

Task-Evoked Pupillary Responses Reveal a Role for Attentional Intensity in Auditory Discrimination

Mark G. Pillai¹, Cody A. Mashburn², Jason S. Tsukahara³, Audrey Leroux¹, and Randall W. Engle¹

¹ Department of Psychological and Brain Sciences, Georgia Institute of Technology

² Department of Psychological Science, Kennesaw State University

³ Department of Psychology, University of Miami

Sensory discrimination has long been theorized to reflect a core cognitive function underlying intelligence. Building on this tradition, we examined whether task-evoked pupil responses, as an index of mental effort, relate to attention control (AC) and auditory discrimination ability. Using a large data set ($N = 336$), participants completed three adaptive auditory discrimination tasks (pitch, duration, loudness), a cognitive battery assessing AC, working memory capacity, and fluid intelligence, and pupillometry during task performance. Structural equation modeling confirmed that AC predicted auditory discrimination thresholds and fully mediated the relationship between task-evoked pupil dilation and performance. Trial-level multilevel models revealed that higher task difficulty and errors were associated with larger pupil dilations, especially in demanding tasks. Crucially, individuals with greater AC exhibited steeper increases in pupil dilation with rising task difficulty and were more resilient to performance declines. For high-AC individuals, large pupil responses predicted preserved accuracy; for low-AC individuals, they signaled compensatory effort with diminished success. These results suggest that AC governs both the extent and efficacy of mental effort, determining whether high arousal facilitates or fails to improve performance. This work refines classic resource models by demonstrating that cognitive effort, indexed via pupillometry, supports perception most effectively when guided by strong AC. Findings support a hierarchical model in which AC serves as the functional bridge between arousal, sensory discrimination, and higher order cognition.

Public Significance Statement

When we try to detect faint sounds or listen in noisy environments, success depends not only on sensory acuity but also on how effectively we can focus and sustain attention. This study shows that individuals with stronger attention control are better able to discriminate subtle differences between sounds and exhibit more efficient patterns of mental effort, as reflected in pupil dilation. These findings suggest that attention control plays a critical role in determining when cognitive effort improves perception and when it does not, with implications for understanding individual differences in listening ability and for developing interventions to enhance perceptual performance.

Keywords: attention control, pupillometry, auditory discrimination

Rachel Wu served as action editor.

Mark G. Pillai  <https://orcid.org/0009-0003-9077-8771>

The data set analyzed in the present article was originally collected and reported in Tsukahara et al. (2020); the pupillometry data analyzed here were not previously reported. Preliminary findings from the present analyses were presented as a poster by Mark G. Pillai and Cody A. Mashburn at the 2024 Annual Meeting of the Psychonomic Society. No portion of the data or analyses appearing in this article has been shared prior. Materials, data, and scripts are publicly available at <https://doi.org/10.17605/OSF.IO/F38KS>.

This work was supported by the Office of Naval Research, Grants N00014-21-1-2327 and N00014-17-1-2061, and Naval Research Lab, Grants N00173-20-2-C003 and SA-GAT-NRL0135-2019 awarded to Randall W. Engle; Office of Naval Research, Grant N00014-22-S-F002 awarded to Randall W. Engle and Alexander P. Burgoyne; and Grant P14278 from the Institute for Mental Chronometry to the Georgia Tech Research Corporation.

Mark G. Pillai: conceptualization, formal analysis, visualization, writing—original draft, and editing and review. Cody A. Mashburn: data curation, methodology, formal analysis, and writing—editing. Jason S. Tsukahara: methodology and data collection. Audrey Leroux: formal analysis and methodology. Randall W. Engle:

supervision, funding acquisition, and writing—review and editing.

Mark G. Pillai played a lead role in supervision, validation, and writing—original draft, a supporting role in funding acquisition and project administration, and an equal role in conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, and writing—review and editing. Cody A. Mashburn played a supporting role in writing—original draft and an equal role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, and writing—review and editing. Jason S. Tsukahara played an equal role in conceptualization, data curation, resources, software, and supervision. Audrey Leroux played a supporting role in formal analysis, methodology, resources, validation, and visualization. Randall W. Engle played a lead role in funding acquisition, investigation, and project administration, a supporting role in supervision, validation, and writing—review and editing, and an equal role in methodology.

Correspondence concerning this article should be addressed to Mark G. Pillai, Department of Psychological and Brain Sciences, Georgia Institute of Technology, 654 Cherry Street NW, Atlanta, GA 30332, United States. Email: mpillai31@gatech.edu

Psychologists have long been interested in the relationship between sensory discrimination and higher order cognitive abilities. In the late 19th and early 20th centuries, Galton (1883) speculated that “the more perceptive the senses are of difference, the larger is the field upon which our judgment and intellect can act” (p. 27). For Galton, sensory discrimination was not merely a peripheral capacity but rather a gateway through which the mind encountered the world; sharper perception accordingly implied a richer basis for reasoning and judgment. This intuition led to the establishment of the first laboratory for the measurement of individual differences at the South Kensington Museum in London, where Galton assessed thousands of visitors on tests of sensory acuity, reaction time, and physical characteristics in the hope of quantifying human mental ability (Galton, 1883; Johnson et al., 1985).

Charles Spearman subjected Galton’s hypothesis to more rigorous empirical scrutiny, and the results appeared to vindicate it. Analyzing the correlations among school children’s performance on simple perceptual tests and their academic rankings, Spearman found that the ability to discriminate nearly identical stimuli, such as two tones close in pitch, correlated positively with general intellectual ability. He concluded that “the central Function, whatever it may be, is hardly anywhere more prominent than in the simple act of discriminating two similar tones” (Spearman, 1904, p. 274) and proposed the existence of a general factor, *g*, common to both sensory and intellectual tasks. In this early formulation, the link between sensory discrimination and intelligence was not incidental but theoretically central: Fine perceptual judgment and abstract reasoning were held to draw on the same underlying cognitive resource. Despite the significance of this claim, the nature Spearman’s “central Function” remained unspecified for nearly a century (Evans & Deary, 1994; Jensen, 1998).

Attention Control as Spearman’s Central Function

Although early differential psychology explored the sensory foundations of intelligence, the field subsequently shifted toward studying complex reasoning and the structure of cognitive abilities, leaving Spearman’s original hypothesis largely dormant. More recently, researchers have returned to Spearman’s original question using modern cognitive frameworks (Evans & Deary, 1994). In a series of studies, Tsukahara et al. (2020) directly reexamined the relationship between sensory discrimination and fluid intelligence (*Gf*) in a large adult sample. Their results confirmed a robust association between both visual and auditory discrimination ability and *Gf* but crucially demonstrated that this relationship was fully mediated by attention control (*AC*), defined as the capacity to maintain goal-relevant representations in an active state while resisting distraction and interference (Engle, 2002; Kane & Engle, 2003). When *AC* variance was statistically controlled in a mediation analysis, the direct link between sensory discrimination and *Gf* was substantially attenuated. This pivotal finding suggested that Spearman’s “central Function” was not sensory acuity per se but rather the attentional regulation of sensory processing, as sharper discrimination primarily reflects the deployment of executive attention to the task of perceptual judgment.

Other studies further evidence that individuals with stronger *AC* consistently perform better on working memory capacity (*WMC*) tasks and measures of *Gf*, suggesting that variation in the ability to regulate attention may underlie broad cognitive abilities (Burgoyne et al., 2023; Heitz et al., 2005; Lee & Engle, 2026; Mashburn et al., 2021;

Unsworth et al., 2014, 2021). *AC* is typically understood as comprising *selective* and *sustained* components (Chun et al., 2011; Posner & Petersen, 1990). Selective attention governs the prioritization of goal-relevant stimuli while suppressing irrelevant information. Classic and contemporary models show that salient or habitual responses can automatically capture attention unless overridden by a supervisory system (Egeth & Yantis, 1997; Norman & Shallice, 1986). At the neural level, top-down biasing enhances the gain of neurons tuned to attended features and dampens responses to distractors (Deco & Rolls, 2005; Desimone & Duncan, 1995; Kastner & Ungerleider, 2000). These mechanisms support both everyday tasks, such as focusing on a single voice in a noisy room and fine-grained perceptual discrimination (Cherry, 1953; Shinn-Cunningham, 2008). Sustained attention, in contrast, reflects the capacity to maintain cognitive engagement and effort over time, particularly under demanding conditions (Robertson et al., 1997; Sarter et al., 2001). Many reasoning and working memory tasks depend on this ability to sustain goal-relevant stimuli and processing across extended sequences of operations (Cowan, 2017; Engle, 2002). Effortful attention becomes especially important when tasks are difficult or conflict-laden, drawing on a limited pool of cognitive resources (Kahneman, 1973; Posner & Petersen, 1990). Individual differences in this sustained control predict performance on *WMC* and *Gf* tasks: People who can maintain effort and resist interference tend to achieve higher scores (Cochrane et al., 2019; Kane & Engle, 2003; Unsworth et al., 2014; Voelke & Roebers, 2016).

Auditory Perception and Attention

A secondary finding in Tsukahara et al. (2020) has received less theoretical attention but is of considerable importance: Auditory discrimination tasks loaded more strongly onto the latent general discriminative ability factor than their visual counterparts. In the same study, the visual discrimination tasks presented two line-shaped stimuli simultaneously, whereas the auditory tasks presented two tones sequentially, separated by a brief interstimulus interval. As we argue in the next section, this asymmetry in factor loadings reflects a fundamental property of the auditory modality, rendering auditory discrimination tasks more salient to *AC* into its decomposing into selective and sustained components.

To understand the differential factor loadings of auditory and visual tasks, it is necessary to consider the fundamental nature of each sensory signal. Sound is a time-varying mechanical pressure wave, and the information it carries is almost entirely encoded in the temporal domain. For example, the auditory dimension of pitch is defined by the periodic frequency of waveform oscillation and cannot be deduced from a single instantaneous sample, as the auditory system requires a minimum signal duration before frequency can be resolved (Moore, 2012). The cochlea accordingly performs a running spectral decomposition of incoming sound, functioning as a biological approximation of a short-time Fourier transform in which different regions of the basilar membrane respond preferentially to different frequency components as they unfold over time (Robles & Ruggero, 2001; von Békésy, 1960). Sound localization similarly depends on temporal computation, with the auditory system detecting interaural time differences on the order of microseconds to calculate the direction of a source (Grothe et al., 2010; Jeffress, 1948). Even at the level of scene analysis, the auditory system parses complex acoustic mixtures by exploiting temporal regularities: Components sharing onset times

are grouped into common perceptual streams, while those that are asynchronous or independently modulated are segregated (Bregman, 1990; Darwin & Carlyon, 1995). The organizing cues of auditory scene analysis are therefore predominantly temporal, as binding successive acoustic events into coherent perceptual streams is the auditory counterpart to spatial integration in vision (Griffiths & Warren, 2004).

Vision, by contrast, operates in the spatial domain. The retina is a two-dimensional array of photoreceptors that samples a scene simultaneously across its entire surface, with spatial relationships between activated receptors directly encoding the layout of objects in the environment (Curcio et al., 1990; Masland, 2001). Early visual cortex preserves this arrangement in the form of a retinotopic map in which neighboring neurons respond to neighboring regions of visual space (Tootell et al., 1982; Wandell et al., 2007). A static image therefore contains rich, extractable information; perceptual organization can proceed, in principle, from a single moment of exposure. Scene segmentation involves the grouping of elements into objects and proceeds largely through the analysis of spatial relationships such as proximity, continuity, and common region (Palmer, 1999; Wertheimer, 1923/1938).

