

A solid frame for the window on cognition: Modeling event-related pupil responses

Department of Psychiatry, Psychotherapy, and Psychosomatics,
University of Zurich, Zurich, Switzerland
Neuroscience Center Zurich, University of Zurich,
Zurich, Switzerland

Christoph W. Korn



Department of Psychiatry, Psychotherapy, and Psychosomatics,
University of Zurich, Zurich, Switzerland
Neuroscience Center Zurich, University of Zurich,
Zurich, Switzerland

Dominik R. Bach

Wellcome Trust Centre for Neuroimaging,
University College London, London, United Kingdom



Pupil size is often used to infer central processes, including attention, memory, and emotion. Recent research has spotlighted its relation to behavioral variables from decision-making models and to neural variables such as locus coeruleus activity and cortical oscillations. As yet, a unified and principled approach for analyzing pupil responses is lacking. Here we seek to establish a formal, quantitative forward model for pupil responses by describing them with linear time-invariant systems. Based on empirical data from human participants, we show that a combination of two linear time-invariant systems can parsimoniously explain approximately all variance evoked by illuminance changes. Notably, the model makes a counterintuitive prediction that pupil constriction dominates the responses to darkness flashes, as in previous empirical reports. This prediction was quantitatively confirmed for responses to light and darkness flashes in an independent group of participants. Crucially, illuminance- and nonilluminance-related inputs to the pupillary system are presumed to share a common final pathway, composed of muscles and nerve terminals. Hence, we can harness our illuminance-based model to estimate the temporal evolution of this neural input for an auditory-oddball task, an emotional-words task, and a visual-detection task. Onset and peak latencies of the estimated neural inputs furnish plausible hypotheses for the complexity of the underlying neural circuit. To conclude, this mathematical description of pupil responses serves as a prerequisite to refining their relation to behavioral and brain indices of cognitive processes.

Introduction

Measures of pupil size have long been used to enlighten the understanding of diverse psychological processes (Granholm & Steinhauer, 2004), including attention (Binda, Pereverzeva, & Murray, 2013; Wang & Munoz, 2015; Wierda, van Rijn, Taatgen, & Martens, 2012), perception (Einhäuser, Stout, Koch, & Carter, 2008; Kloosterman et al., 2015), memory (Goldinger & Papesh, 2012; Kafkas & Montaldi, 2011; Qin, Hermans, van Marle, & Fernandez, 2012), and emotion (Bradley, Miccoli, Escrig, & Lang, 2008; Prehn et al., 2013; Preller et al., 2014). Recent research has spotlighted the relationship between pupil size and behavioral variables derived from formal decision-making models (Browning, Behrens, Jocham, Reilly, & Bishop, 2015; Nassar et al., 2012; Preuschoff, 't Hart, & Einhäuser, 2011) and with neural variables such as the locus coeruleus–noradrenergic system in humans (Aston-Jones & Cohen, 2005; Eldar, Cohen, & Niv, 2013; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010) or cortical activity in rodents (McGinley, David, & McCormick, 2015; Reimer et al., 2014) and humans (Yellin, Berkovich-Ohana, & Malach, 2015; Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014). The analysis approaches of pupil measurements currently used in the literature are rather diverse and lack formal specifications. This is in contrast to the formal biophysical models used for neuroimaging analysis (Friston, 2005) and to the recent development of principled approaches for psychophysiological model-

Citation: Korn, C. W., & Bach, D. R. (2016). A solid frame for the window on cognition: Modeling event-related pupil responses. *Journal of Vision*, 16(3):28, 1–16, doi:10.1167/16.3.28.

doi: 10.1167/16.3.28

Received November 3, 2015; published February 19, 2016

ISSN 1534-7362



This work is licensed under a Creative Commons Attribution 4.0 International License.

Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/Journals/JOV/934914/> on 04/04/2016

Task	Initial sample	Final sample	Age (years; $M \pm SD$)	Female/male
Continuous illuminance (5 s)	23	20	24.6 \pm 3.4	12/8
Illuminance flashes (200 ms)	14	14	24.6 \pm 4.8	6/8
Auditory oddball	92	66	24.2 \pm 3.9	40/26
Emotional word	37	27	26.3 \pm 4.6	20/7
Visual detection	21	13	23.2 \pm 3.4	8/5

Table 1. Overview of participants.

ing (Bach, Flandin, Friston, & Dolan, 2009; Bach & Friston, 2013; Paulus, Castegnetti, & Bach, 2016).

Specifically, most pupil-response studies report baseline-corrected averages, or the relation of central variables with pupil size, within specific time windows. Choosing these time windows and baseline correction or regression methods engenders implicit assumptions on how central input generates pupil responses (Bach & Friston, 2013). As these assumptions are not explicitly stated or standardized, such procedures imply a risk of ad hoc definitions. Here we sought to develop an explicit psychophysiological model for the relationship between neural input and pupil size. This model can be empirically tested, and can be applied to constrain the possible causes of observed data. For other psychophysiological measures, using explicit causal models tends to improve the separation of noise from features of interest, thereby increasing statistical sensitivity (Bach & Friston, 2013).

In keeping with these approaches, we harness the concept of linear time-invariant (LTI) systems, in which time series of inputs are convolved with response functions. Assuming such LTI systems, putative inputs can then be inferred from observed data.

We make the assumption that pupillary responses due to illuminance changes and “psychological” inputs impinge on the same peripheral biophysical system of nerve terminals and muscles. Illuminance changes elicit pupil responses via a rather well-described midbrain circuit (McDougal & Gamlin, 2008) comprising two antagonistic systems: Pupil dilation (mydriasis) relies on the radial *M. dilatator pupillae*, which receives sympathetic innervation (via preganglionic neurons from the spinal cord and postganglionic neurons from the superior cervical ganglion). Pupil constriction (miosis) is mediated by parasympathetic innervation of the circular *M. sphincter pupillae*. In the parasympathetic branch, the preganglionic neurons originate in the Edinger-Westphal nucleus within the midbrain and synapse on the postganglionic neurons in the ciliary ganglion. The Edinger-Westphal nucleus indirectly receives illuminance-mediated inputs from the retina and also seems to constitute the target for upstream inputs that are not related to illuminance, such as those arising from the locus coeruleus (McDougal & Gamlin, 2008). Although the precise anatomical route for

nonilluminance inputs has yet to be firmly established, it can be assumed that both illuminance and non-illuminance inputs share a common final (neural and muscular) pathway.

Therefore, it seems plausible to first establish a model of pupil responses to illuminance changes within the framework of LTI systems and to consequently use this model to infer neural input related to other causes. For this purpose, it is not necessary to capture all biophysical details. Our approach thus complements a recent attempt to describe the mechanical properties and the neuromodulatory inputs of the pupillary muscles (Fan & Yao, 2011). The biophysical realism of this model comes at the cost of a substantial number of free parameters (e.g., 10 parameters in the case of the model by Fan & Yao) that need to be inferred from empirical data. Our phenomenological approach identifies impulse response functions of the pupillary system without reference to the underlying mechanism. This approach therefore furnishes a parsimonious model that can be readily applied to various psychophysical experiments.

In sum, our model development follows two main steps. In the first step, we derive a model of pupil responses to illuminance changes within the framework of LTI systems. In the second step, we harness the obtained LTI systems to infer the shape of the inputs that elicit pupil responses in three different perceptual tasks.

Methods

Participants

Healthy participants who were not taking medications or drugs were recruited from the general and student population for five experiments (two illuminance tasks, three perceptual tasks) and received monetary compensation (see Table 1 for details). The a priori criterion for excluding sessions was more than 30% of data points missing (due to blinks, head movements, or fixation breaks; see later). The study, including the form of taking written informed consent, was conducted in accord with the Declaration of

Helsinki and was approved by the competent research ethics committee (Kantonale Ethikkommission Zürich, KEK-ZH-Nr. 2013-0328).

Illuminance tasks

All participants in the two illuminance tasks tested negative for color-vision deficiency (von Broschmann, 2011). The samples in the continuous-illuminance task and the illuminance-flashes task were independent. No participant in the two illuminance tasks reported vision impairments or a history of eye disease, except for two with mild strabismus.

Two participants performed only four instead of five sessions. Three participants were excluded completely because fewer than three sessions remained in the analyses (in the final sample, three participants had three sessions and three had four sessions; all other participants had five sessions). Four participants wore glasses (diopter left: -3.1 ± 3.9 ; diopter right: -2.6 ± 3.6), and five wore contact lenses (diopter left: -1.2 ± 3.4 ; diopter right: -1.7 ± 3.4). The average percentage of missing data points was similar when comparing participants with glasses or with lenses to those without eyesight correction (Wilcoxon rank-sum tests, all $p > 0.2$).

