

Adaptive thresholding increases sensitivity to detect changes in the rate of skin conductance responses to psychologically arousing stimuli in both laboratory and ambulatory settings

Ian R. Kleckner ^{a,*}, Jolie B. Wormwood ^b, Rebecca M. Jones ^c, Eva Culakova ^d, Lisa Feldman Barrett ^{e,f}, Catherine Lord ^{c,g}, Karen S. Quigley ^e, Matthew S. Goodwin ^e

^a University of Maryland Baltimore, Baltimore, MD, USA

^b University of New Hampshire, Durham, NH, USA

^c Weill Cornell Medicine, The Center for Autism and the Developing Brain, White Plains, NY, USA

^d University of Rochester Medical Center, Rochester, NY, USA

^e Northeastern University, Boston, MA, USA

^f Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA

^g Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA



ARTICLE INFO

Keywords:

Electrodermal activity (EDA)

Skin conductance response (SCR)

Statistical power

Sensitivity

Research equity

ABSTRACT

Psychophysologists recording electrodermal activity (EDA) often derive measures of slow, tonic activity—skin conductance level (SCL)—and faster, more punctate changes—skin conductance responses (SCRs). A SCR is conventionally considered to have occurred when the local amplitude of the EDA signal exceeds a researcher-determined threshold (e.g., 0.05 μ s), typically fixed across study participants and conditions. However, fixed SCR thresholds can preferentially exclude data from individuals with low SCL because their SCRs are smaller on average, thereby reducing statistical power for group-level analyses. Thus, we developed a *fixed plus adaptive* (FA) thresholding method that adjusts identification of SCRs based on an individual's SC at the onset of the SCR to increase statistical power and include data from more participants. We assess the utility of applying FA thresholding across two independent samples and explore age and race-related associations with EDA outcomes. Study 1 uses wired EDA measurements from 254 healthy adults responding to evocative images and sounds in a laboratory setting. Study 2 uses wireless EDA measurements from 20 children with autism in a clinical environment while they completed behavioral tasks. Compared to a 0.01, 0.03, and 0.05 μ s fixed threshold, FA thresholding at 1.9% modestly increases statistical power to detect a difference in SCR rate between tasks with higher vs. lower subjective arousal and reduces exclusion of participants by up to 5% across both samples. This novel method expands the EDA analytical toolbox and may be useful in populations with highly variable basal SCL or when comparing groups with different basal SCL. Future research should test for reproducibility and generalizability in other tasks, samples, and contexts.

Impact statements: This article is important because it introduces a novel method to enhance sensitivity and statistical power in analyses of skin conductance responses from electrodermal data.

1. Introduction

Electrodermal activity (EDA) reflects eccrine sweat gland activity controlled by the sympathetic nervous system (Boucsein, 2012; Dawson et al., 2017; Tronstad et al., 2022). Measures of EDA have been used to advance the study of learning, stress, emotion, pain, attention, affective computing, sleep, and clinical conditions such as autism, depression,

and schizophrenia (Tronstad et al., 2022). Although the number, amplitude, and rate of skin conductance responses (SCRs) are traditional electrodermal metrics, there is little consensus on optimal criteria for detecting SCRs (pp. 156, Boucsein, 2012). In the pre-computer era, an SCR was considered to have occurred when a short-term (a few seconds) deviation from the prevailing skin conductance (SC)—or SC at onset—occurred with an amplitude $>0.05 \mu$ s (Edelberg, 1972; Venables

* Corresponding author at: Department of Pain and Translational Symptom Science, PTSS Dept Room 738, 655 W Lombard St., Baltimore, MD 21201, USA.
E-mail address: Ian.Kleckner@UMaryland.edu (I.R. Kleckner).

and Christie, 1980). Modern electrodermal analytic approaches often use lower thresholds, such as 0.03 and 0.01 μS (Boucsein, 2012). Other criteria have also been suggested, such as requiring each SCR amplitude to exceed 0.1% of the SCL at the onset of the SCR (Edelberg, 1972), but these methods have not been further developed, and fixed thresholds have remained standard practice without further scrutiny for over five decades.

Despite the ubiquity of fixed threshold methods in determining SCR occurrence, prior work has shown that the amplitude of SCRs within a person is positively associated with SCL (Boucsein et al., 1984; Venables and Christie, 1980). This suggests that a fixed threshold will result in discrepancies in the probability of identifying SCRs within and across individuals as SCL varies. Fig. 1 shows an example EDA signal wherein larger and more frequent SCRs are apparent using a fixed threshold when the onset SC is high. The figure shows two potential errors that result: counting relatively small SCRs in the context of high SC (lower left inset) and undercounting of more moderate amplitude SCRs in the context of lower SC (lower right inset).

The current work identifies and proposes an additional EDA metric: a *fixed plus adaptive (FA) threshold* that is proportional to the SC at the onset of an SCR. This procedure requires determining a response amplitude percent (RAP) as a threshold for identifying SCRs. RAP is calculated as SCR amplitude divided by SC value at the onset of the SCR (see Fig. 2 for an illustration and definitions of key terms). This FA thresholding procedure addresses the issue that individuals with lower tonic SCL are under-represented in research using EDA because their SCRs are smaller on average (or SCRs are not counted and hence designated as “non-responders”), when using a fixed threshold for

counting SCRs (Boucsein et al., 1984; Venables and Christie, 1980); for discussion, see p. 240 of (Boucsein, 2012). Participant exclusions due to low SCL introduces research inequities (Bradford et al., 2022; Webb et al., 2022) and can, for example, exclude populations with lower SCL on average, such as African Americans (Alexandra Kredlow et al., 2017) and older adults (Bari et al., 2020; Eisdorfer et al., 1980). Moreover, participant exclusion reduces statistical power in group-level analyses which necessitates larger sample sizes, and burdens participants, research staff, and funding sources.

We hypothesized that detecting the number of SCRs or rate of SCRs using FA thresholds would (1) increase sensitivity to detect differences in SCR rate across conditions of lower vs. higher subjective arousal and (2) reduce exclusion of participants due to a lack of denoted SCRs. We focused our analyses on SCR rate and subjective arousal because of their well-studied association when using standardized photo and audio stimuli in American and European participants (Bradley and Lang, 2007; Dawson et al., 2017; Lang et al., 2008). Therefore, we reasoned this would be a good place to start developing a new method, in the spirit of experiment-based calibration in which an experiment-generated manipulation of a construct (here, subjective arousal) is used to see a change in another variable of interest (here, SCR rate) (Bach et al., 2020; Bach et al., 2023). We do not assume a more context-general one-to-one mapping between SCR rate and self-reported arousal. Indeed, in future work, it will be important to identify the mappings from SCR rate to self-reported arousal in a variety of contexts, for example: (1) for individuals who have difficulty identifying or describing their feelings, such as individuals with autism (Kinnaird et al., 2019), (2) for individuals who cannot communicate their feelings, e.g., pain in certain clinical settings

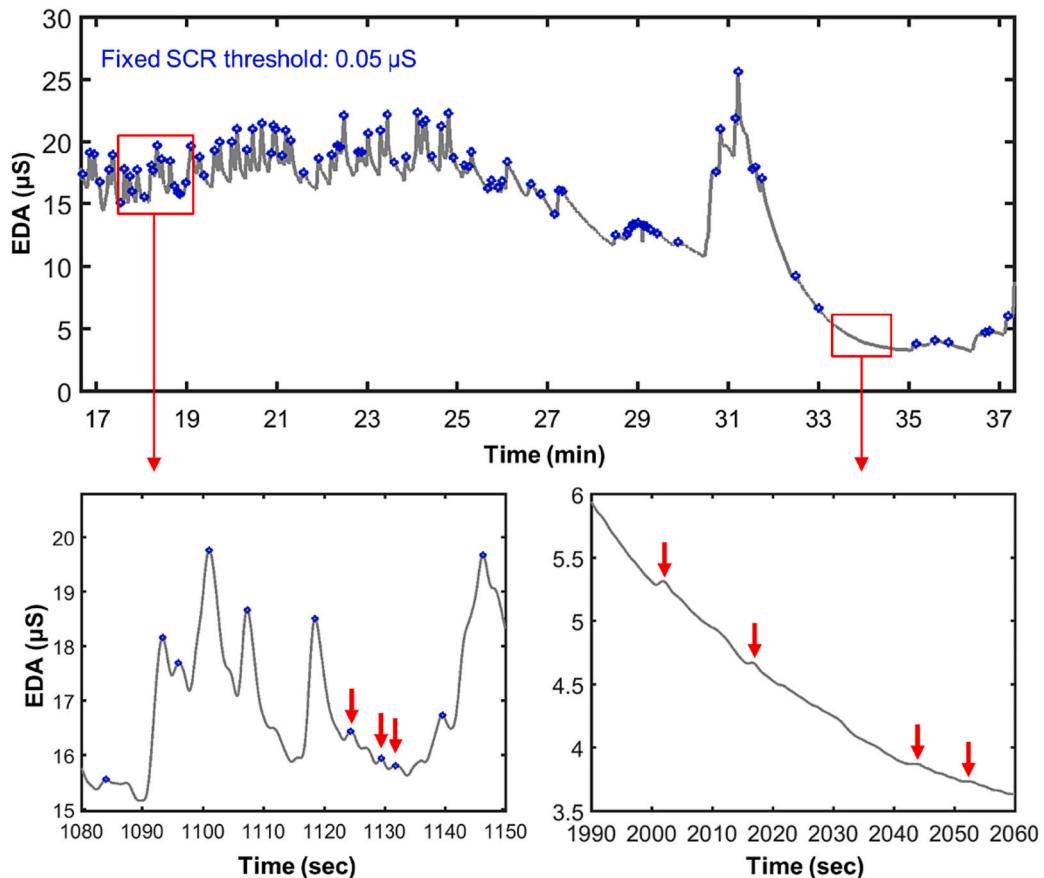


Fig. 1. Example time course showing that higher SC is associated with larger short-term fluctuations in SC. Blue circles indicate rapid SC fluctuations $>0.05 \mu\text{S}$, i.e., SCRs. The bottom left panel shows a portion of data with higher SCL during which there may be some spurious SCRs shown by the red arrows. The bottom right panel shows a portion of data with lower SCL in which deflections that could be potential SCRs were not detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

