

Response trajectories and autonomic reactivity during adaptive human–computer interaction

Maciej Lachowicz

maciej.lachowicz@ideis.pl

DSW Ideis University

Grzegorz Żurek

Wroclaw University of Health and Sport Sciences

Waldemar Andrzejewski

Wroclaw University of Health and Sport Sciences

Article

Keywords: human–computer interaction, adaptive tasks, response trajectories, autonomic reactivity, pulse rate variability, electrodermal activity

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11261977/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Adaptive computerized tasks alter subsequent response conditions, complicating the interpretation of slowing and errors. We examined autonomic reactivity and response trajectories in 31 participants completing an adaptive cognitive task; photoplethysmography-derived pulse rate variability (PRV) passed final quality control (QC) in 26. Heart-rate (HR) change correlated positively with reaction-time (RT) slope ($r = 0.452$, 95% CI [0.116, 0.695]), but the permutation result did not meet the five-test Holm criterion ($p = 0.070$). PRV indices RMSSD and SDNN decreased from baseline to task by 8.82 and 18.98 ms, respectively. No association between PRV change and RT slope was detected (both Holm-adjusted $p = 0.974$). In a separate eight-test family, greater RMSSD reduction was associated with higher accuracy ($r = -0.560$, 95% CI [-0.779, -0.221], Holm-adjusted $p = 0.029$); strict QC retained a similar estimate ($n = 22$, $r = -0.498$) without meeting the corrected threshold. For local slowing, both primary empirical-logit models had Holm-adjusted $p = 0.060$; grouped-binomial sensitivities indicated lower failure and omission odds (adjusted $p = 0.021$ and 0.029). PRV analyses were post hoc.

Introduction

Adaptive computer systems can modify the pace of incoming information and the support available during task execution. Understanding behavior in these environments requires attention to both the user's responses and the system adjustments that shape subsequent demands. Workload varies during goal-directed computer use, making the timing of support relevant to the ongoing interaction [1]. Slower responding may reflect difficulty, greater response caution, or adjustment to a changed presentation rate; distinguishing these possibilities matters when assessing whether users maintain effective performance. In performance-contingent tasks, a response also affects the conditions encountered next. In a Tetris experiment using electroencephalography (EEG) to guide adaptation, changing the thresholds used to trigger adjustments altered the frequency of pace changes and the difficulty reached by users [2]. A useful account of response slowing must therefore consider what the user is trying to preserve and how the adjustment rule changes the opportunity to do so.

Psychological accounts provide complementary reasons to examine response speed alongside response quality. Compensatory-control theory proposes that people can protect task achievement by adjusting effort and attention, potentially maintaining performance at increased cost [3]. Sequential-sampling models distinguish the quality of evidence supporting a decision from the amount of evidence required before responding; in two-choice tasks, these components can produce different combinations of response time and accuracy [4]. These accounts motivate examining whether slowing accompanies less successful responding or greater response accuracy. Under adaptive pacing, however, either pattern must also be interpreted in relation to changes in task conditions. Autonomic recordings offer an additional perspective on the demands experienced during computer use. Their interpretation nevertheless depends on the task and the user's developing proficiency: experimental multitasking research showed that learning altered the sensitivity of some physiological measures to task demand [5]. Electrodermal activity (EDA) can also reflect stress associated with time pressure and evaluation,

alongside cognitive demands [6]. Separating tonic skin conductance level (SCL) from phasic skin-conductance responses (SCRs) permits different components of this activity to be examined [7]. Relating these measures to response trajectories and response quality can help characterize physiological reactivity during interaction, while leaving its underlying psychological interpretation open.

Cardiac responses are particularly relevant to the possibility that continued performance carries a temporal cost. In an on-road experiment with 108 drivers, successive increases in working-memory demand produced incremental increases in heart rate and skin conductance, whereas driving-performance measures did not differentiate those levels as consistently [8]. This dissociation indicates that physiological changes can reveal differences in task demand that are not apparent from an aggregate performance score. In a separate driving-simulator study, increasing working-memory load raised heart rate and lengthened the time required to reach an instructed speed [9]. These were joint effects of experimentally varied load, providing a basis for examining cardiac activation alongside the time needed to respond. Mean HR does not capture beat-to-beat pulse-interval variability. The root mean square of successive differences (RMSSD) summarizes short-term successive interval variation, whereas the standard deviation of accepted pulse intervals (SDNN) summarizes their dispersion within the analyzed window. During cognitive work, reductions in vagally mediated cardiac variability can accompany increased task demand, but these indices do not provide a direct measure of sympathetic activation [10]. When intervals are derived from photoplethysmography rather than an electrocardiogram, they should be described as pulse rate variability (PRV). PPG-derived PRV and ECG-derived heart-rate variability are related but not interchangeable, and agreement can be less stable during cognitive stress or other conditions that alter peripheral hemodynamics [11]. PRV can therefore complement mean HR and EDA while requiring separate methodological interpretation.

Applying this evidence to an adaptive task requires considering change across the response sequence. Performance-contingent pacing can continue to challenge a user as successful responses permit further adjustments. Under a compensatory-control account, preserving task achievement in such conditions may require sustained mobilization while response selection becomes more time-consuming. If greater cardiac reactivity accompanies this demand for continued regulation, participants showing larger increases from baseline may also exhibit a stronger tendency toward slowing. This reasoning motivates a positive association between heart-rate reactivity and reaction-time slope. It extends demand-related findings to a between-person prediction about behavioral development, rather than assuming that a high mean reaction time alone indicates increasing difficulty. The relevant cardiac contrast is reactivity to task engagement, which differs from a progressive rise in heart rate during the task. Simulated-work research illustrates the distinction: in a driving task, heart rate initially increased relative to rest and subsequently declined during continued performance [12]. An elevated task average can therefore coexist with a decreasing physiological trajectory. Relating task-minus-baseline heart rate to ordered response timing addresses whether participants differ in their behavioral course according to overall cardiac reactivity. It does not require heart rate and response time to rise together at each moment.

The second prediction concerns how response quality changes when an individual responds more slowly than usual. Experimental manipulation of response priorities provides evidence for accuracy-oriented response regulation: across random-dot-motion and visual-discrimination tasks, accuracy-focused instructions produced slower and more accurate responses than speed-focused instructions [13]. This supports examining whether local slowing is accompanied by fewer unsuccessful responses, rather than interpreting slowing in isolation. The temporal constraint also matters. In random-dot decisions compared within a common response-time window, accuracy was lower with 800- or 1,000-ms deadlines than in free-response blocks; the shorter deadline also produced more missed responses [14]. Together, these findings link response quality to both the user's response policy and the time constraints imposed by the task.

For an adaptive response sequence, the predicted negative relation between local reaction time and failure rate follows if additional response time accompanies more careful selection under the conditions currently available. Extending the prediction to omissions requires a further consideration. Slower responses need not produce fewer omissions under an unchanged deadline; they could instead miss the available response window. Fewer omissions are plausible when a slower response regime accompanies a better match between the incoming pace and the user's processing capacity. The hypothesis thus concerns the local pattern of speed and quality emerging during adaptation. Comparing bins within the same person, while accounting for stimulus progress, separates that pattern from stable differences in average speed between participants.

Physiologically triggered assistance improved performance in a simulated uninhabited-air-vehicle task [15], and electrodermally guided changes in visual complexity improved working-memory performance in virtual reality [16]. In digital decision games, imposed pacing affected performance, while behavioral indicators showed limited correspondence with physiological measures [17]. These findings motivate examining how physiological and behavioral measurements relate when the task already adapts to the user's responses. Adaptation can also affect outcomes differently: EEG-based adjustment of information-presentation speed reduced perceived temporal demand without significantly improving learning gains, whereas adding a financial incentive improved gains relative to adaptation alone [18].

The need to examine these outcomes together is especially clear in a recent adaptive Stroop study. Performance-dependent changes in trial duration increased perceived workload and produced faster but less accurate responses than nonadaptive pacing [19]. Such condition comparisons establish that the pacing rule can alter the behavioral profile. They leave a complementary question about variation within an already adaptive setting: whether participants with greater autonomic reactivity develop different response trajectories, and whether slower portions of an individual's sequence are accompanied by better response quality. Addressing this question requires separating the two levels of association. A between-person relation between cardiac reactivity and slowing does not determine whether local slowing within a person is helpful, harmful, or unrelated to successful responding.

The research gap addressed here is therefore the joint characterization of physiological reactivity, behavioral change across the task, and local speed–quality covariation under performance-contingent pacing. Average speed can conceal contrasting early and late performance, while an overall accuracy score can conceal local changes in response success. Ordered trajectories make those changes available for analysis. Distinguishing gradual changes from sustained upward departures further clarifies whether slowing follows a broad trend or a more localized change in course. The physiological and behavioral summaries contribute different information to this assessment, without requiring them to identify a common underlying mechanism.

The aim of this study was to characterize how autonomic reactivity relates to response trajectories and how response timing covaries with response quality during performance-adaptive human–computer interaction. We distinguished between-person differences in task-related reactivity from within-person variation across the response sequence.

RQ1. Is greater cardiac reactivity associated with a stronger tendency toward response slowing across the adaptive task?

H1. Participants with greater task-minus-baseline increases in heart rate will exhibit more positive reaction-time slopes across normalized stimulus progress.

RQ2. Within participants, is locally slower responding associated with lower rates of unsuccessful responses?

H2. Bins with longer median reaction times relative to the participant’s own mean bin median will have lower failure and omission rates, after accounting for normalized stimulus progress.

Additional exploratory objectives were to examine associations of mean, tonic, and phasic electrodermal reactivity with response trajectories, and to assess whether individual trajectories showed sustained upward departures from a linear trend. A post hoc extension additionally quantified PPG-derived PRV using RMSSD as the primary PRV measure and SDNN as a complementary measure, and examined their task-related changes in relation to the same existing behavioral summaries. These objectives extend the analysis beyond average response speed while distinguishing gradual slowing from possible changes in its trajectory.

Methods

Participants and ethics

Participants were recruited on a rolling basis from the university student population during the testing period. Inclusion criteria were age 18–27 years, self-reported good general health, absence of contraindications to the study procedures, and written informed consent. The age restriction was

intended to limit age heterogeneity, given reported age-related differences in cognitive performance from early adulthood [20].