In the illuminance-flashes task, three participants wore glasses (diopter left/right: 0.3 ± 2.0) and three wore contact lenses (diopter left: -3.4 ± 1.3 ; diopter right: -3.6 ± 1.1). For two participants, one session was excluded because more than 30% of data points were missing.

Auditory-oddball task

For the intertrial interval (ITI) of 1 s, seven participants had a sampling rate of 250 Hz; in the others it was 500 Hz. All data were down-sampled to 250 Hz. For the ITI of 2 s, six participants had a sampling rate of 1000 Hz; in the others it was 500 Hz. All data were down-sampled to 500 Hz.

Emotional-words task

All participants but two were native German speakers; these two participants were fluent in German and had learned it before the age of 5. Participants were excluded if one of the two sessions had more than 30% missing data points.

Visual-detection task

In six participants, the sampling rate was 250 Hz, while in the others it was 500 Hz. All data were down-sampled to 250 Hz.

Apparatus

Testing was performed in a dark, soundproof chamber (with a background illumination of 3.4 lx provided by the camera and the monitor lights). Participants' heads were positioned on a chin rest 70 cm in front of the monitor (Dell P2012H; 20 in., set to an aspect ratio of 5:4; 60-Hz refresh rate). Pupil diameters and gaze direction for both eyes were recorded with an EyeLink 1000 system (SR Research, Ottawa, Canada) at a sampling rate of 500 Hz unless otherwise indicated. We used the nine-point calibration implemented in the EyeLink 1000 software for calibrating gaze direction.

Illuminance levels were determined off-line by fixing a luxmeter (Digital Luxmeter MS-1300, Voltcraft, Hirschau, Germany) to the chin rest at the position of participants' eyes. Thus the total luminous flux entering the eye is the product of these illuminance levels and the measured pupil area. For our model development, we used illuminance E_v rather than luminance L_v , but we also provide all relevant luminance values to allow comparison to previous studies. For consistency, we thus use the term *illuminance* throughout the article.

Experimental setup

Continuous-illuminance task

Each of five sessions started with a resting period of 45 s during which a medium-gray screen was shown (width: 31.13° visual angle, height: 24.91° ; 46.1 cd/m^2 , [128, 128, 128] in the RGB color scheme, 7.3 lx). In each of 24 trials (six per illuminance level), a circle appeared in the screen center on the medium-gray background and disappeared after 5 s (diameter: 16.58° ; see Figure 1). The circles had one of four colors (black: 33.5 cd/m^2 , [0, 0, 0], 5.3 lx; dark gray: 36.7 cd/m^2 , [64, 64, 64], 5.8 lx; light gray: 60.7 cd/m^2 , [191, 191, 191], 9.6 lx; white: 84.1 cd/m^2 , [255, 255, 255], 13.3 lx), followed by a 5-s background screen. By design, illuminance changed both when the circles appeared and when they disappeared. Thus, there were 24 illuminance changes from darker to brighter (appearance of light-gray and white circles; disappearance of dark-gray and black circles) and 24 illuminance changes from brighter to darker (appearance of dark-gray and black circles; disappearance of light-gray and white circles). A red fixation cross (height/width: 1.47°) remained on-screen throughout and was taken into account for establishing the aforementioned illuminance levels. Participants were instructed to fixate this cross, but otherwise no response was required.

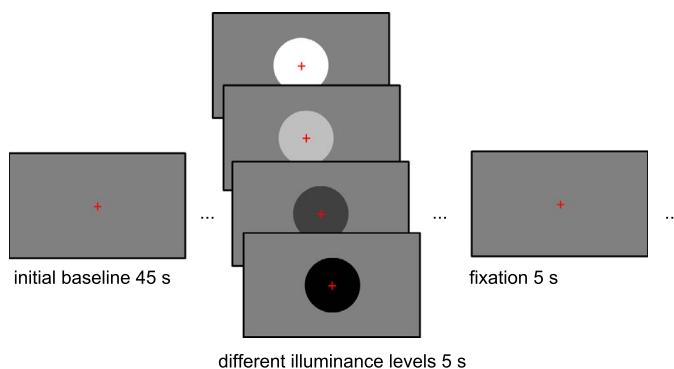


Figure 1. Outline of the illuminance tasks. After a baseline period of 45 s, participants were presented with circles of four different illuminance levels (diameter: 16.58°). In the continuous-illuminance task, these circles remained on screen for 5 s. In the illuminance-flashes task, they remained on screen for 200 ms. Participants were asked to fixate a red cross throughout the task. Both the onset and the offset of the circles elicited pupil responses.

Illuminance-flashes task

The setup was the same as for the continuous-illuminance task, with the sole difference that the circles remained on-screen for 200 ms only. The ITI was 5 s.

Auditory-oddball task

Standard and oddball tones were sine tones (50-ms length; 10-ms ramp; 440 or 660 Hz), delivered via headphones (HD 518, Sennheiser, Wendemark-Wennebostel, Germany) at approximately 60 dB and counterbalanced across participants. During the entire task a red fixation cross (height/width: 1.47°) was presented on a medium-gray background (46.1 cd/m^2 , [128, 128, 128], 7.3 lx). Participants were instructed to fixate and to press a key upon hearing the oddball tone. If participants did not answer in time (i.e., before the next tone) the words “No answer” appeared. Because of the different ITIs, the number of oddballs and standards differed between the three groups (1-s ITI: 30 oddballs, randomized number of standards, mean number = 463.1 ± 8.0 ; 2-s ITI: 40 oddballs, 160 standards; 3-s ITI: 30 oddballs, 120 standards). As intended, participants’ performance was close to ceiling (correctness: $99.9 \pm 0.6\%$). Mean reaction times were similar for the three ITIs (1 s: 395 ± 62 ms; 2 s: 383 ± 80 ms; 3 s: 395 ± 87 ms).

Emotional-words task

Participants were presented with 100 neutral and 100 negative five-letter nouns from the Berlin Affective Word List Reloaded (Võ et al., 2009; see the next subsection for stimulus selection). Each word was

presented for 1 s in the center of the screen using capital letters in gray type (Lucida Console, [128, 128, 128]; height: 0.66° , width: 3.27°) on a gray background (48.7 cd/m^2 , [128, 128, 128]) followed by the letter string “XXXXX,” which was presented instead of a fixation cross for a total of 4 s. Within the accuracy of the luxmeter ($\pm 5\%$), illuminance was constant throughout the task (7.7 lx). Participants were tasked to judge the word as negative or neutral. If they failed to respond in time or pressed a wrong key, the prompt “Please answer” appeared on-screen. As expected, participants judged negative words to be more negative than neutral words (negative words: $70.5\% \pm 11.6\%$; neutral words $9.9\% \pm 8.8\%$), $t(26) = 25.73$, $p < 0.001$.

Emotional-words task: Stimulus selection

We first selected all five-letter nouns with two syllables from the Berlin Affective Word List Reloaded. Based on the list’s norm ratings, neutral words were selected so that valence values lay between -0.5 and $+0.5$ on a 7-point scale from -3 to $+3$ (0.08 ± 0.26) and arousal values lay between 2 and 3 on a 5-point scale from 1 to 5 (2.40 ± 0.24). Negative words were selected so that valence values lay between -3 and -0.75 (-1.48 ± 0.49) and arousal values lay between 2 and 5 (3.39 ± 0.60). As intended, the pair-wise comparisons between the two word sets were significant both for valence and for arousal ratings (all Bonferroni-corrected $ps < 0.001$). Words were randomly split into two matched lists of 50 neutral and 50 negative words each. Assignment of the two lists to the two sessions was counterbalanced across participants, and word presentation within the sessions was randomized.

Visual-detection task

A stream of 10 distractors (digits 0–9) was presented in white font (Arial, [255, 255, 255]; height: 1.00° , width: 0.66°) on a black screen (24.0 cd/m^2 , [0, 0, 0]). Participants were instructed to press a key upon detecting the target stimulus, a red cross that was interspersed in the stream. In total, there were 11 crosses and 599 digits. Stimuli remained on-screen for 200 ms, followed by an ITI of 800 ms. Illuminance was constant throughout the task (3.8 lx) within the accuracy of the luxmeter. Participants’ performance was at ceiling ($100\% \pm 0\%$; RTs: 427 ± 39 ms).

All experiments were programmed in MATLAB using Cogent 2000 (<http://www.vislab.ucl.ac.uk>).