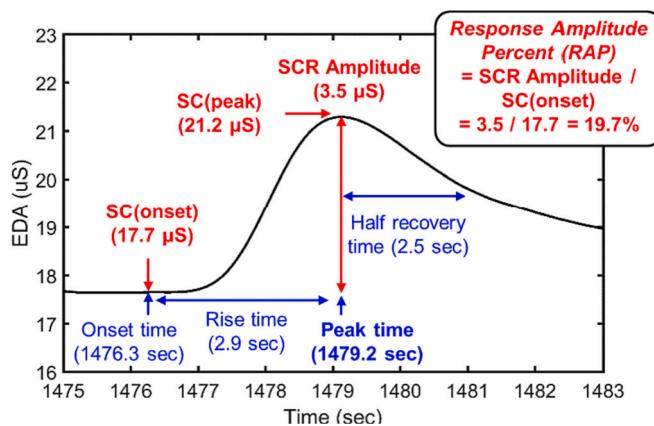


Fig. 2. Each SCR can be characterized by several distinct and measurable features. Example values of a single SCR are in black. Red measurable features are based on EDA level and are in units of μS , and include (1) SC(onset) - the SC at a trough (low point) that precedes a peak (high point); the low and high points are identified computationally by calculating first and second temporal derivatives in widely used peak-detection algorithms; we also refer to SC(onset) as the prevailing SCL, (2) SC(peak) - the SC at the peak that follows a trough, (3) SCR amplitude - the difference in SC values between the peak and its preceding trough. Following the definition of Boucsein et al. (2012), we consider SCL to be the mean of SC over some time period, after ignoring portions of data comprising SCRs. The mean SC is the average SC over some time period that does include SCRs. The blue measurable features are in time units (e.g., seconds), and include (1) onset time - time of the SC(onset), (2) rise time - the time between SC(onset) and SC(peak), (3) peak time - the time of SC(peak), and (4) half recovery time - the time from SC(peak) to the time at which the SC has reduced to a level of $(\text{SC}(\text{peak}) - 0.5 \times \text{SCR amplitude})$. We propose a novel EDA metric: the skin conductance response amplitude percent (RAP), which is the SCR amplitude divided by the SC value at onset (i.e., the trough preceding the SCR peak). One of our main EDA outcomes is SCR rate, which is the number of SCRs per unit time. The Methods section provides more details on how these are calculated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fernandez Rojas et al., 2023), (3) for use in biofeedback therapy, or (4) for affective computing contexts, including immersive video games or virtual reality-based entertainment where content is provided to a user based on their current physiological status. Our proposed FA threshold method is similar to Edelberg's suggestion to use a relative SCR threshold of 0.1% of the baseline SCL (Edelberg, 1972). However, it goes further by accounting for SC at the onset of each SCR, not just the SC at the beginning of an entire recording epoch.

The identification of SCRs is particularly important for outcomes that rely on counting the number of SCRs (e.g., rate of SCRs per unit time), which complement outcomes that rely on the amplitude of SCRs (i.e., average size of detected SCRs). SCR counting methods have been in use for decades (Boucsein, 2012) in studies of emotion (Davydov et al., 2013; Horesh et al., 2022; Kleckner et al., 2017; Troy et al., 2013), sleep (Kurinec et al., 2022), clinical anxiety (Tolin et al., 2023), and schizophrenia (Schell et al., 2005), including the large literature on electrodermal lability (Crider, 2008), which refers to stable and partially heritable (Vaidyanathan et al., 2014) individual differences in the rate of spontaneous non-specific SCRs that occur at rest (i.e., without the use of a specific evocative task or stimulus). SCR rate has shown reasonably high reliability of 0.70 to 0.84 for within-day measurements, and 0.64 across a 2–3-week period (compared to reliability estimates for SCR amplitudes: 0.52–0.69 short term, 0.56 across 2–3 weeks; Schulter and Papousek, 1992). SCR rate has also been suggested to better estimate sympathetic arousal than amplitude-based measures (Bach et al., 2010). It is a particularly useful outcome to quantify non-specific physiological responsiveness and for longer-duration stimuli or tasks like those used herein.

This study aims to evaluate our FA thresholding technique and examine whether it increases participant inclusion and more sensitively detects changes in SCR rate across varying experimental stimuli, tasks, conditions, and samples to get a more complete view of the effects when using FA thresholding in a single manuscript. We analyzed two independent datasets using three widely used traditional fixed SCR threshold levels and our FA thresholding technique to optimize our approach (Study 1) and then replicate and extend its use (Study 2). We also used alternative approaches such as statistical models to account for individual differences in SCR rates (mixed models with random intercepts) and SCR-SCL dependency (covariate adjustment by mean SC). Study 1 explores and optimizes the FA threshold using a dataset from 254 healthy community-recruited adults in a traditional laboratory setting with wired recording equipment who rated their subjective arousal following exposure to standard affective images and sounds. Other data collected from this sample are reported in (Khalaf et al., 2020; Kleckner et al., 2015; Wormwood et al., 2019). Study 2 tests for replication of the FA method using the optimal FA threshold level found in Study 1. The Study 2 dataset comprises 20 children with autism participating in high- and low-arousal social interactions in a clinical setting, while EDA is measured with wireless wrist-worn recording equipment. The initial findings from Study 2 are reported in (Jones et al., 2017a; Jones et al., 2019; Jones et al., 2017b; Jones et al., 2018; Kleckner et al., 2018).

2. Methods

2.1. Study 1

2.1.1. Participants

Two hundred and sixty participants (100 males, 159 females, 1 not reported) were recruited from the Boston area through fliers on college campuses, advertisements on [craigslist.com](#), and in the Metro newspaper. To be eligible, participants had to be 18 to 65 years old, native English speakers, free of skin allergies, sensitive skin, chronic medical conditions including asthma, or a history of cardiovascular illness, stroke, or mental illness, and not taking medications for attention deficit hyperactivity disorder, insomnia, anxiety, high blood pressure, rheumatoid arthritis, epilepsy/seizures, cold/flu, or any autonomically active medications for allergies. Participants were asked to refrain from consuming caffeine, tobacco, diet pills, sleeping pills, and alcohol for 24 h before the session. All participants provided informed consent following the Institutional Review Board of Northeastern University. Six participants were removed due to errors in data collection, yielding a final sample of 254 participants. See Table 1 for demographics.

Table 1
Sample demographics.

	Study 1	Study 2
	Healthy adults wearing wired EDA sensors in a traditional laboratory setting	Children with disorder wearing wireless EDA sensors in a clinical setting
Sample size	254	20
Age (years; mean \pm SD, range)	24.8 ± 10.2 (18–65)	7.2 ± 2.6 (5–13)
Gender	156 female (61.2%)	5 female (25%)
Race and ethnicity		
White, non-Hispanic	143 (56%)	11 (55%)
Asian/Pacific Islander	31 (12%)	1 (5%)
Black, non-Hispanic	31 (12%)	3 (15%)
Other/more than one race	14 (6%)	2 (10%)
Not reported	34 (13%)	3 (15%)
BMI (kg/m^2)	24.1 ± 5.2 (15.2–47.0)	Not available

2.1.2. Affective image and sound task

All Study 1 data reported herein are from affective image and sound tasks. Affective image stimuli were selected from the International Affective Image System (IAPS; Lang et al., 2008) and affective sound stimuli from the International Affective Digitized Sound System (IADS; Bradley and Lang, 2007). Sound and image stimuli were chosen to represent a range of normative subjective valence and arousal ratings across five blocks (Table 2). Stimuli were presented blockwise, and after each block of 10 stimuli, participants rated how pleasant/unpleasant and how activated/de-activated they felt on a 5-point scale using the Self-Assessment Manikin (SAM; Bradley and Lang, 1994). See Table 2 for affect ratings and Supporting Information for more details on this and all other procedures.

