, $r_s(57) = 0.26$, p = .04, q = 0.053.

a short interval may observe memory impairments, as we observed here. Notably, in both of our experiments, the ITI was 1.5 s, which is longer than what is thought to be required for error resolution (Buzzell et al., 2017; Jentzsch & Dudschig, 2009). Our tasks, however, involved processing a new stimulus on every trial, and perhaps this additional task demand extended the period required for resolution. Determining the factors that influence the time course over which error resolution influences performance will be needed to reconcile and predict the divergent effects that errors hold over cognition.

The different findings in experiment 1 and 2 can be used to identify the factors that underlie impairments in memory after errors. In experiment 1, participants were 15% less likely to recognize images presented after errors as compared to those that preceded errors. This difference was only 4% in experiment 2. Thus, task differences led to more severe error resolution costs in experiment 1. Though future work is required to assess these factors, we speculatively identify two likely candidates: inter-stimulus intervals and error frequency. In experiment 1, stimuli were presented throughout the full 1.5 s ITI, without intervening breaks. By contrast, stimuli (and response windows) were only presented for 1 s, with 0.5 s breaks separating each in experiment 2. Thus, there was equal time separating responses in both, however dividing this time into trials and breaks in experiment 2 may have provided dedicated periods for error processing, and preparation for the upcoming trials, hastening error resolution. Alternatively, the longer stimulus presentation in experiment 1 could have provided more opportunity to detect errors and increased their salience. Relatedly, the lower rate of errors in experiment 1 (2% vs. 11% in experiment 2) could have increased their salience; a participant's first few mistakes likely elicit greater arousal as compared to their 20th. Perhaps it is only after people become somewhat accustomed to their fallibility that they can more dispassionately use their mistakes to improve their cognition.

Although we used pupil size to index arousal, we acknowledge that pupil size is an indirect measure of arousal levels. Pupil size is modulated by a variety of factors, including affective responses that accompany arousal (Oliva & Anikin, 2018), cognitive factors like deep memory formation (Kucewicz et al., 2018), cognitive effort (Johansson, Pärnamets, Bjernestedt, & Johansson, 2018) and even time pressure (Gross & Dobbins, 2020). Of note though, in the present study, pupil size was negatively related to subsequent memory formation, suggesting that larger pupils were not driven by factors that would improve task-relevant processing. The negative relationship between pupil size and cognition, combined research showing that errors trigger arousal (Elkins-Brown, Saunders, He, & Inzlicht, 2017; Elkins-Brown, Saunders, & Inzlicht, 2016; Spunt, Lieberman, Cohen, & Eisenberger, 2012), support our interpretation that error-driven increases in dilation reflect arousal. Future research, however, is needed to determine the degree to which error-driven memory modulation depends on arousal. Interestingly, some have argued that errors trigger negative affect which increases arousal, and in turn, facilitates the cognitive control required for resolving errors (Inzlicht et al., 2015; Saunders et al., 2017). We did not measure emotions or ask participants whether they experienced particular feelings after errors as this would have interrupted on-going task performance. Moreover, based on pupil data alone, we could not infer that errors induced participants to experience particular feelings, as various measures of arousal do not necessarily correlate (Barrett & Satpute, 2019; Satpute, Kragel, Barrett, Wager, & Bianciardi, 2019). Therefore, we hope that our findings motivate future research examining how negative affect and cognitive control evolve simultaneously or complement each other after errors, and in turn, affect posterror learning.

To understand how errors shape memory formation, we drew from research on error processing and arousal. We showed that errors transiently impaired subsequent memory formation, and that individual differences in post-error memory decrements correlated with indices of arousal and task-engagement. Moreover, in experiment 2, we observed that although errors led to transient decreases in memory formation, after a longer delay, memory formation was better than pre-error levels. These findings reveal why we sometimes fail to encode new information across time – an important question not only in the lab, but also pedagogical contexts like the classroom. Crucially, our findings also inform theories of error processing, which could be expanded to incorporate how both error detection and resolution affect memory formation across time – a fundamental pre-requisite to adaptive learning.

Acknowledgements

We would like to thank Carissa deMarinis, Inayat Singh, Sabrina Szkolmowski, and Enri Boshti for their help with data collection and Zahra Abolghasem for her help with stimulus development. We are also grateful to Anuya Patil, Johnny Dubois, and Hause Lin for fruitful discussions regarding the study design and analysis. This research was funded by Social Sciences and Humanities Research Council of Canada Insight Program (435171493), Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2016-05), and Brain Canada.

Declaration of competing interest

The authors have no competing interests.

Appendix A. Supplemental Information

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104338.

