iScience, Volume *24*

Supplemental information

Circadian fluctuations in glucocorticoid level predict perceptual discrimination sensitivity

Jonas Obleser, Jens Kreitewolf, Ricarda Vielhauer, Fanny Lindner, Carolin David, Henrik Oster, and Sarah Tune

Supplemental figures

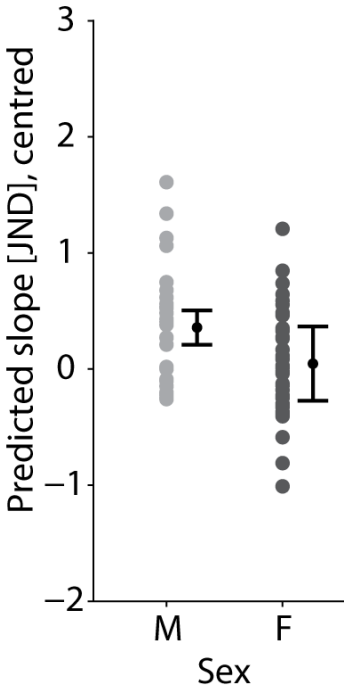

Figure S1. Perceptual discrimination sensitivity as a function of sex, Related to Figure 3.
Coloured dots (light grey, male (M); dark grey, female (F) cohort) show single-subject (N=68) predicted slope values based on the best-fitting linear mixed-effects model. Black dots represent the fixed-effect group-level prediction and 95% CI.

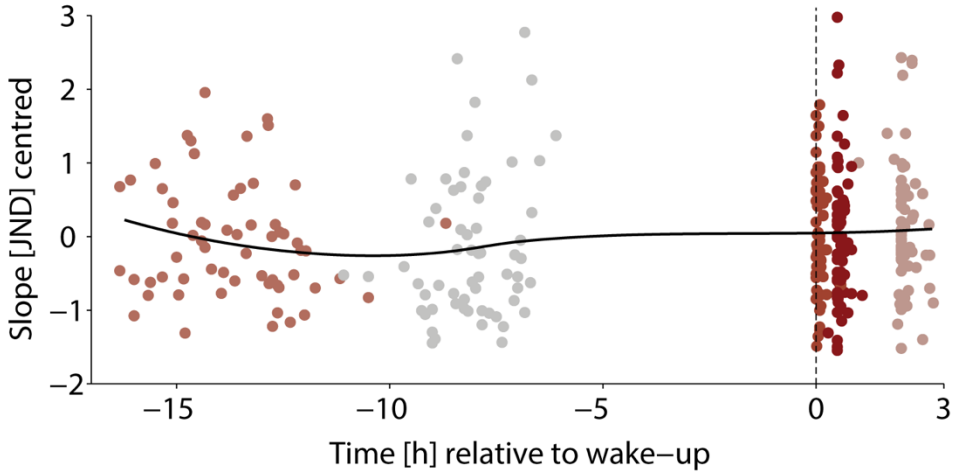

Figure S2. Perceptual discrimination sensitivity does not depend on the time of day,
Related to Figure 3.

Change in individual perceptual sensitivity across five experimental session. Slope values are mean-centred across all N=68 participants. Sessions are grouped by colour and aligned by wake-up time (dashed vertical line). Black curve shows LOESS regression of time.

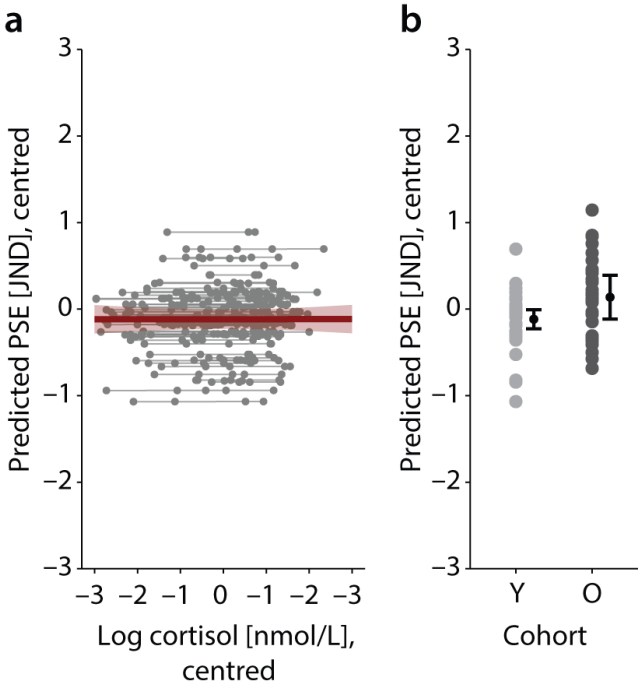

Figure S3. Cortisol dynamics do not impact response bias, Related to Figure 3.