Data preprocessing

Saccades and fixations were detected using the online parsing algorithm implemented in the EyeLink 1000

system, which detects artificial changes in pupil position caused by partial occlusion of the pupil. Further analysis was performed after import of the data into MATLAB (Version R2013a, MathWorks, Natick, MA). We analyzed the time series from the beginning of the first event until 5 s after the last event. Missing data points, which could result from blinking or from brief head movements, were linearly interpolated. Participants were asked to fixate the center of the screen in all tasks. Breaking fixation can distort the pupil size measured by a video-based eye tracker due to the dependence of the measurement on the gaze angle (Hayes & Petrov, 2015). We therefore treated all data points for which x or y gaze positions exceeded an a priori threshold of $\pm 4^\circ$ as missing data points and interpolated them linearly. Including data points for which fixation was broken resulted in quantitatively very similar results. Within each participant, we analyzed the pupil (left or right) for which more data points were available. The pupil diameters of the two eyes were highly correlated (mean Pearson's r across participants in the continuous-illuminance task: 0.98 ± 0.05). Time series were z-scored within each session after interpolating missing values. This accounts for between-subjects variance in overall pupil size (including variance related to corrective lenses).

Additionally, we report pupil diameter values in millimeters for comparability. The pupil diameters are recorded by the EyeLink 1000 system in arbitrary units, which are proportional to true physical diameters (Hayes & Petrov, 2015). To determine the numerical relationship between diameters in arbitrary units and millimeters for our setup, we adapted a procedure described previously (Hayes & Petrov, 2015). Specifically, we printed a series of black circles with diameters ranging from 2.5 to 7 mm (in 0.5-mm increments) on white paper and verified their sizes using calipers. To mimic the corneal reflection, small holes were pinched into the black circles and silver foil added underneath the holes. These artificial pupils of known diameters were measured with the EyeLink 1000 system using the same setup and specifications as for the experiments (i.e., pupil diameter in the ellipse mode). The correlation coefficient was 0.99 and the numerical relationship was diameter (in millimeters) = $0.07 + \text{diameter (in arbitrary units)} / 1325$.

Analyses of mean responses

We extracted data segments following trial onset and averaged these segments first within and then across sessions and participants. In the continuous-illuminance task, these segments were 10 s long (i.e., including both appearance and disappearance of the stimulus). In the other tasks, the segments were 4 s

(auditory-oddball task, visual-detection task) or 5 s (illuminance-flashes task, emotional-words task) long. Grand mean responses were baseline corrected by setting the first data point to zero. Since pupil responses to the oddball and to the standard overlapped in time, we excluded the standards immediately before and after the oddball.

Modeling steady-state pupil size

To obtain the steady-state relationship of pupil size with measured illuminance, we averaged over the last 500 ms of the grand means of the 5-s segments for the five illuminance levels (i.e., means for middle gray resulted from averaging over the 24 fixation periods per session; means for the other illuminance levels resulted from averaging over the six respective trials per session). A previous review article has used a sigmoid function to model steady-state pupil size across a wide range of luminance levels and stimulus areas (Watson & Yellott, 2012). For the intended application of our model, a more restricted range suffices. Within an illuminance range of around 5–14 lx, the data had an exponential (not a sigmoid) shape, which is in qualitative agreement with the model by Watson & Yellot:

$$d(E_v) = C + A \exp(BE_v).$$

Here, d is the z-scored steady-state pupil diameter and E_v is the respective illuminance level in lx (lm/m^2). Parameters were estimated using ordinary least-squares minimization and a Nelder–Mead simplex search algorithm as implemented in the MATLAB function “fminsearch.”

Modeling responses to illuminance changes

We used the grand means from the continuous-illuminance task for deriving empirical response functions (RFs). Since mean responses differed in shape between dilations (illuminance changes from brighter to darker) and constrictions (darker to brighter), the pupil responses could not be fully described by a simple LTI system that only takes a continuous illuminance input. We therefore used a combination of two LTI systems: the first system describes responses to continuous illuminance input and the second system models the difference between constriction and dilation. This second system receives a brief input for any increase in illuminance.

Grand means were split into segments of 5 s, following increases and decreases in illuminance. The first data point was subtracted as a baseline, and data for each segment were divided by the predicted steady-state pupil response for this segment. These scaled grand

means were then averaged separately for dilation and for constriction, and represent the response of the pupillary system to a sudden change in continuous illuminance input. The RF can therefore be obtained from the time derivative of the pupil response (see the Appendix). For all approximations, we chose functional forms upon visual inspection of the time series. Parameters were not intended to reflect biophysical parameters.

To derive the RF of the first LTI system, we approximated the dilation derivative with a gamma probability density function (pdf; using ordinary least-squares minimization and a Nelder–Mead simplex search algorithm):

$$d(t) = c(\theta^k \Gamma(k))^{-1} t^{k-1} \exp(-t/\theta),$$

where d is the z -scored steady-state pupil diameter, t is time, Γ is the gamma function, and k , θ , and c are free parameters. Since the latency of the empirical responses was around 200 ms, we only used data points after the first 200 ms for fitting.

To derive the RF of the second LTI system, we approximated the difference between the dilation and constriction responses with another gamma pdf. To suppress noise, we first approximated the constriction derivative with a gamma pdf (parameters: $k = 2.76$, $\theta = 0.09 \text{ s}^{-1}$, $c = 0.31$) and then used the difference between the predicted dilation and predicted constriction responses to approximate the RF of the second LTI system with a gamma pdf. This second system models the steeper onset and undershoot of the constriction relative to the dilation.

As an alternative formulation of the first LTI system, we approximated the empirical derivative of the dilation with a Gaussian smoothed biexponential function (Bach et al., 2010):

$$d(t) = aN(t) * (E_1(t) + E_2(t)),$$

where $*$ is the convolution operator, $N(t)$ is a centered Gaussian function

$$N(t) = (2\pi\sigma)^{-0.5} \exp(-t^2/2\sigma^2)$$

with standard deviation σ , and $E_1(t)$ and $E_2(t)$ are exponential functions of the form

$$E_1(t) = \exp(-\lambda_1 t) \quad \text{and} \quad E_2(t) = \exp(-\lambda_2 t).$$

The final model will be implemented in the open-source MATLAB toolbox PsPM, which is freely available under the GNU General Public License and obtainable from <http://pspm.sourceforge.net>.

Testing illuminance-response models

Our model was based on grand mean responses from all participants. Thus, we sought to determine how

much variance this model can explain within individual participants. To this end, we computed explained variance per participant. Specifically, we used general linear convolution models (GLMs) using routines in PsPM. The explanatory variables (i.e., the regressors in the design matrix X) were formed by convolving the RFs with the assumed input. That is, the response of the first LTI system, which models the general influence of both dilation and constriction, was predicted by convolving the illuminance time series, scaled by the steady-state model, with the first RF. The second LTI system models the additional contribution of the constriction and was derived by convolving stick functions with unit amplitude at each illuminance change from dark to bright with the second RF. Inverting this GLM yields estimates for the participant-specific amplitude of the assumed input into the LTI systems. Under these parameters, the ratio of explained and observed variance was determined.

To assess baseline variance (i.e., variance in the absence of illuminance input), we calculated the proportion of variance during the 45-s baseline period without illuminance changes at the beginning of each block relative to the variance during the remainder of the block. This 45-s baseline period was not included in the GLMs. Periods with missing data (e.g., due to blinks) were removed before model inversion. We performed individual GLMs for each session and averaged explained variance (and proportion of baseline variance) across sessions and participants.

Additionally, we investigated how different filter settings influenced the proportion of explained variance under the best model. We tested different low- and high-pass frequencies using a unidirectional first-order Butterworth filter. For each filter setting, the RFs were fitted separately as described earlier. Within the range of tested high-pass frequencies (from 0.01 to 0.1 Hz in steps of 0.01 Hz), explained variance was smaller than in the unfiltered data. Across the tested low-pass frequencies (from 0.5 to 3.5 Hz in steps of 0.5 Hz), explained variance remained basically unchanged with respect to the results obtained from unfiltered data. Hence all following analyses were based on unfiltered data.

