Student Conference Proceedings
Vol 2, No 1, 2026, Article ID 2711, 4 Pages

Student Conference Lübeck

Proceedings Article

Flex-Eye-Bility: Dissociable pupil-linked surprise signals during adaptive task switching

Tarek Shafei[a*] · Niels A. Kloosterman[b*]

[a]Study program Psychology - Cognitive Systems, Universität zu Lübeck, Lübeck, Germany
[b]Institute of Psychology, Universität zu Lübeck, Lübeck, Germany
*Corresponding author, email: tarek.shafei@student.uni-luebeck.de; n.kloosterman@uni-luebeck.de

Received 19 February 2026; Accepted 10 September 2026; Published online 27 October 2026

© 2026 Shafei *et al.*; licensee Infinite Science Publishing
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Cognitive flexibility allows rapid behavioral adjustment to unexpected changes in task demands, reflecting distinct effects of surprise magnitude and belief updating. Pupil-linked arousal is known to track surprise, but whether and how it supports our ability to adapt to the environment remains unknown. To address this, we tested how explicitly cued versus implicit task switches shape behavior by tracking computational surprise and pupil-linked arousal in low vs. high volatility environments. Participants (N = 16) performed a probabilistic task-switching paradigm with occasional faulty feedback while pupil size was recorded. Behavior showed improved post-switch accuracy recovery in explicit blocks and during high volatility. A computational Bayesian belief-updating model yielded dissociable surprise signals: outcome surprise about the feedback was associated with poorer post-switch accuracy recovery, whereas Bayesian surprise reflecting belief updating was associated with better recovery. Pupil deconvolution suggests separable early vs. late phasic pupil responses related to outcome and Bayesian surprise, respectively. Together, these results link cognitive flexibility to distinct computational and arousal signatures.

I. Introduction

Cognitive flexibility is a key executive function that enables humans to rapidly adapt behavior when environmental demands change. Task-switching paradigms demonstrate that switching tasks incurs accuracy costs relative to repeating the same task [1]. Phasic arousal mediated by neuromodulatory catecholamine systems might play an important role in the neural computations underlying flexibility, because arousal responses peak at the time of surprising events that call for flexibility [2]. Pupillometry can help quantify these processes, as it provides an applicable, non-invasive window into arousal dynamics [3]. The size of the pupil has been established as an index of surprise-related processes involved in decision making [4], but little is known about the computational role of pupil-linked arousal in cognitive flexibility.

We distinguish two types of flexibility: *explicit* (signaled by cues) and *implicit* (inferred from feedback). In real-world settings (e.g., sports), explicit flexibility occurs when it is clear that the environment has changed (e.g., signals of a coach), whereas implicit flexibility is required when changes are only indirectly indicated (e.g., the opponent becomes fatigued). Few studies have combined both types within the same paradigm, limiting the understanding of how individuals achieve cognitive flexibility in explicit and implicit settings.

Two computationally distinct forms of surprise can support explicit and implicit cognitive flexibility. Assuming that a subject constantly updates their belief about the state of the environment, Outcome Surprise (OS) captures how unexpected a piece of feedback is given that belief (*"Whoa, didn't see that coming!"*). Bayesian surprise (BS), in contrast, reflects the extent to which feedback updates this belief (*"Hmm… maybe the rule switched."*). These signals have been linked to partly distinct cate-

10.18416/SCP.2026.2711
© 2026 Infinite Science Publishing

cholaminergic systems, with OS resembling dopaminergic prediction-error signaling [5] and BS relating more to noradrenergic belief updating [6].

OS and BS can dissociate depending on environmental volatility, for example in terms of the expected frequency of task-rule changes. In low-volatility environments, surprising feedback is typically treated as noise, eliciting high OS but little belief updating (low BS). In contrast, in times of high volatility, surprising outcomes occur more frequently and are therefore less unexpected (low OS), yet remain diagnostic of changes in the underlying state (high BS). Manipulating volatility thus provides a principled way to selectively engage and dissociate these computational forms of surprise.

We tested how flexibility type and volatility shape post-switch behavior, computational surprise, and pupil-linked arousal. We predicted that (i) explicit cueing and low volatility enhance post-switch accuracy, (ii) behavior reflects Bayesian belief updating with dissociable OS and BS dynamics, and (iii) feedback-locked pupil dilation tracks both surprise signals.

II. Materials and Methods

Participants We recruited 17 participants from the University of Lübeck. All participants reported normal or corrected-to-normal vision and provided written informed consent. An a priori power analysis (G*Power 3.1; repeated-measures ANOVA, within factors; 4 measurements) assuming $f = 0.25$, $\alpha = .05$, power = .80, $r = .50$, and $\epsilon = .75$ yielded a required sample size of $N = 29$. Due to feasibility constraints, and after excluding one dataset due to data-quality concerns, we analyzed $N = 16$ (age $M = 22.2$, $SD = 3.7$, 12 female).