Exclusion criteria were neurological or psychiatric disorders, or other conditions that could affect the study procedure or physiological responses; use of medication or substances affecting autonomic function; caffeine or nicotine consumption within 4 h before testing; intensive physical exercise within 12 h before testing; and absence or previous withdrawal of consent. The study was approved by the Ethics Committee of Wroclaw University of Health and Sport Sciences (approval no. 46/2025, 28 January 2026; Supplementary File 1) and was conducted in accordance with the Declaration of Helsinki.

The analytical sample comprised 31 participants with ordered response records and HR/EDA measurements meeting the requirements for temporal analysis. All participants completed the same baseline and cognitive-task protocol. Records required identifiable stimulus positions and agreement with participant-level correct-response and omission counts. Inclusion did not depend on the direction or magnitude of the RT trajectory. Measure-specific availability and PRV QC determined the analysis sets shown in Fig. 1; participant-level PRV QC is reported in Supplementary File 2 (Table S10).

Inference emphasizes effect estimates and confidence intervals. As a sensitivity benchmark, with $N = 31$ a single two-sided Pearson correlation test against zero at $\alpha = 0.05$ has approximately 80% power to detect $|\rho| \approx 0.48$ under a bivariate-normal model. The study therefore had greater sensitivity to moderate-to-large between-participant associations than to smaller effects.

Participant characteristics and questionnaire responses are reported in Table 1. Age was calculated in completed years at questionnaire completion, and body mass index (BMI) was calculated as body mass in kilograms divided by height in metres squared. Pre-task ratings assessed sleep, sleepiness on the Karolinska Sleepiness Scale (KSS; [21]), fatigue, stress, and discomfort. Immediately after the task, participants completed the six NASA Task Load Index (NASA-TLX) subscales on 0–100 scales [22]. Higher ratings indicate greater demand or poorer perceived performance; performance anchors were 0 (perfect) and 100 (failure). The raw NASA-TLX score was the unweighted mean of all six subscales [23], and was available for every participant.

Procedure and adaptive computer task

Participants completed individual sessions in the same laboratory room. Sensors were attached and signal quality was checked before a 5-min seated baseline. Standardized instructions preceded the Vienna Test System Determination Test (DT; SCHUHFRIED GmbH, Mödling, Austria), S2 adaptive long form, lasting approximately 8 min. Recording markers separated baseline, instructions, and task; instruction and transition periods were excluded from phase summaries.

The DT presents successive visual and acoustic stimuli associated with predefined responses [24]. Visual stimuli appeared on a 17.3-inch LCD screen at 1920×1200 pixels, approximately 45 cm from the participant; auditory stimuli were delivered through speakers. Participants responded through the VTS

interface as quickly and accurately as possible. The VTS reports describe a performance-contingent rule under which incorrect or omitted responses lengthen the presentation times of subsequent stimuli during a fixed-duration test. Response quality can therefore alter subsequent response opportunities, and achieved stimulus count is an outcome of this feedback.

Physiological recording, PPG-derived PRV, and electrodermal processing

Electrodermal activity and raw photoplethysmography (PPG) were recorded at 128 Hz using a Shimmer3 GSR system and ConsensysPRO v2.0.23. HR was generated by the built-in PPG-to-HR algorithm. SC electrodes were placed on the proximal phalanges of the index and middle fingers of the non-dominant hand, and the optical pulse sensor on the left earlobe. The recording also contained inter-beat intervals and inertial channels.

Baseline and task HR and mean SC were taken from participant-level phase summaries; reactivity was task mean minus baseline mean. The exported HR interval means were weighted by their durations to check the phase means. Completeness checks for the 464 exported HR summaries are reported in Supplementary File 2 (Table S7).

PPG-derived PRV was calculated independently from the built-in HR summaries by re-detecting systolic pulse peaks in the raw 128-Hz PPG using an Elgendi-based peak-detection pipeline [25]. The default analysis windows were the last 300 s of baseline and the first 300 s of the cognitive task. A documented post hoc extension allowed an actual phase duration from 295 to 300 s when the phase itself was shorter; use of this allowance was flagged. Pulse intervals outside 300–2000 ms or deviating by more than 20% from the local median of 11 intervals were rejected. No interpolation was used.

A phase required at least 200 retained intervals, at least 190 pairs of genuinely adjacent retained intervals, and at least 80% temporal coverage. The original QC limit allowed at most 10% rejected intervals; a documented post hoc extension allowed at most 15% while leaving all interval filters unchanged. RMSSD was calculated only from accepted interval pairs that were adjacent in the original detected sequence, so rejected or missing intervals were not bridged. SDNN was the sample standard deviation of accepted pulse intervals. Primary PRV analyses required both phases to pass final QC. Membership of the strict-QC and designated borderline sensitivity sets is detailed in Fig. 1 and Supplementary File 2 (Table S10).

Tonic and phasic EDA measures were obtained through continuous signal processing. Non-finite values and SC outside 0.2–125 μ S were rejected. Adjacent-sample changes above 0.25 μ S were masked with 1-s margins, and constant runs lasting at least 10 s were excluded. Gaps longer than 1.5 times the median sampling interval separated segments; retained segments lasted at least 30 s. A fourth-order Butterworth low-pass filter at 2 Hz was applied bidirectionally, followed by resampling to 8 Hz within valid segments. Interpolation did not bridge rejected gaps.

Decomposition used the cvxEDA model [26], implemented with a sparse primal–dual solver in SciPy. Parameters were $\tau_0 = 2.0$ s, $\tau_1 = 0.7$ s, tonic spline knots every 10 s, $\alpha = 0.0008$, and $\gamma = 0.01$. Signals were standardized for fitting and restored to μS for quantification. Five-second margins at phase and valid-segment boundaries were excluded. Phase eligibility required confirmed boundaries, at least 80% retained coverage after phase margins, at least 120 s of analyzable data, and non-negative tonic estimates. Signal exclusions were documented explicitly [27].

Tonic SCL was the mean of retained tonic samples. SCR amplitude was the difference between a local phasic peak and its preceding local minimum within the same valid segment and phase. Accepted responses had amplitude ≥ 0.05 μS . SCR frequency was the accepted count divided by analyzable minutes; mean SCR amplitude was the sum of accepted amplitudes divided by their count. A phase without an accepted SCR has zero frequency but undefined mean amplitude. The tonic and phasic measures were checked against the archived sample masks, tonic traces, and event lists. Reactivity was task minus baseline for each of the five original HR/EDA measures and for each PRV metric.

Ordered responses and temporal outcomes

Participant identifiers linked physiological measurements, aggregate DT outcomes, and ordered response series. Individual RT values, stimulus order, and response classifications were obtained from the VTS test reports. RT values were read from native vector-plot coordinates and calibrated to each participant's printed RT scale. Missing RT values were not imputed. Normalized progress was $(\text{position} - 1)/(\text{total stimulus positions} - 1)$, ranging from 0 to 1. It represents relative task order rather than elapsed time.

The primary temporal outcome, termed RT slope, quantified the linear change in response latency across each participant's ordered response series. It was calculated by ordinary least squares, with individual available RT values as the dependent variable and normalized progress as the predictor. The fitted coefficient expresses the estimated change in milliseconds from progress 0 to 1: positive values indicate slowing, negative values indicate faster responding, and values near zero indicate little linear change. RT slope therefore describes change across relative stimulus order, rather than change per unit of elapsed time or the participant's average response latency.

For separate analyses of local trajectories, each ordered series was divided into 20 equal-width progress bins. Bin medians, median absolute deviations (MAD), and response counts described local behavior; the bin midpoint represented progress. These bin summaries were not used to calculate the participant-level RT slope. Additional temporal outcomes were late-third minus early-third median RT, MAD, failure rate, and omission rate.

Aggregate accuracy was correct responses divided by the sum of correct, incorrect, and omitted responses; aggregate omission proportion used the same denominator. At bin level, failures comprised incorrect plotted responses and omissions. The failure rate was their combined count divided by the

number of stimulus positions in the bin. Exported aggregate response counts and plotted stimulus positions have different counting units; consequently, bin failure rate is not interchangeable with one minus aggregate accuracy. Stimulus count and correct-response count were retained as distinct throughput measures.

Statistical analysis

Statistical analyses used jamovi (version 2.7.4) and custom Python scripts. The scripts implemented the permutation procedures, repeated-bin models, and bootstrap change-point tests described below. Continuous variables were summarized using means and standard deviations, with *t*-based confidence intervals for paired physiological changes. Pearson correlations related each physiological change to participant RT slope; individual 95% confidence intervals were calculated using Fisher's transformation. Two-sided participant-label permutation tests used 99,999 draws and the plus-one correction [28]. Holm adjustment [29] covered the five physiological correlations as one family. Spearman correlations provided a rank-based sensitivity check.

Separate repeated-bin models included participant fixed intercepts, normalized progress, and its interaction with each standardized physiological change. The interaction estimates the difference in RT change across full progress per 1 SD higher reactivity. Inference used participant-clustered covariance with the finite-sample correction and a *t* distribution with $G - 1$ degrees of freedom, where G is the number of participants [30]. Holm adjustment covered the five interaction tests. For permutation sensitivity analysis, a linear slope was first fitted to each participant's 20 bin medians; with a common progress grid and equal bin weights, regressing these slopes on standardized reactivity gives the same interaction coefficient as the repeated-bin model. Participant slope values were then randomly reassigned relative to the physiological values 99,999 times, and the absolute observed Pearson correlation was compared with its permutation distribution using the plus-one correction; the statistic was not divided by a robust standard error. The same five-test Holm adjustment was applied, and results from the two inferential procedures were reported separately.

Associations with aggregate accuracy, omission proportion, stimulus count, and correct-response count formed a 20-test family (five physiological predictors by four outcomes). The four late-minus-early temporal outcomes formed another 20-test family. Each family used permutation *p* values and Holm correction. Sensitivity models added the corresponding baseline physiological value, global median RT, accuracy, or stimulus count individually. These adjustments assess dependence on model specification.

PRV analyses were added post hoc and were therefore kept in separate multiplicity families rather than retroactively expanding the original five-predictor HR/EDA families. RMSSD was the primary PRV metric and SDNN was complementary. Baseline and task values and task-minus-baseline changes were summarized on the raw millisecond scale. Paired *t* tests used *t*-based 95% confidence intervals for mean change; Cohen's d_z and 20,000-resample percentile-bootstrap confidence intervals described paired effect size. Holm correction covered the two paired PRV tests. Distribution diagnostics were examined

for raw RMSSD and its paired change; lnRMSSD was analyzed as a sensitivity because logarithmic transformation reduced phase-value skew without changing the substantive inference.