These modality-specific architectures are directly instantiated in the stimulus formats that discrimination paradigms necessarily employ. In auditory frequency discrimination, the two comparison tones must be presented sequentially, as simultaneous presentation would produce acoustic fusion or spectral masking, rendering independent pitch judgments impossible (Plack, 2014). That Tsukahara et al. (2020) presented auditory tones sequentially while visual line stimuli were presented simultaneously therefore reflects genuine modal asymmetry rather than procedural inconsistency. This sequential format imposes a temporal comparison demand that necessarily recruits controlled attentional resources: The first tone's acoustic features must be encoded in working memory and held across the interstimulus interval, where the representation is vulnerable to decay and interference (Deutsch, 1970; Semal et al., 1996), before being compared with the second tone (Cowan, 1984; Pasternak & Greenlee, 2005). The quality of the discrimination judgment is therefore limited by the fidelity of this maintained trace, and the finer the difference to be detected, the more precisely it must be preserved (Carlyon, 2004; Harris, 1952). Attentional shielding is the active protection of a goal-relevant memory trace against interference and is precisely the selective, sustained capacity that AC measures capture (Hasher & Zacks, 1988; Oberauer, 2002).

Pupil Dilation Reflects Effort and Control

As AC operates at multiple processing junctures, it is a natural candidate for linking sensory discrimination with intellectual performance. Crucially, AC is not an all-or-nothing resource but one that scales with task demands; this graded nature has measurable physiological consequences (Alnæs et al., 2014; Botvinick et al., 2001; Kahya et al., 2022; Unsworth & Robison, 2018). Chief among these is pupil dilation: as cognitive load increases and attentional resources are more heavily recruited, the pupils enlarge in proportion to the mental effort required (Beatty & Lucero-Wagoner, 2000; Hess & Polt, 1964; van der Wel & van Steenbergen, 2018). Notably, Kahneman and Beatty (1967) demonstrated that the difficulty of a pitch discrimination task was positively correlated with the amplitude of pupil dilation, establishing an early link between auditory perceptual effort and pupillometry. Subsequent research has confirmed that pupil size is a

reliable gauge of cognitive load and arousal across a wide range of tasks (Beatty, 1982; da Silva Castanheira et al., 2021; Hogervorst et al., 2014; O'Bryan et al., 2024).

The neurophysiological basis for task-evoked pupil dilation lies in the locus coeruleus–norepinephrine (LC–NE) system, a brainstem arousal nucleus that modulates neural gain, attention, and behavioral readiness, with downstream effects on pupil size (Murphy et al., 2014; Sara, 2009; Servan-Schreiber et al., 1990). LC projections release norepinephrine broadly across the cortex, increasing neural excitability and enhancing the signal-to-noise ratio of task-relevant processing (Aston-Jones & Cohen, 2005). Phasic LC bursts amplify processing of goal-relevant stimuli, whereas tonic LC activity sets baseline arousal and follows the classic Yerkes–Dodson law, exhibiting an inverted-U function in which intermediate arousal levels optimize performance (Teigen, 1994; Yerkes & Dodson, 1908). At the cellular level, norepinephrine increases excitability by modulating Ca^{2+} and K^{+} conductance, reducing hyperpolarization and enabling faster neuronal recovery (Foote et al., 1983). At the systems level, NE release promotes coordinated firing and strengthens functional connectivity among task-engaged regions, supporting executive AC (Shenhav et al., 2017). Pupil diameter is closely linked to LC–NE modulation, because LC activation inhibits the parasympathetic Edinger–Westphal nucleus, producing dilation (Joshi et al., 2016; Samuels & Szabadi, 2008). Accordingly, task-evoked pupil dilation provides a real-time index of LC–NE recruitment and neural gain during cognitive effort (da Silva Castanheira et al., 2021).

Pupillary responses have been linked to a broad range of cognitive abilities, though not all associations are equally robust. While evidenced in some studies (Aston-Jones & Cohen, 2005; Tsukahara et al., 2016), large-scale replications have failed to find a reliable relationship between baseline pupil diameter and fluid intelligence (Lorente et al., 2025; Robison & Campbell, 2023). By contrast, the link between task-evoked pupillary responses and WMC is well established. Unsworth and Robison (2017) found that WMC is associated with stable individual differences in tonic and phasic LC–NE activity, such that higher capacity individuals maintain more optimal baseline arousal and recruit more effective phasic noradrenergic responses under increasing task demands. Similarly, higher WMC individuals exhibited larger task-evoked pupil dilations during attention-demanding trials and were more likely to translate increased arousal into accurate performance, whereas lower capacity individuals show smaller and less adaptive pupillary responses (Robison & Unsworth, 2019). These findings are theoretically meaningful given extensive evidence that WMC tasks largely reflect AC processes, particularly the ability to maintain goal-relevant information and resist interference (Engle, 2002; Kane & Engle, 2003; Unsworth et al., 2014). At the neural level, the prefrontal cortex provides direct descending input to the LC, enabling executive systems to modulate the timing and magnitude of noradrenergic release (Arnsten & Goldman-Rakic, 1998; Sara & Bouret, 2012). Prefrontal cortex activity can shape LC firing patterns during goal-directed behavior, regulate phasic LC bursts, and stabilize task engagement under cognitive load (Aston-Jones & Waterhouse, 2016; Breton-Provencher & Sur, 2019; Chandler et al., 2014; Sara & Bouret, 2012). In this framework, AC is not merely a downstream beneficiary of LC–NE arousal but an upstream regulator of when and how the LC fires, forming a bidirectional control loop between executive systems and arousal circuits (Arnsten, 2015; Nieuwenhuis et al., 2011). This architecture predicts that individual differences in AC should manifest in the profile of task-evoked pupil responses:

High-AC individuals show larger, more precisely calibrated phasic dilations as difficulty increases, whereas low-AC individuals show blunted or dysregulated responses (Robison & Unsworth, 2019; Unsworth & Robison, 2017). This regulation account distinguishes effort as a controlled process rather than a simple resource, contrasting with classical capacity models in which pupil dilation reflects only the amount of effort deployed (Kahneman, 1973).

The Present Study

Tsukahara et al. (2020) collected pupil dilation data during auditory discrimination tasks in their second study, providing a direct physiological measure of cognitive effort via the LC–NE system. However, these data remained unexplored in their published findings. The present article leverages these unreported pupillometric data to investigate how task-evoked cognitive effort relates to AC, auditory discrimination ability, and fluid intelligence, providing trial-level evidence about the mechanisms through which AC converts moment-to-moment effort into successful perceptual performance.

We pursue three interconnected aims. First, at the level of latent traits, we test whether task-evoked pupillary response (TEPR) predicts auditory discrimination ability independently of AC or whether AC fully accounts for this relationship, thereby extending the mediation model of Tsukahara et al. (2020) to include a physiological effort index. Second, at the within-trial, between-person level, we examine how pupil dilation tracks momentary fluctuations in task difficulty and whether AC moderates the link between effort and accuracy. Third, we ask whether large pupil responses function as a performance-boosting mechanism or as a marker of compensatory strain and whether this distinction depends on individual differences in AC. Collectively, these analyses aim to clarify the mechanism through which arousal, AC, and sensory discrimination jointly determine intellectual performance by providing a physiological window onto the “central Function” that Spearman identified more than a century ago.

Method

Participants

The present analyses were conducted on $N = 336$ participants drawn from Study 2 of Tsukahara et al. (2020), representing the subset of the original Study 2 sample ($N = 392$) with usable pupillometry data after applying pupil-specific quality criteria (see Pupillometry and TEPR section). Demographic information was collected via self-report during prescreening. Participants ranged in age from 18 to 35 years ($M_{\text{age}} = 24.5$ years, $SD = 4.7$). Gender was self-reported, with the sample including 202 women (60%) and 134 men (40%). Race and ethnicity were self-reported with categories of White, Black or African American, Asian or Pacific Islander, and other (the latter including Native American, Hispanic or Latino, multiracial, and other); the full Study 2 sample comprised 28.6% White, 31.2% Black or African American, 31.0% Asian or Pacific Islander, and 9.2% other. All participants were native English speakers (learned English before age 5) and had not previously participated in studies with this lab. Participants reported no history of hearing or cognitive impairments during prescreening. All participants completed a battery of auditory discrimination tasks and cognitive ability measures. The cognitive battery included measures

of AC, WMC, and Gf. Participants provided informed consent and were compensated for their participation. All procedures were approved by the Georgia Institute of Technology Institutional Review Board (Protocol No. H16409).

Auditory Discrimination Tasks

Participants completed three adaptive auditory discrimination tasks across three acoustic modalities: pitch, duration, and loudness. These tasks measured participants' ability to detect small differences between sequential auditory stimuli. On each trial, participants heard two pure tones presented sequentially, separated by a 500-ms interstimulus interval, and indicated which tone met the task-specific criterion (higher pitch, longer duration, or greater loudness). The intertrial interval was 2 s. Stimuli were presented binaurally through closed-ear headphones (Audio-Technica ATH-M50). Tones used in the pitch and loudness discrimination tasks were 500 ms in duration and synthesized at 44.1 kHz. Across all tasks, first and second tones were correct with equal probability. All tones included a 5-ms linear fade-in and 5-ms fade-out to avoid onset and offset artifacts. Each task consisted of 64 trials divided into four blocks of 16 trials. During tone presentation participants viewed a blank black screen depicting a central fixation “+” to minimize visual distractions. After the second tone, a response screen appeared, and responses were recorded via mouse click.

Classical psychophysical approaches estimate discrimination thresholds by sampling performance across a range of stimulus differences and fitting a psychometric function describing the relationship between stimulus difference and response accuracy. These functions typically take a sigmoidal (logistic) form, in which performance approaches chance levels ($\sim 50\%$) at very small stimulus differences and near-perfect accuracy ($\sim 100\%$) at large differences. Individual differences in perceptual sensitivity are reflected in the location and slope of the transition region of this function. Discrimination thresholds are commonly defined as the stimulus difference corresponding to approximately 75% correct performance, which represents the midpoint between chance and perfect accuracy and lies within the steepest portion of the psychometric slope, where sensitivity differences between individuals are most pronounced.

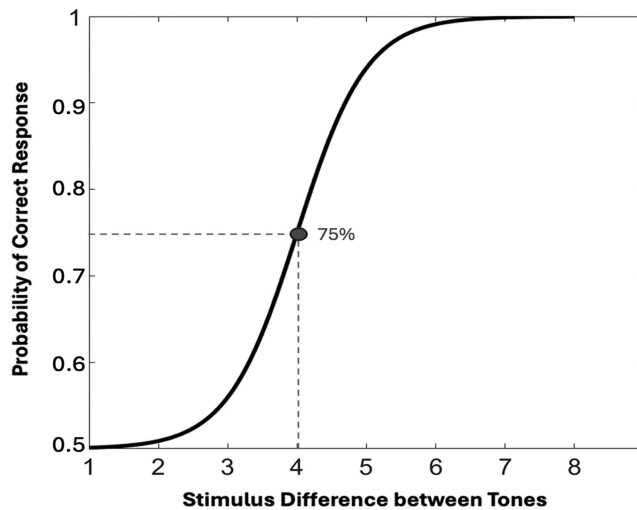
Rather than exhaustively sampling the full psychometric curve, the present study used an adaptive weighted 3:1 up–down staircase procedure (Kaernbach, 1991) to efficiently estimate this threshold. Following a correct response, the difference between the tones was reduced on the next trial (increasing task difficulty); following an incorrect response, the difference was increased (decreasing difficulty). The asymmetric step sizes in the 3:1 procedure cause the staircase to converge on a performance level of approximately 75% correct, corresponding to the participant's discrimination threshold (Figure 1).