Experimental design We used a 2 × 2 within-subject design (Flexibility type, *explicit* vs. *implicit*; Volatility, *low* vs. *high*). Participants completed 8 blocks of 100 trials, covering each of the four conditions twice. Rule switches were rare in low-volatility blocks (hazard rate $h = 0.05$) and frequent in high-volatility blocks ($h = 0.25$).

In each trial, participants performed one of two detection tasks on digit 2, 4, 7, or 9: a judgement of parity (even, yes/no) or of magnitude (larger than 5, yes/no) (Fig. 1A). To avoid stimulus–response confounding between tasks, we used a diagnostic digit set such that the correct response differed between the tasks. The responses were made using keyboard keys `E` and `I` (yes/no mapping counterbalanced). Participants completed brief guided examples and short practice blocks. Stimulus duration was individually titrated during training using a staircase to target ∼85% accuracy to avoid ceiling level performance and then fixed for the main task.

Each trial started with fixation, followed by a cue period (Fig. 1A). In *explicit* blocks, an informative symbol cue (÷ for parity, > for magnitude) was presented briefly and then masked. In *implicit* blocks, a neutral cue ("#") was shown without task information. After a short cue–stimulus interval, the digit appeared for the individually calibrated duration and participants responded via key press. Feedback was presented after the response, followed by an inter-trial interval. Feedback was probabilistic, with incorrect feedback on 20 % of trials to introduce ambiguity about whether errors reflected noise or a rule switch. Stimuli were designed to be isoluminant.

Testing took place in a dimly lit laboratory. Pupil size of both eyes was recorded continuously using an EyeLink 1000 (SR Research) eye tracker. We applied linear blink interpolation, a 4 Hz low-pass Butterworth filter, and Z-scoring across blocks to the pupil data using FieldTrip [7] in MATLAB R2025b (The Mathworks).

Data analysis Behavioral outcomes were accuracy and post-switch vs. post-repeat accuracy recovery ($d\mathrm{Acc}_{\mathrm{post1}} = \mathrm{Acc}_{\mathrm{post1}} - \mathrm{Acc}_{\mathrm{repeat}}$). Effects of Flexibility type and Volatility on these measures were assessed using repeated-measures ANOVAs with planned post-hoc comparisons. P-values were Holm–Bonferroni adjusted within each family of planned tests.

We fitted a Rescorla Wagner (RW) reinforcement model and two Bayesian inference models (BAYES and BAYESCUE) to participants' trial-wise choices. In addition to choice noise and bias, the models estimated either a learning rate about reward (RW) or a hazard-rate parameter capturing the expected rate of rule changes (BAYES and BAYESCUE). BAYESCUE in addition also modeled the explicit task cue by setting belief state in each trial to $p_t = 0.5 \pm 0.45$, as specified by cue_t. Models were fitted using maximum likelihood estimation (MLE). Model comparison relied on the Bayesian information criterion (BIC).

From the fitted belief trajectories, we derived two trial-wise surprise measures: Outcome Surprise (unexpectedness of feedback y_t given the current belief p_t):

$$\mathrm{OS}_t = -\log P(y_t \mid p_t) \tag{1}$$

and Bayesian Surprise (the belief update induced by feedback), defined as the Kullback–Leibler divergence between the posterior belief p_t^{post} and the prior belief p_t^{prior}:

$$\mathrm{BS}_t = p_t^{\mathrm{post}} \log \frac{p_t^{\mathrm{post}}}{p_t^{\mathrm{prior}}} + (1 - p_t^{\mathrm{post}}) \log \frac{1 - p_t^{\mathrm{post}}}{1 - p_t^{\mathrm{prior}}} \tag{2}$$

Model-derived OS and BS were analyzed as a function of flexibility type and volatility.

To isolate signatures of OS and BS in pupil size, we used deconvolution analysis in the Unfold toolbox [8] to estimate linear pupil response kernels with separate, time-expanded event regressors for cue, stimulus, response, and feedback as well as parametric modulators for OS and BS time-locked to feedback, for each condition.