For the primary PRV–trajectory analysis, Pearson correlations related RMSSD and SDNN change to the participant RT slope, using 99,999 participant-label permutations and a separate two-test Holm correction. The corresponding correlations with the 20-bin RT slope and Spearman coefficients were sensitivity analyses. The same four aggregate outcomes used for HR/EDA formed an eight-test PRV family (two PRV changes by four outcomes), and the four late-minus-early temporal outcomes formed a separate eight-test PRV family. Key results were repeated in the strict-QC and designated borderline sensitivity sets (Fig. 1). Descriptive HC3 regressions assessed whether the RMSSD–accuracy estimate changed after adjustment for baseline RMSSD or HR change; these regressions were not used to define the primary inference.

For within-person coupling, each bin’s median RT was centered on that participant’s mean bin median and scaled by the pooled standard deviation of the centered deviations. Failure and omission counts were transformed to empirical logits, $\log[(y + 0.5)/(n - y + 0.5)]$, where y is the event count and n the number of stimulus positions. Separate unweighted least-squares models contained participant fixed intercepts, progress, and standardized RT deviation. Participant-clustered t inference used 30 degrees of freedom and Holm adjustment across the two RT coefficients. These coefficients describe empirical-logit associations rather than fitted event-level odds ratios.

An additional sensitivity analysis for H2 fitted grouped-binomial logistic models to failure and omission counts, with the number of stimulus positions in each bin as the denominator. Both models included participant fixed intercepts, normalized progress, and within-person-centered bin median RT divided by the primary pooled SD (38.27 ms). Participant-cluster sandwich covariance used the same finite-sample correction as the primary analysis, with a t reference distribution on 30 degrees of freedom; the two RT coefficients formed a separate Holm correction family. No continuity correction or additional frequency weights were applied. Details are provided in Supplementary File 2 (Table S6).

The association between RT slope and change in MAD was reported descriptively, with Pearson and Spearman coefficients. Supplementary analyses assessed influence, covariate adjustment, dimensional summaries, internal prediction, and PRV QC sensitivity. All intervals are individual rather than simultaneous. Corrected p values below 0.05 were the inferential threshold within the stated families. Saved outputs for the original analyses and the post hoc PRV analyses are provided in Supplementary File 3 (Statistical Results). Additional estimates and measurement checks are reported in Supplementary File 2 (Tables S1–S13).

Rounding-sensitivity checks are reported in Supplementary File 2 (Table S8).

Directional change-point analysis

Continuous piecewise-linear models were fitted to each participant’s 20 bin medians. Candidate breakpoints ranged from 0.25 to 0.75 in increments of 0.05. A descriptive candidate required a Bayesian

information criterion (BIC; [31]) improvement of at least 6 over a linear model, a positive change in slope, and a positive post-break slope. The estimated breakpoint was treated as an additional model parameter.

Inference for upward departure from linearity used the largest BIC improvement among candidates satisfying both directional conditions. Block resampling preserves local serial dependence [32]. Centered residuals from the fitted linear null model were resampled in circular blocks of three consecutive bins and added to the fitted trajectory. Each of 9,999 bootstrap samples repeated the complete search. Plus-one bootstrap p values were Holm-adjusted across 31 participants. This procedure tests an upward departure from a linear RT trajectory.

Use of generative AI

During preparation of this work, ChatGPT (OpenAI),, was used to support language editing and editorial refinement. The tools were used under the direction of the corresponding author. The authors retain responsibility for the scientific interpretation and final content.

Results

Available observations and physiological reactivity

Participant characteristics and self-reported state and workload are summarized in Table 1.

Table 1
Participant characteristics and self-reported state and workload ($N = 31$).

Characteristic	Mean $\pm$ SD or n (%)	Median [Q1, Q3]	Range
Demographic and anthropometric characteristics			
Age, years	19.55 $\pm$ 1.12	19.00 [19.00, 20.00]	18.00–23.00
Female, n (%)	21 (67.7)	—	—
Male, n (%)	10 (32.3)	—	—
Right-handed, n (%)	31 (100.0)	—	—
Height, cm	172.55 $\pm$ 10.66	169.00 [164.50, 179.00]	160.00–196.00
Body mass, kg	64.56 $\pm$ 13.84	59.00 [55.00, 69.00]	47.00–98.00
BMI, kg/m ²	21.47 $\pm$ 2.58	20.70 [19.65, 23.71]	16.65–26.58
Pre-task self-reported state			
Sleep duration, h	6.29 $\pm$ 1.62	6.00 [6.00, 7.00]	2.00–10.00
Sleep quality, 0–10	6.32 $\pm$ 2.26	7.00 [5.00, 8.00]	1.00–10.00
Sleepiness (KSS), 1–9	4.74 $\pm$ 1.59	5.00 [3.50, 6.00]	2.00–7.00
Physical fatigue, 0–10	3.94 $\pm$ 2.22	4.00 [2.50, 6.00]	0.00–7.00
Mental fatigue, 0–10	4.81 $\pm$ 2.07	5.00 [3.00, 6.00]	1.00–8.00
Stress, 0–10	3.65 $\pm$ 2.60	4.00 [2.00, 6.00]	0.00–9.00
Pain/discomfort, 0–10	1.55 $\pm$ 1.52	1.00 [0.00, 2.00]	0.00–5.00
Post-task subjective workload (NASA-TLX, 0–100)			
Mental demand	68.06 $\pm$ 15.09	70.00 [60.00, 77.50]	30.00–95.00
Physical demand	40.16 $\pm$ 24.44	45.00 [17.50, 57.50]	5.00–90.00
Temporal demand	77.26 $\pm$ 14.48	80.00 [67.50, 85.00]	40.00–100.00
Performance	52.10 $\pm$ 18.52	50.00 [40.00, 65.00]	10.00–90.00
Effort	70.16 $\pm$ 15.99	75.00 [65.00, 80.00]	20.00–90.00
Frustration	60.65 $\pm$ 25.68	65.00 [42.50, 80.00]	10.00–95.00
Raw NASA-TLX (unweighted mean)	61.40 $\pm$ 10.25	60.83 [54.17, 67.92]	40.83–82.50

Values are mean $\pm$ SD, median [Q1, Q3], and observed range unless stated otherwise. BMI, body mass index; KSS, Karolinska Sleepiness Scale; NASA-TLX, NASA Task Load Index. Higher sleep-quality scores

indicate better sleep; higher KSS, fatigue, stress, and discomfort scores indicate greater sleepiness, fatigue, stress, and discomfort, respectively. All characteristics were available for the 31 included participants.

The ordered response data contained 17,320 stimulus positions and 16,304 available RT values, with 973 omissions and 43 positions without a plotted RT. The repeated-bin dataset contained 620 observations. Mean global median RT was 746.23 ± 57.71 ms, and the mean individual RT slope was 8.87 ± 50.31 ms across full progress, ranging from -126.44 to 160.88 ms. The slope range included both increasing and decreasing RT trajectories.

All 62 EDA phases met the quality criteria. Retained coverage was 100% at baseline and 97.44–100% during the task. No baseline SCR met the acceptance criterion for one participant; consequently, the amplitude analysis comprised 30 paired observations and its trajectory model included 600 bins. Final PPG-derived PRV quality control was passed in both phases by 26 participants. Of these, 22 had status OK in both phases; four primary inclusions used the documented post hoc extension for a rejection fraction up to 15% or an actual phase duration of 295–300 s. Physiological phase summaries are presented in Table 2. HR, mean SC, tonic SCL, and SCR frequency increased in every participant; amplitude changes varied in direction. In the PRV sample, RMSSD decreased from 46.13 ± 19.19 to 37.32 ± 16.68 ms (mean change -8.82 ms, 95% CI -14.11 to -3.53 ; Holm-adjusted $p = 0.002$), and SDNN decreased from 62.20 ± 18.59 to 43.22 ± 15.47 ms (mean change -18.98 ms, 95% CI -25.18 to -12.78 ; Holm-adjusted $p < 0.001$).

Table 2
Baseline and task physiology and paired changes.

Measure	<i>n</i>	Baseline Mean $\pm$ SD	Task Mean $\pm$ SD	Change Mean (95% CI)
HR, bpm	31	76.01 ± 9.99	96.68 ± 11.74	20.66 (17.59 to 23.73)
Mean SC, μ S	31	4.13 ± 2.88	8.82 ± 4.28	4.69 (3.81 to 5.56)
Tonic SCL, μ S	31	3.58 ± 2.58	7.91 ± 4.25	4.33 (3.45 to 5.21)
SCR frequency, min^{-1}	31	5.06 ± 3.59	11.83 ± 4.67	6.78 (5.29 to 8.26)
Mean SCR amplitude, μ S	30	0.184 ± 0.087	0.248 ± 0.091	0.064 (0.025 to 0.102)
PPG-derived RMSSD, ms	26	46.13 ± 19.19	37.32 ± 16.68	-8.82 (-14.11 to -3.53)
PPG-derived SDNN, ms	26	62.20 ± 18.59	43.22 ± 15.47	-18.98 (-25.18 to -12.78)

Change = task minus baseline. SCR-amplitude summaries use the same 30 participants in both phases. PPG-derived PRV rows use the 26 participants passing final QC in both phases. CI, confidence interval; PRV, pulse rate variability; RMSSD, root mean square of successive differences; SDNN, standard deviation of accepted pulse intervals; SCL, skin conductance level; SCR, skin-conductance response.

Autonomic reactivity and response trajectories

The five original HR/EDA correlations with RT slope and the separate post hoc PRV correlations are reported in Table 3. HR reactivity was positively associated with RT slope, although its permutation p value exceeded 0.05 after adjustment across the original five-test family. None of the EDA associations met the corrected criterion. No association between PRV change and RT slope was detected: RMSSD $r = -0.060$ (95% CI $[-0.437, 0.335]$) and SDNN $r = -0.143$ (95% CI $[-0.503, 0.258]$); both two-test Holm-adjusted p values were 0.974. The HR association is shown in Fig. 2. Figure 3 shows the paired PRV phase values and the RMSSD-change–RT-slope relation. The analogous 20-bin RT-slope correlations were also near zero (Supplementary File 2, Table S12). Mean-SC and tonic-SCL changes were correlated ($r = 0.945$).

