Psychophysiology

WILEY

PSYCHOPHYSIOLOGY SPR

ORIGINAL ARTICLE OPEN ACCESS

The Interplay Between Tonic and Phasic Pupil Activity and Cognitive Flexibility and Stability

Anna Mini Jos[1] | Andrew Westbrook[2] | Sophia LoParco[3] | A. Ross Otto[1]

[1]Department of Psychology, McGill University, Montréal, Quebec, Canada | [2]Center for Advanced Human Brain Imaging Research, Rutgers University, Piscataway, New Jersey, USA | [3]Department of Bioengineering, McGill University, Montréal, Quebec, Canada

Correspondence: Anna Mini Jos (anna.jos@mail.mcgill.ca)

Received: 6 September 2024 | **Revised:** 12 July 2025 | **Accepted:** 15 July 2025

Funding: This research was supported by an NSERC Discovery Grant, a New Researchers Startup Grant from the Fonds de Recherche du Québec—Nature et Technologies, and an infrastructure award from the Canada Foundation for Innovation (awarded to ARO) and a Fonds de Recherche du Québec—Nature et Technologies Doctoral Fellowship (awarded to AMJ).

Keywords: adaptive gain theory | cognitive control | cognitive effort | cognitive flexibility | cognitive stability | distractor inhibition | phasic pupil | stability flexibility tradeoff | task switching | tonic pupil

ABSTRACT

Previous research has shown that while larger phasic pupillary activity indexes lower switch costs and better performance on a Stroop task, greater tonic pupillary activity indexes greater exploration. However, the direct influence of tonic–phasic pupillary activity on cognitive flexibility and cognitive stability—two control modes that potentially trade off with each other—has not been systematically investigated. We examine these associations using a task that imposes varying requirements on flexibility (task switching) and stability (distractor inhibition). The task included ambiguous trials that captured participants' preference for cognitively flexible performance. Participants ($n = 51$) completed the task with pupillary measurement recording. We find a lower preference to voluntarily switch (lower flexibility preference) in individuals with higher switch costs (lower ability/effort exerted to be flexible) and in individuals with faster RTs on Distractor Inhibition trials (higher stability), indicating a possible trade off between an individual's cognitively stable performance and the preference to be flexible. Examining pupillary data, we show that a larger phasic pupillary response in Task Switch trials is associated with lower switch costs, that is, higher flexibility. Individuals with larger average tonic pupil diameter were less likely to voluntarily switch tasks in ambiguous trials (i.e., lower flexibility preference), contrary to our expectations. Finally, we observed that higher tonic pupillary measures predicted quicker errors on trials measuring cognitive stability and greater overall task disengagement. Taken together, our findings shed light on the differential relationships between phasic pupillary activity and tonic pupil diameter and stable versus flexible modes of cognitive control.

1 | Introduction

An individual's performance on a cognitive task is related not only to acquired skills, but also to adaptability to changes in task demand. This adaptability of behavior favoring the processing of task-relevant information over other sources of competing information is referred to as cognitive control (Cohen and Braver 1996). Cognitive control has been broadly categorized into stable or flexible performance modes (Fröber and Dreisbach 2017; Papadopetraki et al. 2019; Serrien and O'Regan 2019). Cognitive stability refers to the ability to focus on the task at hand, while inhibiting irrelevant environmental distractors—for example, focusing on one conversation at a busy social gathering, or inhibiting (often conflicting) task-relevant information such as word identity in the classic Stroop (1935) task or stimulus location in Flanker-like tasks (Eriksen and Eriksen 1974). In

these situations, task-irrelevant stimuli often conflict with task-relevant stimuli, eliciting incompatible responses (Egner and Hirsch 2005) which necessitate stable maintenance of goals in the face of distracting information. The flexible mode of control, in contrast, entails readily shifting between tasks (or task rules), in order to adapt to relevant environmental changes. In everyday life, flexible control can help us to shift between task sets, for example to switch rules during a card game or code-switching in a bilingual environment. Experimentally, the flexible control mode is most classically operationalized in task switching paradigms, which require rapid shifting between different task rules (Dreisbach and Goschke 2004; Goschke and Bolte 2014; Ionescu 2012; Papadopetraki et al. 2019). The varying demands of controlled behavior are thought to impose an antagonistic relationship between flexible and stable processing modes: at times, one needs to maintain current goals in the face of distracting stimuli, but at other times, one may need to flexibly switch between goals or task sets. According to this view, these two processing modes effectively trade off, with a greater processing flexibility (e.g., being able to flexibly change tasks) gained at the cost of lower stability, and vice versa (Dreisbach and Goschke 2004; Dreisbach et al. 2024).

A considerable body of work in neuroscience has been spurred by Adaptive Gain Theory (Aston-Jones and Cohen 2005), which posits a link between locus coeruleus-norepinephrine (LC-NE) activity, pupil diameter, and stable/flexible performance modes of cognitive processing. Although LC-NE firing rates contain a spectrum of frequencies, its two characteristic modes of activity are often described as phasic (10–15 Hz) and tonic firing modes (1–6 Hz). While the phasic mode is thought to promote stable, focused performance of the task at hand, the tonic mode allows for the influence of other environmental stimuli, thus promoting flexibility (Aston-Jones et al. 1999). According to this view, the LC-NE system adjusts to changes in current task utility by modulating the gain of cortical processing mechanisms underpinning exploitative (or stable) versus exploratory (or flexible) task performance (Fröber et al. 2018). In support of this hypothesis, human studies have found that baseline pupil diameter—a proxy for LC-NE tonic activity—tracked task disengagement as task difficulty increased and, concomitantly, the expected utility of task engagement decreased (Gilzenrat et al. 2010). Supporting the link between LC activity and pupil diameter, non-human primate studies have also observed a strong concordance between LC tonic/phasic firing modes and changes in pupil diameter (Gilzenrat et al. 2010; Joshi et al. 2016; Murphy et al. 2014; Rajkowski et al. 1993).

Intriguingly, a large body of work has observed that larger phasic pupillary responses occur during more effortful tasks (or trial types within a task) in comparison to less effortful tasks/trials (Aminihajibashi et al. 2020; da Silva Castanheira et al. 2021; Rondeel et al. 2015; Van Der Wel and Van Steenbergen 2018), both in situations that favor stable control as well as flexible control modes. For instance, task switching paradigms, which favor flexible control, engender larger phasic pupillary responses on task switch trials in comparison to task repetitions (da Silva Castanheira et al. 2021; Katidioti et al. 2014; Rondeel et al. 2015). A study that allowed participants to voluntarily switch tasks on certain trials, an indication of a more cognitively flexible state that increases the propensity to switch, also revealed larger phasic pupillary dilation when participants decided to switch between tasks (Katidioti et al. 2014). At the same time, in the classic Stroop interference task—for which a stable processing mode is thought to reduce response interference (Goschke and Bolte 2014)—incongruent trials evoked larger phasic pupillary dilation in comparison to congruent trials (Hershman and Henik 2019; Laeng et al. 2011; Rondeel et al. 2015).

Another line of studies, informed by Adaptive Gain Theory, also suggests links between tonic versus phasic pupillary activity and behavioral expression of stable versus flexible control modes. For example, in an auditory oddball task, Gilzenrat et al. (2010) observed that larger tonic pupil diameter was associated with poorer performance. Using the same task, Murphy et al. (2011) revealed an inverted-U relationship between tonic pupil and response time (RT) variability (taking low RT variability as an index of stable cognitive control). This finding aligns with the inverted-U relationship between tonic pupil diameter and task performance predicted by Adaptive Gain Theory—similar to the Yerkes and Dodson (1908) performance-arousal curve—which posits that performance is optimal at intermediate levels of tonic firing and declines on shifting to either extreme (low or high levels) of tonic activity (Aston-Jones et al. 1999). Relatedly, in an "n-armed bandit" reinforcement learning task, larger tonic pupil diameter was associated with greater rates of exploratory choice, interpreted as an indication of the operation of a flexible control state (Jepma and Nieuwenhuis 2011). Similarly, Hayes and Petrov (2016) observed sustained increases in pupil diameter during periods of (presumably flexible) exploration in a Raven's Progressive Matrices task.

Based on these lines of empirical and theoretical work, the extent to which phasic pupillary activity reflects the operation of stable versus flexible processing modes remains unclear. Across studies, larger phasic pupillary dilation has been associated with better performance in tasks favoring flexible control (da Silva Castanheira et al. 2021; Katidioti et al. 2014; Rondeel et al. 2015). At the same time, periods of larger *tonic* pupil diameter are also associated with higher rates of reward-related exploration (Jepma and Nieuwenhuis 2011; Pajkossy et al. 2017) and task disengagement (Gilzenrat et al. 2010). Taken together, these studies suggest that both phasic and tonic pupillary activity might index the operation of a flexible mode of cognitive control. However, according to Adaptive Gain Theory, phasic pupillary activity should reflect task-relevant processing, and the strength of these phasic responses should be greater during periods of higher task performance (Gilzenrat et al. 2010).

An unresolved question concerns whether tonic and phasic pupillary activity map onto (or relate to) distinct control states—that is, whether phasic versus tonic pupillary activity might distinctly index the operation of more stable versus flexible control modes. This ambiguity stems, in part, from the use of behavioral paradigms that singularly favor one putative control mode over another (e.g., task-switching). Accordingly, in the present study, we sought to examine pupil dynamics in a task with varying trial-by-trial demands for stability and flexibility, which permits us to leverage both between- and within-individual differences to examine the extent to which phasic pupillary dilation and tonic pupil diameter are related to the operation of flexible versus stable control modes.