2.1.3. Procedures

All participants completed the affective image and sound tasks during a single experimental session. They also completed other tasks and questionnaires irrelevant to the current investigation.

2.1.4. EDA data collection and processing

After participants washed their hands with warm water only, we recorded EDA from the right palmar surface (i.e., from the thenar and hypothenar eminences) using disposable Ag/AgCl (11 mm diameter) electrodes from Biopac (Goleta, CA). These electrodes contain isotonic paste (0.5% chloride salt) in a 16 mm diameter, 1.5 mm deep well. Before placement, a small amount of isotonic electrode paste (Biopac; 0.5% saline in a neutral base) was added to the electrodes. EDA was sampled at 1000 Hz using BioLab v. 3.0.13 software (Mindware Technologies LTD; Gahanna, OH) and a BioNex 8-Slot Chassis (Model 50-3711-08). Raw skin conductance (in μS) was amplified by a gain of 100, resulting in a full-scale maximum of 500 μS . We used BioLab's default EDA settings, including a low-pass filter with a cutoff of 1 Hz and a moving average filter of 100 samples (0.1 s), which is commonly used to account for a very high sampling rate like that used here.

Skin conductance data were processed using Mindware Technologies EDA software (version 3.0.25) for ten 61.8-s epochs (one for each stimulus block). All epochs were visually inspected for movement artifacts by trained scorers before analyses (i.e., spiking signals which change greatly and quickly compared to the surrounding data, data loss such as abrupt changes to 0 μS , or participant movement from a video recording of the participant); this process did not result in any data loss. The analysis software denotes peaks using a second derivative and requires a minimum peak width of 1/5 of the sample rate to remove higher frequency peaks (i.e., spikes, which are likely noise-related). Then the preceding trough (onset) is identified, and an onset-to-peak change (SCR amplitude) is calculated. The output of interest from this initial analysis was a list of all detected SCRs using a fixed threshold $> 0.01 \mu\text{S}$ and the

Table 2

Mean valence and arousal of standardized images and sounds presented in Study 1.

Block	Stimuli	Category	Mean valence	Mean arousal
1	Images	Pleasant low arousal	7.43	3.73
2	Images	Pleasant high arousal	7.21	6.58
3	Images	Neutral low arousal	5.10	3.27
4	Images	Unpleasant low arousal	3.82	3.56
5	Images	Unpleasant high arousal	2.04	6.67
6	Sounds	Pleasant low arousal	6.70	4.75
7	Sounds	Pleasant high arousal	6.74	5.95
8	Sounds	Neutral low arousal	5.11	4.79
9	Sounds	Unpleasant low arousal	4.16	5.45
10	Sounds	Unpleasant high arousal	2.44	7.18

Mean valence and arousal ratings are from 1 to 9 where 1 is most unpleasant, 9 is most pleasant, and 1 is lowest arousal and 9 is highest arousal. Ratings in the table were obtained by averaging across the 10 stimuli in each block using the ratings in the manuals for the International Affective Image System (Lang et al., 2008) and the International Digitized Sound System (Bradley and Lang, 2007).

following features per SCR: amplitude, SCL value at the onset, and the affective category of the stimulus block during which the peak occurred. The SCR was only considered part of the block if the onset and SCR occurred in the block itself. For each block, we also calculated the mean SC over the entire epoch, regardless of whether it was considered the SCL or an SCR.

2.2. Study 2

2.2.1. Participants

We recruited families of 20 children and adolescents (5 females, age 5–13 years, $M = 8$ years) with a confirmed diagnosis of autism spectrum disorder made by a licensed clinician using the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012) at the Center for Autism and the Developing Brain (CADB). All caregivers provided informed consent, and children and adolescents seven years of age and older also provided assent following the Institutional Review Board at Weill Cornell Medical School.

2.2.2. Social interaction and puzzles tasks

Skin conductance data were collected (see Section 2.2.4) during the administration of a modified version of the Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016). This task consists of a 12-min examiner-child social interaction including two 5-min interactive play segments, each with a standardized set of toys (Play 1 and Play 2) and two 1-minute conversations (Conversation 1 and Conversation 2) conducted in the same order (i.e., Play 1, Conversation 1, Play 2, Conversation 2). See Supporting Information for more details.

Analyses in the present paper focus on Conversation 1 as a high-arousal condition, as it represents face-to-face interactions without toys. We chose the Puzzle task as a low-arousal condition because it requires minimal cognitive demand from participants. We obtained nearly identical results when Conversation 1 (results herein), Conversation 2 (not reported here), and the Pegboard task (not reported here) were selected as the high-arousal condition.

2.2.3. Procedures

Study procedures are described in (Jones et al., 2017b). In brief, there were three 1-hour assessment sessions at CADB: week 1, week 4, and week 8. In each session, participants were provided task instructions, followed by placement of the Q sensor device on participants by a trained researcher as described below before completing the BOSCC tasks (play and conversation), the pegboard fine motor task, the puzzle task, and watching a movie.

2.2.4. Skin conductance data collection and processing

Skin conductance data were obtained via the Q Sensor (Affectiva, Inc.; Waltham, MA) using methods described in Kleckner et al. (2018). The Q Sensor uses two electrodes positioned on the ventral surface of the non-dominant wrist and covered with an athletic sweatband to prevent the wearer from touching, moving, or turning off the device. The Q Sensor remained on the child's wrist for the duration of the visit. Start and end times for each behavioral task and Q Sensor recording start time were recorded. The Q Sensor measures skin conductance, skin surface temperature, and motion data (using a 3-axis accelerometer) at either 16 Hz or 32 Hz (varied across participants). For optimal signal quality, we recorded skin conductance using adhesive, solid conductive hydrogel Ag/AgCl electrodes affixed to the skin (22 mm square; model A10040-5; Vermed; Buffalo, NY).

We performed visual inspection and automated quality assessments for all skin conductance recordings using our freely available software (Kleckner et al., 2018), with all data analyzed that met the following four quality control parameters: (1) range 0.05 to 60 μS ; (2) maximum slope 10 $\mu\text{S}/\text{s}$; (3) temperature range 30–40 °C; and (4) data within 5 s of any data point that met one of the first three rules. We also computed the mean of the magnitude of acceleration, $A_{\text{mag}} = \text{abs}(\sqrt{A_x^2 + A_y^2 + A_z^2})$ –

1) from the 3-axis accelerometers, where 1 is subtracted to remove the effect of gravity, and the absolute value (abs) function is used to remove the impact of directionality.

We identified all skin conductance peaks in the EDA time course using the MATLAB *findpeaks* function, then searched for a preceding trough starting at 1 s before the peak and going back to 3 s before the peak until it was found. The trough was used to calculate an onset-to-peak change (SCR amplitude). The output of interest in the current investigation was a list of all detected SCRs (fixed threshold $> 0.01 \mu\text{S}$) and the following features per SCR: amplitude, SC value at the onset, and the task during which each peak was observed (i.e., Play 1, Conversation 1, Play 2, Pegboard, or Puzzle). Peaks and onsets needed to be within the stimulus condition to be counted as occurring in that condition. We also calculated the mean SC for each task.

2.3. Application of fixed and FA SCR thresholds

In both studies, we calculated RAP (SCR amplitude divided by the SC value at the onset of the SCR; see Fig. 2) to assess the relative size of the SCR to be used in our FA thresholding procedure. We required SCRs (i.e., those greater than a fixed threshold of $0.01 \mu\text{S}$) to have an SCR amplitude percent higher than the selected threshold. Hence, even our smallest SCRs were at least as large as SCRs commonly studied in the literature. To evaluate the effect of each specific threshold, we tested 200 fixed threshold levels between 0.01 and $10 \mu\text{S}$ and 200 FA threshold levels between 0.01% and 10% , both logarithmically spaced.

We used data from Study 1 to explore which FA threshold value maximized the observed effect size in our comparisons of higher vs. lower subjective arousal conditions. Specifically, we used polynomial curve fitting of plots of effect size vs. $\log_{10}(\text{threshold})$ in each of four comparisons: (1) unpleasant high arousal images vs. neutral images; (2) pleasant high arousal images vs. neutral images; (3) unpleasant high arousal sounds vs. neutral sounds; and (4) pleasant high arousal sounds vs. neutral sounds. We obtained the four FA threshold values that maximized the difference in SCR rate for each of the four comparisons. These four were averaged and used as the FA threshold value for replication testing in our Study 2 comparisons of FA vs. fixed thresholding at 0.01 , 0.03 , and $0.05 \mu\text{S}$. We also used exploratory curve fitting in Study 2 to estimate the maxima of the effect size as a function of the range of $\log_{10}(\text{threshold})$ values.

To test for associations between SCR rate and subjective arousal in the evocative images and sounds tasks in Study 1, we ran two types of models. First, to obtain AIC goodness-of-fit values, we ran linear mixed models predicting rated arousal values using SCR rate (from fixed 0.01 , 0.03 , or $0.05 \mu\text{S}$, or FA 1.9%) with a random intercept for each participant. We ran separate models for each SCR rate (3 fixed thresholds, 1 adaptive), each stimulus (images or sounds), and each demographic subgroup (all participants, by age, and by race). To obtain R^2 values, we ran Pearson's correlations between group-averaged rated arousal and SCR rate.