Pitch Discrimination

Participants heard two pure-tone sine waves (e.g., 440 Hz vs. 460 Hz) and identified which was higher in pitch. Difficulty was defined by the absolute frequency difference between tones. The staircase began at 15 Hz; initial steps were $\pm 6/2$ Hz and then $\pm 3/1$ Hz, ranging from 1 to 24 Hz.

Figure 1

Illustration of the Psychometric Function Used to Estimate Auditory Discrimination Thresholds



Note. The sigmoid curve illustrates the relationship between stimulus difference and the probability of a correct response in a two-alternative forced-choice auditory discrimination task. The dashed horizontal line indicates the discrimination threshold (approximately 75% correct), corresponding to the level at which the adaptive staircase procedure converges to estimate participants' perceptual thresholds.

Duration Discrimination

Participants heard two tones differing only in duration (e.g., 500 ms vs. 520 ms) and identified which was longer. Difficulty was defined by the duration difference in milliseconds. The staircase began at 100 ms; initial steps were $\pm 60/20$ ms and then $\pm 30/10$ ms, ranging from 10 to 150 ms.

Loudness Discrimination

Participants heard two tones differing only in intensity (e.g., 65 dB vs. 67 dB) and identified which was louder. Difficulty was defined by the sound pressure level difference in dB. The staircase began at 10 dB; initial steps were $\pm 6/2$ dB and then $\pm 3/1$ dB, ranging from 1 to 13 dB.

AC Tasks

AC was modeled as a latent construct reflecting goal maintenance, interference resolution, and attention regulation under conflict. The three tasks below index complementary facets of the construct:

- The Antisaccade task indexes inhibitory control over reflexive orienting (Kane & Engle, 2003; Unsworth et al., 2004).
- The Sustained Attention to Cue Task indexes proactive goal maintenance in the presence of competing stimuli, consistent with models of proactive control (Braver, 2012; Chun et al., 2011).

- The Visual Arrays task indexes selective encoding and the protection of goal-relevant representations from interference (Fukuda & Vogel, 2009; Vogel et al., 2005).

Antisaccade Task

This task measured the ability to avoid reflexive saccades toward a salient but task-irrelevant cue. Each trial began with a central fixation cross, followed by the brief appearance (100 ms) of a peripheral distractor asterisk to the left or right (12.3° visual angle). Immediately afterward, a target letter ("Q" or "O") appeared for 100 ms on the opposite side of the screen before being masked by a double hashtag (##). Participants reported the identity of the target letter via keyboard. Sixteen practice trials (with extended durations) preceded 48 test trials. Accuracy, or the proportion of correct identifications, served as the performance metric. Higher scores reflect stronger inhibitory control and attentional orienting.

Sustained Attention to Cue Task

This task measured the ability to sustain the focus of attention at a cued spatial location over a variable delay (Tsukahara & Engle, 2022). Each trial began with a 1-s fixation cross, followed by a 750-ms "Get Ready!" prompt accompanied by an auditory beep at the to-be-cued location. A shrinking circle cue was then displayed for approximately 500 ms before disappearing. Participants were required to maintain attention at the now-uncued location across a blank screen for a variable wait time of 0 or 2–12 s (in 500-ms intervals). At the end of the wait time, a target letter (B, P, or R) was briefly displayed in dark-gray font at the center of the cued location for 250 ms among nine nontarget letters and then masked for 300 ms. Participants identified the target letter via mouse click from a response screen. The task comprised three blocks of 22 trials (66 total) with no feedback. Performance was scored as overall accuracy, with higher scores reflecting stronger ability to sustain the focus of attention. This task loads reliably onto a latent AC factor (Draheim et al., 2021; Tsukahara & Engle, 2022).

Selective Visual Arrays Task With Orientation Judgment

This task measured selective encoding of visual information in working memory. On each trial, participants briefly viewed an array of colored squares (set sizes 2–6), followed by a blank retention interval and a probe display. The probe tested whether a specific item's color had changed. Participants responded "same" or "different" via keypress. Performance was scored using Cowan's *K* formula to estimate the number of items stored in working memory. The *K* score at set size 4 (VA4) was used as the primary indicator of attention-guided storage capacity.

Working Memory Capacity

Working memory capacity was assessed using three complex span tasks: the Advanced Operation Span (OSpan), Advanced Symmetry Span (SymSpan), and Advanced Rotation Span (RotSpan). These tasks follow the paradigm established by Unsworth et al. (2005), in which participants alternate between a processing task and the encoding of to-be-remembered items. Scores were calculated using the

partial-credit unit scoring method (Conway et al., 2005), which tallies the number of items recalled in the correct serial position across all trials.

Advanced Operation Span

Participants alternated between solving simple math equations and remembering unrelated letters. Each trial began with an equation, for example, “ $(8/2) + 1 = ?$ ” followed by a single letter to be remembered. This sequence was repeated for a variable set size (3–8). After the final pair, participants recalled the letters in order by selecting them from a matrix. There were 14 trials (two blocks of seven). Due to a programming error, the highest set size (9) was not included, and set size 8 occurred more frequently (four times). The total number of correctly recalled letters in the correct order was used as the task score.

Advanced Symmetry Span

Participants alternated between judging whether a pattern of black and white squares (presented in a 16×16 matrix) was symmetrical along the vertical axis and remembering spatial locations in a 4×4 matrix. Each red-square location was followed by a new symmetry judgment until a variable set size (2–7) was reached. On the recall screen, participants clicked the remembered locations in order. There were 12 trials (two blocks of six), and task score was the total number of spatial locations correctly recalled in serial order.

Advanced Rotation Span

Participants alternated between solving mental rotation problems (identifying whether a rotated letter was mirror-reversed) and encoding directional arrows of varying size and orientation (e.g., small northeast arrow). At recall, participants selected the arrows in the correct order from a set of all possibilities. There were 12 trials (two blocks of six), with set sizes ranging from 2 to 7. Task score was the number of arrows correctly recalled in order.

Fluid Intelligence

Fluid intelligence was modeled as a latent construct reflecting domain-general reasoning ability, including abstract rule induction, relational integration, and flexible problem solving. The three tasks below are standard indicators of Gf, as they require participants to infer underlying patterns or rules and apply them to novel problems, which define fluid reasoning (Carroll, 1993; McGrew, 2009).

Raven’s Advanced Progressive Matrices

Participants viewed 3×3 visual matrices with one element missing in the bottom-right corner. Each matrix followed a logical pattern, and participants selected the best-fitting option from eight possible choices. Eighteen odd-numbered items from the standard set were presented with a 10-min time limit. The score was the total number of correct responses.

Letter Sets

Each item consisted of five letter strings (e.g., “ABCD, EFGH, IJKL, MNOP, QRST”). Four followed a hidden rule, and one deviated. Participants were instructed to identify the rule and choose

the deviant set. They had 10 min to complete 30 problems. Score was the number of correct answers.

Number Series

Participants viewed numeric sequences (e.g., “2, 4, 6, 8, ?”) that followed an implicit logical rule. They selected the correct next number from five options. The task consisted of 15 items with a 5-min time limit. Score was the number of correct completions.

Scores from the three tasks were used to model a latent Gf factor in structural equation modeling analyses, representing domain-general reasoning ability.

Pupillometry and TEPR

Throughout the auditory tasks, participants’ pupil size was continuously recorded using a SMI Red250cm eye tracker with a 250 Hz image-sampling rate. Pupil data were preprocessed using the *pupillometry* R package (Tsukahara, 2022). Blinks were extended by 50 ms in both directions, and their corresponding pupil values were set to missing. Pupil values exceeding a threshold of speed of change were detected and set to missing using the median absolute deviation (Kret & Sjak-Shie, 2019) with a threshold of $n = 10$. The pupil data were then smoothed using a moving-window average with $n = 7$ samples, followed by cubic-spline interpolation for gaps not exceeding 750 ms of continuous missing data. After all preprocessing steps, we assessed the amount of missing data per trial. Trials with greater than 25% missing data were removed from further processing. Trial-level pupil sizes were baseline corrected based on the mean pupil size in the 200 ms preceding the first tone in a trial. Next, we obtained the TEPR for each trial, which we defined as the maximum change in pupil size (in millimeters) from baseline over the remainder of the trial. These trial-level TEPR values served as an index of momentary cognitive effort or arousal evoked by the task. For each participant, we also computed aggregate TEPR measures per task (the average pupil dilation across all trials of a given task).

Analytic Approach

Latent Trait Modeling

We used two latent variable models to analyze between-subject (trait-level) relationships among TEPR, AC, WMC, Gf, and Auditory Discrimination. We first fit a confirmatory factor analysis (CFA) to establish our measurement model, essentially duplicating the structure of Tsukahara et al. (2020). We then extracted Bartlett factor scores from this model for use in subsequent trial-level analyses.

After establishing the measurement model, we next specified a structural model to test the relative, unique prediction of AC and TEPR and for Auditory Discrimination ability. This allowed us to test whether individuals who exhibit greater pupil dilation (larger TEPR) tend to have better discrimination ability only via their superior AC or also via a direct physiological indicator of effort exertion.

Latent variable analyses were conducted using the *lavaan* R package (Rosseel, 2012). Models that were fit were evaluated with robust maximum-likelihood estimation. All latent factors were scaled by fixing one indicator loading or factor variance, and covariances among factors were freely estimated. Fit was assessed according to multiple robust fit indices: comparative fit index (CFI), Tucker–Lewis index (TLI), root-mean-square error of approximation (RMSEA),

standardized root-mean-square residual, and chi-square goodness-of-fit.

Trial-Level Modeling

Next, to capture within-subject (state-level) dynamics and performance on trials of varying difficulty, we conducted multilevel modeling. We modeled trial-by-trial accuracy in the discrimination tasks as a function of momentary pupil dilation and task difficulty while allowing these effects to differ by individual. Specifically, for each separate discrimination task, we fit a generalized linear mixed-effects model (using *lme4* in R) with a binomial logit link for accuracy (correct = 1, incorrect = 0). At Level 1 (within-person), difficulty (the trial's stimulus difference) and TEPR (that trial's peak pupil dilation) were included as continuous predictors of the log-odds of a correct response. Both predictors were person-mean centered to focus on trial-level fluctuations. Each model included random intercepts for participants as well as random slopes for difficulty and TEPR, allowing individuals to vary in overall accuracy and in how strongly difficulty or effort impacted their performance. At Level 2 (between-person), each individual AC factor score was included as a predictor of these random intercepts and slopes. In other words, AC was allowed to moderate the effects of difficulty and TEPR on accuracy (cross-level interactions). This multilevel design tests whether high-AC individuals maintain better performance on difficult, high-effort trials compared to low-AC individuals. Model parameters were estimated via restricted maximum likelihood; effects of predictors were reported as odds ratios with corresponding 95% confidence intervals for interpretability.