Modeling perceptual inputs

On the basis of the illuminance-response model, we aimed at estimating the inputs that elicit pupil changes in three different perceptual tasks. We assumed inputs with the form of gamma pdfs with three free parameters. In brief, we fitted the convolution of the RF with the assumed input to the normalized pupil

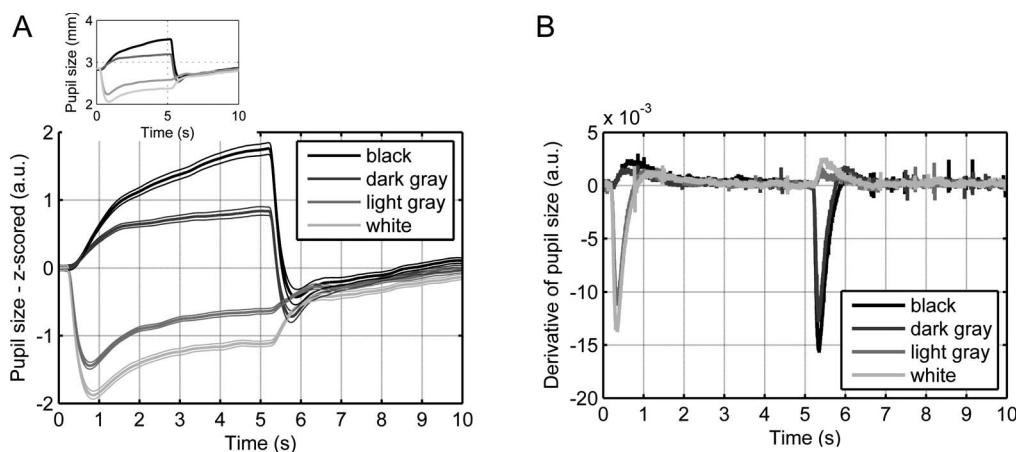


Figure 2. Pupil responses in the continuous-illuminance task. (A) Mean responses show a reliable ordering according to illuminance levels. Thick lines represent mean responses, and thin lines represent standard error of the mean (SEM). For comparability, the inset shows mean responses in millimeters. (B) Derivatives of the mean response indicate a steeper slope for constrictions versus dilations.

responses (using ordinary least-squares minimization and a Nelder–Mead simplex search algorithm).

Results

Pupil responses to illuminance changes

As expected, pupil size was inversely related to illuminance level in the continuous-illuminance task (Figure 2A). The shapes of the pupil responses were rather stereotypical across participants. The onset of dilation and constriction occurred around 200 ms after the appearance or disappearance of the stimuli. The slope of the constriction was steeper than the slope of the dilation (see Figure 2B for the time derivative of the empirical pupil responses). Additionally, the constriction showed an undershoot with a minimum at around 800 ms. Both for constriction and for dilation, pupil

size started to asymptote to steady-state level at around 1500 ms.

Modeling steady-state pupil size

In an intermediate step, we related steady-state pupil size (in arbitrary units) to illuminance (in lux) with an exponential function (parameters: $A = 32.56$, $B = -0.48 \text{ lx}^{-1}$, $C = -1.03$; Figure 3; relating steady-state pupil size in millimeters to luminance in candelas per square meter resulted in the following parameters: $A = 2.5 \text{ mm}$, $B = -1.1 \text{ m}^2/\text{cd}$, $C = 2.9 \text{ mm}$). We also fitted exponential functions to steady-state pupil sizes for each individual. The mean parameter estimates across participants were similar to those obtained from the fit to the grand means, in particular also for B , which determines the curvature ($A = 49.79 \pm 47.29$; $B = -0.50 \pm 0.15 \text{ lx}^{-1}$; $C = -1.05 \pm 0.28$).

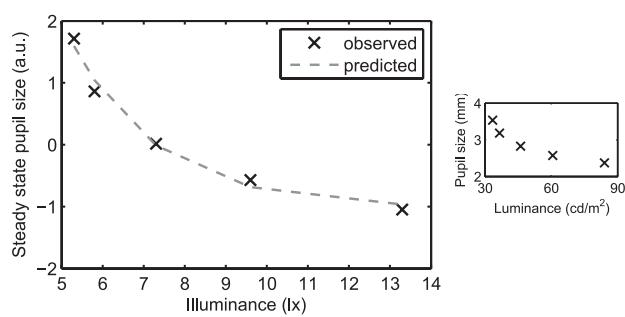


Figure 3. Steady-state pupil model. The relationship between steady-state pupil size and illuminance levels was predicted by an exponential function. For comparability, the inset shows the relationship between observed steady-state pupil sizes in millimeters and luminance levels in candelas per square meter.

Modeling responses to illuminance changes

The difference in shape between dilation and constriction prevented us from modeling them with a single LTI system that takes illuminance level as continuous input. This is physiologically plausible, since dilation and constriction result from two antagonistic muscles with different innervation. Therefore, we approximated pupil responses with a combination of two LTI systems: a first that receives (scaled) illuminance as continuous input and a second that models the difference between dilation and constriction (i.e., the steeper slope and the relative undershoot) and receives a brief input for positive illuminance changes. This arrangement is not supposed to reflect a bio-

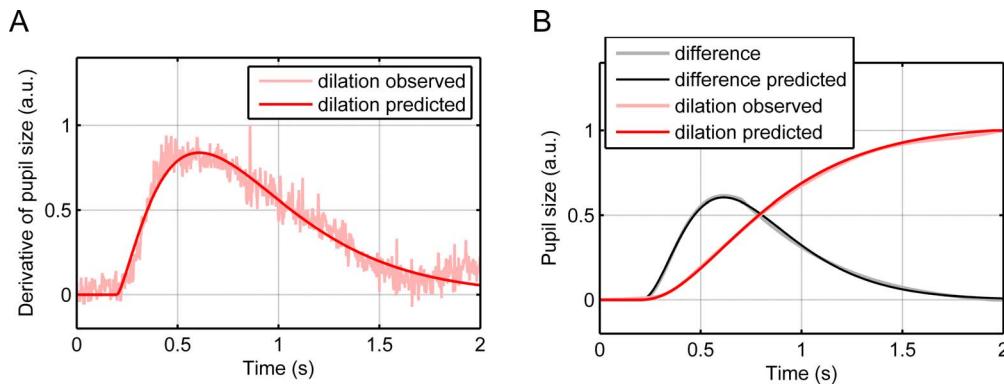


Figure 4. Model of pupil responses to illuminance changes. (A) The derivative of the mean dilation was well predicted by a gamma pdf, which constitutes the RF of the first LTI system. (B) The difference between dilation and constriction was well predicted by another gamma pdf, which constitutes the RF of the second LTI system. For completeness, the observed and predicted dilation according to the first LTI system are plotted.

physical reality, but rather to furnish a parsimonious mathematical description.

The RF for the dilation response could be approximated by a gamma distribution (parameters: $k = 2.40$, $\theta = 0.29 \text{ s}^{-1}$, $c = 0.77$; Figure 4A) and accounts for slow pupil responses to a change in illuminance input. In an alternative approach, we approximated the dilation response with a Gaussian smoothed biexponential function (parameters: $\sigma = 0.27$, $\lambda_1 = 2.04 \text{ s}^{-1}$, $\lambda_2 = 1.48 \text{ s}^{-1}$, $a = 0.004$).

To account for the faster time course and relative undershoot of constriction, we approximated the difference between dilation and constriction with another gamma pdf ($k = 3.24$, $\theta = 0.18 \text{ s}^{-1}$, $c = 0.43$; Figure 4B). This second LTI system receives brief inputs whenever an illuminance increase occurs. We assessed whether illuminance levels scale the input to the second

LTI system. To this end, we used the first LTI system to predict the pupil response in the four conditions that entail a constriction, and analyzed the residuals for each condition. The residual peak amplitudes were very similar for the four illuminance levels (middle gray to white: -1.4 ; middle gray to light gray: -1.2 ; dark gray to middle gray: -1.4 ; black to middle gray: -1.4), suggesting that the undershoot is not scaled by illuminance level or size of illuminance change. Hence, the second LTI system is assumed to receive a unit input whenever an increase in illuminance occurs.

To enhance clarity, we present an example time series for the two LTI systems in the continuous-illuminance task (Figure 5A). The combination of the two systems (scaled by the mean fitted parameter estimates) constitutes the predicted pupil-size trace (Figure 5B).