Based on the literature assessing SCR rates in lower and higher arousal conditions (e.g., Boucsein, 2012; Zimmer, 2000), we expected an effect size of Cohen's d of 0.26 (note: estimating post-hoc power using an expected effect size from the literature has been suggested to be statistically preferable over using effect sizes observed in the current dataset of interest; Lakens, 2022). A post-hoc sensitivity analysis conducted in G*Power (Faul et al., 2009) yielded power estimates of 98.5% for Study 1 ($n = 254$) and 19.7% for Study 2 ($n = 20$), respectively, to detect an effect of this magnitude using paired-samples t -tests.

3. Results

Our first results provide additional evidence that older age and Black race is associated with lower SCL, compared to White race. We next show that skin conductance values positively correlate with SCR amplitudes detected via a fixed threshold of $0.05 \mu\text{S}$, suggesting that SCR

rates can be higher or lower depending on SCL. Next, we use our FA thresholding procedure and compare results to traditional fixed SCR thresholding in analyses of SCR rates in different conditions and compared to subjective arousal. Study 1 data were used to determine an optimal FA threshold, followed by an independent test for replication in Study 2. To streamline the presentation of our key results, we provide detailed statistical models in the Supporting Methods and Results.

3.1. Demographic predictors of SCL (Study 1)

In Study 1, our sample consisted of 254 healthy adults aged 18 to 65 years, on average 24.8 ± 10.2 years, with 61.2% reporting female gender, and race/ethnicity comprising mostly White (56%), followed by other/more than one race (13%), Asian/Pacific Islander (12%), Black (12%), and Hispanic (6%). Tonic SCL varied more across participants than within participants: across-participants SD = 4.2 (range 3.9 – 4.5); within-participant SD = 0.6 (range 0.002 to 7.0). Consistent with the prior literature, we found that tonic SCL was lower with older age (slope $\pm \text{SE} = -0.08 \pm 0.03 \mu\text{S}/\text{year}$, $p = 0.002$) and lower in Blacks compared to non-Hispanic Whites ($-1.46 \pm 0.66 \mu\text{S}$, $p = 0.027$). We also found that tonic SCL was higher in Hispanics compared to non-Hispanic Whites ($+2.96 \pm 0.88 \mu\text{S}$, $p = 0.0009$).

3.2. SCR amplitude is positively correlated with SCL

Consistent with prior work (Boucsein et al., 1984; Venables and Christie, 1980), we found a significant positive association between SCL and SCR amplitude (using a fixed threshold of $0.05 \mu\text{S}$) both within- and across-participants in both studies: Study 1 across-person (223 individuals, $F = 231$, $p < 0.0001$); Study 1 within-person (4976 SCRs, $t = 10.9$, $p < 0.0001$); Study 2 across-person (45 assessments across 20 individuals at 1–3 time points each, $F = 25$, $p = 0.002$); and Study 2 within-person (5137 SCRs, $t = 13.7$, $p < 0.0001$).

3.3. Exploring the effects of SCR threshold level (Study 1)

Next, we compare the rate of SCRs (number of SCRs per minute) in normatively high vs. low psychological arousal conditions and quantify the difference in effect size (Cohen's d) for different fixed and FA thresholds. The fixed thresholding uses SCR amplitude (e.g., $0.05 \mu\text{S}$) whereas FA thresholding uses our derived metric, the RAP, which is the SCR amplitude divided by the SC value at onset preceding the SCR peak (Fig. 2). We used polynomial curve fitting to find the FA threshold value that maximized the observed effect size for comparisons of high arousal vs. neutral conditions.

The plots in Fig. 3a suggest that an optimal threshold can be found that maximizes effect size. The left portion of each FA threshold plot is flat because, at very liberal FA thresholds (to the left of $10^{-1}\%$), the threshold "detects" all SCRs, yielding (as expected) identical results to those with a traditional fixed threshold of $0.01 \mu\text{S}$.¹ With stricter FA thresholds (moving toward the right in the plot), effect sizes increase modestly, starting at $\text{RAP} = 0.1\%$, peaking at approximately $\text{RAP} = 0.5$ – 6% , and then decreasing for larger RAP values. The curve fitting indicates optimal RAP values as 6.3% , 0.6% , 0.3% , and 0.5% for unpleasant images, pleasant images, unpleasant sounds, and pleasant sounds, respectively. This yields an average RAP value of 1.9% , which works reasonably well for all the conditions (vertical red line in Fig. 3a).

In Fig. 3b, we also show the percentage of the sample considered responders exhibiting at least one SCR across all stimulus conditions. At an FA threshold of 1.9% , 93% of the sample are considered responders.

¹ More specifically, a threshold of $\text{RAP} > 10^{-2}\%$ means $\text{SCR} / \text{SC}_{\text{trough}} > 0.0001$, which allows all SCRs with amplitude $> 0.01 \mu\text{S}$ as long as $\text{SC}_{\text{trough}} < 100 \mu\text{S}$, which was always true because the maximum EDA value in the data we analyzed was $60 \mu\text{S}$.

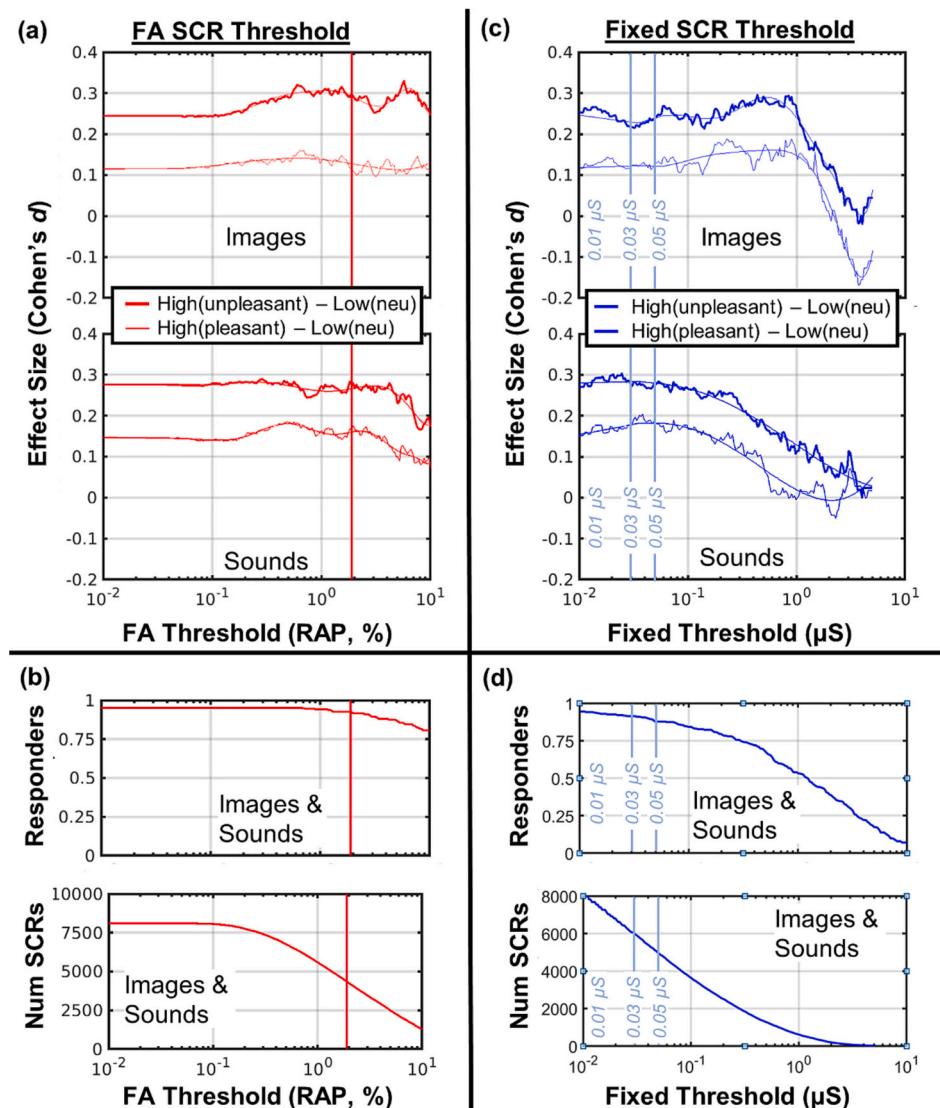


Fig. 3. In the exploratory portion of Study 1, we determined the optimal FA threshold level using curve-fitting to maximize effect sizes comparing the SCR rate during high vs. low arousal images and sounds in 254 healthy adults. Each figure shows two sets of data: unpleasant high arousal stimuli minus neutral low arousal stimuli (thick line) and pleasant high arousal stimuli minus neutral low arousal stimuli (thin lines). The vertical red lines in (a) and (b) show the optimal FA threshold level of 1.9% found by averaging each of the four optimal FA threshold levels. The lower plots show the percentage of the sample that are responders—who exhibited at least one SCR across all conditions at the given threshold—and the number of SCRs across all stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, we show the number of SCRs for all images or sounds. The FA threshold at 1.9% retains approximately 4300 SCRs, which is 53% of the 8100 SCRs $>0.01 \mu$ s (i.e., the original set of SCRs available for FA thresholding) and 86% of the number of the 4976 SCRs $>0.05 \mu$ s.²

For comparison, we also tested a range of fixed SCR threshold levels. The plots in Fig. 3c suggest that increasing fixed thresholds can increase effect sizes, although not for negative sound stimuli. The effect sizes found in these comparisons of higher vs. lower arousal stimuli were typically small (Cohen, 1992), ranging from 0.15 to 0.29 depending on the comparison and fixed threshold level. Fig. 3d shows that the percentage of responders and number of SCRs is quickly reduced upon increasing the fixed threshold level, with 94% responders at 0.01 μ s, 91% responders at 0.03 μ s, and 88% responders at 0.05 μ s.