(a) Response bias (operationalised by the point-of-subjective-equality; PSE) as predicted by cortisol. Predicted non-significant group-level fixed-effect (green slope) with 95% confidence interval (CI) error band is shown along with the estimated subject-specific (N=68) random slopes (thin grey lines) and single-subject, single-session predictions (grey dots). Note that subject-specific random slopes did not improve the model fit and were added for illustrative purposes only.

(b) Change in response bias as a function of age cohort. Coloured dots (light grey, young (Y) cohort; dark grey, older (O) cohort) show single-subject (N=68) predicted PSE values based on the best-fitting linear mixed-effects model. Black dots represent the fixed-effect group-level prediction and 95% CI.

Supplemental tables

Table S1: Predicting Cortisol, Related to Figure 2.

	Cortisol [log nmol/L; z-scored]					
Predictors	*Estimates*	*std. Error*	*CI*	*t*	*p*	*df*
Intercept	-0.01	0.24	-0.81 – 0.80	-0.03	0.979	2.82
Time rel. to wake-up [h; z-scored]	0.57	0.23	-0.06 – 1.20	2.43	0.067	4.31
Time squared [z]	0.11	0.11	-0.12 – 0.34	1.01	0.323	20.94
Time cubic [z]	-0.15	0.07	-0.28 – -0.01	-2.13	**0.035**	128.17
Sleep duration [h; z-scored]	-0.16	0.04	-0.24 – -0.07	-3.57	**0.001**	77.48
Random Effects						
σ^2	0.24					
$\tau_{00\ subj}$	0.06					
$\tau_{00\ session}$	0.29					
ICC	0.59					
N_{subj}	68					
$N_{session}$	5					
Observations	318					
Marginal R^2 / Conditional R^2	0.388 / 0.750					
AIC	553.699					

Table S2: Predicting Perceptual Sensitivity, Related to Figure 3.

	Perceptual Sensitivity [JND; z-scored]					
Predictors	*Estimates*	*std. Error*	*CI*	*t-value*	*p*	*df*
Intercept	-0.024	0.088	-0.199 – 0.152	-0.270	0.7883	64.668
Cortisol [log nmol/L; z-scored]	0.130	0.044	0.043 – 0.217	2.940	**0.0036**	267.356
Age Cohort (Old)	-0.516	0.178	-0.872 – -0.160	-2.896	**0.0051**	65.476
Sex (Female)	-0.360	0.180	-0.718 – -0.001	-2.004	**0.0492**	65.098
Random Effects						
σ^2	0.55					
$\tau_{00\ subj}$	0.39					
ICC	0.42					
N_{subj}	68					
Observations	318					
Marginal R^2 / Conditional R^2	0.089 / 0.470					
AIC	829.669					

Table S3: Predicting Perceptual Sensitivity from State- and Trait-level Cortisol, Related to Figure 3.

	Perceptual Sensitivity [JND; z-scored]					
Predictors	*Estimates*	*std. Error*	*CI*	*t-value*	*p*	*df*
Intercept	-0.024	0.088	-0.201 – 0.153	-0.269	0.7888	63.752
Cortisol within-subject effect [log nmol/L; z-scored]	0.120	0.042	0.038 – 0.202	2.885	**0.0043**	249.107
Cortisol between-subject effect [log nmol/L; z-scored]	0.051	0.088	-0.124 – 0.226	0.580	0.5641	64.762
Age Cohort (Old)	-0.516	0.180	-0.875 – -0.158	-2.875	**0.0055**	64.493
Sex (Female)	-0.360	0.183	-0.726 – 0.006	-1.967	0.0535	63.736
Random Effects						
σ^2	0.55					
$\tau_{00\ subj}$	0.40					
ICC	0.42					
N_{subj}	68					
Observations	318					
Marginal R^2 / Conditional R^2	0.088 / 0.474					
AIC	834.828					

Table S4: Predicting Response Bias, Related to Figure 3.