10.18416/SCP.2026.2711
© 2026 Infinite Science Publishing

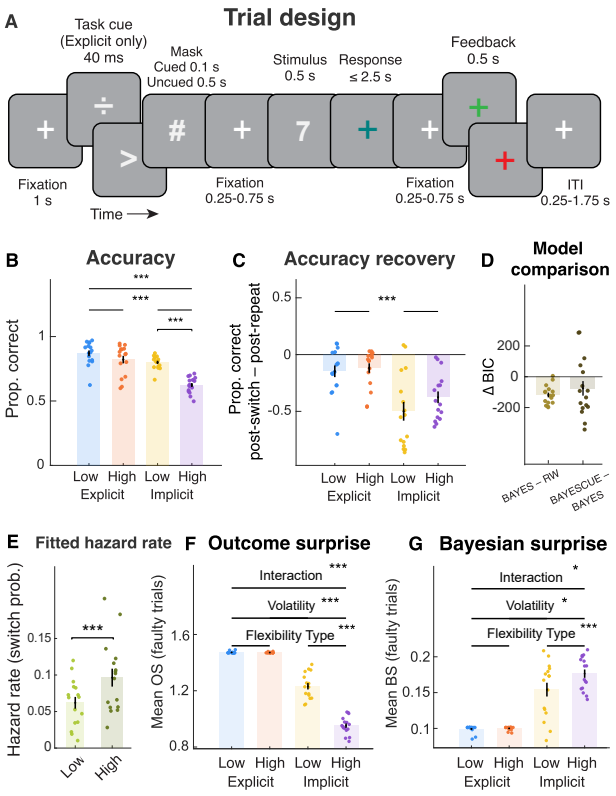


Figure 1: (A) Trial design. (B) Accuracy and (C) post-switch accuracy recovery. (D) Overall model comparison (ΔBIC; lower values indicate better fit). (E) Fitted hazard rate (h) across volatility. (F) Mean Outcome and (G) Bayesian Surprise on faulty-feedback trials. Bars show means (±SEM); dots show participants. Significance markers indicate rmANOVA effects. Asterisks denote Holm–Bonferroni–corrected significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

III. Results and Discussion

Model-free behavior Behavior was shaped by both flexibility type and environmental volatility. Participants showed higher accuracy in explicit vs. implicit blocks (Fig. 1B) ($F(1,15) = 31.72$, $p < .001$, $\eta_p^2 = .68$), and in low vs. high volatility ($F(1,15) = 43.42$, $p < .001$, $\eta_p^2 = .74$). Notably, flexibility type and volatility interacted ($F(1,15) = 20.32$, $p < .001$, $\eta_p^2 = .58$), consistent with volatility modulating how informative feedback is for updating task sets. Post-switch accuracy recovery was better in explicit than implicit blocks (i.e. closer to zero; Fig. 1C) ($F(1,15) = 18.92$, $p < .001$, $\eta_p^2 = .56$) and was better in high vs. low volatility ($F(1,15) = 5.39$, $p < .05$, $\eta_p^2 = .26$), suggesting greater adaptability when switches occurred more often.

Computational modeling Model comparison indicated that BAYESCUE provided a better overall fit than RW and BAYES (Fig. 1D), justifying its use to extract latent belief dynamics and trial-wise surprise. The model fits captured the intended manipulation of environmental volatility, with a higher fitted hazard rate in high- than low-volatility conditions, $t(15) = -4.45$, $p < .001$ (Fig. 1E).

For faulty-feedback trials, both OS (Fig. 1F) and BS (Fig. 1G) showed interactions between flexibility type and volatility (OS: $F(1,15) = 164.58$, $p < .001$, $\eta_p^2 = .92$; BS: $F(1,15) = 8.86$, $p = .009$, $\eta_p^2 = .37$), indicating that the effect of volatility differed between explicit and implicit blocks. Both signals also showed reliable main effects of flexibility and volatility (all $p < .012$). Finally, a within-subject rmANOVA revealed a reliable three-way interaction between flexibility type, volatility and surprise type ($p < .001$), consistent with qualitatively different volatility sensitivity of OS and BS across explicit and implicit modes.

Surprise–performance associations Assuming that belief updating rather than surprise magnitude supports flexible task adjustment, we predicted that higher OS would impair post-switch recovery, whereas higher BS would support it. Indeed, higher OS was associated with poorer post-switch accuracy recovery (Spearman's $\rho = -.86$, $p < .001$). In contrast, higher BS predicted better recovery ($\rho = .75$, $p = .001$; Fig. 2). Partial correlations showed that OS remained associated with recovery when controlling for BS (OS|BS: $r = -.640$, $p = .010$), whereas BS did not remain significantly associated with recovery when controlling for OS (BS|OS: $r = -.157$, $p = .577$).

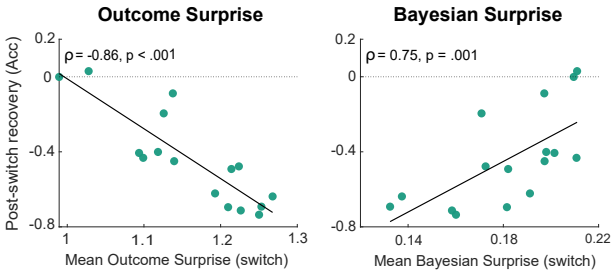


Figure 2: Mean Outcome and Bayesian Surprise (collapsed across implicit blocks) plotted against post-switch accuracy recovery ($dAcc_{post1}$).