Table 3
Physiological changes and participant RT slope.

Predictor	n	r	95% CI for r	Permutation p	Holm p
HR	31	0.452	0.116 to 0.695	0.014	0.070
Mean SC	31	0.126	-0.239 to 0.460	0.503	1.000
Tonic SCL	31	0.172	-0.195 to 0.496	0.359	1.000
SCR frequency	31	-0.094	-0.434 to 0.269	0.623	1.000
Mean SCR amplitude	30	-0.192	-0.516 to 0.181	0.296	1.000
PPG-derived RMSSD	26	-0.060	-0.437 to 0.335	0.765	0.974
PPG-derived SDNN	26	-0.143	-0.503 to 0.258	0.487	0.974

Pearson coefficients; Fisher's z intervals are unadjusted. The original HR/EDA rows form the five-test Holm family used for the planned and original exploratory physiology analyses. The two PPG-derived PRV rows form a separate post hoc two-test Holm family because they use a distinct QC-defined sample. Reactivity is task minus baseline.

Consistent with the participant-level association, the HR–progress interaction was positive in the repeated-bin model (Table 4). Cluster- t inference met the five-test Holm criterion, whereas the participant-label permutation sensitivity did not. EDA interaction intervals included zero. Figure 4 displays the model-implied RT trajectories at different levels of HR reactivity; the lines describe the fitted association and do not represent distinct participant groups.

Table 4
Physiological moderation of RT change across normalized progress.

Predictor	β ms/SD	95% CI	Cluster- <i>t p</i>	Holm cluster- <i>t p</i>	Holm permutation <i>p</i>
HR	18.87	5.78 to 31.95	0.006	0.031	0.148
Mean SC	8.90	-11.00 to 28.80	0.368	1.000	0.935
Tonic SCL	10.65	-9.00 to 30.29	0.277	1.000	0.899
SCR frequency	-2.09	-15.83 to 11.66	0.759	1.000	0.935
Mean SCR amplitude	-7.93	-22.14 to 6.28	0.263	1.000	0.935

Separate models contain participant fixed intercepts, progress, and progress $\times$ standardized reactivity. $N = 31$ and 620 bins, except amplitude ($N = 30$; 600 bins). CI uses participant-clustered t inference. Each p -value column with Holm adjustment covers five predictors; permutation inference is a separate sensitivity procedure.

Response quality and local slowing

Across participants, HR reactivity correlated positively with aggregate accuracy ($r = 0.454$) and negatively with omission proportion ($r = -0.436$). Their permutation p values were 0.012 and 0.014, respectively; the corresponding values after adjustment across the 20 quality/exposure tests were 0.243 and 0.267. No association in that family met the corrected criterion. Full estimates are available in Supplementary File 2 (Table S1).

In the separate eight-test PRV family for the same aggregate outcomes, RMSSD change was negatively associated with accuracy ($r = -0.560$, 95% CI [-0.779, -0.221], permutation $p = 0.004$, Holm-adjusted $p = 0.029$). Because change was calculated as task minus baseline, this direction indicates that a larger RMSSD reduction was associated with higher aggregate accuracy. RMSSD change was positively associated with omission proportion before correction ($r = 0.472$, permutation $p = 0.014$), but not after Holm adjustment ($p = 0.098$). The SDNN-change-accuracy association was $r = -0.423$ (Holm-adjusted $p = 0.183$). No other PRV association in this family met the corrected threshold (Supplementary File 2, Table S12). In the strict QC set ($n = 22$), the RMSSD-accuracy estimate remained negative ($r = -0.498$) but the eight-test Holm-adjusted p value was 0.146; the designated borderline sensitivity gave $n = 27$, $r = -0.558$ and adjusted $p = 0.031$ (Supplementary File 2, Table S13).

Within participants, coefficients for the associations of local RT with failure and omission rates were negative (Table 5). Neither coefficient met the two-test Holm criterion.

In the grouped-binomial sensitivity analysis, a one-SD increase in within-person-centered bin median RT was associated with lower failure odds (OR = 0.906, 95% CI [0.841, 0.975], Holm-adjusted $p = 0.021$) and omission odds (OR = 0.916, 95% CI [0.847, 0.990], Holm-adjusted $p = 0.029$). Both directions agreed with

the primary empirical-logit models, for which both adjusted p values were 0.060. Full estimates are reported in Supplementary File 2 (Table S6).

Table 5
Within-person associations between local RT and response quality.

Outcome	β	95% CI	Cluster- t p	Holm p
Failures	-0.083	-0.158 to - 0.009	0.030	0.060
Omissions	-0.069	-0.139 to 0.002	0.056	0.060

Coefficient per pooled SD of within-person RT deviation (38.27 ms). Failure = incorrect plotted response or omission. Models use empirical logits, participant fixed intercepts, and progress; 31 participants and 620 bins. Holm correction covers the two RT coefficients.

Temporal variability and trajectory changes

Figure 5 shows individual response trajectories alongside HR reactivity and RT slope, with positive and negative slopes among participants with similar HR changes.

More positive RT slopes were associated descriptively with larger late-minus-early changes in MAD ($r = 0.462$, 95% CI [0.129, 0.701]; Spearman $\rho = 0.406$). None of the 20 original HR/EDA associations with late-minus-early temporal outcomes met the Holm criterion. None of the eight PRV associations with those same outcomes met the separate Holm criterion. Individual covariate adjustments retained positive HR trajectory estimates. Complete sensitivity estimates are reported in Supplementary File 2 (Tables S2–S4, S9, S12 and S13).

One participant met the descriptive piecewise criterion, with a breakpoint at normalized progress 0.65 and a BIC improvement of 13.75. The residual-block bootstrap p value was 0.0038 and the 31-test Holm-adjusted value was 0.118. No participant met the multiplicity-adjusted criterion for an upward RT change point (Supplementary File 2, Table S5).

Discussion

The findings address two distinct questions about behavior under performance-contingent pacing and add a post hoc PRV description. Participants with greater task-minus-baseline HR change tended to have more positive RT slopes, whereas locally slower responding tended to accompany lower failure and omission rates. These directions were consistent with the hypotheses, but the evidence remained sensitive to statistical inference. In the final-QC PRV sample, RMSSD and SDNN both decreased from baseline to task. Neither PRV change was associated with RT slope, so the PRV results did not reproduce the specific HR–slowing relation. Instead, a larger RMSSD reduction was associated with higher aggregate accuracy in the corrected primary PRV family; its magnitude remained similar under strict QC, although the corrected threshold was not retained at $n = 22$. The cardiac analyses describe

related observations from overlapping participants, rather than independent replications. Together, the results indicate that mean HR, pulse-interval variability, and response quality carry partially different information without establishing that autonomic adjustment sustained performance or that slowing represented effective compensatory control.

The inferential divergence preceded multiplicity adjustment: unadjusted p values were 0.0062 for cluster- t and 0.0297 for permutation. In this balanced design, both procedures assess the coefficient relating participants' 20-bin RT slopes to standardized HR reactivity, which equals the repeated-bin interaction. Cluster- t inference divides the coefficient by its participant-clustered standard error (6.41 ms/SD) and uses $t(30)$; permutation compares an unstudentized correlation with correlations obtained by randomly reassigning participant labels. The procedures therefore use different uncertainty assumptions: the t reference is approximate, whereas permutation requires exchangeable pairings under its null. Their differing decisions with 31 independent participants warrant cautious interpretation.

Associations were estimated at two levels: the cardiac analysis compared participants, and the local response-quality analysis compared bins within participants. Physiological predictors summarized baseline-to-task changes.

The adaptive controller shapes the setting in which these associations arise: earlier errors and omissions modify subsequent presentation conditions. Adaptive automation and game-difficulty adjustment have improved performance [33, 34], yet participants in a vigilance task performed similarly when schedules were controlled by their own EEG or generated from another person's EEG [35]. Effective scheduling and matching support to an individual's state are therefore distinct achievements. Here, adaptation was driven by response performance, and task-wide physiological reactivity was related retrospectively to behavioral trajectories. The associations characterize behavior within an adaptive environment, where participant regulation and computer-supplied changes in response opportunities can both contribute to performance.

Greater HR reactivity could reflect mobilization accompanying a temporal cost of task engagement. Cardiovascular responses vary with attainable task demands, temporal pressure, and ability [36, 37]. Greater working-memory demand can increase HR while reducing high-frequency heart-period variability, consistent with altered parasympathetic regulation [10]. The present PRV results add a related but non-equivalent observation: PPG-derived RMSSD and SDNN were lower during the task than at baseline, yet no association between their changes and RT slope was detected. A larger RMSSD reduction was instead associated with higher aggregate accuracy. This pattern is compatible with greater autonomic adjustment among participants who maintained more accurate responding, but it does not establish compensatory effort, vagal withdrawal as the sole cause, or a causal benefit of reduced variability. PPG-derived PRV is influenced by peripheral pulse-wave timing and should not be treated as interchangeable with ECG-derived HRV, particularly under cognitive or other non-resting conditions [11].

Task-minus-baseline reactivity must also be distinguished from a physiological trend: HR can decline as fatigue develops [38], even when task-average HR remains above baseline. Slowing and reduced

attentional efficiency in 20-min vigilance [39] and 3-h visual-attention tasks [40] provide comparisons with this approximately 8-min adaptive task, whose RT slope represents normalized stimulus order rather than elapsed time. Fatigue, disengagement, motivation, and autonomic arousal need not change in parallel during prolonged cognitive performance [41, 42, 43]. The present slopes therefore cannot be attributed to any one of these processes.

For H2, longer bin median RT relative to a participant's own average accompanied lower failure and omission rates after accounting for progress. The grouped-binomial sensitivity estimates were $OR = 0.906$ for failures (95% CI [0.841, 0.975], Holm-adjusted $p = 0.021$) and $OR = 0.916$ for omissions (95% CI [0.847, 0.990], Holm-adjusted $p = 0.029$), per one pooled SD (38.27 ms) of within-person-centered RT. Both directions agreed with the primary empirical-logit models, although only the binomial models met the two-test Holm criterion. Binomial likelihood incorporates the denominators and mean–variance relation of event counts; the primary models give equal weight to continuity-corrected empirical logits. This difference indicates dependence on outcome modeling, not independent replication. A more conservative evidence criterion offers one account of slower, more accurate responses [44], although the present task involves multiple stimulus–response mappings. Slowing can also accompany poorer accuracy under greater n -back load [45], and the time–accuracy association varied with difficulty across 29 cognitive-task datasets [46]. The negative coefficients therefore describe local covariation without demonstrating a change in decision threshold or a causal benefit of slowing.