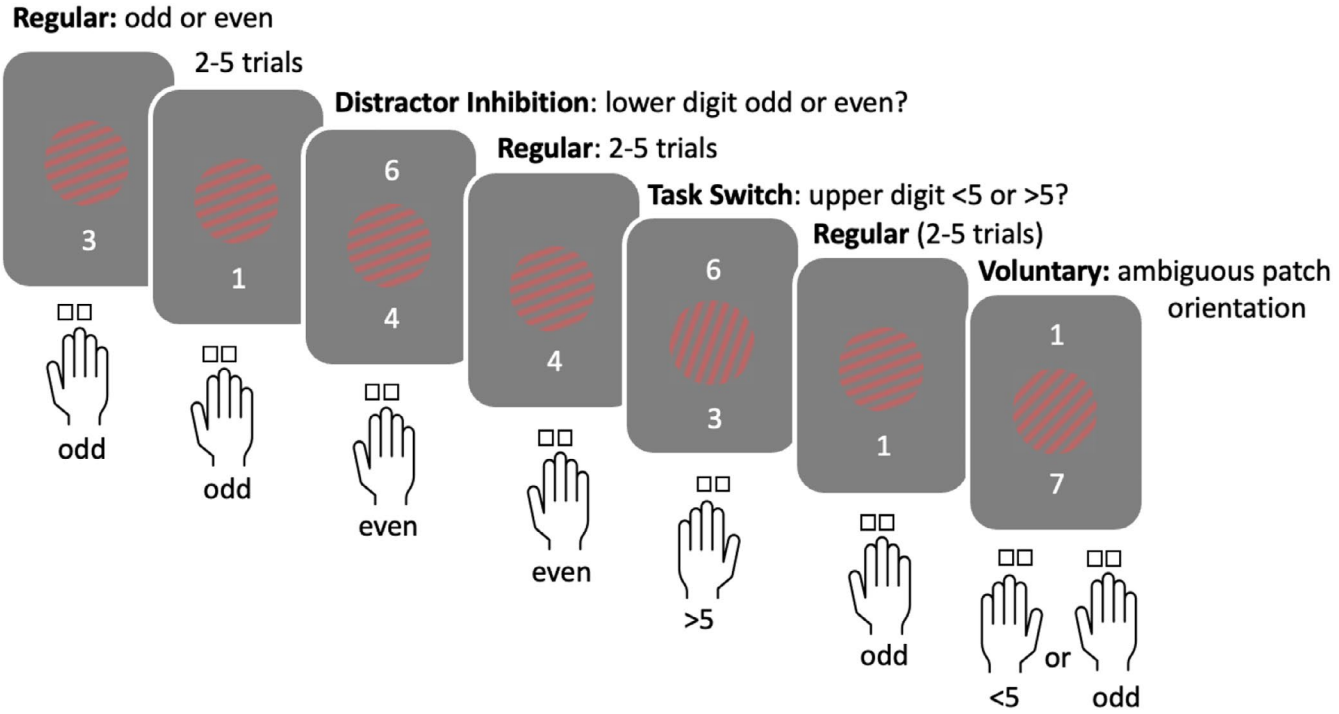

FIGURE 1 | Schematic illustration of the voluntary switch task where participants performed the parity task with no distractors (regular trials) in 70% of the 210 trials. In the remaining trials subjects performed the parity task with distractors (Distractor Inhibition trials), switched to magnitude task (forced Task Switch trials) or chose to perform either task (voluntary trials, cued using an ambiguous stimulus). Task window durations were set as maximum response window of 2.6 s and an inter trial level of 1–3 s during which a Gabor patch was displayed on screen.

To do this, we adapted a previously used cognitive control task for measurement of pupillary activity (Armbruster et al. 2012; Armbruster-Genç et al. 2016). In this paradigm (Figure 1), participants completed four types of trials, for which performance is differentially informative with respect to the use of stable versus flexible control modes: "Regular" and "Distractor Inhibition" trials, which require singular focus on the main subtask, favoring a stable control, as well as task-switching trials that require participants to switch between subtasks, favoring flexible control. These task switches occurred either in a forced ("Task Switch") or spontaneous ("Voluntary Task Switch") fashion—which index an individual's dispositional (or default) tendency to employ the flexible control mode (Fröber and Dreisbach 2017). On Regular and Distractor Inhibition trials, participants perform a (frequently occurring) subtask—for example judging the parity of a number—with the added requirement of inhibiting an irrelevant (e.g., corresponding to a magnitude subtask) on Distractor Inhibition trials. On "Task Switch" or "Voluntary Task Switch" trials participants switch to and perform (a less frequent) subtask (e.g., magnitude judgments), with the former trial type forcing switching between subtasks and the latter affording a choice to voluntary switch between subtasks. On the whole, the paradigm affords characterization of individuals' use of stable processing as a function of performance on Regular and Distractor Inhibition trials, and at the same time we could measure individuals' use of flexible processing by examining task-switching performance, as well as the possible interrelationships between these performance measures and voluntary task switch rates.

Furthermore, we analyzed the relationships between phasic (i.e., task-evoked) and tonic (slow-varying, non-task evoked) pupillary activity and behavioral measures of expression of stable versus flexible processing across the trial types of interest. As previous works find that larger phasic pupillary dilations are typically observed during more demanding task circumstances (Van Der Wel and Van Steenbergen 2018), we predicted larger phasic pupillary dilations on Task Switch, Distractor Inhibition, and Voluntary Switch trials in comparison to Regular trials. We also examined the interrelationships between phasic pupillary dilations on Task Switch and Distractor Inhibition and (behavioral) performance on these respective trials to probe the link between phasic pupillary activity and control modes. Based on our recent findings (da Silva Castanheira et al. 2021), we predicted that individuals with larger phasic pupillary dilations on Task Switch trials would exhibit reduced task switch costs (defined as the difference between RTs on Task Switch and Regular trials), an established index of flexible control (Armbruster et al. 2012).

With respect to tonic pupil diameter, we expected to observe a positive relationship between an individual's average tonic pupil diameter and their overall rate of voluntary task-switching, on the basis of prior work suggesting an association between larger tonic pupil diameter and flexible behavior (Gilzenrat et al. 2010; Jepma and Nieuwenhuis 2011). In addition, beyond considering average tonic pupil diameter as an individual difference, we also examined relationships between momentary shifts in tonic pupil diameter within participants—filtering trial-level fluctuations in pupil size (< 0.05 Hz; Yellin et al. 2015) as well as the impact of recent errors (Gilzenrat et al. 2010; van den Brink et al. 2016)—and task performance across trial types, permitting us to probe possible associations between tonic pupillary activity and expression of stable versus flexible cognitive control modes. For example, previous work has found that a larger tonic (or baseline) pupil diameter predicts higher false alarm rates in

detection tasks (Gilzenrat et al. 2010; van den Brink et al. 2016) in humans, as well as increased susceptibility to distractor interference in non-human primates (Ebitz et al. 2014). Finally, on the basis of prior work observing relationships between tonic pupil diameter and performance variability (Murphy et al. 2011), we examined the association between fluctuations in tonic pupil diameter and variability in task performance.

2 | Methods

2.1 | Participants

Fifty-one healthy adults with normal or corrected-to-normal vision aged between 18 and 35 years (36 females, mean age = 23.0, SD = 3.7) from the McGill University community were recruited for the study. All participants provided informed consent in accordance with the McGill University Research Ethics Board. Individuals with color blindness, a history of a traumatic eye or head injury, learning disorders, psychiatric or psychological disorders, or sleep problems were excluded from the study. Participants were instructed to sleep adequately the night before and avoid caffeine and drugs on the day of each visit. They completed two visits scheduled at least 4 days apart, with most participants having a 1-week interval between sessions. For 13.7% of participants, sessions were scheduled < 4 days or > 7 days apart. Both visits were scheduled at the same time of day, for the majority of participants (with 29.4% participants scheduled at different times of the day), to reduce the influence of circadian rhythmicity of attention (Valdez et al. 2005) on task performance.

On day 1, participants performed the automated Operation Span task (OSPAN); a measure of working memory capacity (Unsworth et al. 2005) followed by the first session of the task-switching paradigm (Session 1). This was succeeded by a filler task with two conditions (not detailed in this paper) and a second session of the voluntary switch task (Session 2). On Day 2, participants performed tasks in the same order, with the exception of OSPAN (Session 3 and 4). Task sessions on both days began with a 1-min baseline pupil recording. Three participants did not complete their second visit. Two participants dropped out of the study as they found the eye-tracker or the seating position uncomfortable. Data from a final sample of 46 subjects were analyzed.

2.2 | Stability/Flexibility Task

Our main task was based on the voluntary task-switching paradigm used by Armbruster et al. (2012) and Armbruster-Genç et al. (2016), which we adapted for pupillary recording. We changed the task cue from the original design to reduce luminance-driven changes in pupillometric responses. On each trial, a digit between 1 and 9 (excluding 5) is shown above or below a central Gabor patch (see Figure 1). Participants were required to perform one of two subtasks on each trial—judge the magnitude or parity of the number. In the parity subtask, participants were required to indicate whether the digit on the screen is odd or even. In the magnitude subtask, participants were required to indicate whether the digit is less than or greater than 5. The subtask was indicated by the orientation (horizontal or vertical) of the Gabor patch. The vertical Gabor patch was at an angle of 75° while the horizontal Gabor patch was at an angle of 15°. For instance, for a sample participant, a horizontal Gabor patch would require them to focus on the digit above the patch and perform the parity subtask. The orientation of the Gabor patch (horizontal or vertical), the spatial mapping of tasks (parity/magnitude task above or below the patch), as well as the response mapping (using left/right hand for parity/magnitude task) were randomized and counterbalanced across participants.

Following Armbruster et al. (2012), participants used the index and middle fingers on either hand (keys 'D', 'F' or 'J', 'K') to respond to parity and magnitude subtasks with a maximum response window of 2.6 s. The intertrial interval varied randomly between 1 and 3 s during which a noisy Gabor patch (with both horizontal and vertical lines) was presented on screen.

The task involved four different trial types—*Regular, Task Switch, Distractor Inhibition*, and *Ambiguous* trials. On *Regular* trials, participants saw one digit on the screen, above or below the Gabor patch, and performed the parity subtask. On the other three trial types, two digits appeared on screen, one above and one below the central Gabor patch. On *Distractor Inhibition* trials, participants were required to perform the parity subtask (as in the Regular trials) but simultaneously ignore the second digit, which pertained to the magnitude subtask (as learned during preceding practice blocks). On *Task Switch* trials, participants were required to attend to the second digit and perform the magnitude subtask while ignoring the first digit (which pertains to the parity subtask). Finally, on *Ambiguous* trials, the orientation of the Gabor patch was displayed at an angle between the two learned orientations, creating a perceptually ambiguous task cue. On these trials, participants could choose to perform either the magnitude or parity subtask. Participants did not have prior practice on Ambiguous trials, and no instructions or mention of these trials were made before the main task.

The first session on each day (i.e., Session 1 and 3) began with four practice blocks.

The first and second practice blocks each consisted of 10 trials of the parity and magnitude subtasks, respectively. The third block consisted of 10 trials each of the parity and magnitude subtasks, randomly intermixed, thereby requiring task switching. During the first three blocks, for each trial, participants were shown one digit on the screen. The fourth block consisted of 30 trials of the parity and magnitude subtasks, with two digits shown on the screen during magnitude trials, thereby requiring task switching and inhibition. Participants were required to achieve an accuracy of 80% on each practice block (Regular and non-Regular trials) to progress to the main task.