² This 86% is not a strict subset of the SCRs found at 0.05 μ s because FA thresholding detects SCRs smaller than 0.05 μ s as long as the SCL is low enough (below 2.6 μ s for an FA threshold of 1.9%).

Next, we explored whether commonly used statistical methods, such as mixed modeling and covariate adjustment by mean SCL, influenced the difference between FA and fixed thresholds. FA thresholding at 1.9% yields equivalent performance compared to traditional fixed thresholding in terms of effect size and goodness-of-fit even when controlling for within- and between-participant differences in SCL using mixed modeling (Figs. S2 and S3; see Supporting Methods for details). Therefore, we did not observe a statistical drawback to using this distinct SCR criterion, which enabled us to include a larger portion of our sample. We obtained identical conclusions when excluding non-responders.

3.4. FA SCR threshold increases correspondence of SCR rate and self-reported arousal (Study 1)

This analysis explores how FA thresholding affects the ability to assess the relationship between SCR rate and subjective arousal in the evocative images and sounds tasks in Study 1. We hypothesized that the SCR rate would more closely track participant arousal ratings using FA

thresholding at 1.9% (value obtained from the prior analysis) compared to traditional fixed thresholding at 0.01, 0.03, or 0.05 μS . Specifically, we predicted that linear models would show increased power (t ratio), superior goodness-of-fit, and smaller standard errors using FA thresholding than fixed thresholding.

As expected, Fig. 4a shows a positive association between SCR rate and rated arousal in response to evocative images and sounds. Moreover, it suggests that FA thresholding (red) performs slightly better than fixed thresholding at 0.01, 0.03, or 0.05 μS because the FA thresholding plots exhibit a stronger, more monotonic linear relationship than fixed thresholding, as indicated by higher R^2 values for images (96% vs. 74–84% for fixed thresholding) and sounds (89% vs. 63–78% for fixed thresholding) and lower AIC values for images (4303 vs. 4317–4319 for fixed thresholding) and sounds (3966 vs. 3970–3971 for fixed thresholding). This stronger linear relationship with adaptive thresholding is important because a linear relationship between two variables gives greater sensitivity to detect or infer a change in one variable (Y) when measuring the other (X) across the entire range of measured values because a line's slope is constant.³

Using mixed modeling, we see that FA thresholding has a slight advantage over fixed thresholding in modeling the positive association between the rate of SCRs and rated arousal from evocative sounds or images, based on larger t ratios than fixed thresholding at 0.05 μS (advantage of $M \pm SE = 0.77 \pm 0.22$ units on the t ratio; $p = 0.002$; Table 3) and at 0.03 μS (advantage of 1.02 ± 0.26 units on the t ratio; $p = 0.009$; Table S3; Fig. S3), after statistically accounting for which type of mixed model and covariates were used, stimulus type (sounds, images), and participants (all, only responders).

Finally, we explored how age and race related to the effectiveness of the thresholding method based on a linear mapping between SCR rate and rated arousal. In Fig. 4b with young adults, adaptive thresholding is comparable or slightly more sensitive than fixed thresholding for images (R^2 of 95% vs. 79–88%, AIC of 3469 vs. 3480–3482) and much more sensitive for sounds (R^2 of 81% vs. 43–65%, with comparable AIC of 3185 vs. 3185–3187). In Fig. 4c with adults at least 30 years old, adaptive thresholding is much more sensitive than fixed thresholding for images (R^2 of 99% vs. 43–83%, with comparable AIC of 825 vs. 825–826) and comparable or less sensitive for sounds (R^2 of 41% vs. 29–55%, with comparable AIC of 3185 vs. 3185–3187). In Fig. 4d with White participants, adaptive thresholding is comparable for images (R^2 of 90% vs. 80–84%, with better AIC of 2389 vs. 2399–2402) and comparable for sounds (R^2 of 84% vs. 87–91% with slightly worse AIC of 2140 vs. 2134–2138). Finally, in Fig. 4e with Black participants, adaptive thresholding is much more sensitive for images (R^2 of 65% vs. 17–22%, with slightly better AIC 545 vs. 547). For sounds, the adaptive thresholding yields a less unexpected result in that all the fixed thresholding methods show an unexpected negative correlation between SCR rate and rated arousal, whereas adaptive thresholding produces a small and likely non-meaningful negative correlation ($R^2 = 16\%$ vs. 50–59%, with comparable AIC of 496 vs. 495–496).

3.5. Independently testing for replication comparing FA thresholding vs. fixed thresholding (Study 2)

Next, we quantitatively assess FA thresholding by directly comparing results from FA thresholding at 1.9%—the value obtained from Study 1—and fixed thresholding at 0.05 μS and 0.03 μS —two of the most commonly used values in the literature (Boucsein, 2012). First, as expected, the FA thresholding at 1.9% indicates that the high-arousal condition has a higher SCR rate ($t = 2.22$, $p = 0.041$; Table S5 Model A). By comparison, fixed thresholding at 0.05 μS yields smaller and non-

significant effects ($t = 1.19$, $p = 0.251$; with 0.03 μS threshold $t = 1.26$, $p = 0.227$). These results suggest that analyses using a fixed threshold may lead to false negative results (type II errors), as the face-to-face conversation in an autistic sample likely induces greater arousal than the puzzle task based on its greater duration of eye contact in this sample (Jones et al., 2017b) and prior demonstrations that eye contact increases SCL in children with autism spectrum disorder (Kylliainen and Hietanen, 2006).

When we analyzed the t ratios across several models and conditions, we found that FA thresholding at 1.9% yielded larger t ratios than fixed thresholding at 0.05 μS (advantage of 1.24 ± 0.14 units on the t ratio; $p < 0.0001$; Table 4) and fixed thresholding at 0.03 μS (advantage of 1.09 ± 0.13 units on the t ratio; $p < 0.0001$; Table S6) after adjusting for the type of mixed model and covariates used, high-arousal task (Conversation 1 or Play 2), session (1,2,3), and participants (all, responders only). These conclusions also held when considering participants' movement (which can affect the EDA data), as we also found the superiority of FA thresholding when adjusting by per-condition mean acceleration readings from the Q Sensor and when comparing the low-arousal puzzle task to a high-arousal play task, which involved more movement (Model D in Fig. S5).

3.6. Exploring the effects of SCR threshold level (Study 2)

Fig. 5 explores Study 2 data to find an optimal FA threshold level. Our results suggest superiority of FA thresholding compared to traditional fixed thresholding across a broader range of FA threshold levels (not just 1.9%). Compared to Study 1 results, Fig. 5a shows that FA thresholding yields a larger increase in effect size with a pattern that replicates over the three sessions (over the course of eight weeks). Fig. 5b also shows that these increases in effect size from $10^{-1}\%$ to $10^{0.2}\%$ do not result from removing participants from the analysis but rather by removing some SCRs from the analysis, again consistent with the results from Study 1. Specifically, using an FA threshold of 1.9% retains of 1582 (56%) of the original 2817 SCRs $> 0.01 \mu\text{S}$ in Session 1, another 1482 (66%) in Session 2, and 1033 (54%) in Session 3. This is approximately 84% of the number of SCRs detected using a fixed threshold of 0.05 μS .

Finally, we explored how effect size related to increasing fixed threshold levels in Study 2. Fig. 5c illustrates that for Session 1 and 2, the effect size was fairly stable across threshold levels $10^{-2} \mu\text{S}$ to $10^{-1} \mu\text{S}$, and for Session 3, there was a rapid decrease in effect size from 0.18 to 0 spanning $10^{-2} \mu\text{S}$ to approximately $10^{-1} \mu\text{S}$. Fig. 5d shows that these changes come at the undesirable expense of removing a few participants from the analysis. The effect sizes found in these comparisons of higher vs. lower arousal conditions were typically small (Cohen, 1992), ranging from 0.17 to 0.38 depending on the session number and fixed threshold level, and one effect size of -0.02 , which is effectively zero (Session 3).