Pacing and response history provide alternative explanations. When unsuccessful responses reduce the subsequent presentation rate, relief from temporal competition could improve later response quality while response timing changes. Slower presentation does not necessarily lengthen measured RT, so the coefficient's sign alone cannot identify this mechanism. Slowing can also reflect orienting to infrequent events or perceptual and motor adjustments after errors [47, 48]; with sufficiently long interstimulus intervals, larger error-related pupil responses predicted both slowing and greater subsequent accuracy [49]. Because the present bins combine observations before and after errors, they cannot distinguish caution preceding an error from orienting, recovery, or controller adjustment following it.

The EDA findings distinguish responsiveness to task conditions from information about behavioral trajectories. HR, mean SC, tonic SCL, and SCR frequency increased in every participant, yet EDA–trajectory associations did not meet corrected thresholds. Tonic activity and SCR counts responded to simulation demands [50], whereas SCR frequency showed no clear relation to working-memory accuracy in another task [51]. A changing Tetris task also showed limited differentiation in electrocardiography- and EDA-based physiological stress despite changes in load and performance [52]. Event-related research linked errors to larger SCRs and cardiac deceleration, with SCR magnitude covarying with post-error slowing [53]. Task-wide SCR frequency and mean accepted amplitude pool such events with different elicitors and consequences; mean-SC and tonic-SCL changes also overlapped strongly here ($r = 0.945$). Aggregation may therefore obscure temporal relationships, without establishing an unobserved EDA–performance association.

Response variability and change points provided additional descriptions of the behavioral sequence. Prior work links variability, momentary slowing, and reduced processing of external events to fluctuations in attention [54, 55, 56]. The present RT-slope–MAD association involved two summaries derived from the same response series and did not independently index deterioration. One participant met the descriptive piecewise criterion, but no upward change point survived multiplicity adjustment. Twenty-bin trajectories may be insensitive to brief or controller-interrupted changes.

For adaptive human–computer interaction, measurement, prediction, and intervention require separate evaluation. EEG workload classification accuracy declined across days [57], and a personalized multimodal workload algorithm that initially appeared promising failed a subsequent test of task-response matching [58]. A between-person HR association cannot itself define an individual warning threshold. Physiological information must improve decisions beyond recent responses and current task conditions, with intervention effectiveness assessed against a separately defined outcome. Synchronized physiological and behavioral measurements, together with the computer's actual adjustments, would clarify what physiology adds to interpreting adaptive interaction.

Limitations

The sample size limited precision for between-participant associations, particularly for small effects and analyses subject to multiplicity correction. PRV had a smaller final-QC sample ($n = 26$; strict $n = 22$) than the full analytical sample, which further reduced precision. The predominantly young and female sample limits generalization to broader populations.

The observed trajectories combine participants' responses with changing response opportunities under adaptive pacing.

The physiological predictors were phase-level contrasts. PRV covered the first 300 s of the task, whereas the HR and EDA summaries used data across the full task phase. Comparisons among their associations therefore concern different summary windows. The PRV analyses, the QC extension from a 10% to 15% rejected-interval limit, and the 295–300 s short-window allowance were post hoc. The extensions were documented before the final analyses. The strict-QC sensitivity showed that inference for the RMSSD–accuracy association varied with the PRV sample.

RT observations were extracted from VTS report plots, with numerical consistency assessed through vector-coordinate and rounding checks.

Inference for the HR trajectory association differed between cluster-based and permutation procedures. The change-point analysis was exploratory and based on 20-bin trajectories.

Future research

Future research should prioritize synchronized recording of event-level physiology, behavior, and controller state. This temporal distinction matters because neural changes can precede errors [59], while

error-related cardiac deceleration can depend on conscious error detection [60]. Sustained deterioration should be defined independently, for example as repeated failures or persistent loss of accuracy over a predefined interval.

Comparisons of fixed, yoked, and adaptive pacing would quantify the contribution of system adjustments to performance maintenance. Fixed-demand blocks and counterbalanced or programmed difficulty schedules would help separate workload from time-on-task effects and show how response trajectories develop when errors do not automatically slow presentation.

Candidate physiological markers should then be evaluated prospectively for predictive value beyond recent errors, RT variability, and current difficulty. Individual baselines, standardized practice, repeated sessions, and independent participants would support this evaluation, using warning lead time, false alarms, missed deterioration, and recovery as outcomes. Randomized comparisons of assistance, breaks, and unchanged task conditions could establish whether acting on a warning improves performance. The improvement in sustained attention obtained with closed-loop neurofeedback relative to yoked feedback illustrates the value of testing intervention benefits directly [61].

Conclusions

RT trajectories varied across participants during the adaptive computerized task. Greater HR reactivity was associated with more positive RT slopes, but evidence depended on the inferential procedure and did not consistently meet corrected thresholds. PPG-derived RMSSD and SDNN decreased from baseline to task; no association between either change and RT slope was detected. Greater RMSSD reduction was associated with higher aggregate accuracy in the primary PRV sample, while strict QC retained a similar estimate without meeting the corrected threshold. EDA associations with temporal outcomes did not meet corrected thresholds. Locally slower responding accompanied lower failure and omission rates, although corrected inference depended on outcome modeling. These results distinguish task-related autonomic reactivity from temporal response change; they do not establish physiological compensation, a transition to performance breakdown, or an individual warning threshold.

Declarations

Data availability

Supplementary File 2 (Tables S1–S13) reports additional estimates, PRV QC, sensitivity analyses, and measurement checks. Saved outputs for the original analyses and the post hoc PRV analyses are provided in Supplementary File 3 (Statistical Results). Ordered-response and EDA-interval datasets are available from the corresponding author, subject to institutional and ethical requirements.

Author contributions

ML: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. GZ: Data curation, Formal analysis, Visualization, Writing – original draft. WA: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Funding

The experimental work was supported by institutional resources of the Wroclaw University of Health and Sport Sciences. No commercial sponsor contributed to study design, analysis, or interpretation.

Competing interests

The authors declare no competing interests.

Code availability

Ordered-response and EDA processing code is available from the corresponding author, subject to institutional and ethical requirements.

References

1. Bailey, B. P. & Iqbal, S. T. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput.-Hum. Interact.* **14**, 21 (2008). <https://doi.org/10.1145/1314683.1314689>
2. Ewing, K. C., Fairclough, S. H. & Gilleade, K. Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop. *Front. Hum. Neurosci.* **10**, 223 (2016). <https://doi.org/10.3389/fnhum.2016.00223>
3. Hockey, G. R. J. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biol. Psychol.* **45**, 73–93 (1997). [https://doi.org/10.1016/S0301-0511\(96\)05223-4](https://doi.org/10.1016/S0301-0511(96)05223-4)
4. Ratcliff, R. & McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008). <https://doi.org/10.1162/neco.2008.12-06-420>
5. Fairclough, S. H., Venables, L. & Tattersall, A. The influence of task demand and learning on the psychophysiological response. *Int. J. Psychophysiol.* **56**, 171–184 (2005). <https://doi.org/10.1016/j.ijpsycho.2004.11.003>
6. Setz, C. et al. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans. Inf. Technol. Biomed.* **14**, 410–417 (2010). <https://doi.org/10.1109/titb.2009.2036164>
7. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010). <https://doi.org/10.1016/j.jneumeth.2010.04.028>

8. Mehler, B., Reimer, B. & Coughlin, J. F. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Hum. Factors* **54**, 396–412 (2012). <https://doi.org/10.1177/0018720812442086>
9. Unni, A., Ihme, K., Jipp, M. & Rieger, J. W. Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Front. Hum. Neurosci.* **11**, 167 (2017). <https://doi.org/10.3389/fnhum.2017.00167>
10. Gianaros, P. J., Van der Veen, F. M. & Jennings, J. R. Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working-memory tasks: Implications for the cortical and subcortical regulation of cardiac autonomic activity. *Psychophysiology* **41**, 521–530 (2004). <https://doi.org/10.1111/1469-8986.2004.00179.x>
11. Xu, S., Liu, H., Liu, Z., Su, P. & Gu, Z. Accuracy of photoplethysmography-derived pulse rate variability compared with electrocardiography-derived heart rate variability: A systematic review and meta-analysis. *Sensors* **26**, 5192 (2026). <https://doi.org/10.3390/s26165192>
12. Stuiver, A. & Mulder, B. Cardiovascular state changes in simulated work environments. *Front. Neurosci.* **8**, 399 (2014). <https://doi.org/10.3389/fnins.2014.00399>
13. Katsimpokis, D., Hawkins, G. E. & van Maanen, L. Not all speed-accuracy trade-off manipulations have the same psychological effect. *Comput. Brain Behav.* **3**, 252–268 (2020). <https://doi.org/10.1007/s42113-020-00074-y>
14. Karşilar, H., Simen, P., Papadakis, S. & Balci, F. Speed accuracy trade-off under response deadlines. *Front. Neurosci.* **8**, 248 (2014). <https://doi.org/10.3389/fnins.2014.00248>
15. Wilson, G. F. & Russell, C. A. Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Hum. Factors* **49**, 1005–1018 (2007). <https://doi.org/10.1518/001872007x249875>
16. Chiossi, F., Turgut, Y., Welsch, R. & Mayer, S. Adapting visual complexity based on electrodermal activity improves working memory performance in virtual reality. *Proc. ACM Hum.-Comput. Interact.* **7**, 196 (2023). <https://doi.org/10.1145/3604243>
17. Steinrücke, J., Veldkamp, B. P. & de Jong, T. Determining the effect of stress on analytical skills performance in digital decision games towards an unobtrusive measure of experienced stress in gameplay scenarios. *Comput. Hum. Behav.* **99**, 144–155 (2019). <https://doi.org/10.1016/j.chb.2019.05.014>
18. Beauchemin, N. et al. Enhancing learning experiences: EEG-based passive BCI system adapts learning speed to cognitive load in real-time, with motivation as catalyst. *Front. Hum. Neurosci.* **18**, 1416683 (2024). <https://doi.org/10.3389/fnhum.2024.1416683>
19. Jaiaue, T., Cooke, A. & Gallicchio, G. Differential contributions of inhibition and task engagement to workload: Evidence from a novel adaptive Stroop task. *Ergonomics* <https://doi.org/10.1080/00140139.2026.2702629> (2026).
20. Salthouse, T. A. When does age-related cognitive decline begin? *Neurobiol. Aging* **30**, 507–514 (2009). <https://doi.org/10.1016/j.neurobiolaging.2008.09.023>