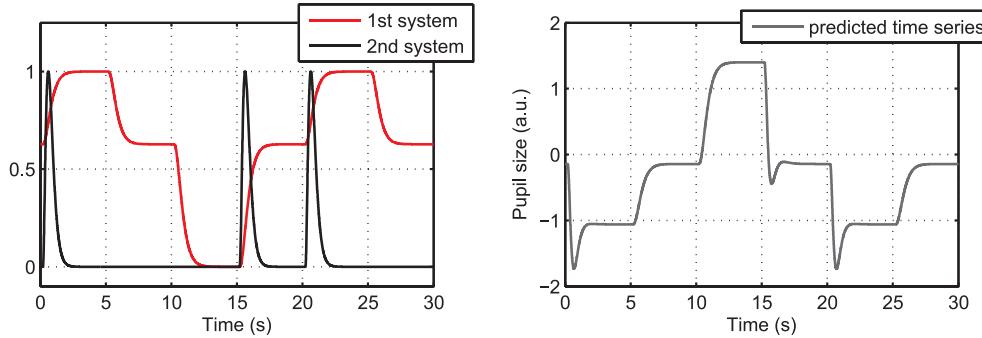


Figure 5. Illustration of the model for an example time series. (A) The time series is taken from the continuous-illuminance task, in which luminance changes every 5 s. The first LTI system captures the overall changes in pupil size due to varying illuminance levels as well as the shape of the dilation response. The second LTI system models the steeper onset as well as the undershoot of the constrictions that occur for positive illuminance changes. In this example time series, illuminance increases at 0, 15, and 20 s, and therefore the “peaks” of the second LTI system only occur following these time points. Please note that the two LTI systems are defined to be positive for higher illuminance inputs. Thus, fitted parameter estimates into both systems are negative, because higher illuminance inputs result in smaller pupil size. (B) The sum of the two LTI systems depicted in (A) (and the intercept) according to the mean fitted parameter estimates constitutes the predicted pupil time series. Since both parameter estimates are negative, the traces in (B) are changed in sign in relation to the traces in (A).

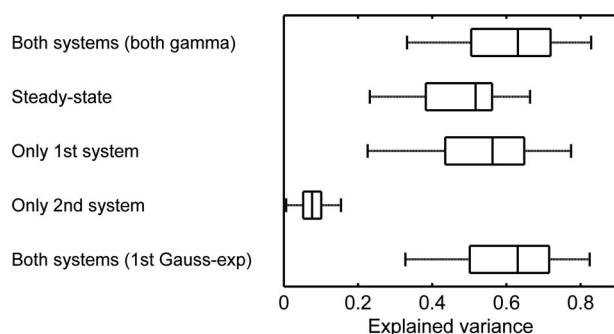


Figure 6. Model accuracy. The full model with both LTI systems explained around 60% of the variance. A pure steady-state model, as well as models with only one of the two systems, explained considerably less variance. A model in which the first LTI is described by a Gaussian smoothed biexponential function explained the same of amount of variance as the model in which both RFs are described by gamma pdfs.

Model accuracy

In the next step, we asked how much variance our model explained. Across participants, the full model explained $60.8\% \pm 13.6\%$ of the variance in the data (Figure 6; Table 2), and hence considerably more variance than a simple steady-state pupil model that used the scaled illuminance time series as a predictor or a model that included only either the first or the second LTI system. Using a Gaussian smoothed biexponential function as the RF for the first LTI system resulted in similar proportions of explained variance.

The explained variance can be seen in proportion to the variance during the 45-s baseline period without illuminance changes, which was $40.9\% \pm 14.8\%$ of the variance during illuminance changes and therefore close to the residual variance of the full model. This suggests that the full model explains approximately all variance elicited by illuminance changes, while residual variance stems from other processes.

Model validation

To validate our model, we assessed whether it predicts the counterintuitive pupil contractions in response to darkness flashes that have been observed previously (Barbur, Harlow, & Sahraie, 1992). To illustrate this prediction, we simulated a brief input of both a light- and

a dark-gray circle with the two LTI systems under different relative contributions of the two systems. Trivially, the model predicts a constriction after a light-gray circle (Figure 7A). Importantly, it also predicts the previously observed constriction following a darkness flash (Figure 7B). (For this illustration, the first LTI system was weighted by -1 and the second by -1.5 .) This is because the second LTI system, which models the steeper slope of constrictions relative to dilations, receives brief inputs to positive illuminance changes regardless of the ensuing duration of the illuminance input. In contrast, the first LTI system receives continuous (positive and negative) illuminance inputs but has a shallower slope. Loosely speaking, for a short flash of a dark stimulus, the input to the first system is not long enough to achieve a noticeable dilation. But the offset of the dark stimulus entails a positive illuminance change that elicits a sufficient input to the second system.

To test the prediction empirically, we presented an independent sample with circles of five different illuminance levels that were flashed on the screen for 200 ms (Figure 8A). In line with previous results (Barbur et al., 1992) and the predictions of our model, pupil traces were dominated by constrictions. On average, the GLM using the two LTI systems explained $23.5\% \pm 9.6\%$ of the variance in the data (Figure 8B; Table 2). In contrast, the simple steady-state model that included just the scaled illuminance input as a regressor explained only a negligible variance proportion. A model that included only the second LTI explained most of the variance in the task, while the first LTI system alone explained almost no variance.

Modeling perceptual inputs

In the previous sections, we derived a combination of LTI systems that describes pupillary responses to illuminance changes. In the following, we use this model to estimate the shape of the neural inputs generated by nonilluminance processes. This approach makes the assumption of a common final pathway in the peripheral pupillary system.

Auditory oddballs elicit reliable pupil responses and are not confounded by illuminance changes (Hong, Walz, & Sajda, 2014; Murphy, Robertson, Balsters, & O'Connell, 2011). Participants attended to standard and oddball tones with an ITI of either 1, 2, or 3 s. For

Task	Both LTI systems ^a	Steady-state input	First LTI system only	Second LTI system only	Both LTI systems ^b
Continuous illuminance (5 s)	60.8 ± 13.6	47.4 ± 11.6	53.4 ± 14.0	7.9 ± 3.7	60.5 ± 13.5
Illuminance flashes (200 ms)	23.5 ± 9.6	0.2 ± 0.1	3.3 ± 2.2	12.2 ± 7.5	—

Table 2. Proportion of explained variance in illuminance tasks (%; $M \pm SD$). Notes: ^aBoth fitted with gamma probability density function. ^bFirst fitted with Gaussian smoothed bi-exponential function.

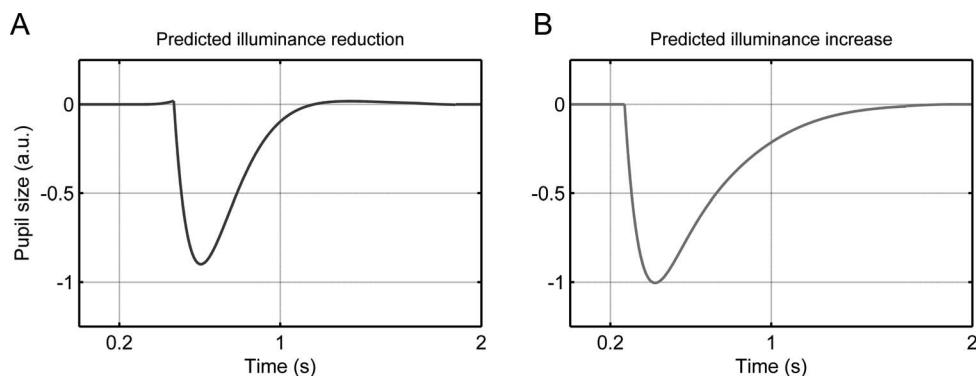


Figure 7. Model predictions for brief luminance inputs. (A) For an illuminance increase of 200 ms duration, the model trivially predicts a constriction. (B) Crucially, the model also predicts a constriction following a brief illuminance reduction. Please note the different onset latency in (B) versus (A).

all three ITIs, the pupil dilated reliably more in response to oddballs compared to standards in a time window of roughly 0.5–2 s after stimulus onset (Figure 9A through C). Standards also modulated pupil size in an ITI-dependent manner. Pupil responses to the oddball were thereby influenced by the responses to the standards that followed the oddball.

To fit the input relating to the oddball response alone, we subtracted the grand means of the pupil time series for standards from those for oddballs and then baseline-corrected the resulting traces at 200 ms after stimulus onset, because at this time point the inflection to the oddball response occurred. Since oddballs induced pupil dilation, we modeled these responses with the first LTI system. For all three ITIs, the input could be fitted well with gamma pdfs (Figure 9D, E). These had their peaks at biologically plausible time points—i.e., within the order of magnitude of around

300 ms in which oddball responses in the electroencephalograph are assumed to occur (Friedman, Cycowicz, & Gaeta, 2001).

Emotional words constitute a second type of stimulus that elicits pupillary responses independent of illuminance inputs (Bayer, Sommer, & Schacht, 2011; Võ et al., 2008). Negative nouns led to a slightly more pronounced and longer dilation than neutral nouns (Figure 10A). These responses could be well estimated by gamma inputs into the first LTI system (Figure 10B, C).