In a complementary set of multilevel models, we also treated TEPR as an outcome to examine how task difficulty and performance affect pupil dilation. These linear mixed-effects models predicted trial-level pupil dilation (a continuous outcome) from trial difficulty and accuracy (correct vs. incorrect), with AC included as a Level 2 moderator. This analysis served to confirm that pupil responses indeed increase with task demands and to explore whether high-AC individuals exhibit different pupillary patterns under challenge. All multilevel analyses used an $\alpha = .05$ significance criterion, and where relevant, we report the intraclass correlation (ICC) to quantify the proportion of variance between versus within individuals. Additional analyses on reaction time, as well as full model outputs and supplementary materials, are provided in the accompanying attachment (see Figures A3 and A4).

Transparency and Openness

The present analyses were conducted on a data set originally collected and reported in Tsukahara et al. (2020); behavioral discrimination thresholds and cognitive ability data were previously reported, whereas the pupillometric analyses presented here are novel.

In accordance with the seven Transparency and Openness Promotion guidelines: (a) Citation standards: All data, materials, and software used are cited in the text and reference list. (b) Data transparency: De-identified participant data are publicly available on the Open Science Framework (Pillai et al., 2026; <https://doi.org/10.17605/OSF.IO/F38KS>). (c) Analytic code transparency: All R scripts used to produce the reported analyses are publicly available at the same Open Science Framework repository. (d) Research materials

transparency: Full task descriptions, stimulus parameters, and presentation procedures are provided in the Method section. (e) Design and analysis transparency: All model specifications, random-effects structures, and fit criteria are reported in the Method section and Appendix. (f) Preregistration of studies: No aspects of the study were preregistered. (g) Preregistration of analysis plans: The present analyses were not preregistered. Analyses were conducted in R using the *lme4*, *lavaan*, and *sjPlot* packages.

Results

Confirmatory Factor Analysis

All variables were approximately normally distributed (skew and kurtosis within acceptable bounds; see Table A1 for all descriptives in Appendix). Before estimating the structural paths, we conducted a CFA to evaluate the measurement model. The CFA model specified five latent variables: TEPR, Auditory Threshold, AC, WMC, and Gf. Each latent factor was indicated by its respective three manifest variables. No cross-loadings were allowed. Factor variances were fixed to 1 for model identification, and covariances among all latent variables were freely estimated. The measurement model achieved an acceptable fit to the data, $\chi^2(80) = 104.62, p = .033$; CFI = 0.96; TLI = 0.94; RMSEA = 0.066. All observed indicators loaded significantly on their intended latent factors (standardized loadings ranged from 0.52 to 0.87), consistent with the factorial structure derived from Tsukahara et al. (2020). In particular, the three auditory discrimination task scores loaded strongly on a common Auditory Discrimination factor (loadings 0.61–0.70), confirming that they tap a shared ability. The three pupil dilation measures (average TEPR in each task) likewise defined a coherent latent TEPR factor (loadings 0.55–0.72), indicating stable individual differences in task-evoked pupillary response across tasks. The latent AC factor was well-defined by its cognitive task indicators (loadings 0.60–0.69). Finally, standardized loadings ranged from .63 to .81 for WMC and from .71 to .77 for Gf. Much of these results are redundant with those reported by Tsukahara et al. One novelty is that TEPRs were significantly correlated with each of the other latent variables, such that those who exhibited larger TEPRs also tended to rank highly on the other latent variables (all $ps < .02$). However, while significant, these correlations tended to be small in magnitude, with absolute values averaging only $r = .21$ (Figure 2).

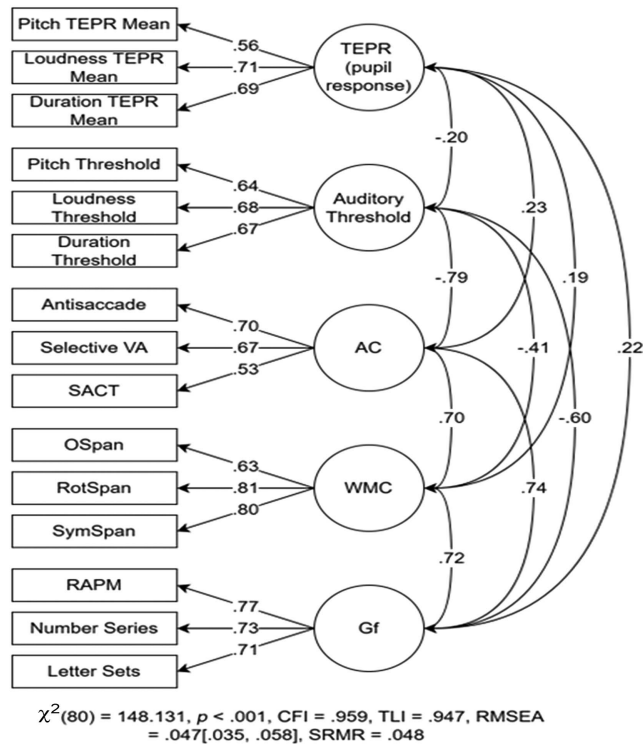
Structural Equation Model

Following our CFA, we specified a structural model to test whether AC and effort (TEPR) independently predict auditory discrimination ability. Regression paths were drawn from TEPR and AC to the Auditory Threshold factor, and the correlation between AC and TEPR was freely estimated (Figure 3).

As hypothesized, AC emerged as a key intermediary between physiological effort and performance (Figure 1). The structural equation model showed good fit, $\chi^2(24) = 61.24, p < .001$; CFI = 0.94; RMSEA = 0.063. Consistent with the earlier CFA, AC was a strong predictor of auditory discrimination ability; the path from AC to the Auditory Discrimination factor was significantly negative ($\beta = -0.80, p < .001$), indicating that higher AC is associated with lower (better) discrimination thresholds. Importantly, the path from TEPR to Auditory Discrimination was near zero and nonsignificant

Figure 2

Confirmatory Factor Analysis Relating TEPR, Auditory Discrimination Thresholds, and Cognitive Ability Constructs



Note. Circles represent latent variables and rectangles represent observed indicators. Auditory Threshold reflects participants' discrimination thresholds across the same three auditory domains. AC was defined by Antisaccade, Selective VA, and the SACT. WMC was measured by Operation Span, Rotation Span, and Symmetry Span tasks. Gf was indicated by RAPM, Number Series, and Letter Sets. VA = Visual Arrays; RAPM = Raven's Advanced Progressive Matrices; AC = attentional control; TEPR = task-evoked pupillary responses; SACT = Sustained Attention to Cue Task; WMC = working memory capacity; Gf = fluid intelligence; OSpan = advanced operation span; RotSpan = advanced rotation span; SymSpan = advanced symmetry span; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual.

($\beta = -0.03, n.s.$). Thus, there was no evidence that individual differences in pupil dilation translated directly into better or worse perceptual thresholds except via AC.

These results support a full mediation model, as individuals who exert higher effort (as indexed by larger pupils) achieve superior auditory discrimination performance by virtue of their stronger AC. In summary, at the latent trait level, task-evoked pupillary response is positively linked to attentional control, and attentional control in turn underpins better sensory discrimination. This mirrors the findings of Tsukahara et al. (2020), who reported that AC bridges the association between basic sensory discrimination and higher order cognition. While the structural equation modeling approach confirms broad trait-level relationships, it does not reveal how momentary fluctuations in effort relate to performance on a trial-by-trial basis or how these effects depend on task demands. We therefore turned to multilevel modeling to examine within-person dynamics across varying difficulties.

Multilevel (Trial-Level) Results

To examine trial-by-trial predictors of auditory discrimination performance across tasks, we fit a multilevel model with trials nested within subjects and task (pitch, duration, loudness) included as a within-subject factor. For robustness and interpretive context, we also ran separate models for each task individually. These single-task models, reported in Appendix, produced consistent results that reinforce the generalizability of the nested findings.

TEPR Predicted by Difficulty, AC, and Their Interaction

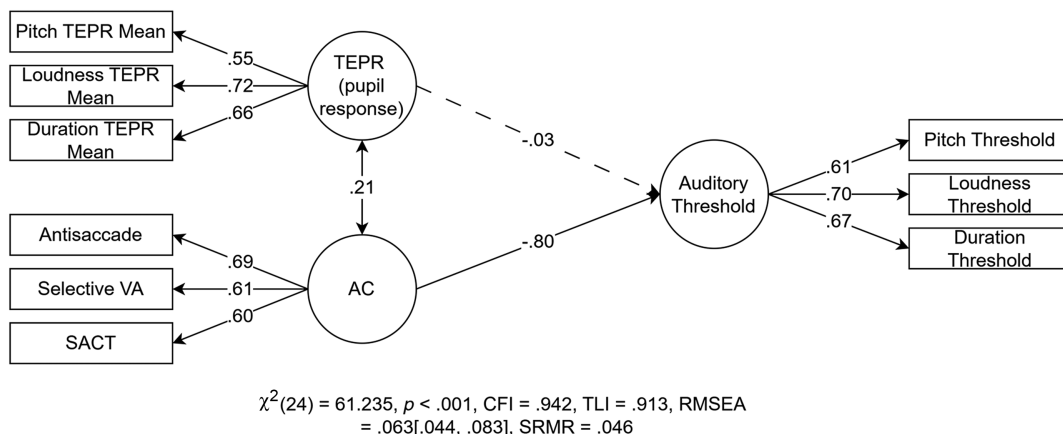
We first estimated an unconditional random-effects model with TEPR as the outcome to quantify the proportion of variance attributable to between- versus within-person differences. The ICC was approximately 0.08, indicating that while pupil dilation varied considerably from trial to trial, there was modest stable between-subject variance in TEPR. To examine how pupil dilation reflects task-evoked effort, we modeled TEPR as the outcome of trial-level Difficulty, individual differences in AC, and their interaction. As expected, more difficult trials evoked significantly larger pupil responses across tasks ($p < .001$). Crucially, AC moderated this effect: Participants higher in AC showed steeper increases in pupil dilation as difficulty increased. This Difficulty $\times$ AC interaction was significant in all three tasks (pitch: $b = 0.021, p < .001$; duration: $b = 0.018, p < .001$; loudness: $b = 0.012, p = .004$), suggesting that high-AC individuals mobilized greater physiological effort when task demands were high (Figure 4).

Task-Specific Moderation of Pupil Dilation

To determine whether the interaction between trial difficulty and AC on pupil dilation varied across tasks, we compared a model with a full three-way interaction (Difficulty $\times$ AC $\times$ Task) to a model without it. The likelihood ratio test showed a (marginally) non-significant improvement in fit when including the three-way interaction, $\chi^2(2) = 5.96, p = .051$. This suggests that the combined influence of difficulty and AC on TEPR may differ slightly by task, though the overall interaction pattern held across tasks. In line with this, follow-up models showed that the Difficulty $\times$ AC interaction remained significant across pitch, duration, and loudness, but the slope was steeper in pitch and duration, indicating more effort mobilization by high-AC individuals under increasing difficulty in those tasks (for overall effects of task on accuracy and TEPR, see Figure A1; see Figure 5).