21. Åkerstedt, T. & Gillberg, M. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* **52**, 29–37 (1990). <https://doi.org/10.3109/00207459008994241>
22. Hart, S. G. & Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload* (eds Hancock, P. A. & Meshkati, N.) 139–183 (North-Holland, 1988). [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
23. Hart, S. G. NASA-Task Load Index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **50**, 904–908 (2006). <https://doi.org/10.1177/154193120605000909>
24. SCHUHFRIED. DT: Determination Test. Vienna Test System. <https://marketplace.schuhfried.com/en/DT> (n.d.).
25. Elgendi, M., Norton, I., Brearley, M., Abbott, D. & Schuurmans, D. Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. *PLoS ONE* **8**, e76585 (2013). <https://doi.org/10.1371/journal.pone.0076585>
26. Greco, A., Valenza, G., Lanata, A., Scilingo, E. P. & Citi, L. cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Trans. Biomed. Eng.* **63**, 797–804 (2016). <https://doi.org/10.1109/TBME.2015.2474131>
27. Kleckner, I. R. et al. Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. *IEEE Trans. Biomed. Eng.* **65**, 1460–1467 (2018). <https://doi.org/10.1109/TBME.2017.2758643>
28. Phipson, B. & Smyth, G. K. Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, 39 (2010). <https://doi.org/10.2202/1544-6115.1585>
29. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979). <https://www.jstor.org/stable/4615733>
30. Cameron, A. C. & Miller, D. L. A practitioner's guide to cluster-robust inference. *J. Hum. Resour.* **50**, 317–372 (2015). <https://doi.org/10.3368/jhr.50.2.317>
31. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978). <https://doi.org/10.1214/aos/1176344136>
32. Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989). <https://doi.org/10.1214/aos/1176347265>
33. Prinzl, L. J. III, Freeman, F. G., Scerbo, M. W., Mikulka, P. J. & Pope, A. T. Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Hum. Factors* **45**, 601–614 (2003). <https://doi.org/10.1518/hfes.45.4.601.27092>
34. Liu, C., Agrawal, P., Sarkar, N. & Chen, S. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *Int. J. Hum.-Comput. Interact.* **25**, 506–529 (2009). <https://doi.org/10.1080/10447310902963944>
35. Mikulka, P. J., Scerbo, M. W. & Freeman, F. G. Effects of a biocybernetic system on vigilance performance. *Hum. Factors* **44**, 654–664 (2002). <https://doi.org/10.1518/0018720024496944>

36. Backs, R. W. & Seljos, K. A. Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *Int. J. Psychophysiol.* **16**, 57–68 (1994).
[https://doi.org/10.1016/0167-8760\(94\)90042-6](https://doi.org/10.1016/0167-8760(94)90042-6)
37. Richter, M., Friedrich, A. & Gendolla, G. H. E. Task difficulty effects on cardiac activity. *Psychophysiology* **45**, 869–875 (2008). <https://doi.org/10.1111/j.1469-8986.2008.00688.x>
38. Matuz, A. et al. Enhanced cardiac vagal tone in mental fatigue: Analysis of heart rate variability in Time-on-Task, recovery, and reactivity. *PLoS ONE* **16**, e0238670 (2021).
<https://doi.org/10.1371/journal.pone.0238670>
39. Lim, J. et al. Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *Neuroimage* **49**, 3426–3435 (2010).
<https://doi.org/10.1016/j.neuroimage.2009.11.020>
40. Boksem, M. A. S., Meijman, T. F. & Lorist, M. M. Effects of mental fatigue on attention: An ERP study. *Cogn. Brain Res.* **25**, 107–116 (2005). <https://doi.org/10.1016/j.cogbrainres.2005.04.011>
41. Hopstaken, J. F., van der Linden, D., Bakker, A. B. & Kompier, M. A. J. A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology* **52**, 305–315 (2015).
<https://doi.org/10.1111/psyp.12339>
42. Hopstaken, J. F., van der Linden, D., Bakker, A. B. & Kompier, M. A. J. The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol.* **110**, 100–106 (2015).
<https://doi.org/10.1016/j.biopsycho.2015.06.013>
43. Gergelyfi, M., Jacob, B., Olivier, E. & Zénon, A. Dissociation between mental fatigue and motivational state during prolonged mental activity. *Front. Behav. Neurosci.* **9**, 176 (2015).
<https://doi.org/10.3389/fnbeh.2015.00176>
44. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
<https://doi.org/10.1037/0033-295x.85.2.59>
45. Hogervorst, M. A., Brouwer, A.-M. & van Erp, J. B. F. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci.* **8**, 322 (2014). <https://doi.org/10.3389/fnins.2014.00322>
46. Domingue, B. W. et al. Speed–accuracy trade-off? Not so fast: Marginal changes in speed have inconsistent relationships with accuracy in real-world settings. *J. Educ. Behav. Stat.* **47**, 576–602 (2022). <https://doi.org/10.3102/10769986221099906>
47. Notebaert, W. et al. Post-error slowing: An orienting account. *Cognition* **111**, 275–279 (2009).
<https://doi.org/10.1016/j.cognition.2009.02.002>
48. Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M. & Ullsperger, M. Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *J. Neurosci.* **31**, 1780–1789 (2011). <https://doi.org/10.1523/jneurosci.4299-10.2011>
49. Murphy, P. R., van Moort, M. L. & Nieuwenhuis, S. The pupillary orienting response predicts adaptive behavioral adjustment after errors. *PLoS ONE* **11**, e0151763 (2016).
<https://doi.org/10.1371/journal.pone.0151763>

50. Wulvik, A. S., Dybvik, H. & Steinert, M. Investigating the relationship between mental state (workload and affect) and physiology in a control room setting (ship bridge simulator). *Cogn. Technol. Work* **22**, 95–108 (2020). <https://doi.org/10.1007/s10111-019-00553-8>
51. Sazuka, N., Katsumata, K., Komoriya, Y., Oba, T. & Ohira, H. Association of brain–autonomic activities and task accuracy under cognitive load: A pilot study using electroencephalogram, autonomic activity measurements, and arousal level estimated by machine learning. *Front. Hum. Neurosci.* **18**, 1272121 (2024). <https://doi.org/10.3389/fnhum.2024.1272121>
52. Dybvik, H., Erichsen, C. K., Snider, C. & Steinert, M. fNIRS, EEG, ECG, and GSR reveal an effect of complex, dynamically changing environments on cognitive load, affective state, and performance, but not physiological stress. *Front. Hum. Neurosci.* **19**, 1459653 (2025). <https://doi.org/10.3389/fnhum.2025.1459653>
53. Hajcak, G., McDonald, N. & Simons, R. F. To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology* **40**, 895–903 (2003). <https://doi.org/10.1111/1469-8986.00107>
54. Esterman, M., Noonan, S. K., Rosenberg, M. & DeGutis, J. In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* **23**, 2712–2723 (2013). <https://doi.org/10.1093/cercor/bhs261>
55. Weissman, D. H., Roberts, K. C., Visscher, K. M. & Woldorff, M. G. The neural bases of momentary lapses in attention. *Nat. Neurosci.* **9**, 971–978 (2006). <https://doi.org/10.1038/nn1727>
56. Smallwood, J., Beach, E., Schooler, J. W. & Handy, T. C. Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *J. Cogn. Neurosci.* **20**, 458–469 (2008). <https://doi.org/10.1162/jocn.2008.20037>
57. Christensen, J. C., Estep, J. R., Wilson, G. F. & Russell, C. A. The effects of day-to-day variability of physiological data on operator functional state classification. *Neuroimage* **59**, 57–63 (2012). <https://doi.org/10.1016/j.neuroimage.2011.07.091>
58. Teo, G., Matthews, G., Reinerman-Jones, L. & Barber, D. Adaptive aiding with an individualized workload model based on psychophysiological measures. *Hum.-Intell. Syst. Integr.* **2**, 1–15 (2020). <https://doi.org/10.1007/s42454-019-00005-8>
59. Eichele, T. et al. Prediction of human errors by maladaptive changes in event-related brain networks. *Proc. Natl Acad. Sci. USA* **105**, 6173–6178 (2008). <https://doi.org/10.1073/pnas.0708965105>
60. Wessel, J. R., Danielmeier, C. & Ullsperger, M. Error awareness revisited: Accumulation of multimodal evidence from central and autonomic nervous systems. *J. Cogn. Neurosci.* **23**, 3021–3036 (2011). <https://doi.org/10.1162/jocn.2011.21635>
61. deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A. & Turk-Browne, N. B. Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* **18**, 470–475 (2015). <https://doi.org/10.1038/nn.3940>