Detecting visual target stimuli in a stream of distractors also elicited reliable pupil dilation that was unrelated to illuminance changes (Figure 11A). The time course of this pupil response was similar to that of illuminance changes. Indeed, the estimation of a gamma input into the first LTI system showed that the neural input into the pupil system occurred at the same time as an illuminance input (Figure 11B), which

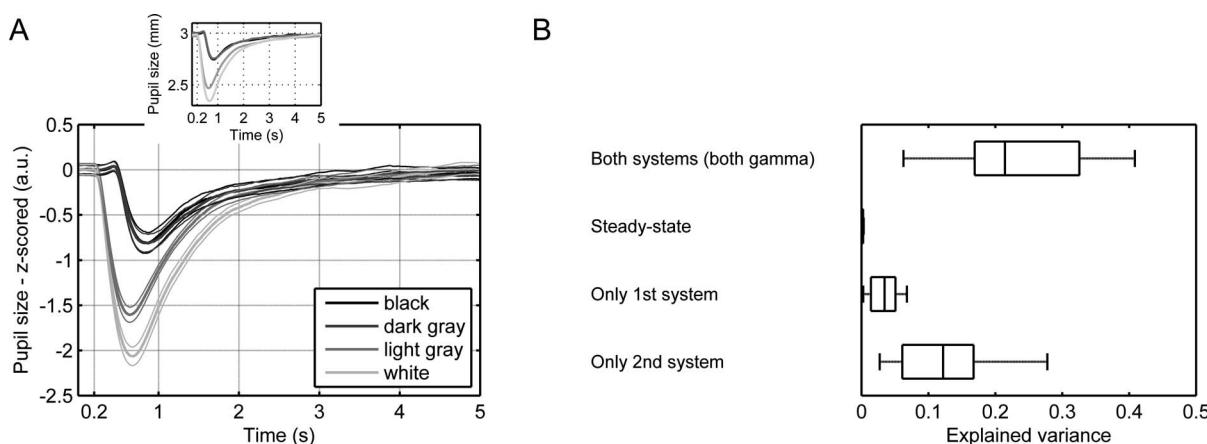


Figure 8. Model validation in the illuminance-flashes task. (A) Mean pupil responses were dominated by the constrictions following the onset of bright stimuli and the offset of dark stimuli. Thick lines represent mean responses, and thin lines standard error of the mean (SEM). For comparability, the inset shows mean responses in millimeters. (B) The second LTI system explained relatively more variance than the first LTI system.

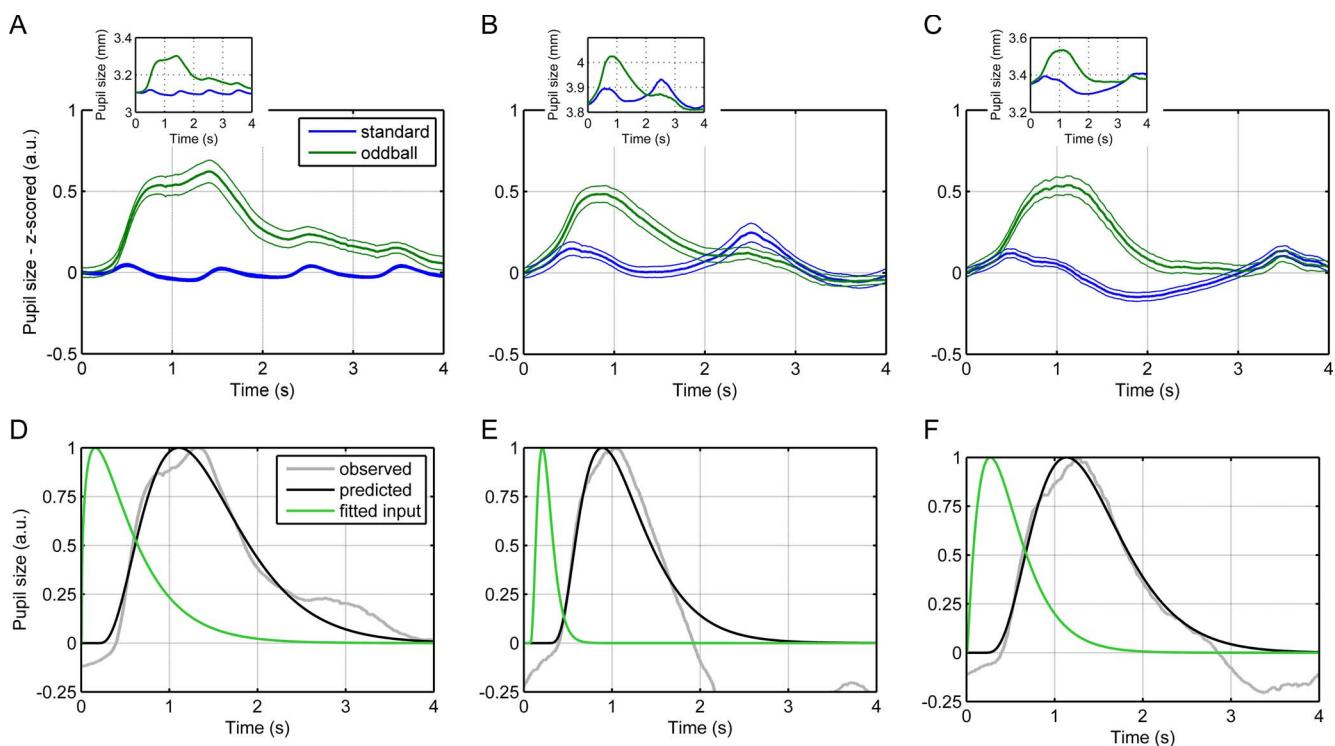


Figure 9. Auditory-oddball task. (A–C) Mean pupil responses to oddballs and standards in experiments with 1, 2, and 3 s, respectively. Thick lines represent mean responses, and thin lines standard error of the mean (SEM). For comparability, the inset shows mean responses in millimeters. (D–F) Fitted inputs of the oddball responses.

suggests a neural processing stream that has a similar latency as the illuminance response itself.

Discussion

In this work, we specify a psychophysiological model for the temporal evolution of pupil responses, based on two LTI systems. These LTI systems were modeled according to the empirical relationship between pupil

size and illuminance. The model explained more variance in the data than a baseline model, which took only steady-state pupil size into account, and explained all variance elicited by illuminance changes. In line with previous empirical findings (Barbur et al., 1992), the model makes the counterintuitive prediction that both light flashes and darkness flashes elicit pupil constriction. This prediction was confirmed quantitatively in an independent data set. Importantly, the model could be used to derive nonilluminance inputs into the pupillary system from three different perceptual tasks. The

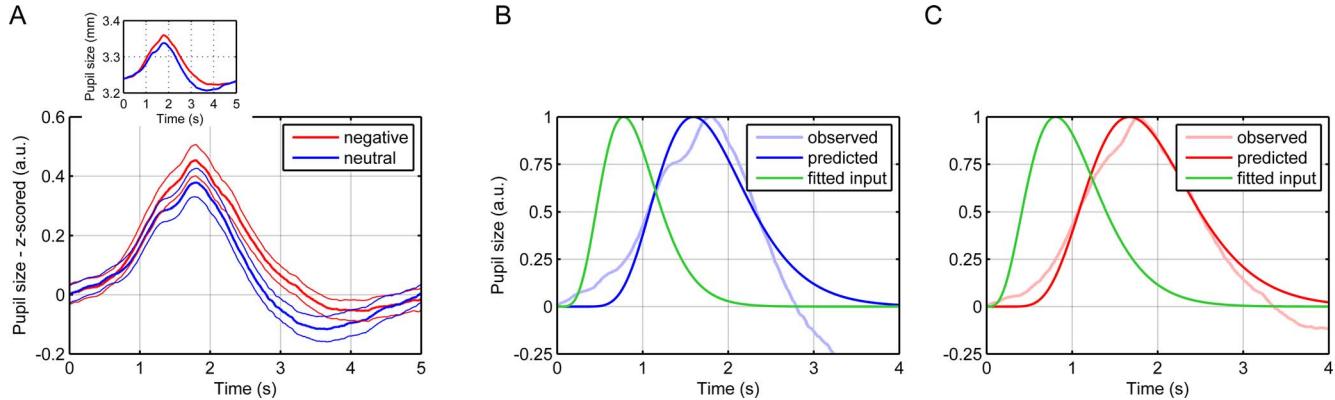


Figure 10. Emotional-words task. (A) Mean pupil responses to negative and neutral nouns. Thick lines represent mean responses, and thin lines standard error of the mean (SEM). For comparability, the inset shows mean responses in millimeters. (B) Fitted input of the responses to neutral words. (C) Fitted input of the responses to negative words.