TEPR Predicted by Accuracy

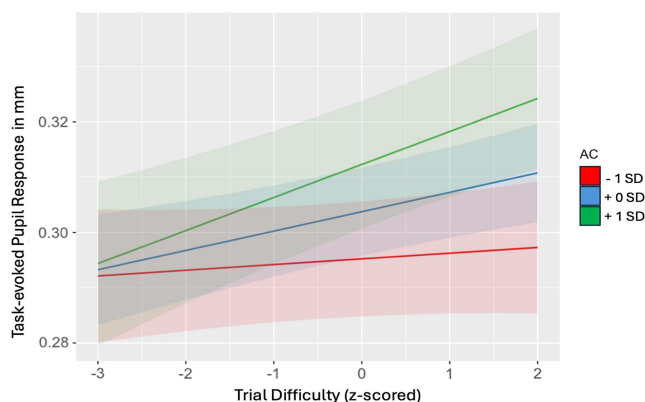
Next, we assessed whether TEPR varied as a function of trial outcome (correct vs. incorrect). Errors were associated with significantly larger pupil dilations in the pitch ($b = -0.017, p < .001$) and duration tasks ($b = -0.013, p = .002$), but not in loudness ($b = -0.006, p = .11$). This pattern suggests that trials requiring greater cognitive effort were more likely to result in errors, particularly in the more demanding discrimination conditions. In other words, increased pupil dilation appeared to index moments in which participants were exerting greater effort to resolve difficult perceptual judgments but were not always successful in doing so. This interpretation is consistent with the idea that TEPR reflects effort

Figure 3*Mediational Equation Model Relating TEPR, AC, and Auditory Discrimination Thresholds*

Note. This model tests whether AC accounts for the association between TEPR and auditory discrimination thresholds. Higher TEPR reflects greater task-evoked pupil dilation across pitch, loudness, and duration discrimination tasks. Auditory Threshold reflects perceptual discrimination thresholds across the same auditory domains. Standardized estimates are shown. The dashed path indicates a nonsignificant direct association between TEPR and auditory threshold when AC is included in the model. TEPR = task-evoked pupil responses; SACT = Sustained Attention to Cue Task; AC = attention control; VA = Visual Arrays; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual.

mobilization under challenging conditions rather than successful processing per se (Figure A2).

We then estimated an unconditional logistic model for trial accuracy. The ICC was approximately 0.06 overall, indicating that most variance in performance occurred within individuals across trials rather than between participants. In single-task models, ICC values were ≈ 0.15 for pitch and ≈ 0.01 for duration and loudness,

Figure 4*Predicted Task-Evoked Pupil Response as a Function of z-Scored Trial Difficulty and AC*

Note. Lines represent model-predicted values from the multilevel model at low (-1 SD), mean, and high ($+1$ SD) levels of AC. Trial difficulty was z-scored across all auditory discrimination tasks prior to analysis, such that higher values indicate relatively more difficult trials across pitch, loudness, and duration conditions. Shaded bands represent 95% confidence intervals around the predicted values. AC = attention control. See the online article for the color version of this figure.

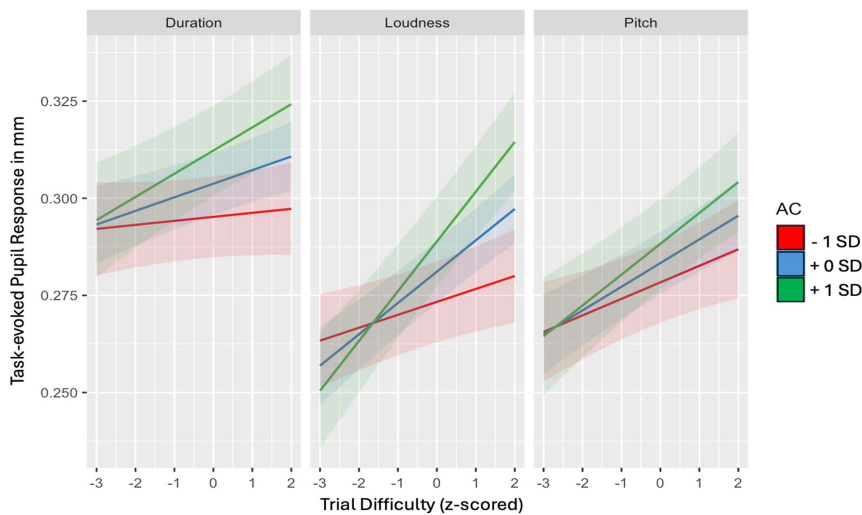
supporting the use of trial-level predictors in subsequent models and indicating that fluctuations in task difficulty meaningfully contributed to trial-to-trial variation in performance.

Difficulty had a robust negative effect on accuracy across all tasks ($\beta = -0.78, OR = 0.46, 95\% CI [0.45, 0.47], p < .001$), indicating that performance declined systematically as trials became more difficult. AC positively predicted performance ($\beta = 0.42, OR = 1.52, 95\% CI [1.38, 1.68], p < .001$), suggesting that individuals higher in AC were generally more accurate across tasks. A main effect of task further indicated that pitch discrimination was more difficult than duration ($OR = 0.59, 95\% CI [0.54, 0.65], p < .001$), whereas loudness did not differ significantly from duration ($OR = 1.04, p = .17$). Together, these results confirm that both trial-level difficulty and individual differences in AC played important roles in shaping performance in the auditory discrimination tasks (see Figure 6).

Effect of Pupil Dilation (TEPR) on Accuracy

Trial-level pupil dilation was negatively associated with accuracy ($\beta = -0.15, OR = 0.86, 95\% CI [0.79, 0.93], p = .003$). Larger dilations were more likely to occur on trials that were ultimately answered incorrectly, suggesting that TEPR captured effort under strain but not necessarily success. The relationship between both difficulty and TEPR with performance was significantly moderated by AC. The $AC \times Difficulty$ interaction ($\beta = -0.32, OR = 0.73, 95\% CI [0.72, 0.74], p < .001$) indicated that high-AC individuals experienced smaller drops in performance as task difficulty increased. The $AC \times TEPR$ interaction ($\beta = -0.20, OR = 0.82, 95\% CI [0.76, 0.89], p < .001$) showed that the detrimental effect of high pupil dilation on accuracy was reduced with higher AC (Figure 7).

Figure 5
Task-Specific Predicted Accuracy as a Function of Trial Difficulty and AC



Note. Lines represent model predictions at low (-1 SD), mean, and high ($+1$ SD) levels of AC. Trial difficulty was z-scored across all auditory discrimination tasks prior to analysis. Shaded bands represent 95% confidence intervals around the predicted values. AC = attention control. See the online article for the color version of this figure.

Discussion

The present findings underline the central function of AC in linking mental effort to perceptual performance.¹ At the latent trait level, we replicated and extended prior work by Tsukahara et al. (2020), demonstrating that AC is strongly associated with auditory discrimination ability. Specifically, individuals' task-evoked pupillary responses showed significant but modest positive correlations with all other latent variables (AC, auditory discrimination, WMC, Gf; mean $|r| \approx .21$). However, in our structural equation model, AC emerged as fully accounting for the relationship between task-evoked pupil response and auditory discrimination performance. TEPR had no direct link to auditory discrimination performance when AC was considered. In our mediation model, the path from the latent TEPR factor to auditory discrimination was near zero and nonsignificant, whereas AC had a large beneficial effect on performance. In other words, people with larger pupil responses tended to perform better on auditory discrimination only if they also scored high in AC. This pattern suggests that AC is the central function through which mental effort translates into successful performance. Those with greater ability to focus and sustain goal-directed attention not only tend to have better sensory acuity but also exhibit heightened effort across tasks, which further improves accuracy.

At the trial-by-trial level, our multilevel analyses further illustrate how moment-to-moment effort allocation and difficulty interact with individual differences in AC to determine performance. As hypothesized, increasing task difficulty elicited larger pupillary dilations on average, confirming that the TEPR tracked trial-level mental effort or arousal in response to challenge. Crucially, this effect was moderated by AC: Participants high in AC showed steeper increases in pupil size as trials became more difficult, whereas those lower in AC showed a blunted increase in pupillary

response to rising demands. Additionally, higher difficulty greatly reduced accuracy overall. However, high-AC individuals were significantly more resilient to this decline, as their performance dropped less sharply as difficulty increased compared to low-AC individuals. We also found that trials with larger TEPRs were somewhat more likely to end in an incorrect answer. This negative association between pupil dilation and accuracy was especially apparent in the more challenging tasks (pitch and duration discrimination), where errors tended to be accompanied by larger dilations. Crucially, AC moderated this pupil–performance link, as the negative effect of large dilations on accuracy was significantly attenuated among high-AC individuals. In summary, although larger task-evoked pupil dilations generally signaled high-effort and high-difficulty trials that often led to errors, those participants with superior AC were better able to maintain accuracy even on trials that evoked strong pupillary responses. Hence, AC provides individuals with a greater capacity to convert effort into successful performance, especially under taxing conditions.

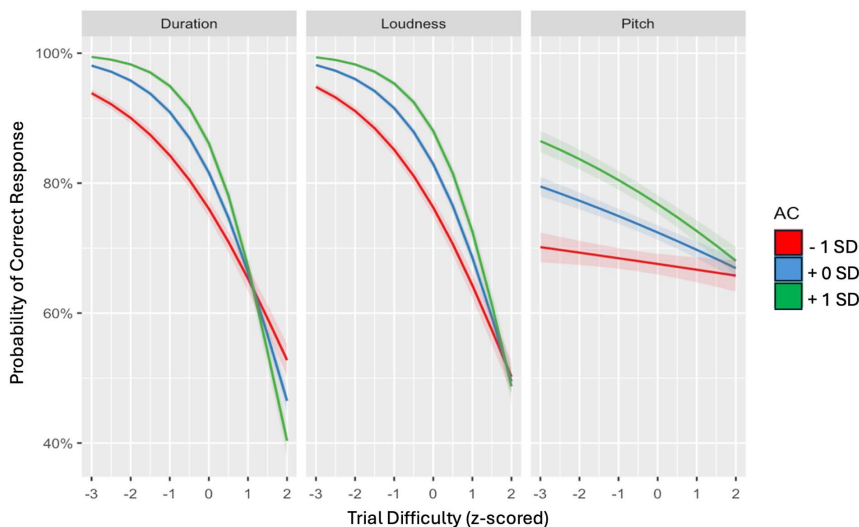
Effort as Compensatory Versus Effort as Boost: The Role of AC

Our results speak to a fundamental question about mental effort: When does increased effort act as a helpful boost to performance, and when is it merely a sign of being overtaxed (a compensatory response)? The present data suggest that the answer hinges on AC ability. For individuals with lower AC, a large pupil dilation on a

¹ Although the primary analyses focus on pupil dilation, models using reaction time as the outcome produced patterns consistent with the pupil results, indicating that high-AC individuals not only showed greater dilations and better accuracy but also responded more quickly (see Appendix).