4. Discussion

This work evaluates a novel fixed plus adaptive SCR thresholding technique (FA thresholding) based on the SCR peak amplitude divided by SC value at onset (response amplitude percent, or RAP, see Fig. 2). This method complements fixed thresholding of the absolute SCR amplitude, which is traditionally used in the literature. Our method is an extension of a prior relative SCR thresholding procedure (Edelberg, 1972) where we instead propose dynamic thresholding relative to the prevailing SC immediately before each SCR, rather than based on a single SCL at the start of a recording. We demonstrate that using this FA thresholding procedure to compute SCR rates yields equivalent or modestly larger effect sizes in tests comparing normatively high vs. low subjective arousal conditions in two distinct datasets: 254 healthy young adults wearing wired EDA sensors in a laboratory while viewing or listening to standardized evocative images and sounds; and 20 children with autism spectrum disorder wearing wireless EDA sensors in a

³ By comparison, exponential, sigmoidal, or parabolic relationships which all lose sensitivity to change in X or Y at the steepest or shallowest sloping regions, respectively, similar to a floor or ceiling effect.

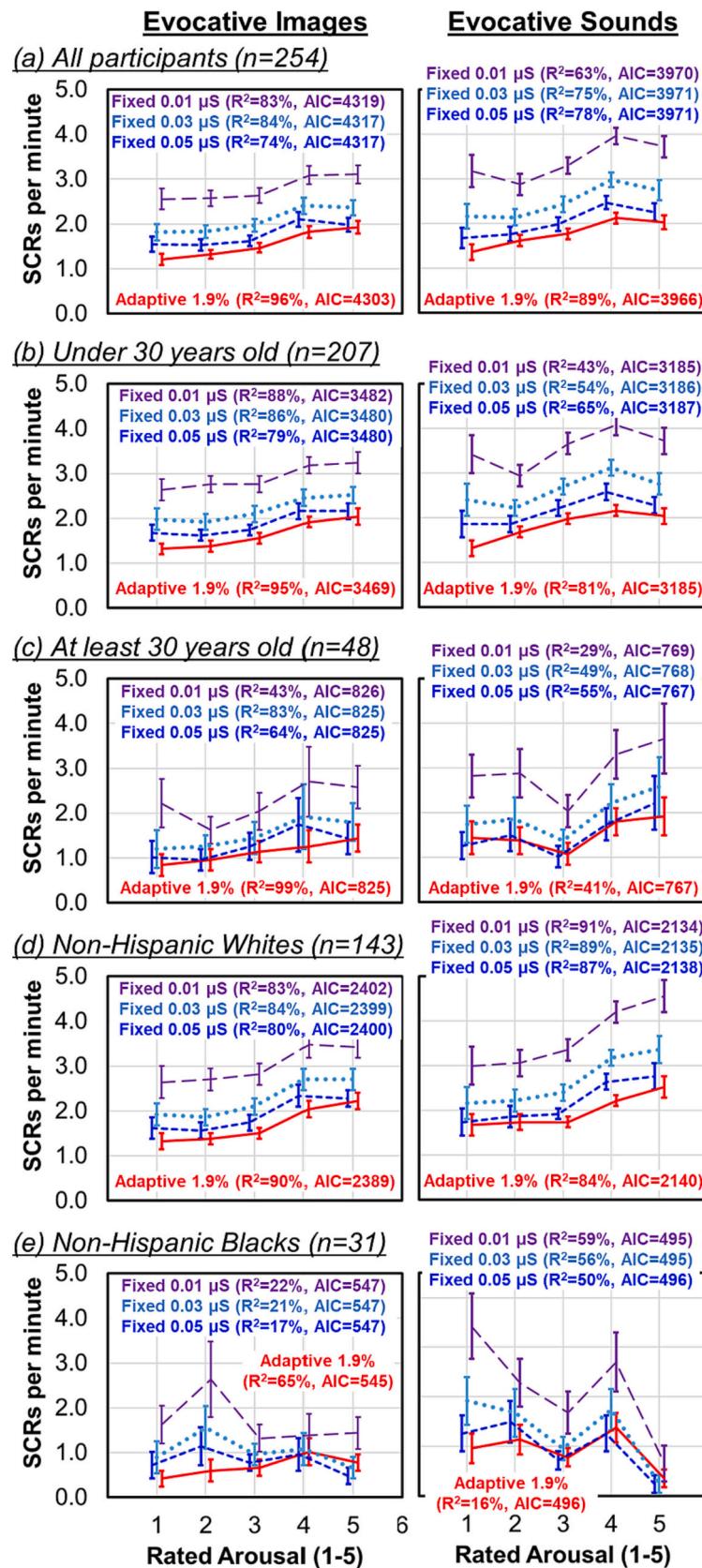


Fig. 4. FA thresholding (red solid line) yields slightly smaller error bars, higher goodness of fit (R^2) in nearly all cases, and a more linear and monotonic relation between SCR rate and arousal ratings compared to traditional fixed thresholds of 0.01 μ S (purple line at top), 0.03 μ S (teal dotted line), and 0.05 μ S (blue dashed line). Each of the 254 participants contributed five data points (X, Y pairs) to each subplot (i.e., 5 arousal ratings plus EDA data). The error bars are standard errors. The x-axis values were offset for display purposes only (not analysis) to avoid overlapping error bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Study 1 statistical comparison of power (t-ratios) using FA thresholding at 1.9% vs. fixed thresholding at 0.05 μ S using linear regression to model SCR rate using Arousal Rating.

Term	Estimate	SE	DF	t ratio	p-Value	Lower 95%	Upper 95%
Intercept	2.65	0.27	18	9.92	<0.0001	2.09	3.21
Method: FA 1.9% fixed 0.05 μ S	0.77	0.22	18	3.54	0.0024*	0.31	1.23
Model: A – C	-0.11	0.27	18	-0.41	0.6883	-0.67	0.45
Model: B – C	0.18	0.27	18	0.66	0.5172	-0.38	0.74
Task: Sounds – Images	-0.44	0.22	18	-2.01	0.0594	-0.90	0.02
Participants: All – Responders	1.70	0.22	18	7.8	<0.0001	1.24	2.16

Larger estimates indicate greater statistical power (t-ratio) associated with that variable's level.

Source data in Supporting methods Table S1.

Identical conclusions were obtained using a fixed threshold of 0.03 μ S, see Table S3.

Model A SCR Rate = Arousal(fixed) + intercept(random).

Model B SCR Rate = Arousal(fixed) + Mean SC(fixed) + intercept(random).

Model C SCR Rate = Arousal(fixed) + Mean SC(random) + intercept(random).

Table 4

Study 2 statistical comparison of FA thresholding at 1.9% vs. fixed thresholding at 0.05 μ S using linear regression to model SCR rate using condition (high arousal conversation or play minus low arousal puzzle).

Term	Estimate	SE	t ratio	p-Value	Lower 95%	Upper 95%
Intercept	1.43	0.20	7.26	<0.0001	1.04	1.83
Method: FA 1.9% - Fixed 0.05 μ S	1.24	0.14	8.89	<0.0001	0.96	1.52
Model: A – D	-0.73	0.17	-4.26	<0.0001	-1.07	-0.39
Model: B – D	0.22	0.17	1.27	0.2084	-0.12	0.56
Tasks: (Convo1-Puzzle) – (Play2-Puzzle)	-0.15	0.14	-1.05	0.2997	-0.43	0.13
Participants: All – Responders	-0.37	0.14	-2.66	0.0099	-0.65	-0.09
Session: 1–3	1.13	0.17	6.6	<0.0001	0.79	1.47
Session: 2–3	0.50	0.17	2.9	0.0051	0.15	0.84

Larger estimates indicate greater statistical power (t-ratio) associated with that variable's level.

Source data in Table S4.

Identical conclusions were obtained using a fixed threshold of 0.03 μ S, see Table S6.

Model A SCR Rate = Condition(fixed) + intercept(random).

Model B SCR Rate = Condition (fixed) + Mean SC(fixed) + intercept(random).

Model D SCR Rate = Condition (fixed) + Mean SC(fixed) + Mean Acceleration (fixed) + intercept(random).

clinical setting engaging in a high-arousal social interaction and a low-arousal puzzle task with measures obtained three times across eight weeks. We show that the FA threshold allows researchers to retain more participants' data and thus only modestly increase statistical power and research equity. These results suggest a way to address challenges researchers face when analyzing EDA data and the inclusion or exclusion of study participants in the analysis phase (e.g., the fact that participants with lower SC levels are more likely to be excluded; Lonsdorf et al., 2019). Our method also may help conduct research in a more equitable manner, which is a more broadly and recently recognized issue (Bradford et al., 2022; Webb et al., 2022).

Several results and methods support these conclusions. First, Study 1 data were used in an exploratory manner to suggest an FA threshold level that works well for several conditions, which we found to be 1.9%. Study 2 data were used to independently test for replication and extension of Study 1 by comparing FA thresholding at 1.9% to traditional fixed thresholding ranging from 0.01 to 0.05 μ S in a very different sample and setting. Second, our findings suggest superiority of FA thresholding over fixed thresholding for several analytical frameworks including linear mixed modeling with or without adjusting for mean SC (Figs. S2, S3, S4, and S5). Third, in an independent evaluation exploring the relationship between ratings of fear and rate of SCRs (McVeigh et al.,

2023), an optimal RAP threshold of 3% was identified using the FA threshold (after testing values from 1 to 10%), consistent with potential optimal values ranging 0.5% to 6% found here across the different conditions and studies.