Figures

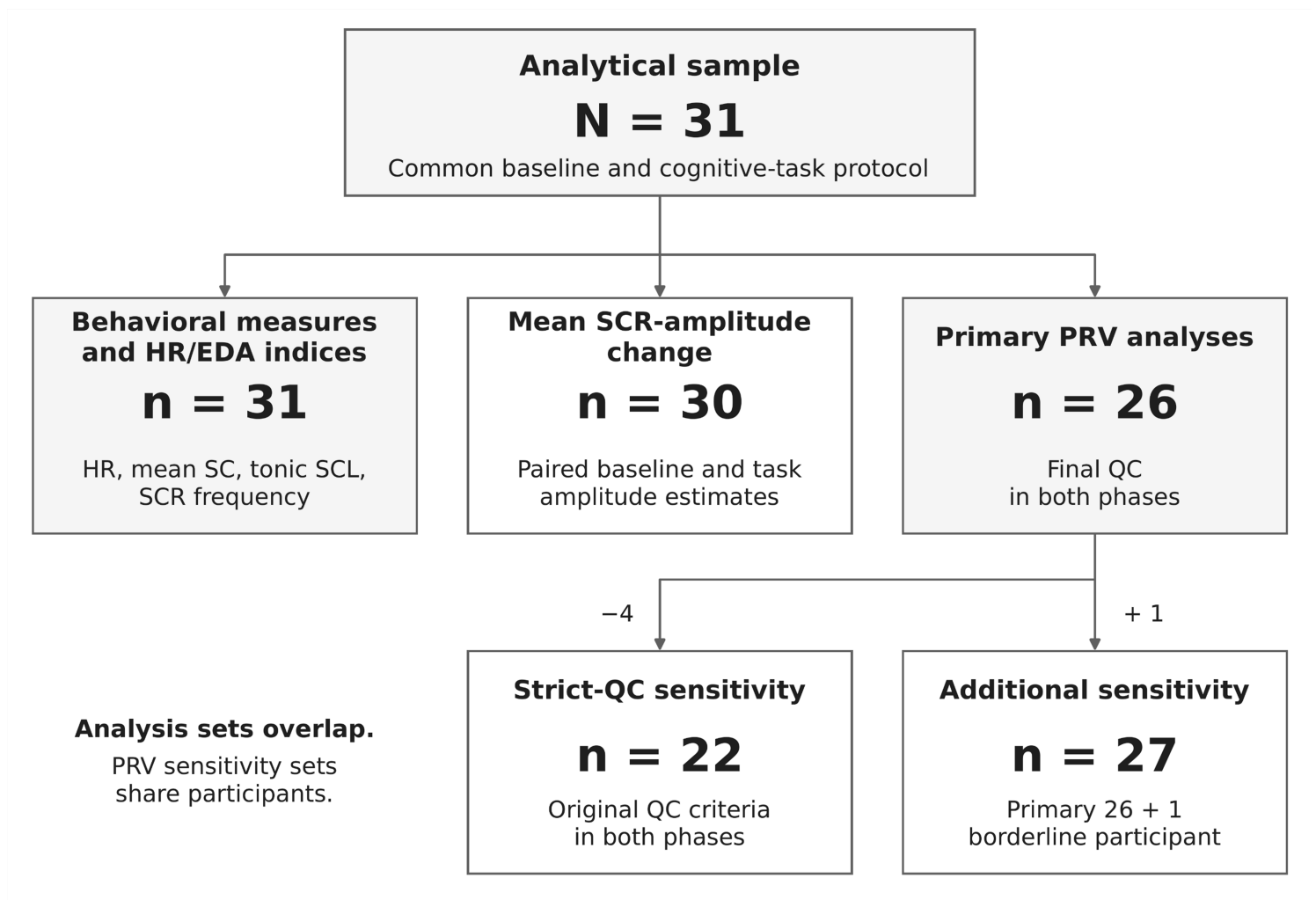


Figure 1

Participant inclusion in the analysis sets. The analytical sample comprised 31 participants who completed the same baseline and cognitive-task protocol. All 31 contributed to behavioral, HR, mean SC, tonic SCL and SCR-frequency analyses. Mean SCR-amplitude change was available for 30 participants. Primary PRV analyses included 26 participants meeting final QC in both phases. Of these, 22 met the original QC criteria and four qualified under documented post hoc extensions: a rejected-interval fraction of up to 15% in three participants (one task phase and two baseline phases), or a 295–300 s baseline window in one participant. The strict-QC sensitivity included the 22 participants meeting the original criteria. The $n = 27$ sensitivity added one borderline participant using candidate PRV values, with a baseline rejected-interval fraction of 17.7%. Analysis sets overlap. HR, heart rate; EDA, electrodermal activity; SC, skin conductance; SCL, skin conductance level; SCR, skin conductance response; PRV, pulse rate variability; QC, quality control.

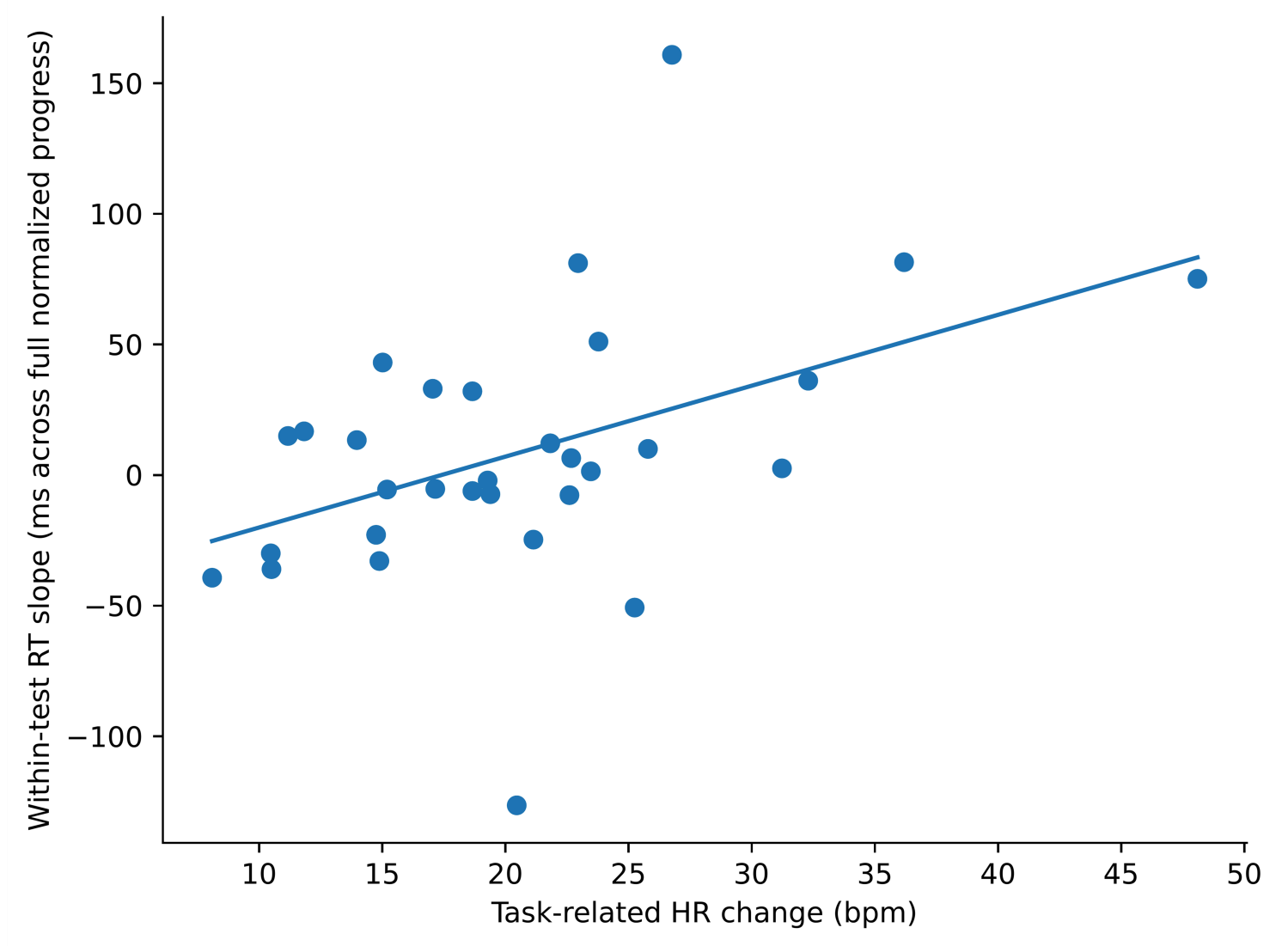


Figure 2

Task-minus-baseline HR change and participant RT slope. Each point represents one participant ($N= 31$); the line is the unadjusted least-squares fit. Positive slopes indicate increasing RT over normalized stimulus progress. The five-predictor permutation family gives Holm-adjusted $p = 0.070$ for HR.