	PSE [point of subjective equality; z-scored]					
Predictors	*Estimates*	*std. Error*	*CI*	*t-value*	*p*	*df*
Intercept	0.019	0.099	-0.178 – 0.215	0.189	0.8509	66.036
Cortisol [log nmol/L; z-scored]	0.001	0.041	-0.080 – 0.082	0.022	0.9823	261.837
Age Cohort (Old)	0.445	0.197	0.052 – 0.839	2.260	**0.0271**	66.035
Random Effects						
σ^2	0.47					
$\tau_{00\ subj}$	0.55					
ICC	0.54					
N_{subj}	68					
Observations	318					
Marginal R^2 / Conditional R^2	0.046 / 0.564					
AIC	803.468					

Table S5. Overview of prescription medication reported by cohort of middle-aged and older participants, Related to Figure 2.

Purpose	Medication	N
Asthma medication	Alvesco (Ciclesonid)	1
Hypertension	Amlodipin	5
	Atacant	3
	Bisoporol	5
	Candesartan	1
	Enalapril	1
	Hydrochlorothiazide	1
	Metopropol	1
	Ramipril	4
	Valsartan	1
	Vocado40	1
Lipid-lowering medication	Simvastatin	1
Non-steroidal anti-inflammatory medication	Aspirin	1
	Naproxen	1
	Salofalk	1

Transparent methods

Participants

Seventy-five participants took part in this study, acquired in two waves (younger participants in April–May 2018, older participants in April–May 2019). The participants were recruited through the database of the Department of Psychology at the University of Lübeck, using the online recruiting system ORSEE (Greiner, 2015). The cohort of younger participants consisted of 37 university students (24 females, mean age 22.6, SD = 2.58, age range 19–30 years). The cohort of middle-aged and older participants consisted of 38 persons (19 females, mean age 60.6, SD = 5.98, age range 50–70 years); 16 of them were retired.

All participants were screened to avoid any history of disorders that could have impacted their GC balance, such as neurological or psychiatric disorders as well as any known metabolic diseases. Furthermore, none showed a BMI over 30 kg/m^2 or had been working in shifts. None reported any known hearing disorders, severe current hearing loss, or a persistent tinnitus. Note, however, that participants with mild age-related hearing loss were not excluded from the cohort of older participants due to its high prevalence in this age group.

In the cohort of younger participants, none took any medication that could have influenced their GC balance, including medication for asthma- or allergy treatment, systemic immunosuppressants or antihypertensives.

In the cohort of older participants, more lenient inclusion criteria with respect to medication applied (see Table S5). Here, participants who took any type of antihypertensives were still included to allow for a representative sample of older adults.

Written informed consent was collected from all participants according to procedures approved by the Research Ethics Committee of the University of Lübeck. Listeners were paid 25–30 € or received course credit for their participation in the experiment.

Experimental protocol

On the first day, participants came to the laboratory between 4pm and 6pm for the first session, lasting about one and a half hours. A maximum of four participants conducted the first session on a given day. The session started with an adaptive tracking procedure that measured auditory pitch thresholds (see section Psychoacoustic testing for details). Participants were then asked to complete three questionnaires on their general medical history, their chronotype (Horne and Ostberg, 1976), and their momentary sleepiness (assessed using the Karolinska Sleepiness Scale; Akerstedt and Gillberg, 1990). The scale consists of three items: (1) sleepiness during the last 10 minutes (nine steps on a Likert-Scale), (2) the current state with relaxation on one end and tension on the other end of a visual analogue scale and (3) the current fatigue (visual analogue scale).

Next, participants received detailed instructions for the subsequent measurements. Each session included taking a saliva sample and performing a challenging pitch discrimination task in a browser-based online study (Labvanced, Osnabrück), followed by the sleepiness questionnaire. According to their auditory pitch threshold, participants were assigned to an experimental group, designed to yield equivalent difficulties of the pitch discrimination task (see *Assessment of pitch discrimination thresholds* below), and provided with an individual link, which gave them exactly five times access to the online task.

Finally, participants completed the first session: taking a saliva sample first (see Saliva cortisol collection for details) and performing the online pitch discrimination task secondly before they were sent home. Throughout all sessions, participants in the younger cohort used their own technical devices (laptop and headphones) whereas participants in the older cohort used their own headphones for all experimental sessions but were provided with computers for the first session due to their lack of portable computers. Usage of participants' own equipment ensured that the acoustic properties of the pitch discrimination task remained constant across sessions and, whenever possible, that the experiment could be adequately performed with the participants' personal equipment.