The FA threshold level that maximized effect sizes from Study 1 was RAP > 1.9%, resulting in approximately 85% of the peaks considered SCRs using a traditional fixed threshold of 0.05 μ S, or 50% of the peaks considered SCRs at 0.01 μ S. The FA threshold of 1.9% categorized slightly more subjects as responders than the fixed threshold at 0.05 μ S (93% vs. 88% in Study 1, 78% vs. 75% in Study 2) because FA thresholding includes smaller SCRs that occur at low SCLs that would otherwise be excluded using a typical fixed threshold. Altogether, the FA threshold method identifies an SCR metric that, in these experimental contexts, more closely relates to self-reported arousal than the SCR metrics traditionally identified in these types of analyses.

Our FA thresholding procedure may be particularly relevant to electrodermal datasets that demonstrate a large range in SCL (e.g., Lykken et al., 1966), including data collected in ambulatory settings wherein greater changes in sympathetic nervous system activity can occur and more sources of noise are present (Doberenz et al., 2011; Goodwin et al., 2008; Adams et al., 2017; Wilhelm and Grossman, 2010). Indeed, our results revealed greater beneficial effects of FA thresholding in our sample of children with autism spectrum disorder who moved freely during the tasks wearing wrist-based devices (Study 2) compared to the data from seated healthy young adults in a standardized laboratory experiment (Study 1). Moreover, FA thresholding can help increase research equity by not systematically excluding data from participants from demographic groups that have lower SCL on average (Bradford et al., 2022; Webb et al., 2022) such as African Americans (Alexandra Kredlow et al., 2017) and older adults (Bari et al., 2020; Eisendorfer et al., 1980). Finally, because we observed more variation in SCL between participants than within participants, an approach involving scaling each participant's data time series in an appropriate way may also prove fruitful, though this requires further empirical investigation.

This study has several notable strengths. First, we had a large sample size—254 participants in Study 1 and 20 participants with three longitudinal measures across eight weeks in Study 2, which increases the reliability of our findings. Second, because our results were consistent across two very distinctive studies and samples, our conclusions should have better generalizability across diverse settings (college laboratory and outpatient clinic), populations (healthy young adults and children with autism spectrum disorder), tasks (standard evocative images and sounds, face-to-face social interactions, and puzzle games), recording hardware (traditional wired laboratory EDA sensors with standard isotonic electrodes and wireless EDA sensors with high-conductivity gelled electrodes), and other study-related factors that likely changed across the eight weeks of Study 2 (e.g., exact sensor placement, mood, hydration, temperature, humidity, task practice, habituation effects,

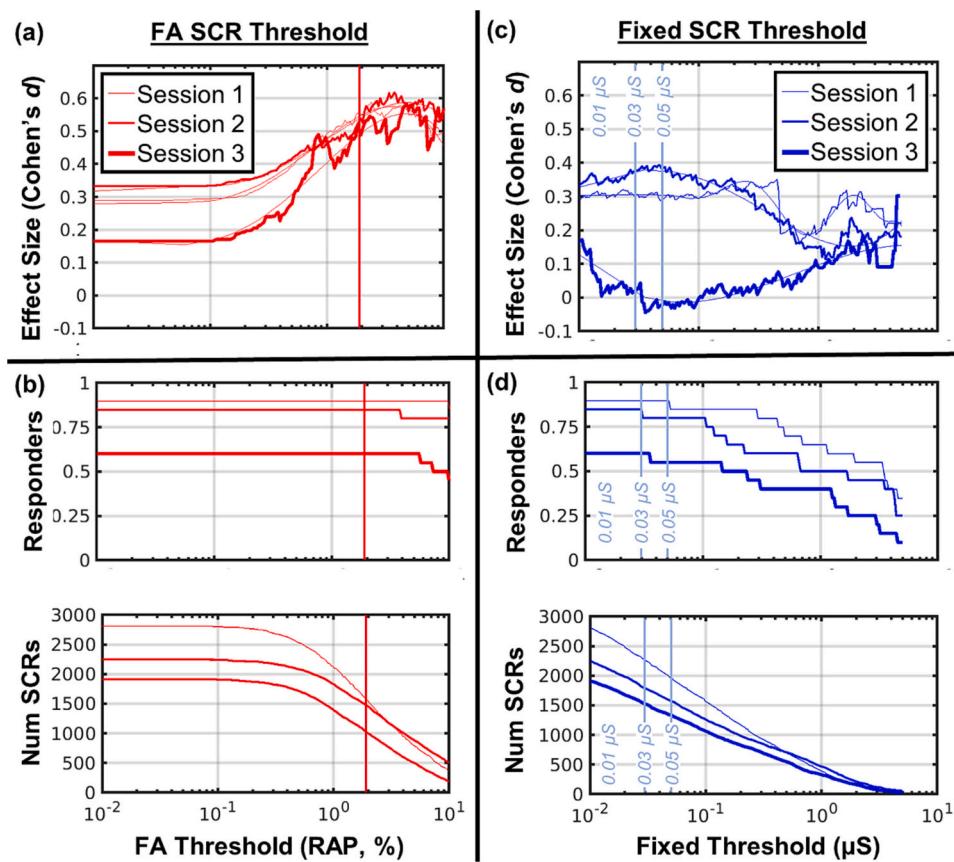


Fig. 5. In Study 2, among 20 children with autism spectrum disorder, FA thresholding (a, red) yields higher effect sizes than traditional fixed thresholding (c, blue) in comparisons between high vs. low arousal tasks: a face-to-face conversation vs. puzzle. Results are consistent with and extend results from Study 1, suggesting consistency over the three sessions over eight weeks (thin, thicker, and thickest lines). The vertical red line in (a) shows the optimal threshold value 1.9% found in the exploratory portion of Study 1. Study 2 was treated as a test for replication and the effect size curves here did not influence or inform the optimization of FA threshold values found in Study 1. The lower plots in (b) and (d) show the percentage of responders who exhibited at least one SCR across all conditions at the given threshold and the number of SCRs across all stimuli. For the exploratory portion of Study 2, each effect size trace was curve-fitted to determine the FA threshold value that maximizes the effect size. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

etc.). Third, our approach is computationally simple because it relies on only two EDA metrics—SCR amplitude and SC onset value—that are typically available in EDA analysis software packages. Moreover, because of its simplicity, our method can easily be used in real-time analyses of EDA data, wherein complex across-participant statistical methods such as mixed modeling and covariate adjustment are often not feasible. Finally, FA thresholding increased the number of responders in our dataset, which is helpful since handling of non-responders is already an impediment to research using EDA (Alexandra Kredlow et al., 2017).

The present results suggest that FA thresholding only modestly increases statistical power to detect differences across people and conditions in several ways. First, our method removes the relatively smallest SCRs that would otherwise be detected using a traditional fixed threshold. In statistical terms, FA thresholding better accounts for large between-person variance, as evidenced by the fact that mixed modeling estimates of between-person covariance were lower when using an FA threshold compared to a fixed threshold (Table S7). Second, FA thresholding appears to partially address the correlation between SCR rate and SCL because covariate adjustment for mean SCL explained less variance in models that already used FA thresholding compared to models that used fixed thresholding (i.e., FA thresholds yielded slightly smaller increases in power by going from Model A to Model B than fixed thresholds did when going from Model A to Model C; Figs. S2, S3, S4).

Notwithstanding, this study has some limitations. First, there is variability across studies in the optimal FA threshold. Based on Study 1 data, we found the largest effect sizes on average at RAP = 1.9%, but this

number varied by condition (Study 1 range 0.5–6% and Study 2 range 2–5%). Nevertheless, our results show that even if the RAP value is not perfectly optimized for a given dataset, it can still match or outperform fixed thresholding at 0.01, 0.03, or 0.05 μS . Second, the FA threshold requires selecting a fixed threshold to build the initial list of detected peaks before applying a second criterion for the relative size of the peak. Here we chose 0.01 μS as that minimal threshold because it was the smallest value typically used in the literature, but future research could test other values or other methods altogether to identify peaks of interest. Third, future work should assess the reproducibility and generalizability of FA thresholding vs. fixed thresholding in other samples, contexts, psychological tasks, and different EDA recording devices (Kleckner et al., 2021). Our two studies were quite different in methodologies, so it is unclear which aspects may or may not translate between these two datasets. Notwithstanding, this method appears to meet or exceed the performance of fixed thresholds. Fourth, this work focused on relationships between SCR rate and self-reported arousal in a specific, empirically well-defined context. However, SCR rate cannot be assumed to relate in a one-to-one way with self-reported arousal in all tasks or contexts. Further, EDA outcomes such as SCR rate are also related to orienting, attention, salience, etc. and have been used in clinical applications such as seizure detection, autism, and biofeedback (Dindo and Fowles, 2008; Tronstad et al., 2022). Future work will require determining the mapping of these psychological phenomena with SCR rate (and other SCR-related outcomes) using the adaptive threshold approach. Finally, because our main approach was to obtain a simple

measure of the rate of SCRs that is less dependent on SCL, it would be interesting in future work to assess whether FA thresholding remains superior to fixed thresholding even in cases where SCL and the number of SCRs are uncorrelated.