The main task block consisted of 210 trials. 75% of trials were *Regular* trials. Every 2 to 5 Regular trials were followed by either a *Task Switch, Distractor Inhibition*, or *Ambiguous* trial. Each Task Switch, Distractor Inhibition, or Ambiguous trial was followed by a Regular trial. While response times (RTs) on cued (or forced) correct Task Switch trials were taken as an individual's ability to be flexible, the proportion of chosen task switches on Ambiguous trials—in which participants chose to complete the magnitude subtask instead of

the more frequent parity subtask—was taken as indicative of an individual's disposition (or preference) for flexible control (Armbruster et al. 2012). In addition, correct RTs and accuracy on Distractor Inhibition trials were indicative of the expression of a stable processing mode.

The Gabor patch and the digit colors were isoluminant with respect to the background to avoid luminance-driven pupillary change in pupil diameter. The task was presented using PsychoPy (Version 1.85.2), with the pupillary recording annotated with stimuli and response onsets and offsets. Each session of the main task was completed by participants in approximately 12 min.

2.3 | Pupillary Data Recording

All task sessions for all participants were conducted in the same experiment room with the ambient room lighting set at a constant low level between 100 and 103 cd/m^2 throughout data collection. The positions of the chin rest, testing computer, and eyetracker were kept constant across subjects. We used a 24-in. monitor set to a resolution of 1280 × 1024 pixels with brightness maintained at a luminance level of 22 cd/m^2 (measured at the distance of the eye-tracker chin rest). Luminance measurements were obtained using a ColorCAL MKII colorimeter (Cambridge Research Systems, Rochester, Kent, United Kingdom). Participants' right pupil diameter was measured throughout the experiment using an EyeLink 1000 eye tracker (SR Research, Osgoode, ON) at a sampling rate of 250 Hz.

2.4 | Behavioral Data Analysis

Data from the voluntary switch task from the four study sessions were included in all analyses, with the addition of nuisance variables to control for session number, the condition of the filler task, and counterbalanced task order. Ambiguous trials in which participants chose to repeat the parity subtask were considered *Voluntary Repeat* trials, while ambiguous trials in which participants chose to (switch and) perform the magnitude subtask were considered *Voluntary Switch* trials. Individual preference to voluntarily switch was measured as the percentage voluntary switches on ambiguous trials. RTs on correct trials were log-transformed to mitigate skew. Switch costs were defined as (the log of) the difference in median RTs between correct Task Switch trials and correct Regular repeat trials.

2.5 | Pupillary Data Analysis

Pupillary data were preprocessed in MATLAB (Version 2020a). First, eye blinks were detected and corrected using linear interpolation. As previous studies demonstrate that stimulus-evoked (i.e., phasic) pupillary responses are typically observable below 4 Hz (Hoeks and Levelt 1993; Nakayama and Shimizu 2004), we applied a low-pass (4 Hz) Butterworth filter to phasic pupillary responses (from stimulus onset of each trial to 200 ms before the stimulus onset of the next). Next, following previous work, we removed low-frequency drifts by applying a high-pass Butterworth filter of 0.2 Hz (Knapen et al. 2016). In line with prior research, pupil diameters were z-scored within each session to ensure comparability across sessions (de Gee et al. 2014; Nassar et al. 2012; Urai et al. 2017) and baseline corrected on a trial-by-trial basis by subtracting the average pupil diameter in the 200 ms before stimulus onset (da Silva Castanheira et al. 2021).

Phasic pupillary dilations were calculated as the average dilation in the window between 900 to 1900 ms after stimulus onset (Figure 3A,B show the time course of phasic pupillary response and average phasic dilation). The time window of interest was chosen to capture the order of peaks of phasic pupillary dilations across all trial types (Figure 3A), as well as time periods with the largest number of observations (which were dependent on participants' RTs). Across all trial types, 90%–93% of trials have complete pupillary time course data in this window (900–1900 ms).

Baseline tonic pupil was measured as the average diameter in the 200 ms before stimulus onset. A power spectrum analysis (following (Joshi et al. 2016; Nakayama and Shimizu 2004; Peysakhovich et al. 2015) using the PSD function in MATLAB (MathWorks Inc.)) of the 1-min pupillary recording taken prior to the task indicated that the highest recorded frequency of baseline tonic pupil was 3 Hz. Accordingly, we applied a low-pass Butterworth filter of 3 Hz to the pre-trial baseline tonic pupil to remove any spillover effects of phasic pupil dilation on tonic pupil. No high-pass filter was applied on baseline tonic pupil data. For within-participant trial-level analyses of tonic pupil, we *z*-scored pre-trial baseline tonic pupil within each session. For moving average analyses of tonic pupil, based on spectral power distribution of filtered tonic pupil diameter, we used a window of 10 trials. To analyze session-level tonic pupil, we *z*-scored, across sessions and participants, the session-level average baseline tonic pupil.

Trial-level tonic and phasic pupil data requiring more than 50% interpolation in the time window of interest were removed from further analysis, following previous work (McGowan et al. 2019). To avoid the influence of task novelty on pupil diameter, we also excluded the first ten trials of each session (which contained only Regular trials). To account for any remaining linear influence of time on task on pupillary variables of interest, trial number was included as a covariate in linear regressions. The last trial was excluded as the complete pupillary response was not captured in the eye tracker recording. Pupillary recordings from 7 sessions (across 5 participants) were excluded from analyses due to technical issues arising from eye-tracker calibration recordings or excessive head movement. Trials with response times less than or greater than 3 standard deviations away from the mean, as well as trials with no responses, were excluded from all analyses.

2.6 | Inferential Statistics

We estimated regressions using Bayesian multilevel models, using the brms package for the R programming language

(Bürkner 2017). For each model, we ran 4 chains (Bürkner 2018), each with 1000 warmup samples and a minimum 2000 iterations (for a few models more iterations were run to get reliable estimates, that is, R-hat less than 1.01 for each estimate). To analyze binary outcomes (e.g., trial-level voluntary switching) we used multilevel logistic regressions. We used weakly informative standard normal priors. Our regression models also included nuisance variables representing the trial number (to capture practice effects), the two conditions of the filler task, and whether the voluntary switch task session was preceded by the OSPAN measurement (which was binary coded). To account for the observed nonlinear influence of time on task on tonic pupil diameter, the square root of trial number was also included in tonic pupil analyses. Bayesian regression analysis provides estimates of model parameters along with measures of uncertainty based on posterior distributions, rather than point estimates. The coefficient estimates we report represent the mean of the posterior distribution and the standard error, which reflects the uncertainty in the posterior distribution. The 95% HPDI-L and HPDI-U (Highest Posterior Density Intervals; Lower and Upper Bounds) represent the range within which the true parameter value lies with 95% probability, given the data and model assumptions. HPDIs quantify uncertainty based on posterior probability.

3 | Results

3.1 | Task Performance

We first examined accuracies and response times (RTs) for correct responses across the five trial types (see Figure 2A,B and Table S1 for descriptive statistics). On Regular trials, participants indicated the parity of the single digit presented, while on Task Switch trials, with two digits on screen, participants were required to respond to the magnitude task. On Distractor Inhibition trials, participants responded to the parity task while ignoring a second digit, corresponding to the magnitude task. On Ambiguous trials, participants either responded to the parity task (yielding a *Voluntary Repeat* trial) or the magnitude task (yielding a *Voluntary Switch* trial).

Participants were slower to respond on Distractor Inhibition ($b = 0.42$, 95% HPD = [0.38, 0.46]), Task Switch ($b = 0.49$, 95% HPD = [0.45, 0.54]), Voluntary Repeat ($b = 0.46$, 95% HPD = [0.39, 0.52]) and Voluntary Switch trials ($b = 0.62$, 95% HPD = [0.58, 0.67]; Table S2) compared to Regular trials. On ambiguous trials, we also found that participants were slower to respond on Voluntary Switch versus Voluntary Repeat ($b = 0.18$, 95% HPD = [0.12, 0.24]), mirroring previous examinations (Arrington and Logan 2004). In comparison to Regular trials, participants were less accurate on Task Switch ($b = -0.97$, 95% HPD = [−1.50, −0.45]), Distractor Inhibition ($b = -1.00$, 95% HPD = [−1.30, −0.70]), and Voluntary Repeat trials ($b = -0.37$, 95% HPD = [−0.66, −0.06]; Table S3).

Examining voluntary switch rates on ambiguous trials, we found that, on average, participants switched on 50.64% of trials (SD = 38.45%), suggesting we had sufficient variability to examine voluntary switch rate as an individual difference. Participants with higher task switch costs—measured as the difference between (forced) Task Switch RTs and Regular Trials RTs—exhibited lower voluntary switch rates ($b = -1.10$, 95% HPD = [−1.15, −0.69]) (see Figure 2C and Table S4). Furthermore, we observed that participants who were faster on (correct) Distractor Inhibition trials exhibited a lower voluntary switch rate ($b = 0.01$, 95% HPD = [0.003, 0.01]) (see Figure 2D and Table S5). These contrasting relationships between spontaneous switch rates and task switch costs as well as Distractor Inhibition RT are further examined in the Discussion section.

We did not find evidence for a trade-off in performance between (forced) Task Switch trial RTs and Distractor Inhibition trial RTs, but rather observed a positive relationship between RTs across participants, such that individuals who were slower on (correct) Task Switch trials were also slower on Distractor Inhibition trials ($b = 0.23$, 95% HPD = [0.10, 0.34]). Together, these patterns of behavioral results suggest that the trade-off between putative flexible and stable control modes manifested in the relationship between indices of stable versus flexible performance and voluntary task switch rates, rather than the relationship between indices of stable versus flexible performance.

3.2 | Phasic Pupillary Dilation

Across trial types, we observed that the time window between 900 and 1900 ms after stimulus onset allowed for differentiation of trial types based on the ordinal pattern of phasic pupillary dilations (Figure 3A; see Methods). This window was used to calculate the phasic pupillary dilations for each trial type (Figure 3B; see Methods for phasic pupillary dilation calculation details). We examined the difference in phasic pupillary dilations across trial types using pupillary measures in mm, in line with the guidelines published by the Society for Psychophysiological Research (Steinhauer et al. 2022).