All other measurements were conducted at home, scheduled at certain times of day relative to the participants' sleep–wake cycle: Session 2 had to be performed just before going to sleep, Session 3 immediately after waking up (participants were instructed to place the equipment, or at least the Salivette tube for the saliva sample, next to their bed), Session 4 30 minutes, and Session 5 about 120 minutes after awakening. To assess compliance and to gather information about the time of events, participants recorded the starting time of each session as well as the activities that they were engaged in between two consecutive sessions in a time protocol. Additionally, they were asked to maintain their typical sleeping and wake-up times, which they had recorded for the last two weeks.

Saliva cortisol collection

Salivary cortisol level was measured to deduce the amount of unbound cortisol in blood (Kirschbaum and Hellhammer, 1994). To capture a comprehensive cycle of cortisol secretion as used in former studies (Dijckmans et al., 2017; Evans et al., 2011; Lee et al., 2007), including the characteristic morning rise, a saliva sample was collected at each single experimental session. As described above, sessions were scheduled according to the individual participant's wake-up time. Following instructions and the collection of a first saliva sampling in the lab session, participants were provided with a saliva self-collection pack containing four Salivette Cortisol tubes (Sarstedt, Nümbrecht, Germany), pre-labelled with participant code and number of session, and written instructions. For a correct usage, the Salivette dental swab from the correctly labelled Salivette had to be chewed until fully saturated and then be put back into the tube. Saliva samples were then stored in the participants' own freezer until they were brought back or picked up after one to seven days, together with the time protocol and stored in the freezer of the Department of Psychology.

To avoid bias, participants were asked not to smoke, eat, drink (except water) or brush their teeth 30 minutes before sampling.

All saliva samples (180 from the younger cohort and 185 from the older cohort) were analysed at the Biochemical Laboratory of the Technical University Dresden. The fraction of free cortisol in saliva (salivary cortisol) was determined using a time-resolved immunoassay with fluorometric detection (for detailed method see Dressendorfer et al., 1992) and reported back to the authors in the unit of measurement, nmol/l, to 1-decimal precision.

Psychoacoustic testing

Assessment of pitch discrimination thresholds. In the first session, we assessed individual participants' pitch discrimination thresholds (i.e., their so-called just-

noticeable differences; JNDs) using a weighted one-up, one-down method (Kaernbach, 1991). On each trial, participants heard two pure tones. Each tone had a duration of 100 ms with a silence period of 25 ms between tones. The first tone always had a frequency of 1 kHz; the frequency of the second tone differed from that of the first tone by delta f. The participants were asked to indicate via button press which of the two tones had the higher frequency. The next trial started 750 ms after the participants' response. Responses were self-paced. No feedback was given.

The assessment of pitch discrimination thresholds comprised five staircases per participant. Each staircase started with a delta f of 100 cents (i.e., one semitone). In the first phase, delta f was increased by a factor of 2.25 following an incorrect response and was decreased by the cube root of 2.25 following a correct response. Hence, the magnitude of upward steps was three times larger than the magnitude of downward steps, estimating approximately 75%-correct on the psychometric function. In the second phase, we used a factor of 1.5 and cube root of 1.5 for up- and down-steps, respectively. Each staircase was terminated after the twelfth reversal; there were four reversals in the first phase and eight reversals in the second phase. The threshold in each staircase was defined as the arithmetic mean of delta fs visited on all second-phase reversal trials. Finally, individual JNDs were defined as the average of thresholds across all five staircases per participant.

Assessment of psychometric curves. In each of the five sessions, participants performed a pitch discrimination task in a browser-based online study (Labvanced, Osnabrück). This task was similar to the assessment of pitch discrimination thresholds, which was completed in the first session only (see above): on each trial, participants heard two pure tones which differed in frequency and were asked to indicate which tone had the higher pitch. Here, however, we used a method of constant stimuli to assess participants' individual pitch sensitivity. In each session, participants completed 148 trials, comprising seven stimulus levels relative to their individual pitch discrimination threshold (JND). This means that participants were assigned to different groups based on their individual thresholds to ensure similar difficulty levels across participants. We considered five different groups: 5ct, 10ct, 15ct, 20ct, and 25ct. Participants were assigned to the group closest to their individual JND (e.g., a participant with a JND of 7.5ct would be assigned to the 10ct group, while a participant with a JND of 7.4ct would be assigned to the 5ct group).