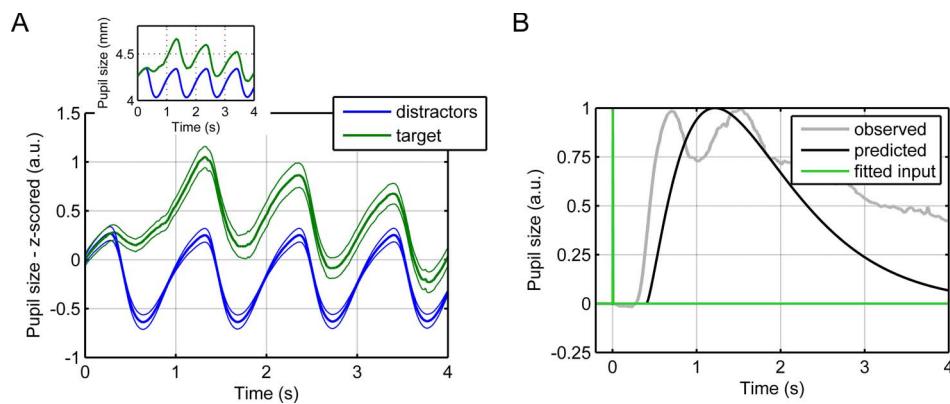


Figure 11. Visual-detection task. (A) Mean pupil responses to distractor and target stimuli. Thick lines represent mean responses, and thin lines standard error of the mean (SEM). For comparability, the inset shows mean responses in millimeters. (B) Fitted input of the responses to the target stimulus.

temporal evolution of this input furnished novel and interesting hypotheses on the underlying neural circuits. Auditory oddball input elicited neural input into the pupil with a similar latency as event-related potentials in electroencephalography (Friedman et al., 2001), underlining the plausibility of the approach. Responses to emotional words, probably requiring more complex neural processes (Bayer et al., 2011; Võ et al., 2008), occurred later. This was in contrast to the neural input to visual targets in a distractor stream. While these constitute oddball events just as in the auditory experiment, we designed the targets to entail high visual salience. Neural input into the pupil system then had the same latency as illuminance inputs, and therefore occurred much earlier than the response to auditory oddballs. It is therefore plausible to assume that the detection of these visual targets involved a neural circuit with complexity similar to that of the illuminance circuit. Hence, an analysis of estimated neural input can provide insight into the underlying neural circuits in relation to the well-known illuminance circuit. Crucially, such conclusions could not be drawn from merely analyzing the pupil traces in cognitive tasks on their own.

In an intermediate step, we took into account the fact that steady-state pupil size relates nonlinearly to illuminance. An exponential function provided a good mapping from illuminance levels to predicted steady-state responses. This exponential function served the purpose of scaling continuous illuminance time series in order to model the time evolution of pupil responses. The exponential function is in qualitative agreement with a previous review article (Watson & Yellott, 2012) that developed a relationship between steady-state pupil size and corneal flux density (i.e., the product of stimulus luminance and area) across a wide range of conditions (including monocular versus binocular illumination and age of the perceiver). The simple

exponential function used here served a more specific aim. We intended to provide a parsimonious description of the pupillary dynamics in conditions that are commonly employed in psychophysiological experiments. Furthermore, we used z-scored pupil sizes (in arbitrary units) for determining response functions because analyzing psychophysiological data often involves normalizing within participants.

LTI systems, which are commonplace in engineering, have a number of useful properties and are widely applied for analyzing neural and peripheral data. Our model modifies a simple LTI system to suit the biophysics of the pupil. The different shapes of positive and negative illuminance inputs relative to baseline precluded modeling both changes with a single system. We expected such a difference, since dilation and constriction are mediated by two antagonistic muscles, innervated respectively by sympathetic and parasympathetic afferents. For simplicity, we approximated this pattern with a combination of two LTI systems such that for each of the two, the linearity assumption holds approximately. The first system specifies responses to continuous illuminance input. The second system receives an instantaneous input and describes the undershoot of the constriction that cannot be accounted for by the first system. Based on our data, we took the parsimonious assumption of a constant input to the second system. This appears reasonable under a conjecture that the steeper and more pronounced response to positive illuminance changes serves to protect the retina from damage (McDougal & Gamlin, 2008). However, the precise relationship between illuminance changes and the initial part of the constriction response should be tested across a wider range of illuminance changes.

LTI systems have the important property that the response to two inputs is the sum of the responses to the individual inputs. This linearity assumption will

break down at the limits of the pupil's dynamic range, which lie roughly at 1.5 to 8 mm (McDougal & Gamlin, 2008). This would in practice only be a problem if one tested at a fast event rate (i.e., ITIs <5 s). Also, it is likely that the time-invariance assumption will not hold when testing at very high or low illuminance levels, where the pupil is already at the limits of its dynamic range.

Although the specific derivation of the two LTI systems was initially based on mathematical parsimony and not on principled physiological reasons, the model made a prediction for the relative importance of the two systems which is in line with previous reports (Barbur et al., 1992). Our data validated the model by showing that brief illuminance changes were mostly described by inputs to the second system. Furthermore, this showed that the model generalized to short inputs (of 200 ms), which may be important given the short stimulus duration in many perceptual tasks.

It has recently been shown that different pathways are involved in the responses of the pupil to luminous stimuli (e.g., rod-, cone-, or melanopsin-mediated pathways; Park & McAnany, 2015). Here, we did not aim at separating different illuminance-related pathways, but we did aim at a general approach. We acknowledge that the current model relies on overall illuminance (total flux arriving on the eye) and does not capture effects related to eye movements or the contrast and the spatial distribution of the light falling on the retina. For example, a small bright and a large dark stimulus of the same overall illuminance are treated the same by the model. This assumption is motivated by the well-described relationship between steady-state pupil size and corneal flux density (Watson & Yellott, 2012). Pupil dynamics may at least to some degree depend on the spatial distribution of the incoming light. For example, inverting a checkerboard (i.e., changing the white squares to black and vice versa) triggers pupil constriction at constant illuminance (Mathôt, Melmi, & Castet, 2015). This phenomenon can possibly be explained by the faster time course of constrictions versus dilations, under an assumption that signals for controlling constriction and dilation are independently computed at different points in retinal space and then combined rather than computed from summed flux density across the retina. Alternatively, it may be induced by attentional effects similar to the ones reported here for oddball responses. Our modeling approach may be able to arbitrate between these possible explanations. Overall, it is an empirical question whether a more fine-grained characterization of illuminance-related pupillary dynamics can provide even more leverage in explaining pupil responses in psychophysiological experiments.

Ultimately, our goal is to characterize nonilluminance inputs into the pupillary system. The combina-

tion of two LTI systems, which we derived for illuminance inputs, can serve two purposes towards that goal. First, it can be used to remove illuminance-related variance from data obtained in tasks that entailed illuminance changes. Second, and more important, neural input unrelated to illuminance can be estimated in relation to the timing of the illuminance-related input. For three perceptual tasks, we were able to recover plausible neural inputs into the first system, thus furnishing insights into the neural circuits mediating these responses. Specifically, the estimated inputs portend that pupil responses to emotional words and auditory oddballs result from cortical influences, while responses to visual targets may originate more proximally (e.g., in the brain stem). Hence, an important application of our model is that it allows inference regarding the level of processing from which the inputs to the pupil arise. Therefore, our illuminance-based model differs from a previous model that directly inferred an LTI system from empirically observed "attentional pulses" and consequently uses this system to deconvolve pupil time series (Hoeks & Levelt, 1993; Wierda et al., 2012). This earlier approach assumes a common shape of all attention-related pupil signals. In contrast, we capitalize on the notion of a common final pathway and are thus able to harness illuminance-related dynamics.

Perceptual and cognitive tasks typically modulate pupillary dynamics on a rather short time scale (i.e., in the range of seconds). It has been shown that pupil responses also vary on longer time scales (i.e., in the range of hours). For example, circadian rhythm (as measured by subjective sleepiness and salivary melatonin levels) modulates pupil responses to luminance inputs (Münch, Léon, Crippa, & Kawasaki, 2012). Relatedly, task-induced fatigue reduces stimulus-evoked pupil responses (Hopstaken, van der Linden, Bakker, & Kompier, 2015). Our LTI-based model could be easily used to quantify whether these effects reflect different amplitudes of the same response functions or whether the shapes of the response functions themselves change. Elucidating these effects may inform our understanding of the noradrenergic system, since circadian rhythms and fatigue depend in part on noradrenaline (Hopstaken et al., 2015).

In sum, we provide an explicit psychophysiological model for peripheral pupil responses. This is a prerequisite for specifying the exact neural inputs that drive cognitive pupil responses. If the pupil is a window into cognition, we hope our model furnishes a solid frame.