References

- Aarts, K., De Houwer, J., & Pourtois, G. (2013). Erroneous and correct actions have a different affective valence: Evidence from ERPs. *Emotion (Washington, D.C.)*, 13(5), 960–973. https://doi.org/10.1037/a0032808.
- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604. 135709.
- Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, 693, 9–18. https://doi.org/10.1016/j.neulet.2017.07.045.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss. v067.i01.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.
- Buzzell, G. A., Beatty, P. J., Paquette, N. A., Roberts, D. M., & McDonald, C. G. (2017). Error-induced blindness: Error detection leads to impaired sensory processing and lower accuracy at short response-stimulus intervals. *Journal of Neuroscience*, 37(11), 2895–2903. https://doi.org/10.1523/JNEUROSCI.1202-16.2017.
- Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, 377(6547), 295–296. https://doi.org/10.1038/ 377295a0.
- Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, 4(4), 410–421. https:// doi.org/10.1006/ccog.1995.1048.
- Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, 21(7), 294–299. https://doi.org/10.1016/ S0166-2236(97)01214-9.
- Carp, J., & Compton, R. J. (2009). Alpha power is influenced by performance errors. Psychophysiology, 46(2), 336–343. https://doi.org/10.1111/j.1469-8986.2008. 00773.x.
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences, 18(8), 414–421. https://doi.org/10.1016/j.tics.2014.04. 012.
- Clewett, D., Sakaki, M., Nielsen, S., Petzinger, G., & Mather, M. (2017). Noradrenergic mechanisms of arousal's bidirectional effects on episodic memory. *Neurobiology of Learning and Memory*, 137, 1–14. https://doi.org/10.1016/j.nlm.2016.10.017.
- Clewett, D. V., Huang, R., Velasco, R., Lee, T.-H., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 38*(6), 1558–1574. https://doi.org/ 10.1523/JNEUROSCI.2097-17.2017.
- Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic response. *NeuroImage*, 27(4), 885–895. https://doi.org/10.1016/j.neuroimage.2005.05.047.
- Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience: The Official Journal of the Society* for Neuroscience, 31(5), 1780–1789. https://doi.org/10.1523/JNEUROSCI.4299-10. 2011.
- Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. Frontiers in Psychology, 2. https://doi.org/10.3389/fpsyg.2011.00233.
- deBettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2018). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review*, 25(2), 605–611. https://doi. org/10.3758/s13423-017-1309-5.
- Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion (Washington, D.C.)*, 1(1), 5–11.
- Elkins-Brown, N., Saunders, B., He, F., & Inzlicht, M. (2017). Stability and reliability of error-related electromyography over the corrugator supercilii with increasing trials. *Psychophysiology*, 54(10), 1559–1573. https://doi.org/10.1111/psyp.12902.
- Elkins-Brown, N., Saunders, B., & Inzlicht, M. (2016). Error-related electromyographic activity over the corrugator supercilii is associated with neural performance monitoring. *Psychophysiology*, 53(2), 159–170. https://doi.org/10.1111/psyp.12556.
- Gross, M. P., & Dobbins, I. G. (2020). Pupil dilation during memory encoding reflects time pressure rather than depth of processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000818.
- Hajcak, G., Mcdonald, N. J., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, 40(6), 895–903. https://doi.org/10.1111/1469-8986.00107.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679.

- Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, 68(1), 1–17. https://doi. org/10.1007/s00426-003-0132-y.
- Hoshikawa, Y., & Yamamoto, Y. (1997). Effects of Stroop color-word conflict test on the autonomic nervous system responses. *American Journal of Physiology - Heart and Circulatory Physiology, 272*(3), H1113–H1121. https://doi.org/10.1152/ajpheart. 1997.272.3.H1113.
- Houtman, F., & Notebaert, W. (2013). Blinded by an error. Cognition, 128(2), 228–236. https://doi.org/10.1016/j.cognition.2013.04.003.
- Hurlemann, R., Hawellek, B., Matusch, A., Kolsch, H., Wollersen, H., Madea, B., ... Dolan, R. J. (2005). Noradrenergic modulation of emotion-induced forgetting and remembering. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(27), 6343–6349. https://doi.org/10.1523/JNEUROSCI.0228-05. 2005.
- Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. Trends in Cognitive Sciences, 19(3), 126–132. https://doi.org/10.1016/j.tics. 2015.01.004.
- Jentzsch, I., & Dudschig, C. (2009). Short article: Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, 62(2), 209–218. https://doi.org/10.1080/ 17470210802240655.
- Johansson, R., Pärnamets, P., Bjernestedt, A., & Johansson, M. (2018). Pupil dilation tracks the dynamics of mnemonic interference resolution. *Scientific Reports*, 8(1), 4826. https://doi.org/10.1038/s41598-018-23297-3.
- Joshi, S., Li, Y., Kalwani, R., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028.
- Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, 16(4), 213–218. https://doi.org/10.1111/j.1467-8721.2007.00506.x.
- Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31(8), 1169–1180. https://doi.org/10.3758/BF03195800.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303(5660), 1023–1026. https://doi.org/10.1126/science.1089910.
- Krebs, R. M., Boehler, C. N., De Belder, M., & Egner, T. (2015). Neural conflict-control mechanisms improve memory for target stimuli. *Cerebral Cortex (New York, N.Y.*: 1991), 25(3), 833–843. https://doi.org/10.1093/cercor/bht283.
- Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., ... Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*. https://doi.org/10.1038/s41598-018-23197-6.
- Lin, H., Saunders, B., Hutcherson, C. A., & Inzlicht, M. (2018). Midfrontal theta and pupil dilation parametrically track subjective conflict (but also surprise) during intertemporal choice. *NeuroImage*, 172, 838–852. https://doi.org/10.1016/j.neuroimage. 2017.10.055.
- Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". Law and Human Behavior, 11(1), 55–62. https://doi.org/10.1007/BF01044839.
- Maier, M. E., Ernst, B., & Steinhauser, M. (2019). Error-related pupil dilation is sensitive to the evaluation of different error types. *Biological Psychology*, 141, 25–34. https:// doi.org/10.1016/j.biopsycho.2018.12.013.
- Maier, M. E., Yeung, N., & Steinhauser, M. (2011). Error-related brain activity and adjustments of selective attention following errors. *NeuroImage*, 56(4), 2339–2347. https://doi.org/10.1016/j.neuroimage.2011.03.083.
- Marco-Pallarés, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience*, 20(9), 1595–1610. https://doi.org/10.1162/jocn.2008.20117.
- Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *The Behavioral and Brain Sciences*, 39, Article e200. https://doi.org/10. 1017/S0140525X15000667.
- Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 6(2), 114–133. https://doi.org/10.1177/1745691611400234.
- McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, 27(1), 1–28. https://doi.org/10.1146/annurev.neuro.27.070203.144157.
- McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, 87(1), 179–192. https://doi.org/10.1016/j.neuron.2015.05.038.
- Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154. https://doi.org/10.1002/hbm.22466.
- Murphy, P. R., van Moort, M. L., & Nieuwenhuis, S. (2016). The pupillary orienting response predicts adaptive behavioral adjustment after errors. *PLoS One*, 11(3), Article e0151763. https://doi.org/10.1371/journal.pone.0151763.
- Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal* of Vision, 13(2), 11. https://doi.org/10.1167/13.2.11.
- Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Posterror slowing: An orienting account. *Cognition*, 111(2), 275–279. https://doi.org/10. 1016/j.cognition.2009.02.002.
- Núñez Castellar, E., Kühn, S., Fias, W., & Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: ERP support. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 270–278. https://doi.org/10.3758/CABN.10.2.270.
- Oliva, M. J. R., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion

A. Decker, et al.

recognition in human vocalizations. *Scientific Reports*. https://doi.org/10.1038/s41598-018-23265-x.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods, 162(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11. 017.

- Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, 29(4), 727–743. https://doi.org/10.1080/14640747708400645.
- Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., & Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, 84(2), 355–362. https://doi.org/10.1016/j.neuron.2014.09.033.
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7. https://doi.org/10.1038/ ncomms13289.
- Sakaki, M., Fryer, K., & Mather, M. (2014). Emotion strengthens high-priority memory traces but weakens low-priority memory traces. *Psychological Science*, 25(2), 387–395. https://doi.org/10.1177/0956797613504784.
- Satpute, A. B., Kragel, P. A., Barrett, L. F., Wager, T. D., & Bianciardi, M. (2019). Deconstructing arousal into wakeful, autonomic and affective varieties. *Neuroscience Letters*, 693, 19–28. https://doi.org/10.1016/j.neulet.2018.01.042.
- Saunders, B., Lin, H., Milyavskaya, M., & Inzlicht, M. (2017). The emotive nature of conflict monitoring in the medial prefrontal cortex. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 119, 31–40. https://doi.org/10.1016/j.ijpsycho.2017.01.004.
- Sommer, T., Gläscher, J., Moritz, S., & Büchel, C. (2008). Emotional enhancement effect of memory: Removing the influence of cognitive factors. *Learning & Memory*, 15(8), 569–573. https://doi.org/10.1101/lm.995108.
- Spunt, R. P., Lieberman, M. D., Cohen, J. R., & Eisenberger, N. I. (2012). The

phenomenology of error processing: The dorsal ACC response to stop-signal errors tracks reports of negative affect. *Journal of Cognitive Neuroscience*, 24(8), 1753–1765. https://doi.org/10.1162/jocn_a_00242.

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. Current Directions in Psychological Science, 22(6), 430–436. https://doi.org/10.1177/ 0963721413498893.

Team, R. C. (2013). R: A language and environment for statistical computing.

- van Steenbergen, H., & Band, G. P. H. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience*, 7, 215. https://doi.org/10. 3389/fnhum.2013.00215.
- Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *Journal of Neuroscience*, 35(20), 7866–7877. https:// doi.org/10.1523/JNEUROSCI.0454-15.2015.
- Verguts, T., Notebaert, W., Kunde, W., & Wühr, P. (2011). Post-conflict slowing: Cognitive adaptation after conflict processing. *Psychonomic Bulletin & Review*, 18(1), 76–82. https://doi.org/10.3758/s13423-010-0016-2.
- Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, J. A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 86(3), 740–754. https://doi.org/10.1016/j.neuron.2015.03.028.
- Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, 49(7), 1831–1842. https://doi.org/10. 1016/j.neuropsychologia.2011.03.007.
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, 42(3), 671–684. https://doi.org/10.3758/BRM.42.3.671.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931.