The stimulus levels were approximately -3, -1.5, -0.5, 0, 0.5 1.5, and 3 JNDs. This choice of stimulus levels allowed us to sample the linear part of the logistic function (slope), while also capturing its asymptotes (Herbst and Obleser, 2019; Waschke et al., 2019). Note that a stimulus level of zero JND means that the two tones on a given trial had the same frequency of 1 kHz. Hence, there was no correct response for this stimulus level. Each stimulus level was presented 21 times per session. We additionally included one dummy trial at the beginning of each session. The response in this trial was excluded from the analysis; however, inclusion of this dummy trial allowed us to present the stimulus levels using a type-1 index-1 sequence (Finney and Outhwaite, 1956). Type-1 index-1 sequences control for potential carry-over effects by first-order counterbalancing. This means that each stimulus level has the same probability to occur after each other stimulus level, including itself.

In each session, we calculated the proportion of 'second tone higher' responses per stimulus level and fitted a logistic function to the data using the

Palamedes toolbox version 1.7.0 (Prins and Kingdom, 2018) in MATLAB (MathWorks, Natick, Massachusetts, USA; R2017b). We fitted three parameters: The point of subjective equality (PSE; i.e., the point where subjects reported 'second tone higher' in 50% of trials), the slope at the PSE (i.e., our measure of perceptual sensitivity), and the lapse rate (i.e., the lower asymptote). The guess rate (i.e., the higher asymptote) was fixed at 1 minus the guess rate, which resulted in symmetric asymptotes of the psychometric fit.

Data sets from eight individual sessions did not follow a psychometric curve and no fit was possible. Additionally, we excluded fits with extreme slopes (i.e., larger than 5) as well as flat psychometric curves. Based on these criteria, six participants produced less than two usable fits. All data from these participants were therefore excluded from further analyses.

The data from one participant in the younger cohort who reported to follow an unusually shifted sleep-wake cycle were excluded prior to analysis. The data of six participants in the older cohort were excluded from analysis because they either dropped out of the study after the first session (N=3), or because of missing or unusable data in more than three sessions (N=3; see details on psychoacoustic testing below).

The final sample consisted of N=68 individuals and, in sum, we used 318 of a possible maximum of 340 observations in the statistical analyses.

Statistical analysis

We used linear mixed-effect models to investigate how circadian fluctuations in salivary cortisol level influence perceptual sensitivity. To this end, we first investigated how cortisol expression levels change throughout the day by modelling increasingly complex trajectories via the inclusion of polynomial regressors of different orders. We also tested for changes in cortisol levels as a function of sleep duration, age cohort (young/old), and sex (male/female). In the main analysis, we then modelled the influence of momentary cortisol levels on auditory perceptual discrimination sensitivity, expressed as the slope of the psychometric function. We also tested for the impact of time (expressed relative to the individual wake-up time), age cohort, sex, sleep duration, pitch group, sleepiness (assessed using the Karolinska Sleepiness Scale), and response bias (expressed as the point of subjective equality on the psychometric curve, PSE).

Estimation and selection of linear mixed-effect models (Gaussian distribution, identity link function) followed an iterative model fitting procedure (Alavash et al., 2019; Tune et al., 2018). We started with an intercept-only null model including subject-specific random intercepts and added fixed-effects terms in a stepwise procedure following their conceptual importance. Main effects were added prior to higher-order interaction terms. Lastly, we tested whether the inclusion of a session-specific random intercept or subject-specific random slopes for time-varying within-subject effects would improve model fit. Change in model fit was assessed via likelihood ratio tests on models (re-fit with maximum-likelihood estimation for comparison of fixed effects).

We used deviation coding for categorical predictors. Single-subject observations with unusually high cortisol levels of above 60 nmol/L were discarded. Cortisol levels were log-transformed and as all other continuous variables z-scored prior to modelling. To facilitate interpretation, in the visual presentation of model results, we transformed the continuous variables back to their original units.

An additional control analysis included two separate predictors for the influence of cortisol on perceptual sensitivity to tease apart within- and between-subject effects of cortisol on behaviour. Mean cortisol levels per subject captured the trait-like, between-subject effect while the state-like, within-subject effect was modelled by the session-by-session deviation from this subject-level mean (Bell et al., 2019).

In a second control analysis, we performed a causal mediation analysis (Imai et al., 2010) to formally test the possibility of cortisol only mediating a daytime effect on perceptual sensitivity. We estimated the direct, indirect (mediated) and total effect of the cubic trend of time on perceptual sensitivity using the same set of covariate regressors in the mediation and outcome model. We calculated 95 % quasi-Bayesian confidence intervals using 5,000 replications.