Mirroring the observed ordinal pattern of RTs, participants exhibited larger phasic pupillary dilations on Task Switch trials compared to Regular trials ($b = 0.82$, 95% HPD = [0.64, 1.00]; Table S6A), replicating previous observations (da Silva Castanheira et al. 2021; Rondeel et al. 2015). Further, we observed significantly larger phasic pupillary dilations on Distractor Inhibition trials ($b = 0.75$, 95% HPD = [0.57, 0.92]), Voluntary Repeat trials ($b = 0.81$, 95% HPD = [0.62, 1.00]) and Voluntary Switch trials ($b = 0.96$, 95% HPD = [0.70, 1.21]) compared to Regular trials. This pattern of phasic pupillary dilations suggests that processing demands for these trial types were greater than on Regular trials. Finally, echoing the finding that (forced) task switches imposed greater processing demands than task repetitions, we found that on Ambiguous trials (where a participant was free to switch or repeat tasks), phasic pupillary dilations were larger for Voluntary Switch versus Voluntary Repeat trials ($b = 0.14$, 95% HPD = [0.001, 0.27]).

Additionally, to examine the potential influence of the pupillary transient around the motor response (de Gee et al. 2014; Hupe et al. 2009), we examined differences in *response-locked* phasic pupillary dilations across trials. Our analyses yield the same pattern of phasic pupillary dilations across trial types as with stimulus-locked phasic pupillary dilations (see Tables S6A and S6B) indicating a plausibly low influence of the motor response on the differences in phasic pupillary dilation.

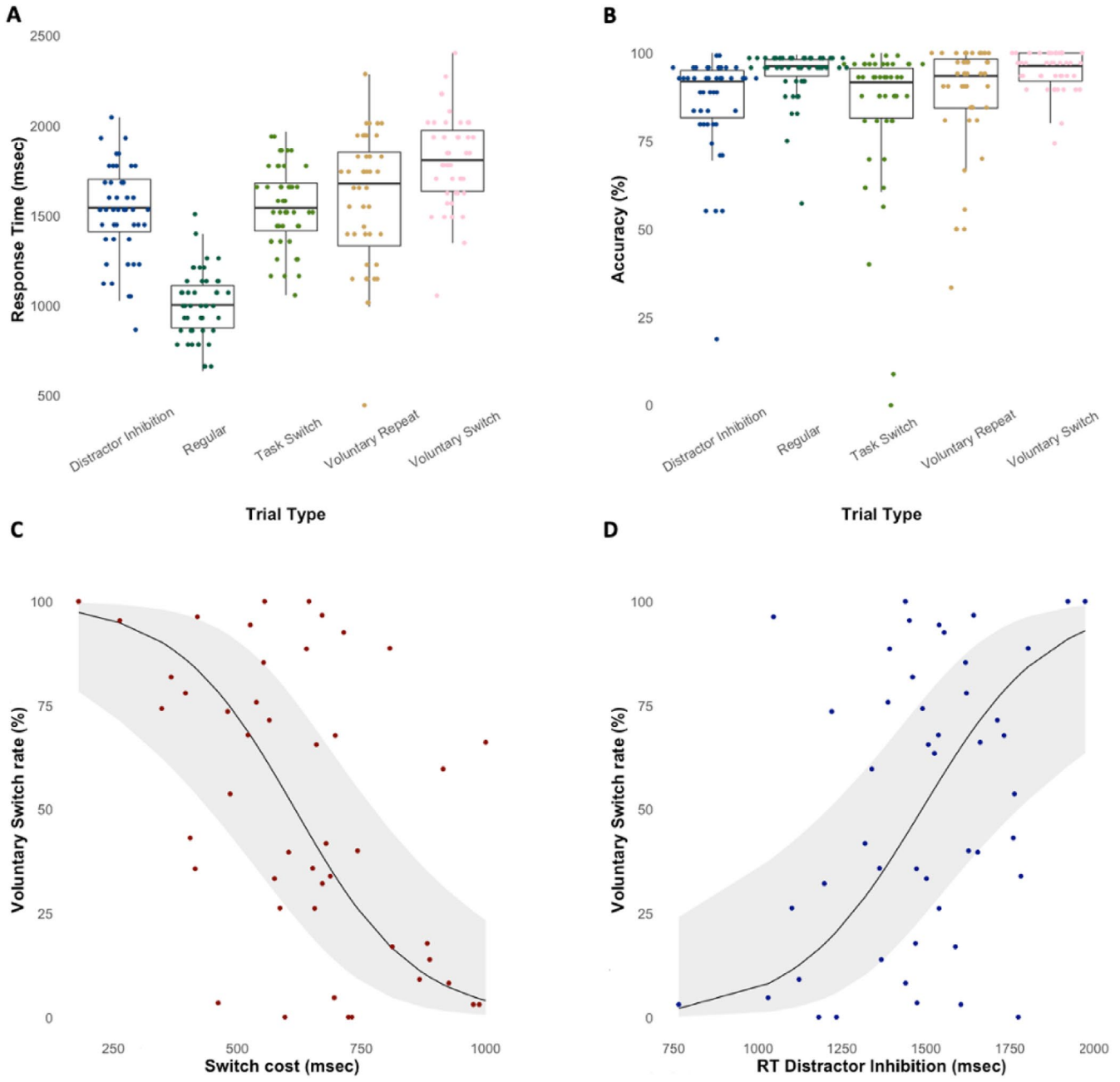

FIGURE 2 | Mean response times (A) and accuracy (B) on correct trials across trial types. Each dot represents mean subject level RT/accuracy averaged across the four sessions. Within each box, middle lines denote median values and boxes extend from the 25th to the 75th percentile of the distribution of values. Error bars represent 95% confidence intervals. (C) Negative association observed between voluntary switch rate on Ambiguous trials and individual switch cost, calculated as the difference in median RT on forced Task Switch and Regular repeat trials. (D) Positive association observed between voluntary switch rate and RT on Distractor Inhibition trials. (C, D) Within each box, gray band represents 95% confidence interval of the best fit line. The intercept and slope were estimated using Bayesian logistic regressions. The intercept indicates the log-odds of voluntary switching when task switch cost/Distractor Inhibition RT is zero. The slope indicates the change in log-odds of voluntary switching with 1 msec change in task switch cost/Distractor Inhibition RT. For ease of interpretation, the plots reflect subject- level estimates.

3.3 | Relationships Between Phasic Pupillary Dilations and Task Performance

We next investigated the extent to which phasic pupillary dilations related to behavioral indices of flexible versus stable processing modes. We first examined the relationship between phasic pupillary dilations on (forced) Task Switch trials and task switch costs, computed as the RT difference between Task Switch and Regular trials per session, finding a significant predictive relationship between individuals' average (per session) phasic pupillary dilations on task switch trials and task switch costs (see Figure 4). Specifically, we observed that larger average phasic pupillary dilations on Task Switch trials predicted smaller RT switch costs ($b = -0.14$, 95% HPD = [−0.26, −0.01]; Table S7). Applying the same analysis approach, we also examined whether average phasic pupillary dilations on Distractor

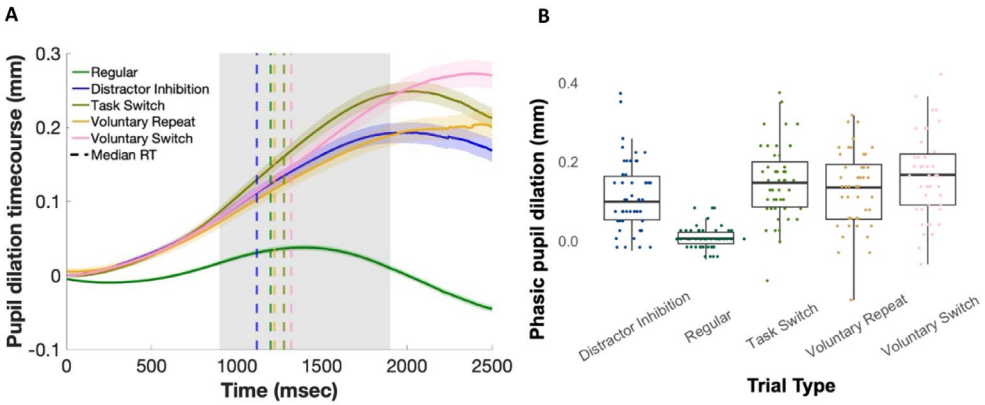

FIGURE 3 | (A) Time course of average phasic pupil from stimulus onset to 2 s after, by trial type (after removing trials with RTs <> 3SDs from the mean). Median RTs for each trial type are depicted as dashed vertical lines. The shaded area represents the standard error of mean across participants. (B) Average phasic pupillary dilation in the 900 msec to 1900 msec window, by trial type. Each dot represents mean subject level phasic pupillary dilation averaged across the four sessions. Within each box, middle lines denote median values and boxes extend from the 25th to the 75th percentile of the distribution of values. Error bars represent 95% confidence intervals.

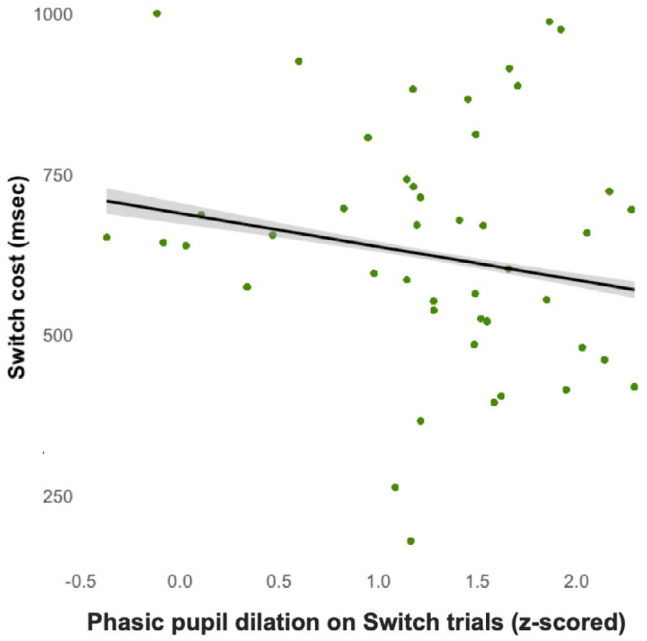

FIGURE 4 | Negative relationship between task switch cost and phasic pupil dilation on correct Task Switch trials. Gray band represents 95% confidence interval of the best fit line. The intercept and slope were estimated using Bayesian linear regression. The intercept indicates session-level task switch costs when session-level phasic pupil dilation on correct Task Switch trials is at its mean value. The slope indicates the change in session-level task switch costs with 1 standard deviation increase in session-level phasic pupil dilation on correct Task Switch trials. For ease of interpretation, the plots reflect subject-level estimates.