Figure 6

Task-Specific Predicted Task-Evoked Pupil Responses as a Function of Trial Difficulty and AC



Note. Predicted values from the multilevel model are shown separately for duration, loudness, and pitch discrimination tasks. Lines represent model predictions at low (-1 SD), mean, and high ($+1$ SD) levels of AC. Trial difficulty was z-scored across all auditory discrimination tasks prior to analysis. Shaded bands represent 95% confidence intervals around the predicted values. AC = attention control. See the online article for the color version of this figure.

difficult trial seems to reflect compensatory attempt to meet task demands that often proves insufficient. These individuals show relatively smaller increases in effort as difficulty rises, and when they do exhibit pronounced dilations, it is frequently on trials they still get wrong. In essence, their effort spikes may indicate that they are struggling or approaching the limits of their capacity. Conversely, for high-AC individuals, increased effort more often functions as a performance-boosting mechanism compared to low-AC. High-AC participants systematically ramp up their effort in response to greater difficulty (as evidenced by their steeper Difficulty $\times$ AC pupillary slopes), and this mobilization of effort helps them preserve accuracy on those very trials. In their case, large TEPRs are less tightly coupled with failure. Instead, they likely reflect effective recruitment of cognitive resources that enables superior task engagement and goal maintenance under high demand. This pattern of findings aligns with theories of attention resource allocation which propose that having greater control allows one to deploy attentional resources more flexibly and efficiently when challenges arise. It also coincides with our results that, at an interindividual level, greater phasic pupillary responses can correlate with better performance if accompanied by strong AC. Thus, AC appears to be the critical factor that determines whether high effort yields dividends (boosting performance) or simply signifies a difficulty-induced strain (compensatory effort).

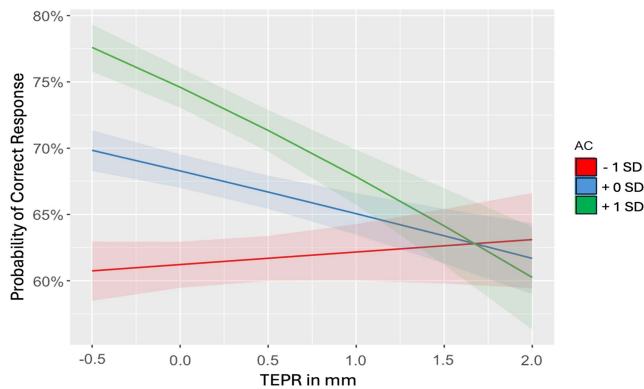
It is worth noting that these effects were observed across three different auditory discrimination tasks (pitch, duration, loudness), suggesting a domain-general pattern. That said, we found some nuance in how strongly effort was linked to performance in each task. The moderating influence of AC on the difficulty–pupil relationship was somewhat larger in the pitch and duration tasks than in loudness discrimination, and pupil-linked performance decrements were most evident in the pitch task. Pitch discrimination was overall the most

difficult task for our sample, which may have amplified the contribution of AC. Participants likely needed maximal focus and thus greater effort to make fine pitch judgments. Loudness discrimination, in contrast, might have been constrained more by basic sensory sensitivity (and perhaps approached ceiling for some listeners), leaving less room for upregulation of attentional resources to make a difference (Jestadt et al., 1977; Schneider & Parker, 1990). Although we must be cautious in overinterpreting task differences, these patterns reinforce the notion that attentional control and effort investment may play a larger performance-enhancing role in tasks that heavily tax cognitive resources, as opposed to tasks that are either easier or limited by raw sensory capacity. Future research could explore this by systematically varying task demand and sensory modality to see when boosting effort via AC most improves outcomes.

Our finding that pupil dilations were larger on error trials, especially for the high-demand pitch and duration tasks, is also significant. Intuitively, this result fits with the idea that near-threshold, high-effort trials often reflect moments of struggle or uncertainty that sometimes end in incorrect responses. This could indicate that when participants are pushed to their perceptual limits, they exert more effort (hence bigger pupil dilations) but may still fall short. Interestingly, this effect was absent in the easier loudness task, reinforcing the notion that pupil-linked effort is most pronounced as individuals approach the boundary of their abilities. Importantly, however, high-AC individuals were better at avoiding errors despite those high-effort trials. Another interpretation is that high AC enables more precise or timely allocation of effort exactly when and where it is needed during a trial, for example, by sustaining focus on the critical stimulus features or by quickly engaging strategies to parse difficult discriminations. Lower AC individuals might also feel the surge of arousal on hard trials (hence some dilation), but perhaps this effort comes either too

Figure 7

Predicted Probability of Correct Responses as a Function of TEPR and AC



Note. Lines show predicted probabilities of correct responses from the logistic mixed-effects model at low (-1 SD), mean, and high ($+1$ SD) levels of AC. Higher TEPR were associated with lower accuracy overall, but this relationship varied as a function of AC. Specifically, individuals with higher AC showed a steeper decline in performance as TEPR increased, whereas individuals with lower AC showed relatively little change in performance across levels of TEPR. Shaded bands represent 95% confidence intervals around the predicted probabilities. AC = attention control; TEPR = task-evoked pupil responses. See the online article for the color version of this figure.

late, too diffusely, or cannot be effectively translated into the fine attentional operations required, resulting in an error despite the high arousal. In practical terms, mental effort without guidance from AC may amount to “throwing energy at the problem” inefficiently, whereas effort guided by strong control is more like a laser-focused boost to performance.

Integration With Prior Research

By highlighting the mediating role of AC between effort and performance, our study adds a new trial-by-trial lens to the literature on how basic sensory discrimination skills connect to broader cognitive abilities. Tsukahara et al. (2020) characterized AC as “the missing link” between sensory discrimination and intelligence. We not only replicate their core finding that attentional control underpins individual differences in sensory acuity, but we also extend it by introducing a physiological measure of effort into the picture. To our knowledge, this is the first evidence that individuals who reliably exhibit larger task-evoked pupil dilations are those who also tend to score higher on AC, working memory, and fluid intelligence. This suggests that the propensity to invest mental effort in challenging tasks is itself a stable trait that aligns with cognitive capacity. However, our structural models indicate that simply exerting more effort (as indexed by pupil size) does not guarantee better perceptual performance unless one has the attentional control to direct that effort effectively. In a sense, AC is the moderator at both the trait and state levels: It governs the long-term correlation between effortfulness and ability, and it governs the short-term efficacy of effort on a trial-by-trial basis.

Our findings can be framed in the context of classic and contemporary views of cognitive energetics. Kahneman’s (1973) capacity

model of attention proposed effort as a limited resource that can be allocated to demanding tasks, and individuals differ in their available capacity or their threshold for investing effort. More recent theories, such as the adaptive gain theory of the LC–NE system (Aston-Jones & Cohen, 2005), suggest that the brain’s arousal system modulates the balance between exploiting current goals and exploring alternatives. As task-evoked pupil dilations are often interpreted as a readout of LC–NE activity, reflecting a surge in arousal or effort mobilization in response to task demands, our results imply that high AC individuals may possess a more responsive or optimally tuned arousal system: When faced with difficulty, they exhibit a larger phasic arousal response (larger pupil), which in turn is harnessed to sustain performance (consistent with an adaptive gain mechanism). Low-AC individuals, on the other hand, either fail to sufficiently engage the arousal system in time (smaller pupil response to difficulty) or the arousal they do muster does not translate into improved focus, potentially reflecting a less efficient coupling between the LC–NE surge and the cognitive control of attention.

Limitations and Future Directions

A key limitation of the present study is the use of pupillometry as the sole index of mental effort. Pupil dilation provides only a coarse, slow-varying measure of arousal and cannot disentangle the precise temporal stages of processing during each trial. Our findings leave open questions about when during the decision-making process attentional control and effort exert their influence. Does the critical benefit of effortful focus occur at early sensory encoding, during the comparison and decision phase, or perhaps in monitoring and adjusting responses? To answer such questions, future studies would benefit from integrating methods with high temporal resolution, such as electroencephalography, alongside pupillometry. For instance, recording auditory event-related potentials in tandem with pupil responses could allow us to map effort-related changes onto specific processing stages.

Although our analysis revealed that incorrect trials were associated with larger pupil dilations, this relationship may be confounded by the temporal dynamics of the pupil response relative to behavioral outcomes. Task-evoked pupil dilation is a slow signal with a latency of approximately 200–500 ms after cognitive events, and it typically peaks 1- to 2-s poststimulus (Beatty & Lucero-Wagoner, 2000; Mathôt et al., 2018). In many cases, this latency overlaps with or follows the behavioral response, making it problematic to treat accuracy as a predictor of pupil size. In other words, a post hoc variable like accuracy may not temporally precede the pupil response it is purported to predict. This undermines causal interpretation and raises the possibility that increased pupil dilation observed on error trials may instead reflect increased prerresponse uncertainty or pretrial vigilance state, not the error per se (de Gee et al., 2014; van Kempen et al., 2019). Indeed, prior work shows that pupil dilation scales with the subjective difficulty of a decision and the level of response conflict, regardless of whether the outcome is correct (Einhäuser et al., 2010; Urai et al., 2017). Thus, while errors are associated with higher TEPRs, this may not reflect posterror arousal or monitoring but rather effort mobilization or uncertainty prior to response, which coincidentally co-occurs more often on trials that end in error.

One promising approach to address this temporal confound would be to examine the auditory N1 and P300 components as neural markers

of, respectively, early perceptual encoding and later decision-related processing. The N1 is a frontocentral waveform that peaks around 100 ms after an auditory stimulus and is sensitive to attention and discrimination demands (Näätänen & Picton, 1987). In difficult auditory tasks, a larger N1 might indicate that a listener is successfully applying attentional gain to boost the representation of critical stimulus features (e.g., the subtle difference in pitch or duration between two tones). If high-AC individuals indeed deploy more efficient early attention, we would predict they might show enhanced N1 amplitudes on high-difficulty trials, relative to low-AC individuals. Simultaneously measuring pupil size could reveal whether trials with greater pupil-linked effort also show enhanced N1, linking phasic arousal to heightened sensory gain. On the other hand, if effort is being applied more at the decision/evaluation stage, we would expect links with the P300 component. The P300 (particularly its P3b subcomponent) has traditionally been associated with processes like stimulus evaluation, context updating, and working memory updating (Donchin & Coles, 1988; Polich, 2007). In auditory discrimination tasks, one usually observes a P3b after the second stimulus when the listener evaluates the difference and makes a decision; the amplitude of P3b tends to increase with task relevance and perhaps subjective difficulty. Recent theoretical perspectives have reinterpreted P3b as a manifestation of the brain's evidence accumulation process rather than a discrete event marker. For example, Twomey et al. (2015) presented data that the P3b in decision tasks reaches a consistent peak just before a response and that its buildup rate varies with difficulty, mirroring the predictions of sequential sampling (drift-diffusion) models. In this view, a large P3 amplitude on a given trial might signify a prolonged or vigorous accumulation of evidence to reach a decision threshold. A compelling direction our lab is currently undertaking is to determine how pupil dilation dynamics align with the time course of decision-related neural activity and whether AC moderates these neural signatures in the same way it moderates behavioral performance and pupil indices. By concurrently measuring electroencephalography and pupillary responses, one could test whether high-AC individuals exhibit different N1/P3 patterns than low-AC individuals when facing difficulty. Such multimodal approaches would help disentangle whether the effort indexed by the pupil is allocated to perceptual enhancement, sustained effort mobilization, decision formation, or a combination of all.