Keywords: *psychological model, response function, general linear convolution model, auditory oddball, emotional words, visual detection*

Acknowledgments

We thank Boris Quednow for providing the EyeLink equipment and for inspiring discussions, Katrin Preller for support with the EyeLink equipment, and Samuel Gerster for continuing technical support. Giuseppe Castegnetti, Athina Tzovara, Nicolas Hofer, and Banan Al-Nasery helped with data collection and discussions. Matthias Staib and Saurabh Khemka provided helpful suggestions. This work was funded by the University of Zurich and by the Swiss National Science Foundation (320030_149586/1). The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust (091593/Z/10/Z). The authors declare no competing financial interests. The research was designed by CWK and DRB, and performed by CWK; both authors analyzed the data and wrote the manuscript.

Commercial relationships: none.

Corresponding author: Christoph W. Korn.
Email: christoph.korn@uzh.ch.

Address: Department of Psychiatry, Psychotherapy, and Psychosomatics, University of Zurich, Zurich, Switzerland.

References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450, doi:10.1146/annurev.neuro.28.061604.135709.
- Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods*, 184(2), 224–234, doi:10.1016/j.jneumeth.2009.08.005.
- Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2010). Modelling event-related skin conductance responses. *International Journal of Psychophysiology*, 75(3), 349–356.
- Bach, D. R., & Friston, K. J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, 50(1), 15–22, doi:10.1111/j.1469-8986.2012.01483.x.
- Barbur, J. L., Harlow, A. J., & Sahraie, A. J. (1992). Pupillary responses to stimulus structure, colour and movement. *Ophthalmic & Physiological Optics*, 12, 137–141.
- Bayer, M., Sommer, W., & Schacht, A. (2011). Emotional words impact the mind but not the body: Evidence from pupillary responses. *Psychophysiology*, 48(11), 1554–1562, doi:10.1111/j.1469-8986.2011.01219.x.
- Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to bright surfaces enhances the pupillary light reflex. *The Journal of Neuroscience*, 33(5), 2199–2204, doi:10.1523/JNEUROSCI.3440-12.2013.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607, doi:10.1111/j.1469-8986.2008.00654.x.
- Browning, M., Behrens, T. E., Jocham, G., Reilly, J. X. O., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, 18(4), 590–596, doi:10.1038/nn.3961.
- Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences, USA*, 105(5), 1704–1709, doi:10.1073/pnas.0707727105.
- Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, 16(8), 1146–1153.
- Fan, X., & Yao, G. (2011). Modeling transient pupillary light reflex induced by a short light flash. *IEEE Transactions on Biomedical Engineering*, 58(1), 36–42.
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25(4), 355–373, doi:10.1016/S0149-7634(01)00019-7.
- Friston, K. J. (2005). Models of brain function in neuroimaging. *Annual Review of Psychology*, 56, 57–87.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, 10(2), 252–269, doi:10.3758/CABN.10.2.252.
- Goldinger, S. D., & Papse, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, 21(2), 90–95, doi:10.1177/0963721412436811.
- Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional pro-

- cesses. *International Journal of Psychophysiology*, 52(1), 1–6, doi:10.1016/j.ijpsycho.2003.12.001.
- Hayes, T. R., & Petrov, A. A. (2015). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, E-pub ahead of print, doi:10.3758/s13428-015-0588-x.
- Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods*, 25(1), 16–26, doi:10.3758/BF03204445.
- Hong, L., Walz, J. M., & Sajda, P. (2014). Your eyes give you away: Prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics. *PLoS ONE*, 9(3), e91321, doi:10.1371/journal.pone.0091321.
- Hopstaken, J. F., van der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, 110, 100–106, doi:10.1016/j.biopsych.2015.06.013.
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *The Quarterly Journal of Experimental Psychology*, 64(10), 1971–1989, doi:10.1080/17470218.2011.588335.
- Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, 41(8), 1068–1078.
- Mathôt, S., Melmi, J.-B., & Castet, E. (2015). Intra-saccadic perception triggers pupillary constriction. *PeerJ*, 3, e1150, doi:10.7717/peerj.1150.
- McDougal, D., & Gamlan, P. (2008). Pupillary control pathways. In A. Basbaum, A. Kaneko, G. M. Shepherd, et al. (Eds.), *The senses: A comprehensive reference* (pp. 521–536). San Diego, CA: Academic Press.
- McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, 87(1), 179–192.
- Münch, M., Léon, L., Crippa, S. V., & Kawasaki, A. (2012). Circadian and wake-dependent effects on the pupil light reflex in response to narrow-bandwidth light pulses. *Investigative Ophthalmology & Visual Science*, 53(8), 4546–4555. [PubMed] [Article]
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532–1543, doi:10.1111/j.1469-8986.2011.01226.x.
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, 15(7), 1040–1046.
- Park, J. C., & McAnany, J. J. (2015). Effect of stimulus size and luminance on the rod-, cone-, and melanopsin-mediated pupillary light reflex. *Journal of Vision*, 15(3):13, 1–13, doi:10.1167/15.3.13. [PubMed] [Article]
- Paulus, P., Castegnetti, G., & Bach, D. R. (2016). Modelling event-related heart period responses. *Psychophysiology*, E-pub ahead of print, doi:10.1111/psyp.12622.
- Prehn, K., Kazzer, P., Lischke, A., Heinrichs, M., Herpertz, S. C., & Domes, G. (2013). Effects of intranasal oxytocin on pupil dilation indicate increased salience of socioaffective stimuli. *Psychophysiology*, 50(6), 528–537, doi:10.1111/psyp.12042.
- Preller, K. H., Herdener, M., Schilbach, L., Stämpfli, P., Hulka, L. M., Vonmoos, M., & Quednow, B. B. (2014). Functional changes of the reward system underlie blunted response to social gaze in cocaine users. *Proceedings of the National Academy of Sciences, USA*, 111(7), 2842–2847.
- Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, 5, 115, doi:10.3389/fnins.2011.00115.
- Qin, S., Hermans, E. J., van Marle, H. J. F., & Fernandez, G. (2012). Understanding low reliability of memories for neutral information encoded under stress: Alterations in memory-related activation in the hippocampus and midbrain. *The Journal of Neuroscience*, 32(12), 4032–4041, doi:10.1523/JNEUROSCI.3101-11.2012.
- Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., & Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, 84(2), 355–362.
- Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538, doi:10.3758/BRM.41.2.534.
- Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophys-*

- iology*, 45(1), 130–140, doi:10.1111/j.1469-8986.2007.00606.x.
- von Broschmann, D. (2011). *Tafeln zur Prüfung des Farbensinnes: Begründet und weiterentwickelt von Stilling/Hertel/Velhagen* (34th ed.). Stuttgart: Thieme.
- Wang, C.-A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, 33, 134–140.
- Watson, A. B., & Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *Journal of Vision*, 12(10):12, 1–16, doi:10.1167/12.10.12. [PubMed] [Article]
- Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences, USA*, 109(22), 8456–8460, doi:10.1073/pnas.1201858109.
- Yellin, D., Berkovich-Ohana, A., & Malach, R. (2015). Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. *NeuroImage*, 106, 414–427.
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76–86, doi:10.1016/j.neuroimage.2014.06.069.

Appendix: Mathematical rationale for approximating pupil responses with LTI systems

To motivate our approach, we derive here a well-known result from linear systems theory. Inputs f are convolved with impulse response functions (RFs) g to obtain the predicted time series (in continuous time). To derive the impulse RFs from observed responses

elicited by a sudden change in illuminance, we have to perform the inverse operation (i.e., taking the time derivative of the observed response).

Recall that the definition of the convolution is

$$(f*g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)g(t - \tau) d\tau.$$

Here, t is time and τ is a dummy variable. Let f be the input into the LTI system and g be its characteristic impulse RF. According to the differentiation rules for convolution, the derivative of the convolution of f and g equals the convolution of g and the derivative of f :

$$\frac{d}{dt}(f*g)(t) = \frac{df}{dt}*g.$$

Combining these two equations gives

$$\frac{d}{dt}(f*g) = \int_{-\infty}^{\infty} \frac{df(\tau)}{d\tau} g(t - \tau) d\tau.$$

A sudden change in illuminance can be described with a step RF

$$f(t) = H(t) = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases}$$

with

$$\frac{d}{dt}H(t) = \delta(t),$$

where $\delta(t)$ is a Dirac delta function. According to the definition of a delta function, inserting this yields

$$\begin{aligned} \frac{d}{dt}(f*g)(t) &= \int_{-\infty}^{\infty} \frac{d}{dt}H(\tau)g(t - \tau) d\tau \\ &= \int_{-\infty}^{\infty} \delta(\tau)g(t - \tau) d\tau = g(t). \end{aligned}$$

This shows that the derivative of the response to a step input is the RF g .