Inhibition trials predicted interference costs on these trials (computed as the per-session RT difference between Distractor Inhibition and Regular trials), but we did not find a relationship ($b = 11.91$, 95% HPD = [−36.43, 58.72]; Table S8).

Finally, to understand the robustness of the observed individual differences analyses, we conducted a post hoc power analysis of our individual difference analysis examining the relationship between phasic pupillary dilations on switch trials, and RT switch costs (reported in Table S7). We ran 200 simulations of the Bayesian model with 4 chains each with 2000 samples. We used a Bayesian version of a well-established frequentist power estimation approach (Gelman and Hill 2006; Chapter 20). The computed power based on 95% HPD interval excluding zero in simulated datasets is 1—indicating all simulations exclude zero. The interpretation of this computed power is analogous to traditionally computed frequentist power. The power based on posterior probability exceeding 0.95 is 0.14, and exceeding 0.80 is 0.71. In other words, in 71% of the simulations, this posterior probability exceeded 0.80. In summary, while these simulations suggest the effect estimate (posterior mean) of this individual difference analysis is intrinsically variable, the 95% HPD interval of all stimulated datasets excludes zero, suggesting that given the assumed effect size, our study design is sufficiently powered to detect it.

3.4 | Relationships Between Trial-Level Tonic Pupil Diameter and Task Performance

As previous works suggest an association between higher tonic activity and behavioral measures of flexibility (Jepma and Nieuwenhuis 2011) as well as task disengagement (Gilzenrat et al. 2010), we examined trial-level tonic pupil diameter (in mm) and its relation to trial-level voluntary task switching, expecting to observe a positive association (Jepma and Nieuwenhuis 2011). Contrary to our expectations, there was no association between baseline tonic pupil diameter on Ambiguous trials and participants' likelihood of making a voluntary task switch ($b = -0.30$, 95% HPD = [−0.66, 0.04]; Table S9).

Past studies have also observed that a larger trial-level tonic pupil diameter was evoked by errors in preceding trials, with the errors resulting in a reorientation of attention to the task at hand (Unsworth et al. 2018). Mirroring these findings, in our task, we observed higher trial-level tonic pupil post error on the previous trial ($b = 10.86$, 95% HPD = [5.08, 16.29]) and two trials prior ($b = 4.66$, 95% HPD = [0.003, 9.10]; Figure 5 and Table S10). This post-error increase in tonic pupil was also higher for the more difficult trial types in comparison to Regular trials (Table S11).

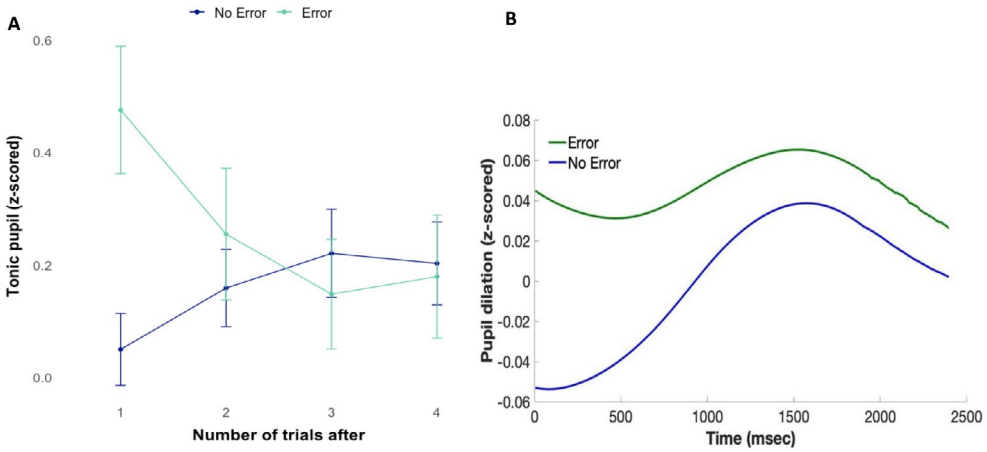

FIGURE 5 | (A) Higher tonic pupil diameter was observed post errors on previous trial and two trials prior. Error bars represent 95% confidence intervals. (B) Greater phasic pupil dilation post error on previous trial.

Larger baseline tonic pupil was also associated with slower RTs on Regular trials ($b = 0.02$, 95% HPD = [0.01, 0.02]; S11).

Urai et al. (2017) reported larger phasic pupil dilation after errors within the same trial. We observed larger phasic pupil dilation on trials following error trials (Figure 5B). However, the sustained elevation of tonic pupil diameter two trials after an error challenges the possibility of a simple carry-over effect of phasic pupil dilation on tonic pupil diameter. The delay of 3–5.5 s between the occurrence of an error two trials prior to the current trial and tonic pupil diameter on the current trial signifies a gradual carryover, which seems conceptually indistinguishable from tonic pupillary response.

3.5 | Relationships Between Fluctuations in Tonic Pupil Diameter and Task Performance

We also considered, in an exploratory analysis, whether putatively spontaneous fluctuations in tonic pupil diameter might predict performance across different trial types (van den Brink et al. 2016). Examination of the filtered tonic pupil diameter showed fluctuations which roughly corresponded to a window of approximately 10 trials. Accordingly, for this analysis of tonic pupillary fluctuations, we calculated a moving average of tonic pupil diameter across 10 trials (see Figure S1). In examining attentional lapses, we focused on Distractor Inhibition trials as these place the highest demands on stability, presumably rendering them most sensitive to these lapses (Ebitz et al. 2014). We observed that larger moving average tonic pupil diameter was not associated with changes in accuracy on Distractor Inhibition trials ($b = -0.00021$, 95% HPD = [−0.0004, 0.0001]; Table S12), or RT on correct Distractor Inhibition trials ($b = 0.000017$, 95% HPD = [−0.000001, 0.000038]); Table S13; we found the same relationships between moving average tonic pupil diameter and RT and accuracy for all other trial types. However, we did find that higher moving average of tonic pupil diameter was associated with faster Distractor Inhibition errors ($b = -0.00011$, 95% HPD = [−0.000168, −0.000003]) (Figure 6 and Table S14).

Additionally, we analyzed the relationship between moving average tonic pupil diameter and behavioral variability (similar to

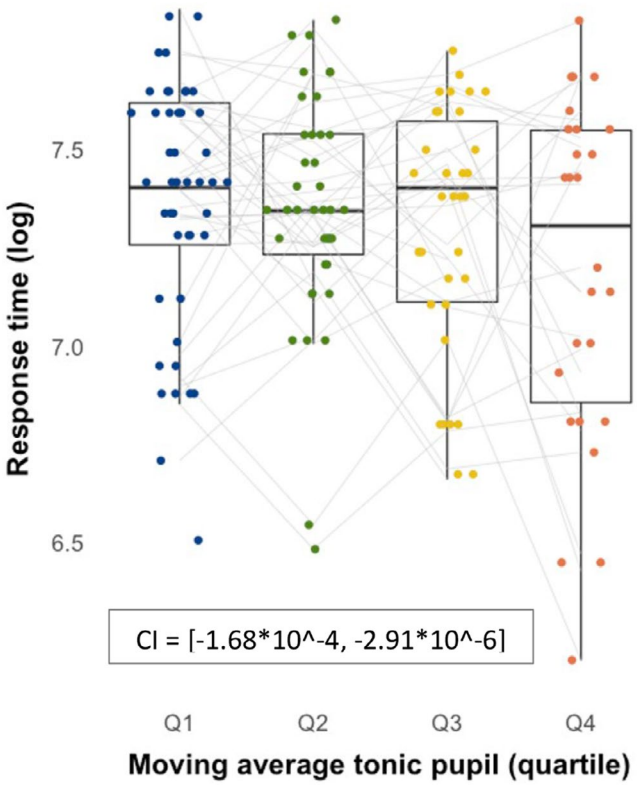

FIGURE 6 | Faster errors were observed on Distractor Inhibition trials with higher moving average tonic pupil diameter. Each dot represents mean subject level RT averaged across the four sessions within each moving average tonic quartile. Within each box, middle lines denote median values and boxes extend from the 25th to the 75th percentile of the distribution of values. Error bars represent 95% confidence intervals.

Murphy et al. 2011), which we operationalized here as the moving standard deviation in Accuracy as well as RT (in a 10-trial window, mirroring the tonic pupillary window). We observed, ordinally, that variability in accuracy and RT was largest in periods with larger moving average tonic pupil diameter (Figure 7). Statistically, we found that although moving average tonic pupil diameter was not associated with RT variability ($b = 0.05$, 95%

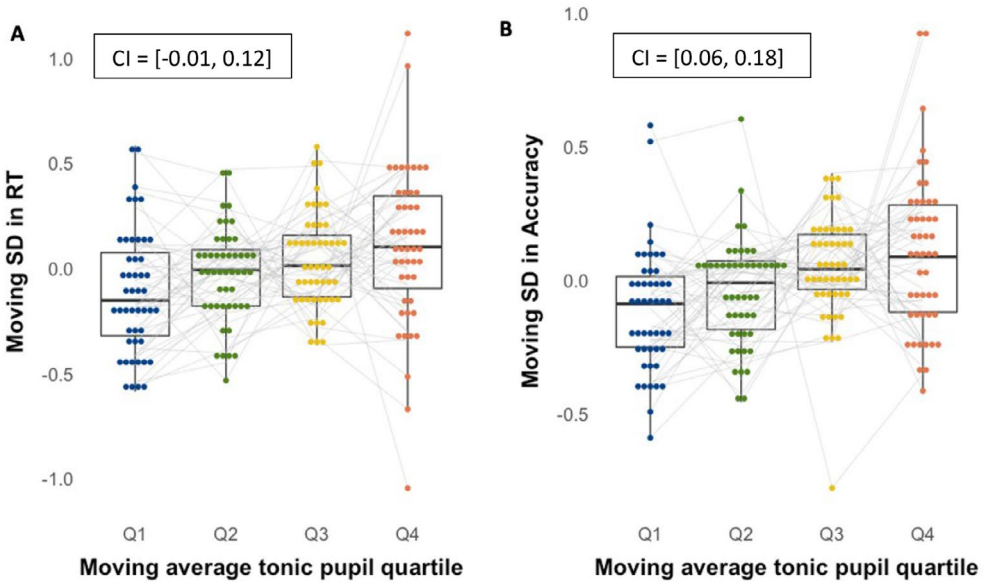

FIGURE 7 | (A) Moving SD of RTs did not vary with moving average tonic pupil. (B) Higher Moving SD in Accuracy was observed at higher moving average tonic pupil. Each dot represents mean subject level moving SD in RT/accuracy. Within each box, middle lines denote median values and boxes extend from the 25th to the 75th percentile of the distribution of values. Error bars represent 95% confidence intervals.