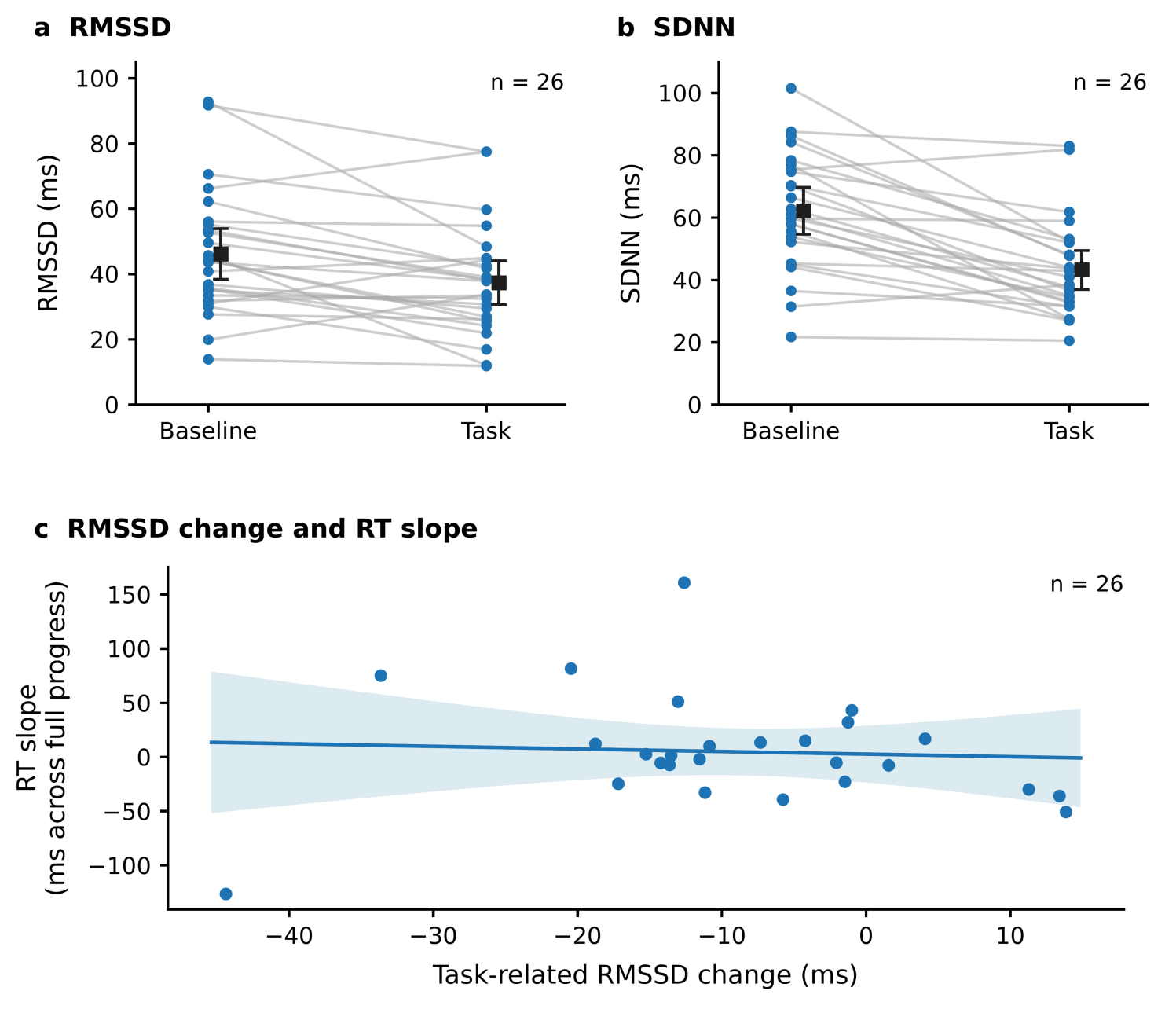


Figure 3

PPG-derived pulse rate variability during the adaptive task and its relation to RT slope. (a) Paired participant RMSSD values at baseline and during the task; filled squares and error bars show the mean and 95% CI. (b) Corresponding SDNN values. (c) Task-minus-baseline RMSSD change and participant RT slope. Points represent the 26 participants passing final PRV QC; the line is the unadjusted least-squares fit and shading is its 95% confidence band. The RMSSD-change–RT-slope association was $r = -0.060$ (95% CI $[-0.437, 0.335]$; two-test Holm-adjusted permutation $p = 0.974$).

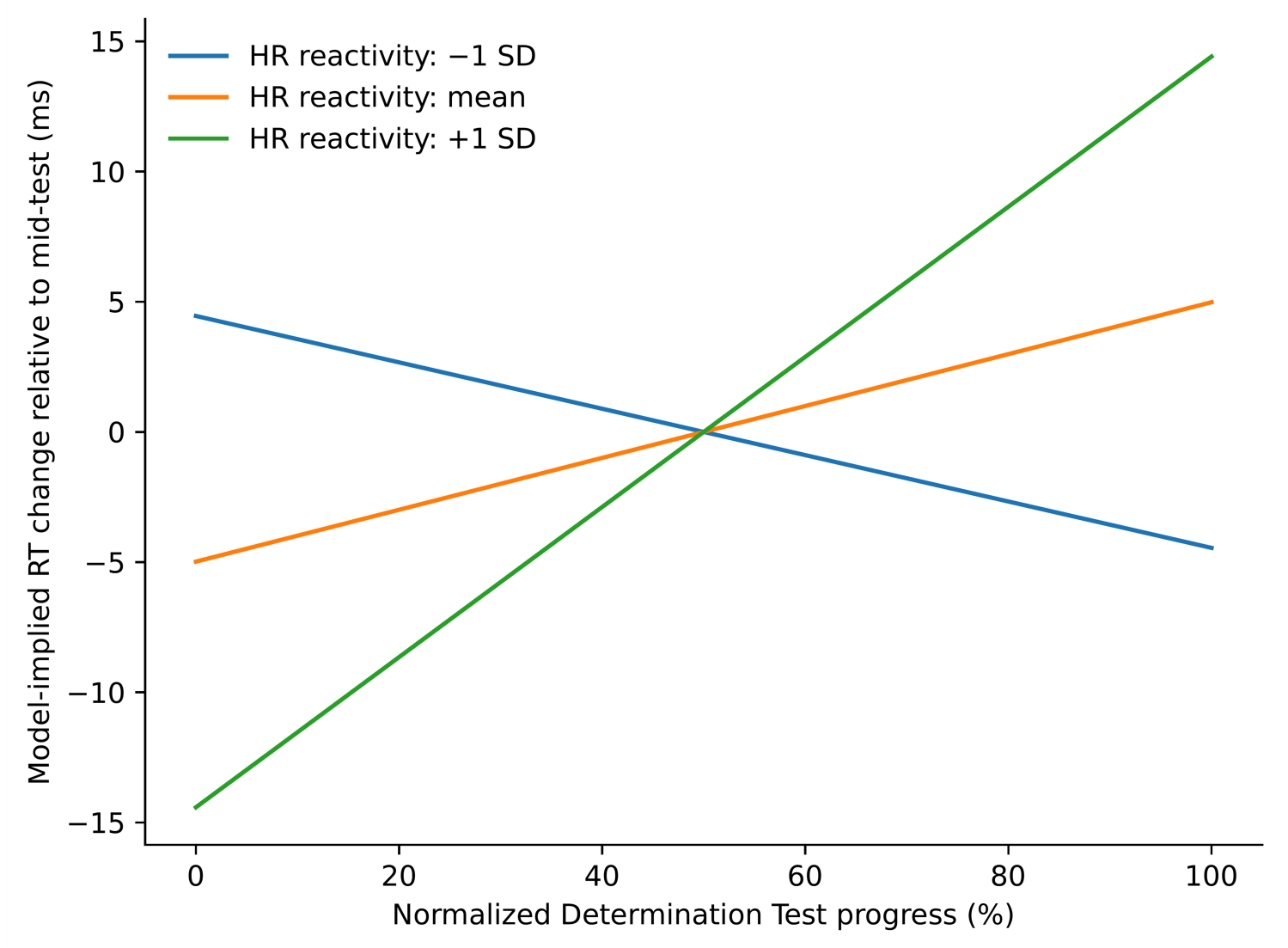


Figure 4

Model-implied RT change at mean HR reactivity and at ± 1 SD, centered at mid-task for display. The progress $\times$ HR interaction is 18.87 ms/SD (cluster- $t_{95\%}$ CI 5.78–31.95). Holm-adjusted p values are 0.031 for cluster-inference and 0.148 for participant-label permutation. The lines represent model predictions, not participant categories.

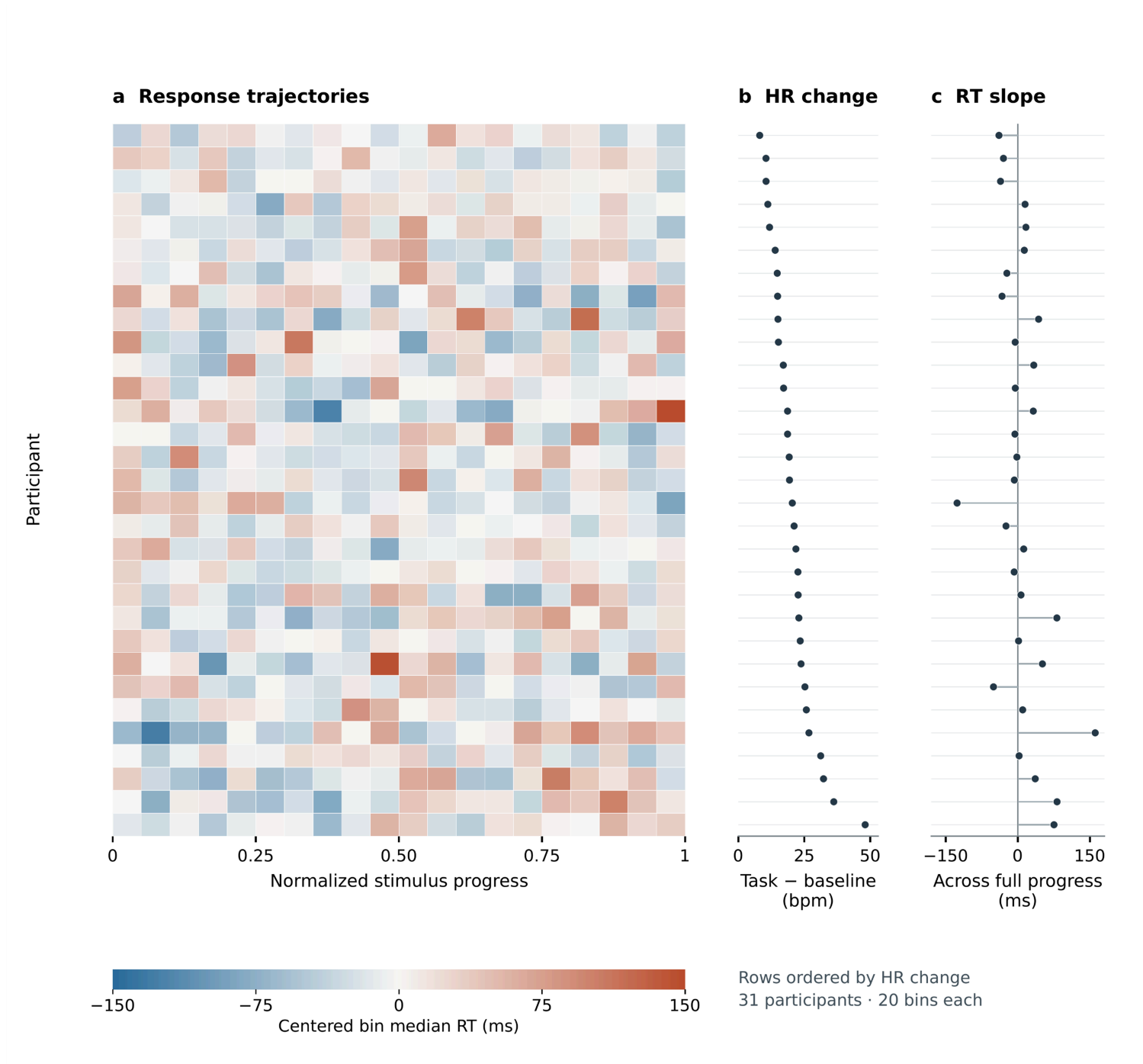


Figure 5

Participant trajectories ordered by increasing HR reactivity. (a) Each cell shows a bin median RT minus the participant's unweighted mean of 20 bin medians. Blue/orange denote locally faster/slower responses; the common ± 150 -ms scale involves no clipping, smoothing, or interpolation. (b) Task-minus-baseline HR change. (c) RT slope estimated across full normalized progress using the individual available RTs rather than bin medians. Each row represents one participant ($N = 31$). Progress represents stimulus order, not elapsed time.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile2ScientificReportsSupplementaryTables.docx](#)
- [SupplementaryFile1EthicalApproval.pdf](#)
- [SupplementaryFile3ScientificReportsStatisticalResults.pdf](#)