Pupil deconvolution We applied deconvolution to the pupil time series to dissociate feedback-evoked pupil responses related to OS and BS from overlapping event-related signals (stimuli, motor responses). The resulting kernels were temporally separable, with BS peaking later than OS following feedback (Fig. 3). This may align with different computational roles reflected in the pupil of early, dopaminergic and later, noradrenergic neuromodulation in outcome surprise and belief updating, respectively.

10.18416/SCP.2026.2711
© 2026 Infinite Science Publishing

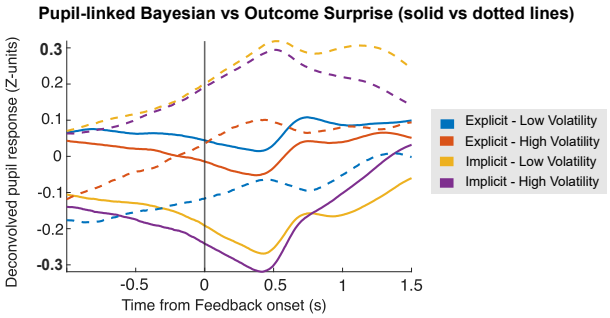


Figure 3: Deconvolution estimates of feedback-locked pupil response kernels for Outcome Surprise and Bayesian Surprise.

Discussion We found a clear dissociation between cue-guided (explicit) and feedback-based (implicit) flexibility, and evidence that volatility shifts which surprise type matters for behavior. Explicit blocks showed higher accuracy and faster post-switch recovery, consistent with externally instructed task appearance that reduces reliance on noisy feedback [1]. In implicit blocks, where participants must infer rule switches from ambiguous outcomes, belief updating becomes central. Consequently, the Bayesian hazard-rate model best accounted for choices and recovered higher hazard-rate estimates under high volatility, consistent with volatility-dependent updating [2]. Critically, OS and BS diverged, predicting poorer and better post-switch recovery, respectively. This suggests that BS captures diagnostic updating that may support adaptation, while unsigned surprise can be disruptive. However, only the OS association remained significant when controlling for the other surprise measure. Pupil deconvolution converged with this account by revealing temporally separable feedback-locked components related to OS and BS, supporting the view that pupil-linked arousal reflects multiple computations rather than a single "surprise" signal [3, 4, 6]. Together, these results motivate causal tests to determine whether modulating arousal-related systems can selectively enhance BS-linked adaptation, e.g., using targeted electrical brain stimulation.

Our sample size was modest ($N = 16$), which limits power for smaller effects and increases uncertainty in correlational surprise–behavior links. We are continuing data collection toward the planned sample size of $N \approx 29$.

IV. Conclusion

We combine a carefully controlled volatility-switching paradigm with computational modeling and pupil deconvolution to isolate latent surprise computations that drive adaptive control. This approach suggests that pupil-linked arousal can be decomposed into temporally separable components related to outcome surprise and belief updating. Computational pupillometry therefore provides a principled bridge from behavior to neuromodulatory mechanisms beyond traditional descriptive approaches.

Acknowledgments

The work has been carried out at the Institute of Psychology, Universität zu Lübeck and was supervised by Niels Kloosterman.

Author's statement

Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee. ChatGPT (OpenAI) was used as a supporting tool for code development and troubleshooting, including parts of the computational modelling workflow and the formulation of Wilkinson-style model equations for the pupil deconvolution analysis. It was additionally used for limited language editing and clarification of methodological descriptions. All AI-assisted outputs were independently checked, tested, and revised by the authors. The authors take full responsibility for the final manuscript, analyses, and interpretations.

References

[1] S. Monsell. Task switching. *Trends in Cognitive Sciences*, 7(3):134–140, 2003, doi:10.1016/S1364-6613(03)00028-7.

[2] M. R. Nassar. Toward a computational role for locus coeruleus/norepinephrine arousal systems. *Current Opinion in Behavioral Sciences*, 59:101407, 2024, doi:10.1016/j.cobeha.2024.101407.

[3] S. Joshi, Y. Li, R. M. Kalwani, and J. I. Gold. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1):221–234, 2016, doi:10.1016/j.neuron.2015.11.028.

[4] N. A. Kloosterman, T. Meindertsma, A. M. Van Loon, V. A. F. Lamme, Y. S. Bonneh, and T. H. Donner. Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, 41(8):1068–1078, 2015, doi:10.1111/ejn.12859.

[5] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997, doi:10.1126/science.275.5306.1593.

[6] S. Joshi and J. I. Gold. Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, 24(6):466–480, 2020, doi:10.1016/j.tics.2020.03.005.

[7] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:1–9, 2011, doi:10.1155/2011/156869.

[8] B. V. Ehinger and O. Dimigen. Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, 7:e7838, 2019, doi:10.7717/peerj.7838.

© 2026 Infinite Science Publishing