HPD = [−0.01, 0.12]; Table S15), moving average tonic pupil diameter was positively associated with variability in accuracy (taking the SD of accuracy over a 10-trial window; $b = 0.12$, 95% HPD = [0.06, 0.18]; Table S16).

3.6 | Relationship Between Average Session-Level Tonic Pupillary Activity and Behavioral Performance

Finally, we expected that individual differences in tonic pupil diameter would be positively associated with individuals' rates of voluntary task switching. To probe this relationship, we examined per-participant rates of voluntary task switching on Ambiguous trials, which we regressed against average tonic pupil diameter *z*-scored across all participants and sessions. We observed a negative relationship between average tonic pupil diameter and voluntary switch rates (see Figure 8). Statistically, a Bayesian mixed-effects regression revealed that participants with larger average (per-session) tonic pupil diameter were significantly less likely to voluntarily switch tasks on Ambiguous trials ($b = -0.33$, 95% HPD = [−0.61, −0.09]; see Table S17). We also observed that higher (session-level) average tonic pupil diameter during a session was associated with slower RTs ($b = 0.03$, 95% HPD = [0.02, 0.04]; see Table S18) and lower accuracy ($b = -0.36$, 95% HPD = -0.49, −0.23]; see Table S19) on Regular trials.

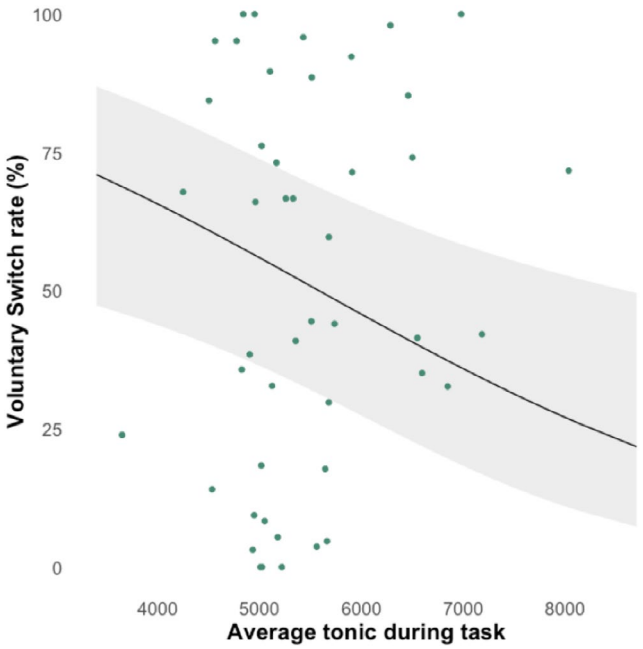

FIGURE 8 | Relationship between voluntary switch rate and average tonic pupil diameter for each subject. Gray band represents 95% confidence interval of the logistic regression slope. The intercept and slope were estimated using Bayesian logistic regression. The intercept indicates the log-odds of voluntary switching when session-level tonic pupil diameter is zero. The slope indicates the change in log-odds of voluntary switching with unit increase in session-level tonic pupil diameter. For ease of interpretation, the plots reflect subject-level estimates.

4 | Discussion

A body of previous research has examined the link between phasic and tonic pupillary activity and cognitive control modes. While larger phasic pupillary activity has been shown to index lower switch costs (da Silva Castanheira et al. 2021) and quicker responses on congruent Stroop trials (Rondeel et al. 2015), greater tonic pupil diameter has also been observed to index greater exploration (Gilzenrat et al. 2010; Jepma and Nieuwenhuis 2011). Here, we investigated the extent to which tonic and phasic pupil activity relate to cognitive stability and flexibility, using a task that allows us to characterize both task switching and distractor inhibition costs. Additionally, we also capture participants' voluntary task switching preferences, to index their cognitive flexibility.

Examining individual differences in task performance, we replicated previous work finding that individuals with larger task switch costs are less likely to voluntarily switch tasks on ambiguous trials (Armbruster et al. 2012). A possible interpretation of this finding is that individuals implicitly or explicitly account for their cost of task switching while choosing to be cognitively flexible or stable, that is, choosing which task to perform on ambiguous trials. It also indicates that voluntary switch rates positively index individuals' cognitive flexibility.

Perhaps more interestingly, we also observed a previously untested relationship between RTs on Distractor Inhibition trials and voluntary switch rates, finding that individuals who made slower responses on distractor inhibition trials—presumably indexing lower cognitive stability—exhibited larger voluntary switch rates. This finding suggests that voluntary switch rates negatively index individuals' cognitive stability.

Taken together, the above two observed behavioral relationships suggest that voluntary switch rates differentially index cognitive flexibility and cognitive stability. Moreover, these differential predictive relationships highlight a potential trade-off between voluntary task-switching preference and flexible versus stable performance, as a higher voluntary switch rate is observed in individuals with lower switch costs and longer Distractor Inhibition RT.

Echoing past work, our analysis revealed larger phasic pupillary dilations in Task Switch, Distractor Inhibition, Voluntary Repeat, and Voluntary Switch trials in comparison to Regular trials, suggesting the greater task demands exerted by these trials. It is plausible that the infrequent occurrence of Task Switch, Distractor Inhibition, and Voluntary trials (relative to Regular trials) engendered a "surprise" response which manifested as larger phasic pupillary dilations on these trials. However, the observation that individuals exhibiting larger phasic pupillary dilations on Task Switch trials exhibited smaller RT switch costs—which conceptually replicates our previous findings (da Silva Castanheira et al. 2021; Katidioti et al. 2014; Rondeel et al. 2015)—is difficult to explain with a surprise account of phasic pupillary dilations. Similarly, Rondeel et al. (2015) found that larger pupil dilations on congruent trials were associated with quicker responses, suggesting that greater effort investment on these trials resulted in faster RTs. However, we do not observe a similar relationship between RTs on Distractor Inhibition trials and phasic pupillary dilations. Along similar lines, Rondeel et al. saw a positive relationship between the Stroop Effect in RT and pupillary Stroop Effect (difference in pupil dilation between incongruent and congruent trials). The authors interpret these results to signify the relationship between pupil dilation and task difficulty, rather than task performance. However, dovetailing our findings, Hong et al. (2022) showed that larger phasic pupillary responses were associated with a neural measure of response inhibition—as indexed by decreased BOLD signals in the postcentral gyrus and superior parietal lobule. The varied findings across the three studies point towards a lack of a clear association between phasic pupil and distractor inhibition performance. In our study, the absence of a relationship between response times and phasic pupillary dilations in Distractor Inhibition trials alludes to the possibility that, unlike for trials requiring cognitive flexibility, in trials requiring stability, phasic dilation indexes task difficulty over performance.

We hypothesized that individuals with larger tonic pupil diameter, as indexed by average tonic pupil diameter during a session, would display a higher voluntary switch rate, based on earlier work showing higher exploration at higher tonic (Gilzenrat et al. 2010; Hong et al. 2022; Jepma and Nieuwenhuis 2011). Somewhat unexpectedly, we find no association between trial level tonic pupil diameter and voluntary switching, as well as the opposite relationship between voluntary switch rate and average tonic pupil diameter. The latter finding might be driven by the difference between exploration, that is potentially rewarding, and voluntary switching, which, as operationalized in our task, was not rewarding. Additionally, as shown in Figure 3, we observed a larger phasic pupillary dilation in Voluntary Switch trials in comparison to Voluntary Repeat trials, indicating the higher effort required to switch on Ambiguous trials. If our study participants lie on greater than optimal arousal levels on the Yerkes-Dodson inverted-U curve, a higher tonic pupil diameter, that is, a greater arousal, would be associated with lower task performance or greater task engagement. As such, a possible reason for the observed lower voluntary switching at larger average tonic pupil diameter might be that we are capturing task performance at higher than optimal arousal levels in our analysis.

Furthermore, we observed that an error on the previous trial and an error two trials before were followed by larger tonic pupil diameter on the current trial. This observation echoes previous work finding an association between slower RTs on a prior trial and greater tonic activity on the current trial (Unsworth et al. 2018), potentially indicating an intentional reorientation of attention after task disengagement. Unsworth et al. (2018) showed that trials with longer response times were followed by quicker response times, larger tonic pupil diameter, and larger phasic pupillary response on the current trial, suggesting an increase in attention. Corroborating these results, Weissman et al. (2006) showed that slower RTs were associated with reduced activity in the inferior frontal gyrus and increased activity in areas of the default mode network, reflecting greater focus on task-irrelevant internal thoughts and lower sustained attention. The observed increase in tonic pupil diameter in our study, therefore, indicates a possible increase in alertness and arousal and greater attention to the current task after an error.

Our analyses also revealed that higher moving average tonic pupil is associated with quicker errors on Distractor Inhibition trials, signaling attentional lapses or greater mind wandering/off-task brain activity (Unsworth et al. 2010). This result is also evocative of prior work observing a higher false alarm/error rate and longer correct response times in detection tasks on trials with larger baseline pupil diameters (Gilzenrat et al. 2010; Konishi et al. 2017; Rajkowski et al. 1993), especially in the presence of distractors (Ebitz et al. 2014). This association also corroborates Hong et al.'s finding of a negative relationship between baseline tonic pupil diameter and diminished P3 ERP and BOLD responses associated with task performance. The observed relationship also aligns with the Yerkes–Dodson performance-arousal curve (Murphy et al. 2014), wherein higher than optimal levels of tonic pupil diameter predict impaired task performance. However, one interpretational difficulty in assessing concordance between data and predictions of nonlinear (e.g., inverted-U shaped) accounts lies in characterizing where participants lie on the spectrum of possible tonic pupil diameter.

Finally, we also found that higher moving average tonic pupil diameter is associated with higher variability in accuracy, but not RT. Murphy et al. (2011) showed a U-shaped relationship between quintiles of tonic pupil diameter and RT variability (measured as the RT coefficient of variation). However, our analyses reveal an increasing variability in accuracy during periods with greater tonic pupillary activity. Along with the previous finding of quicker errors on Distractor Inhibition trials at higher moving average tonic pupil diameter, our finding of higher variability in accuracy at higher moving average tonic pupil diameter supports the overall idea of impaired task performance at higher than optimal tonic pupil levels.