Constraints on Generality

The present findings are based on a sample of healthy young adults and tasks involving simple auditory discrimination of pure tones. Consequently, the conclusions may not generalize to populations with hearing impairments, neurological conditions, or substantially different age distributions. Similarly, the auditory tasks used here involve sequential comparisons of relatively simple acoustic features and may not fully capture perceptual processing in more complex auditory environments such as speech perception in noise. Future research should examine whether the relationship between AC, task-evoked pupil responses, and perceptual discrimination extends to other sensory modalities, age groups, and real-world listening contexts.

Conclusion

The present findings demonstrate that AC plays a central role in linking cognitive effort to perceptual performance. Task-evoked pupil

dilation indexed moment-to-moment fluctuations in effort, but this effort translated into successful auditory discrimination primarily among individuals with stronger AC. At both the trait and trial levels, AC determined not only the degree to which effort was mobilized under increasing task demands but also whether that effort supported performance or reflected compensatory strain. These results support a hierarchical account in which AC functions as a regulatory mechanism that governs the relationship between arousal and cognition, refining classical resource models by emphasizing the controlled, rather than purely capacity-limited, nature of mental effort.

References

- Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, 14(4), Article 1. <https://doi.org/10.1167/14.4.1>
- Arnsten, A. F. T. (2015). Stress weakens prefrontal networks: Molecular insults to higher cognition. *Nature Neuroscience*, 18(10), 1376–1385. <https://doi.org/10.1038/nn.4087>
- Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive function in monkeys: Evidence for a hyperdopaminergic mechanism. *Archives of General Psychiatry*, 55(4), 362–368. <https://doi.org/10.1001/archpsyc.55.4.362>
- 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>
- Aston-Jones, G., & Waterhouse, B. (2016). Locus coeruleus: From global projection system to adaptive regulation of behavior. *Brain Research*, 1645, 75–78. <https://doi.org/10.1016/j.brainres.2016.03.001>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge University Press.
- 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. <https://doi.org/10.1037/0033-295X.108.3.624>
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press. <https://doi.org/10.7551/mitpress/1486.001.0001>
- Breton-Provencher, V., & Sur, M. (2019). Active control of arousal by a locus coeruleus GABAergic circuit. *Nature Neuroscience*, 22(2), 218–228. <https://doi.org/10.1038/s41593-018-0305-z>
- Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, 152(8), 2369–2402. <https://doi.org/10.1037/xge0001408>
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, 8(10), 465–471. <https://doi.org/10.1016/j.tics.2004.08.008>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Chandler, D. J., Gao, W.-J., & Waterhouse, B. D. (2014). Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cortices. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), 6816–6821. <https://doi.org/10.1073/pnas.1320827111>

- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5), 975–979. <https://doi.org/10.1121/1.1907229>
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, 62, 73–101. <https://doi.org/10.1146/annurev.psych.093008.100427>
- Cochrane, A., Simmering, V., & Green, C. S. (2019). Fluid intelligence is related to capacity in memory as well as attention: Evidence from middle childhood and adulthood. *PLOS ONE*, 14(8), Article e0221353. <https://doi.org/10.1371/journal.pone.0221353>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. <https://doi.org/10.3758/BF03196772>
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, 96(2), 341–370. <https://doi.org/10.1037/0033-2909.96.2.341>
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 24(4), 1158–1170. <https://doi.org/10.3758/s13423-016-1191-6>
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4), 497–523. <https://doi.org/10.1002/cne.902920402>
- da Silva Castanheira, K., LoParco, S., & Otto, A. R. (2021). Correction to: Task-evoked pupillary responses track effort exertion: Evidence from task-switching. *Cognitive, Affective, & Behavioral Neuroscience*, 21(1), 265. <https://doi.org/10.3758/s13415-020-00852-y>
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). Academic Press. <https://doi.org/10.1016/B978-012505626-7/50013-3>
- de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the United States of America*, 111(5), E618–E625. <https://doi.org/10.1073/pnas.1317557111>
- Deco, G., & Rolls, E. T. (2005). Attention, short-term memory, and action selection: A unifying theory. *Progress in Neurobiology*, 76(4), 236–256. <https://doi.org/10.1016/j.pneurobio.2005.08.004>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, 168(3939), 1604–1605. <https://doi.org/10.1126/science.168.3939.1604>
- Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(3), 357–427. <https://doi.org/10.1017/S0140525X00058027>
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, 150(2), 242–275. <https://doi.org/10.1037/xge0000783>
- Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48, 269–297. <https://doi.org/10.1146/annurev.psych.48.1.269>
- Einhäuser, W., Koch, C., & Carter, O. L. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*, 4, Article 18. <https://doi.org/10.3389/fnhum.2010.00018>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Evans, R. B., & Deary, I. J. (1994). Sensory discrimination and intelligence: Postmortem or resurrection? *American Journal of Psychology*, 107(1), 95–115. <https://doi.org/10.2307/1423292>
- Foot, S. L., Bloom, F. E., & Aston-Jones, G. (1983). Nucleus locus ceruleus: New evidence of anatomical and physiological specificity. *Physiological Reviews*, 63(3), 844–914. <https://doi.org/10.1152/physrev.1983.63.3.844>
- Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience*, 29(27), 8726–8733. <https://doi.org/10.1523/JNEUROSCI.2145-09.2009>
- Galton, F. (1883). *Inquiries into human faculty and its development*. Macmillan. <https://doi.org/10.1037/14178-000>
- Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11), 887–892. <https://doi.org/10.1038/nrn1538>
- Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3), 983–1012. <https://doi.org/10.1152/physrev.00026.2009>
- Harris, Z. S. (1952). Discourse analysis: A sample text. *Language*, 28(4), 474–494. <https://doi.org/10.2307/409683>
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60041-9](https://doi.org/10.1016/S0079-7421(08)60041-9)
- Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61–77). Sage Publications. <https://doi.org/10.4135/9781452233529.n5>
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611), 1190–1192. <https://doi.org/10.1126/science.143.3611.1190>
- Hogervorst, M. A., Brouwer, A.-M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, 8, Article 322. <https://doi.org/10.3389/fnins.2014.00322>
- Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41(1), 35–39. <https://doi.org/10.1037/h0061495>
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 92–104. <https://doi.org/10.1037/0096-1523.3.1.92>
- Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, 40(8), 875–892. <https://doi.org/10.1037/0003-066X.40.8.875>
- Joshi, S., Li, Y., Kalwani, R. M., & 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>
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227–229. <https://doi.org/10.3758/BF03214307>
- Kahneman, D. (1973). *Attention and effort*. Prentice Hall.
- Kahneman, D., & Beatty, J. (1967). Pupillary changes in two memory tasks. *Science*, 156(3773), 1180–1182. <https://doi.org/10.3758/BF03210302>
- Kahya, M., Gouskova, N. A., Lo, O. Y., Zhou, J., Cappon, D., Finnerty, E., Pascual-Leone, A., Lipsitz, L. A., Hausdorff, J. M., & Manor, B. (2022). Brain activity during dual-task standing in older adults. *Journal of Neuroengineering and Rehabilitation*, 19(1), Article 123. <https://doi.org/10.1186/s12984-022-01095-3>
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132(1), 47–70. <https://doi.org/10.1037/0096-3445.132.1.47>
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341. <https://doi.org/10.1146/annurev.neuro.23.1.315>
- Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, 51(3), 1336–1342. <https://doi.org/10.3758/s13428-018-1075-y>

- Lee, Y., & Engle, R. (2026). Beyond working memory capacity: Attention control as the underlying mechanism of cognitive abilities. *Journal of Intelligence*, 14(2), Article 22. <https://doi.org/10.3390/jintelligence14020022>
- Lorente, P., Ruuskanen, V., Mathôt, S., Crespo, A., & Radl, J. (2025). No evidence for association between pupil size and fluid intelligence among either children or adults. *Psychonomic Bulletin & Review*, 32(4), 1795–1802. <https://doi.org/10.3758/s13423-025-02644-2>
- Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2021). Individual differences in attention control: Implications for the relationship between working memory capacity and fluid intelligence. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: The state of the science* (pp. 175–207). Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0007>
- Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, 4(9), 877–886. <https://doi.org/10.1038/nn0901-877>
- Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, 50(1), 94–106. <https://doi.org/10.3758/s13428-017-1007-2>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.). Brill.
- 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>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>
- Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, 48(2), 162–175. <https://doi.org/10.1111/j.1469-8986.2010.01057.x>
- Norman, D. A., & Shallice, T. (1986). Attention to action. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1–18). Plenum Press. https://doi.org/10.1007/978-1-4757-0629-1_1
- O'Bryan, S. R., Price, M. M., Alquist, J. L., Javis, T., & Scolari, M. (2024). Changes in pupil size track self-control failure. *Experimental Brain Research*, 24, 1–13. <https://doi.org/10.1007/s00221-024-06781-3>
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. <https://doi.org/10.1037/0278-7393.28.3.411>
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2), 97–107. <https://doi.org/10.1038/nrn1603>
- Pillai, M. G., Mashburn, C. A., Tsukahara, J. S., Leroux, A., & Engle, R. W. (2026). *Task-evoked pupillary responses reveal a role for attentional intensity in auditory discrimination* [Data set]. Open Science Framework. <https://doi.org/10.17605/OSF.IO/F38KS>
- Plack, C. J. (2014). *The sense of hearing* (2nd ed.). Routledge.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42. <https://doi.org/10.1146/annurev.ne.13.030190.000325>
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6), 747–758. [https://doi.org/10.1016/S0028-3932\(97\)00015-8](https://doi.org/10.1016/S0028-3932(97)00015-8)
- Robison, M. K., & Campbell, S. (2023). Baseline pupil diameter does not correlate with fluid intelligence. *Psychonomic Bulletin & Review*, 35(5), 1988–2001. <https://doi.org/10.3758/s13423-023-02273-7>
- Robison, M. K., & Unsworth, N. (2019). Pupillometry tracks fluctuations in working memory performance. *Attention, Perception, & Psychophysics*, 81(2), 407–419. <https://doi.org/10.3758/s13414-018-1618-4>
- Robles, L., & Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiological Reviews*, 81(3), 1305–1352. <https://doi.org/10.1152/physrev.2001.81.3.1305>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function Part II: Physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current Neuropharmacology*, 6(3), 254–285. <https://doi.org/10.2174/157015908785777193>
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10(3), 211–223. <https://doi.org/10.1038/nrn2573>
- Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, 76(1), 130–141. <https://doi.org/10.1016/j.neuron.2012.09.011>
- Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews*, 35(2), 146–160. [https://doi.org/10.1016/S0165-0173\(01\)00044-3](https://doi.org/10.1016/S0165-0173(01)00044-3)
- Schneider, B., & Parker, S. (1990). Does stimulus context affect loudness or only loudness judgments? *Perception & Psychophysics*, 48(5), 409–418. <https://doi.org/10.3758/bf03211584>
- Semal, C., Demany, L., Ueda, K., & Hallé, P.-A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America*, 100(2), 1132–1140. <https://doi.org/10.1121/1.416298>
- Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science*, 249(4971), 892–895. <https://doi.org/10.1126/science.2392679>
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40, 99–124. <https://doi.org/10.1146/annurev-neuro-072116-031526>
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186. <https://doi.org/10.1016/j.tics.2008.02.003>
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology*, 15(2), 201–292. <https://doi.org/10.2307/1412107>
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4(4), 525–547. <https://doi.org/10.1177/0959354394044004>
- Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575), 902–904. <https://doi.org/10.1126/science.7134981>
- Tsukahara, J. S. (2022). *pupillometry: Preprocessing pupil size data* (R package) [Computer software]. <https://dr-jt.github.io/pupillometry/>
- Tsukahara, J. S., & Engle, R. W. (2022). *Sustaining the focus of attention and how it relates to performance in complex cognitive tasks*. PsyArXiv. <https://doi.org/10.31234/osf.io/wd5kz>
- Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, & Psychophysics*, 82(7), 3445–3478. <https://doi.org/10.3758/s13414-020-02044-9>

- Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology*, 91, 109–123. <https://doi.org/10.1016/j.cogpsych.2016.10.001>
- Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, 42(1), 1636–1643. <https://doi.org/10.1111/ejn.12936>
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, 71, 1–26. <https://doi.org/10.1016/j.cogpsych.2014.01.003>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/bf03192720>
- Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Are individual differences in attention control related to working memory capacity? A latent variable mega-analysis. *Journal of Experimental Psychology: General*, 150(7), 1332–1357. <https://doi.org/10.1037/xge0001000>
- Unsworth, N., & Robison, M. K. (2017). A locus coeruleus–norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review*, 24(4), 1282–1311. <https://doi.org/10.3758/s13423-016-1220-5>
- Unsworth, N., & Robison, M. K. (2018). Tracking working memory maintenance with pupillometry. *Attention, Perception, & Psychophysics*, 80(2), 461–484. <https://doi.org/10.3758/s13414-017-1455-x>
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1302–1321. <https://doi.org/10.1037/0278-7393.30.6.1302>
- Urai, A., Braun, A., & Donner, T. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, 8(1), Article 14637. <https://doi.org/10.1038/ncomms14637>
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. <https://doi.org/10.3758/s13423-018-1432-y>
- van Kempen, J., Loughnane, G. M., Newman, D. P., Kelly, S. P., Thiele, A., O'Connell, R. G., & Bellgrove, M. A. (2019). Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. *eLife*, 8, Article e42541. <https://doi.org/10.7554/eLife.42541>
- Voelke, A. E., & Roebbers, C. M. (2016). Sustained attention and its relationship to fluid intelligence and working memory in children. *Journal of Educational and Developmental Psychology*, 6(1), 131–139. <https://doi.org/10.5539/jedp.v6n1p131>
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438(7067), 500–503. <https://doi.org/10.1038/nature04171>
- von Békésy, G. (1960). *Experiments in hearing*. McGraw-Hill.
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, 56(2), 366–383. <https://doi.org/10.1016/j.neuron.2007.10.012>
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.), *A source book of gestalt psychology* (pp. 71–88). Routledge; Kegan Paul. <https://doi.org/10.1037/11496-005> (Original work published 1923)
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. <https://doi.org/10.1002/cne.920180503>

Appendix

Supplemental Scores and Figures Pertaining to Descriptives and Reaction Time Analyses

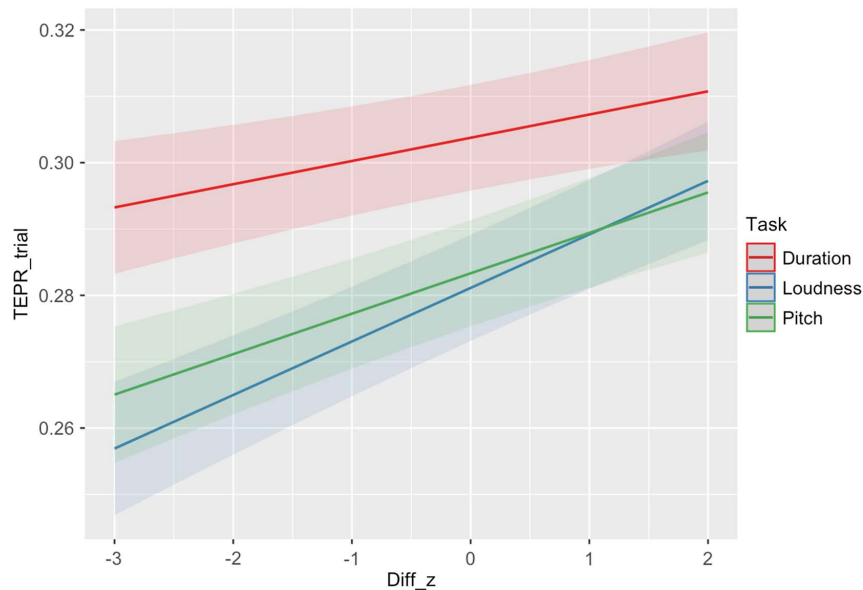
Table A1

Task Descriptives: Attention Control, Working Memory Capacity, and Fluid Intelligence

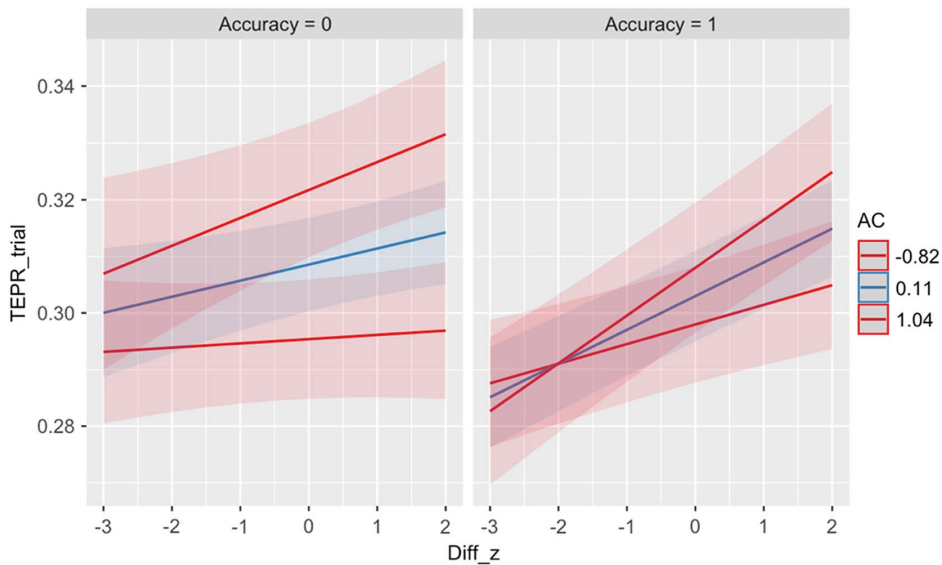
Construct: Task	<i>M</i> (<i>SD</i>)	Minimum–maximum	Skew	Kurtosis
Attention				
Antisaccade	0.8 (0.2)	0.5–0.9	–0.27	–1.50
SACT	0.6 (0.1)	0.4–0.8	–0.01	–0.89
VAorient_S	2.7 (0.6)	1.6–3.7	0.14	–0.10
WMC (ACC)				
OSpan	32.7 (6.6)	20.0–41.0	–0.38	–0.51
SymSpan	19.0 (4.3)	12.0–27.0	0.13	–0.20
RotSpan	15.8 (3.0)	12.0–22.0	0.87	0.36
Gf (ACC)				
RAPM	10.8 (3.5)	6.0–16.0	0.51	–1.30
Number series	11.1 (1.9)	8.0–14.0	0.15	–0.67
Letter sets	18.0 (4.2)	10.0–23.0	–0.99	0.26

Note. SACT = Sustained Attention to Cue Task; VAorient_S = Visual Arrays Orientation; WMC = working memory capacity; ACC = accuracy scores as indicators; OSpan = advanced operation span; RotSpan = advanced rotation span; SymSpan = advanced symmetry span; Gf = fluid intelligence; RAPM = Raven's Advanced Progressive Matrices.

(Appendix continues)

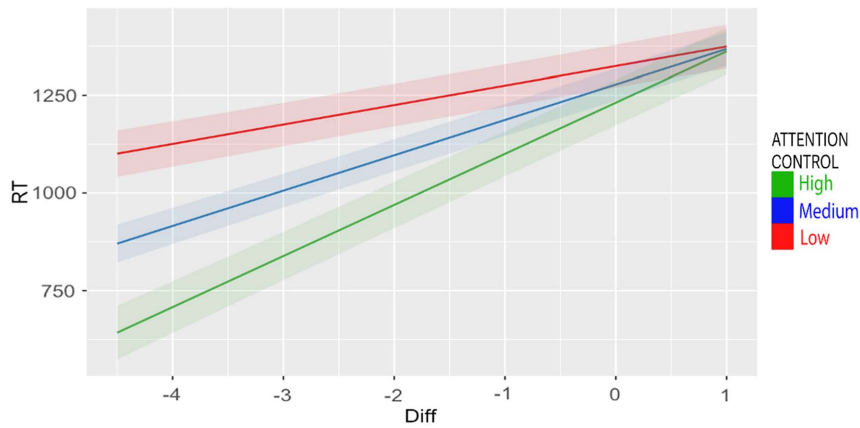
Figure A1*Task Effect on Evoked Pupil Response Moderated by Attention Control*

Note. TEPR = task-evoked pupil responses; Diff_z = z-scored difference between tones in Hertz, milliseconds, and decibels. See the online article for the color version of this figure.

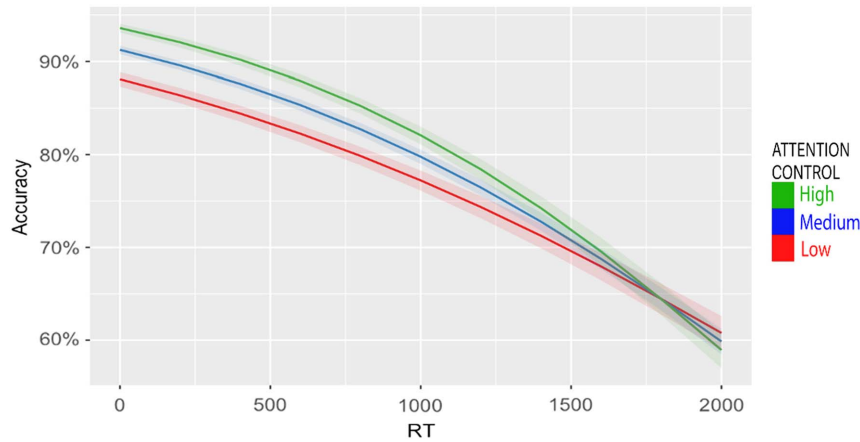
Figure A2*Effect of Accuracy on Evoked Pupil Response Moderated by AC*

Note. AC = attention control; TEPR = task-evoked pupil responses; Diff_z = z-scored difference between tones in Hertz, milliseconds, and decibels. See the online article for the color version of this figure.

(Appendix continues)

Figure A3*Effect of Difficulty on Reaction Time Moderated by Attention Control*

Note. Diff = difference between Tones in Hertz, decibels or milliseconds; RT = reaction time. See the online article for the color version of this figure.

Figure A4*Effect of Reaction Time on Evoked Accuracy Moderated by Attention Control*

Note. RT = reaction time. See the online article for the color version of this figure.

Received August 22, 2025
 Revision received May 28, 2026
 Accepted June 5, 2026 ■