We report p-values for individual model terms that were derived using the Kenward-Roger approximation for degrees of freedom (Luke, 2017)**.** As goodness-of-fit measures, we report R^2 (marginal and conditional R^2; taking into account only fixed or fixed and random effects, respectively) along with the Akaike information criterion (AIC) (Nakagawa et al., 2017). To facilitate interpretation of (non-)significant effects, we also calculated the Bayes factor (BF) based on the comparison of Bayesian information criterion (BIC) values as proposed by Wagenmakers (Wagenmakers, 2007). Throughout we report log Bayes Factors, with a log BF of 0 representing equal evidence for and against the null hypothesis; log BFs with a positive sign indicating relatively more evidence for the alternative hypothesis than the null hypothesis, and vice versa. Magnitudes > 1 are taken as moderate, > 2.3 as strong evidence for either of the alternative or null hypotheses, respectively. All analyses were performed in R (version 3.6.1) using the lme4 (Bates et al., 2015), mediation (Tingley et al., 2014)**,** and sjPlot (Lüdecke, 2020) packages.

Supplementary References

Akerstedt, T., and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. Int J Neurosci *52*, 29-37.

Alavash, M., Tune, S., and Obleser, J. (2019). Modular reconfiguration of an auditory control brain network supports adaptive listening behavior. Proc Natl Acad Sci U S A *116*, 660-669.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software *67*, 1-48.

Bell, A., Fairbrother, M., and Jones, K. (2019). Fixed and random effects models: making an informed choice. Quality \& Quantity *53*, 1051--1074.

Dijckmans, B., Tortosa-Martínez, J., Caus, N., González-Caballero, G., Martínez-Pelegrin, B., Manchado-Lopez, C., and Clow, A. (2017). Does the diurnal cycle of cortisol explain the relationship between physical performance and cognitive function in older adults? European Review of Aging and Physical Activity *14*, 6.

Dressendorfer, R.A., Kirschbaum, C., Rohde, W., Stahl, F., and Strasburger, C.J. (1992). Synthesis of a cortisol-biotin conjugate and evaluation as a tracer in an immunoassay for salivary cortisol measurement. J Steroid Biochem Mol Biol *43*, 683-692.

Evans, P.D., Fredhoi, C., Loveday, C., Hucklebridge, F., Aitchison, E., Forte, D., and Clow, A. (2011). The diurnal cortisol cycle and cognitive performance in the healthy old. International Journal of Psychophysiology *79*, 371–377.

Finney, D.J., and Outhwaite, A.D. (1956). Serially balanced sequences in bioassay. Proc R Soc Lond B Biol Sci *145*, 493-507.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. Journal of the Economic Science Association *1*, 114--125.

Herbst, S.K., and Obleser, J. (2019). Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. NeuroImage *203*, 116198.

Horne, J.A., and Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. International Journal of Chronobiology *4*, 97–110.

Imai, K., Keele, L., and Tingley, D. (2010). A General Approach to Causal Mediation Analysis. Psychological methods *15*, 309--334.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. Perception & psychophysics *49*, 227-229.

Kirschbaum, C., and Hellhammer, D.H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. Psychoneuroendocrinology *19*, 313–333.

Lee, B.K., Glass, T.A., McAtee, M.J., and Wand, G.S. (2007). Associations of Salivary Cortisol With Cognitive Function in the Baltimore Memory Study. Arch Gen Psychiatry *64*, 810–818.

Lüdecke, D. (2020). Extracting, Computing and Exploring the Parameters of Statistical Models using R. Journal of Open Source Software *5*, 2445.

Luke, S.G. (2017). Evaluating significance in linear mixed-effects models in R. Behavior Research Methods *49*, 1494--1502.

Nakagawa, S., Johnson, P.C.D., and Schielzeth, H. (2017). The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. Journal of the Royal Society Interface *14*, 20170213.

Prins, N., and Kingdom, F.A.A. (2018). Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. Front Psychol *9*, 1250.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation : R Package for Causal Mediation Analysis. Journal of Statistical Software *59*.

Tune, S., Wostmann, M., and Obleser, J. (2018). Probing the limits of alpha power lateralisation as a neural marker of selective attention in middle-aged and older listeners. Eur J Neurosci *48*, 2537-2550.

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review *14*, 779-804.

Waschke, L., Tune, S., and Obleser, J. (2019). Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. eLife *8*.