The observed association between average tonic pupil diameter and voluntary task switching, in addition to the associations between moving average tonic pupil diameter and behavioral variability, as well as error RTs on DI trials, could reflect, in addition to lower flexibility preference and attentional lapses, off-task mind wandering. Although we mostly consider voluntary task switching as a marker of the operation of flexible (vs. stable) control, a lower preference to switch tasks, as observed at higher tonic pupil diameter could also be indicative of greater distracted states of attention (Unsworth and Robison 2016). Our results linking phasic pupil with task switching cost and distraction inhibition RT highlight that while phasic pupil dilation indexes flexible performance, it does not index improvements in stable performance. It is also plausible that phasic pupil dilation is reflecting the overall higher requirement of flexibility in this experimental design—involving switches between various trial types. However, as noted, the link between greater phasic pupillary activity and reduced (behavioral) task switch costs has been observed in previous work employing tasks with an equal need for stable and flexible control (da Silva Castanheira et al. 2021; Katidioti et al. 2014; Rondeel et al. 2015). These prior findings render the interpretation that greater phasic pupil dilation is associated with higher cognitively flexible, versus stable, performance, more likely.

While our study focuses on the association between LC-NE activity and pupillary activity based on past work (Joshi et al. 2016), newer work has called into question whether pupillary responses provide a valid real-time readout of LC activity (Megemont et al. 2022), finding that the pupil-LC relationship is dynamically modulated by brain states. Furthermore, pupillary responses might also reflect the influence of acetylcholine (Loewenfeld and Lowenstein 1993; Mridha et al. 2021; Reimer et al. 2016), 5HT (Cazettes et al. 2021) and orexin (Grujic et al. 2023). Further work is needed to understand the LC and non-LC contributions to pupillary dilation, especially as they relate to behavioral indices of cognitive control.

Our relatively small sample size of 51 participants may lead to considerable instability in regression estimates, making it difficult to interpret the results (Schönbrodt and Perugini 2013). Additionally, the individual difference scores examined in this study may exhibit low reliability, which further complicates the interpretation of the associated correlations and regression coefficients (Ackerman and Hambrick 2020; Hedge et al. 2018). Future research with substantially larger samples is therefore needed to assess the robustness and reliability of the present findings (For reliability estimates and descriptive statistics of regression measures see Table S20).

Overall, our findings shed initial light on how phasic and tonic pupillary activity differentially relate to the operation of stable versus flexible cognitive control modes. We observe that phasic pupil dilation indexes flexible versus stable task performance. Additionally, larger tonic pupil in our task was associated with decreased flexibility and increased task disengagement. While the association between phasic pupil dilation and cognitive control modes dovetails existing findings in the literature (Rondeel et al. 2015), the associations between tonic pupil and task performance add to our current understanding of the relationship between LC-NE activity and operation of flexible versus stable control modes (Aston-Jones et al. 1999).

Author Contributions

Anna Mini Jos: conceptualization, data curation, formal analysis, methodology, writing – original draft. **Andrew Westbrook:** formal analysis, supervision, writing – review and editing. **Sophia LoParco:** methodology, software. **A. Ross Otto:** conceptualization, funding acquisition, methodology, supervision, writing – review and editing.

Acknowledgments

We sincerely thank Winda Hao, Pelin Denecli, Stephany Dumas, Mikayla Schkolne and Gabriela Rey for their help with data collection. We also appreciate the time and effort of all participants. Additionally, we are grateful to Kevin da Silva-Castanheira for valuable discussions on pupillary analyses.

Ethics Statement

Approval was obtained from the McGill University Research Ethics Board. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent

Informed consent was obtained from all individual participants included in the study. Participants signed informed consent regarding publishing their data.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data and materials for the experiments are available upon request and, upon acceptance, all raw data will be made publicly available on the Open Science Foundation (www.osf.io) website. This experiment was not pre-registered.

References

Ackerman, P. L., and D. Z. Hambrick. 2020. "A Primer on Assessing Intelligence in Laboratory Studies." *Intelligence* 80: 101440. https://doi.org/10.1016/j.intell.2020.101440.

Aminihajibashi, S., T. Hagen, O. A. Andreassen, B. Laeng, and T. Espeseth. 2020. "The Effects of Cognitive Abilities and Task Demands on Tonic and Phasic Pupil Sizes." *Biological Psychology* 156: 107945. https://doi.org/10.1016/j.biopsycho.2020.107945.

Armbruster, D. J. N., K. Ueltzhöffer, U. Basten, and C. J. Fiebach. 2012. "Prefrontal Cortical Mechanisms Underlying Individual Differences in Cognitive Flexibility and Stability." *Journal of Cognitive Neuroscience* 24, no. 12: 2385–2399. https://doi.org/10.1162/jocn_a_00286.

Armbruster-Genç, D. J. N., K. Ueltzhöffer, and C. J. Fiebach. 2016. "Brain Signal Variability Differentially Affects Cognitive Flexibility and Cognitive Stability." *Journal of Neuroscience* 36, no. 14: 3978–3987. https://doi.org/10.1523/JNEUROSCI.2517-14.2016.

Arrington, C. M., and G. D. Logan. 2004. "The Cost of a Voluntary Task Switch." *Psychological Science* 15, no. 9: 610–615. https://doi.org/10.1111/j.0956-7976.2004.00728.x.

Aston-Jones, G., and J. D. Cohen. 2005. "An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance." *Annual Review of Neuroscience* 28, no. 1: 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709.

Aston-Jones, G., J. Rajkowski, and J. Cohen. 1999. "Role of Locus Coeruleus in Attention and Behavioral Flexibility." *Biological Psychiatry* 46, no. 9: 1309–1320. https://doi.org/10.1016/S0006-3223(99)00140-7.

Bürkner, P. C. 2017. "brms: An *R* Package for Bayesian Multilevel Models Using *Stan*." *Journal of Statistical Software* 80, no. 1: 1–28. https://doi.org/10.18637/jss.v080.i01.

Bürkner, P.-C. 2018. "Advanced Bayesian Multilevel Modeling With the R Package brms." *R Journal* 10, no. 1: 395. https://doi.org/10.32614/RJ-2018-017.

Cazettes, F., D. Reato, J. P. Morais, A. Renart, and Z. F. Mainen. 2021. "Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size." *Current Biology* 31, no. 1: 192–197.e4. https://doi.org/10.1016/j.cub.2020.09.090.

Cohen, J. D., and T. S. Braver. 1996. "A computational approach to prefrontal cortex, cognitive control and schizophrenia: Recent developments and current challenges." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 351, no. 1346: 1515–1527.

da Silva Castanheira, K., S. LoParco, and A. R. Otto. 2021. "Task-Evoked Pupillary Responses Track Effort Exertion: Evidence From Task-Switching." *Cognitive, Affective, & Behavioral Neuroscience* 21: 592–606. https://doi.org/10.3758/s13415-020-00843-z.

de Gee, J. W., T. Knapen, and T. H. Donner. 2014. "Decision-Related Pupil Dilation Reflects Upcoming Choice and Individual Bias." *Proceedings of the National Academy of Sciences* 111, no. 5: E618–E625. https://doi.org/10.1073/pnas.1317557111.

Dreisbach, G., and T. Goschke. 2004. "How Positive Affect Modulates Cognitive Control: Reduced Perseveration at the Cost of Increased Distractibility." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30, no. 2: 343–353. https://doi.org/10.1037/0278-7393.30.2.343.

Dreisbach, G., S. Musslick, and S. Braem. 2024. "Flexibility and Stability Can Be Both Dependent and Independent." *Nature Reviews Psychology* 3, no. 9: 636. https://doi.org/10.1038/s44159-024-00348-3.

Ebitz, R. B., J. M. Pearson, and M. L. Platt. 2014. "Pupil Size and Social Vigilance in Rhesus Macaques." *Frontiers in Neuroscience* 8: 100. https://doi.org/10.3389/fnins.2014.00100.

Egner, T., and J. Hirsch. 2005. "The Neural Correlates and Functional Integration of Cognitive Control in a Stroop Task." *NeuroImage* 24, no. 2: 539–547. https://doi.org/10.1016/j.neuroimage.2004.09.007.

Eriksen, B. A., and C. W. Eriksen. 1974. "Effects of Noise Letters Upon the Identification of a Target Letter in a Nonsearch Task." *Perception & Psychophysics* 16, no. 1: 143–149. https://doi.org/10.3758/BF03203267.

Fröber, K., and G. Dreisbach. 2017. "Keep Flexible—Keep Switching! The Influence of Forced Task Switching on Voluntary Task Switching." *Cognition* 162: 48–53. https://doi.org/10.1016/j.cognition.2017.01.024.

Fröber, K., L. Raith, and G. Dreisbach. 2018. "The Dynamic Balance Between Cognitive Flexibility and Stability: The Influence of Local Changes in Reward Expectation and Global Task Context on Voluntary Switch Rate." *Psychological Research* 82, no. 1: 65–77. https://doi.org/10.1007/s00426-017-0922-2.

Gelman, A., and J. Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942.

Gilzenrat, M. S., S. Nieuwenhuis, M. Jepma, and J. D. Cohen. 2010. "Pupil Diameter Tracks Changes in Control State Predicted by the Adaptive Gain Theory of Locus Coeruleus Function." *Cognitive, Affective, & Behavioral Neuroscience* 10, no. 2: 252–269. https://doi.org/10.3758/CABN.10.2.252.

Goschke, T., and A. Bolte. 2014. "Emotional Modulation of Control Dilemmas: The Role of Positive Affect, Reward, and Dopamine in Cognitive Stability and Flexibility." *Neuropsychologia* 62: 403–423. https://doi.org/10.1016/j.neuropsychologia.2014.07.015.

Grujic, N., A. Tesmer, E. Bracey, D. Peleg-Raibstein, and D. Burdakov. 2023. "Control and Coding of Pupil Size by Hypothalamic Orexin Neurons." *Nature Neuroscience* 26, no. 7: 1160–1164. https://doi.org/10.1038/s41593-023-01365-w.

Hayes, T. R., and A. A. Petrov. 2016. "Pupil Diameter Tracks the Exploration–Exploitation Trade-off During Analogical Reasoning and Explains Individual Differences in Fluid Intelligence." *Journal of Cognitive Neuroscience* 28, no. 2: 308–318. https://doi.org/10.1162/jocn_a_00895.

Hedge, C., G. Powell, and P. Sumner. 2018. "The Reliability Paradox: Why Robust Cognitive Tasks Do Not Produce Reliable Individual Differences." *Behavior Research Methods* 50, no. 3: 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

Hershman, R., and A. Henik. 2019. "Dissociation Between Reaction Time and Pupil Dilation in the Stroop Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45, no. 10: 1899–1909. https://doi.org/10.1037/xlm0000690.

Hoeks, B., and W. J. M. Levelt. 1993. "Pupillary Dilation as a Measure of Attention: A Quantitative System Analysis." *Behavior Research Methods, Instruments, & Computers* 25, no. 1: 16–26. https://doi.org/10.3758/BF03204445.

Hong, L., H. He, and P. Sajda. 2022. "Pupil-Linked Phasic Arousal Relates to the Reduction of Response Inhibition: Inferences From a Simultaneous Pupillometry and EEG-fMRI Study." https://doi.org/10.1101/2022.08.22.504728.

Hupe, J. M., C. Lamirel, J. Lorenceau, et al. 2009. "Pupil Dynamics During Bistable Motion Perception." *Journal of Vision* 9, no. 7: 10–10. https://doi.org/10.1167/9.7.10.

Ionescu, T. 2012. "Exploring the Nature of Cognitive Flexibility." *New Ideas in Psychology* 30, no. 2: 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001.

Jepma, M., and S. Nieuwenhuis. 2011. "Pupil Diameter Predicts Changes in the Exploration–Exploitation Trade-Off: Evidence for the Adaptive Gain Theory." *Journal of Cognitive Neuroscience* 23, no. 7: 1587–1596. https://doi.org/10.1162/jocn.2010.21548.

Joshi, S., Y. Li, R. M. Kalwani, and J. I. Gold. 2016. "Relationships Between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex." *Neuron* 89, no. 1: 221–234. https://doi.org/10.1016/j.neuron.2015.11.028.

Katidioti, I., J. P. Borst, and N. A. Taatgen. 2014. "What Happens When We Switch Tasks: Pupil Dilation in Multitasking." *Journal of Experimental Psychology: Applied* 20, no. 4: 380–396. https://doi.org/10.1037/xap0000031.

Knapen, T., J. W. de Gee, J. Brascamp, S. Nuiten, S. Hoppenbrouwers, and J. Theeuwes. 2016. "Cognitive and Ocular Factors Jointly Determine Pupil Responses Under Equiluminance." *PLoS One* 11, no. 5: e0155574. https://doi.org/10.1371/journal.pone.0155574.

Konishi, M., K. Brown, L. Battaglini, and J. Smallwood. 2017. "When Attention Wanders: Pupillometric Signatures of Fluctuations in External Attention." *Cognition* 168: 16–26. https://doi.org/10.1016/j.cognition.2017.06.006.

Laeng, B., M. Ørbo, T. Holmlund, and M. Miozzo. 2011. "Pupillary Stroop Effects." *Cognitive Processing* 12, no. 1: 13–21. https://doi.org/10.1007/s10339-010-0370-z.

Loewenfeld, I. E., and O. Lowenstein. 1993. *The Pupil: Anatomy, Physiology, and Clinical Applications*. Iowa State University Press; Wayne State University Press.

McGowan, A. L., M. C. Chandler, J. W. Brascamp, and M. B. Pontifex. 2019. "Pupillometric Indices of Locus-Coeruleus Activation Are Not Modulated Following Single Bouts of Exercise." *International Journal of Psychophysiology* 140: 41–52. https://doi.org/10.1016/j.ijpsycho.2019.04.004.

Megemont, M., J. McBurney-Lin, H. Yang, et al. 2022. "Pupil Diameter Is Not an Accurate Real-Time Readout of Locus Coeruleus Activity." *eLife* 11: e70510. https://doi.org/10.7554/eLife.70510.

Mridha, Z., J. W. De Gee, Y. Shi, et al. 2021. "Graded Recruitment of Pupil-Linked Neuromodulation by Parametric Stimulation of the Vagus Nerve." *Nature Communications* 12, no. 1: 1539. https://doi.org/10.1038/s41467-021-21730-2.

Murphy, P. R., R. G. O'Connell, M. O'Sullivan, I. H. Robertson, and J. H. Balsters. 2014. "Pupil Diameter Covaries With BOLD Activity in Human Locus Coeruleus." *Human Brain Mapping* 35, no. 8: 4140–4154. https://doi.org/10.1002/hbm.22466.

Murphy, P. R., I. H. Robertson, J. H. Balsters, and R. G. O'connell. 2011. "Pupillometry and P3 Index the Locus Coeruleus-Noradrenergic Arousal Function in Humans: Indirect Markers of Locus Coeruleus Activity." *Psychophysiology* 48, no. 11: 1532–1543. https://doi.org/10.1111/j.1469-8986.2011.01226.x.

Nakayama, M., and Y. Shimizu. 2004. Frequency Analysis of Task Evoked Pupillary Response and Eye-Movement. 6.

Nassar, M. R., K. M. Rumsey, R. C. Wilson, K. Parikh, B. Heasly, and J. I. Gold. 2012. "Rational Regulation of Learning Dynamics by Pupil-Linked Arousal Systems." *Nature Neuroscience* 15, no. 7: 1040–1046. https://doi.org/10.1038/nn.3130.

Pajkossy, P., Á. Szőllősi, G. Demeter, and M. Racsmány. 2017. "Tonic Noradrenergic Activity Modulates Explorative Behavior and Attentional Set Shifting: Evidence From Pupillometry and Gaze Pattern Analysis." *Psychophysiology* 54, no. 12: 1839–1854. https://doi.org/10.1111/psyp.12964.

Papadopetraki, D., M. I. Froböse, A. Westbrook, B. B. Zandbelt, and R. Cools. 2019. "Quantifying the Cost of Cognitive Stability and Flexibility." [Preprint] *Neuroscience*. e743120. https://doi.org/10.1101/743120.

Peysakhovich, V., M. Causse, S. Scannella, and F. Dehais. 2015. "Frequency Analysis of a Task-Evoked Pupillary Response: Luminance-Independent Measure of Mental Effort." *International Journal of Psychophysiology* 97, no. 1: 30–37. https://doi.org/10.1016/j.ijpsycho.2015.04.019.

Rajkowski, J., G. Aston-Jones, and P. Kubiak. 1993. "Correlations Between Locus Coeruleus (LC) Neural Activity, Pupil Diameter and Behavior in Monkey Support a Role of LC in Attention." *Society for Neuroscience – Abstracts* 19: 974.

Reimer, J., M. J. McGinley, Y. Liu, et al. 2016. "Pupil Fluctuations Track Rapid Changes in Adrenergic and Cholinergic Activity in Cortex." *Nature Communications* 7, no. 1: 13289. https://doi.org/10.1038/ncomms13289.

Rondeel, E. W. M., H. van Steenbergen, R. W. Holland, and A. van Knippenberg. 2015. "A Closer Look at Cognitive Control: Differences in Resource Allocation During Updating, Inhibition and Switching as Revealed by Pupillometry." *Frontiers in Human Neuroscience* 9: 494. https://doi.org/10.3389/fnhum.2015.00494.

Schönbrodt, F. D., and M. Perugini. 2013. "At What Sample Size Do Correlations Stabilize?" *Journal of Research in Personality* 47, no. 5: 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Serrien, D. J., and L. O'Regan. 2019. "Stability and Flexibility in Cognitive Control: Interindividual Dynamics and Task Context Processing." *PLoS One* 14, no. 7: e0219397. https://doi.org/10.1371/journal.pone.0219397.

Steinhauer, S. R., M. M. Bradley, G. J. Siegle, K. A. Roecklein, and A. Dix. 2022. "Publication Guidelines and Recommendations for Pupillary Measurement in Psychophysiological Studies." *Psychophysiology* 59, no. 4: e14035. https://doi.org/10.1111/psyp.14035.

Stroop, J. R. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18: 643–662.

Unsworth, N., R. P. Heitz, J. C. Schrock, and R. W. Engle. 2005. "An Automated Version of the Operation Span Task." *Behavior Research Methods* 37, no. 3: 498–505. https://doi.org/10.3758/BF03192720.

Unsworth, N., T. S. Redick, C. E. Lakey, and D. L. Young. 2010. "Lapses in Sustained Attention and Their Relation to Executive Control and Fluid Abilities: An Individual Differences Investigation." *Intelligence* 38, no. 1: 111–122. https://doi.org/10.1016/j.intell.2009.08.002.

Unsworth, N., and M. K. Robison. 2016. "Pupillary Correlates of Lapses of Sustained Attention." *Cognitive, Affective, & Behavioral Neuroscience* 16, no. 4: 601–615. https://doi.org/10.3758/s13415-016-0417-4.

Unsworth, N., M. K. Robison, and A. L. Miller. 2018. "Pupillary Correlates of Fluctuations in Sustained Attention." *Journal of Cognitive Neuroscience* 30, no. 9: 1241–1253. https://doi.org/10.1162/jocn_a_01251.

Urai, A. E., A. Braun, and T. H. Donner. 2017. "Pupil-Linked Arousal Is Driven by Decision Uncertainty and Alters Serial Choice Bias." *Nature Communications* 8: 14637. https://doi.org/10.1038/ncomms14637.

Valdez, P., C. Ramírez, A. García, J. Talamantes, P. Armijo, and J. Borrani. 2005. "Circadian Rhythms in Components of Attention." *Biological Rhythm Research* 36, no. 1–2: 57–65. https://doi.org/10.1080/09291010400028633.

van den Brink, R. L., P. R. Murphy, and S. Nieuwenhuis. 2016. "Pupil Diameter Tracks Lapses of Attention." *PLoS One* 11, no. 10: e0165274. https://doi.org/10.1371/journal.pone.0165274.

Van Der Wel, P., and H. Van Steenbergen. 2018. "Pupil Dilation as an Index of Effort in Cognitive Control Tasks: A Review." *Psychonomic Bulletin & Review* 25, no. 6: 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.

Weissman, D. H., K. C. Roberts, K. M. Visscher, and M. G. Woldorff. 2006. "The Neural Bases of Momentary Lapses in Attention." *Nature Neuroscience* 9, no. 7: 971–978. https://doi.org/10.1038/nn1727.

Yellin, D., A. Berkovich-Ohana, and R. Malach. 2015. "Coupling Between Pupil Fluctuations and Resting-State fMRI Uncovers a Slow Build-Up of Antagonistic Responses in the Human Cortex." *NeuroImage* 106: 414–427. https://doi.org/10.1016/j.neuroimage.2014.11.034.

Yerkes, R. M., and J. D. Dodson. 1908. "The Relation of Strength of Stimulus to Rapidity of Habit Formation." *Journal of Comparative Neurology and Psychology* 18: 459–482. https://doi.org/10.1002/cne.920180503.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.