We make the following suggestions for the use of FA thresholding. First, we reiterate standard suggestions for analyses of SCR data, whereby one uses a rigorous quality assessment procedure including visual inspection and removal of obvious movement artifacts (e.g., via our freely available automated quality assessment procedure for EDA data; Kleckner et al., 2018), followed by use of an FA threshold at 1.9% for laboratory recordings or 3.3% for ambulatory conditions with features similar to those used here. However, researchers can determine their own optimal FA threshold using a small independent pilot study per the approach we used in Study 1 when using paradigms with very different task or sample features than those tested here. After applying an FA threshold, we also suggest that researchers use mixed modeling to increase statistical power and further address between-person variance (Kristjansson et al., 2007), including controlling for mean SCL as a nuisance covariate when modeling SCR rates because this further increased (or maintained) statistical power in our data (i.e., compare Models A and B, which adds the covariate, in Figs. S2, S3, and S4). However, if mixed modeling is not possible (e.g., real-time EDA processing), FA thresholding will likely help even more than traditional fixed thresholding.

The results herein suggest that FA SCR thresholding is a simple and useful method that expands the EDA analytical toolbox. This novel method demonstrates modest increases in sensitivity to skin conductance differences between higher vs. lower arousal conditions and enables the inclusion of a greater proportion of the study sample. While the sensitivity benefits of our method were small in the datasets presented here, the benefits of FA thresholding are likely to be most pronounced in datasets with large variations in basal SCL. Moreover, this work is important as a starting point for future research to continue to explore and optimize the detection of SCRs. Future research should test for reproducibility and generalizability of FA thresholding in other populations, contexts, and psychological tasks and work toward optimizing its use to obtain larger and more consistent improvements in the measurement of SCRs.

Funding/support

This research was supported in part by the Simons Foundation (336363 and 568330 to MSG, IRK, CL, and RMJ), the National Institute of Nursing Research (NR013500 to MSG), the National Institute on Deafness and Other Communication Disorders (P50 DC013027 to MSG), the National Cancer Institute (K07CA221931 to IRK and R25 CA102618 and UG1 CA189961 to support IRK and 1U01CA193632-01A1 to LFB), National Institute of Mental Health (R01MH113234 to LFB), and grants from the U.S. Army Research Institute for the Behavioral and Social Sciences (W911NF-16-1-0191 to KSQ and JBW and W5J9CQ-12-C-0049 to LFB). The views, opinions, and findings contained in this paper are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents.

Declaration of competing interest

Catherine Lord receives royalties for the diagnostic instruments, the ADOS and the ADI-R, all proceeds that were related to this study are donated to charity. The other authors declare no potential conflicts of interest.

Acknowledgements

The authors would like to acknowledge Lauren Sears and Justin Kopec for their help with data collection on Study 1, Caroline Carberry

and Amarelle Hamo for their help with data collection on Study 2, editorial and statistical assistance from Dr. Amber Kleckner, and editorial and scientific input from Dr. James Heathers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijpsycho.2023.112280>.

References

- Adams, Z.W., McClure, E.A., Gray, K.M., Danielson, C.K., Treiber, F.A., Ruggiero, K.J., 2017. Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. *J. Psychiatr. Res.* 85, 1–14.
- Alexandra Kredlow, M., Pineles, S.L., Inslitch, S.S., Marin, M.F., Milad, M.R., Otto, M.W., Orr, S.P., 2017. Assessment of skin conductance in African American and Non-African American participants in studies of conditioned fear. *Psychophysiology* 54 (11), 1741–1754. <https://doi.org/10.1111/psyp.12909>.
- Bach, D.R., Friston, K.J., Dolan, R.J., 2010. Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *Int. J. Psychophysiol.* 76 (1), 52–55.
- Bach, D.R., Melinščák, F., Fleming, S.M., Voelkle, M.C., 2020. Calibrating the experimental measurement of psychological attributes. *Nat. Hum. Behav.* 4 (12), 1229–1235.
- Bach, D.R., Sporrer, J., Abend, R., Beckers, T., Dunsmoor, J.E., Fullana, M.A., Gamer, M., Gee, D.G., Hamm, A., Hartley, C.A., Herringa, R.J., Jovanovic, T., Kalisch, R., Knight, D.C., Lissek, S., Lonsdorf, T.B., Merz, C.J., Milad, M., Morriss, J., Phelps, E. A., Pine, D.S., Olsson, A., van Reekum, C.M., Schiller, D., 2023. Consensus design of a calibration experiment for human fear conditioning. *Neurosci. Biobehav. Rev.* 148, 105146. <https://doi.org/10.1016/j.neubiorev.2023.105146>.
- Bari, D.S., Yacoob Aldosky, H.Y., Martinsen, Ø.G., 2020. Simultaneous measurement of electrodermal activity components correlated with age-related differences. *J. Biol. Phys.* 46, 177–188.
- Boucsein, W., 2012. *Electrodermal Activity*. Springer, New York.
- Boucsein, W., Baltissen, R., Euler, W., 1984. Dependence of skin conductance reactions and skin resistance reactions on previous level. *Psychophysiology* 21, 212–218.
- Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., Dawson, M.E., Society for Psychophysiological Research Ad Hoc Committee on Electrodermal, M., 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49 (8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>.
- Bradford, D.E., DeFalco, A., Perkins, E.R., Carbajal, I., Kwasa, J., Goodman, F.R., Jackson, F., Richardson, L.N.S., Woodley, N., Neuberger, L., Sandoval, J.A., Huang, H.J., Joyner, K.J., 2022. Whose signals are being amplified? Toward a more equitable clinical psychophysiology. *Clin. Psychol. Sci.* <https://doi.org/10.1177/21677026221112117>.
- Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25 (1), 49–59.
- Bradley, M.M., Lang, P.J., 2007. The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual. Technical Report B-3 (Retrieved from Gainesville, FL).
- Cohen, J., 1992. Quantitative methods in psychology: a power primer. *Psychol. Bull.* 112, 1155–1159.
- Crider, A., 2008. Personality and electrodermal response lability: an interpretation. *Appl. Psychophysiol. Biofeedback* 33, 141–148.
- Davydov, D.M., Luminet, O., Zech, E., 2013. An externally oriented style of thinking as a moderator of responses to affective films in women. *Int. J. Psychophysiol.* 87 (2), 152–164.
- Dawson, M.E., Schell, A.M., Filion, D.L., 2017. The electrodermal system. In: Cacioppo, J. T., Tassinary, L.G., Berntson, G.G. (Eds.), *Handbook of Psychophysiology*, 4th ed. University Press, Cambridge, pp. 217–243.
- Dindo, L., Fowles, D.C., 2008. The skin conductance orienting response to semantic stimuli: significance can be independent of arousal. *Psychophysiology* 45 (1), 111–118.
- Doberenz, S., Roth, W.T., Wollburg, E., Maslowski, N.I., Kim, S., 2011. Methodological considerations in ambulatory skin conductance monitoring. *Int. J. Psychophysiol.* 80 (2), 87–95. <https://doi.org/10.1016/j.ijpsycho.2011.02.002>.
- Edelberg, R., 1972. Electrical activity of the skin: its measurement and uses in psychophysiology. In: Greenfield, N.S., Sternbach, R.A. (Eds.), *Handbook of Psychophysiology*. Holt, Rinehart, & Winston, New York, pp. 367–418.
- Eisendorfer, C., Doerr, H.O., Follette, W., 1980. Electrodermal reactivity: an analysis by age and sex. *J. Hum. Stress.* 6 (4), 39–42.
- Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41 (4), 1149–1160.
- Fernandez Rojas, R., Hirachan, N., Brown, N., Waddington, G., Murtagh, L., Seymour, B., Goecke, R., 2023. Multimodal physiological sensing for the assessment of acute pain. *Front. Pain Res.* 4, 1150264.
- Goodwin, M.S., Velicer, W.F., Intille, S.S., 2008. Telemetric monitoring in the behavior sciences. *Behav. Res. Methods* 40 (1), 328–341.
- Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., Lord, C., 2016. Measuring changes in social communication behaviors: preliminary development of

- the Brief Observation of Social Communication Change (BOSCC). *J. Autism Dev. Disord.* 46 (7), 2464–2479. <https://doi.org/10.1007/s10803-016-2782-9>.
- Horesh, D., Milstein, N., Tomashin, A., Mayo, O., Gordon, I., 2022. Pre-pandemic electrodermal activity predicts current COVID-related fears: household size during lockdown as a moderating factor. *Stress* 25 (1), 22–29.
- Jones, R.M., Carberry, C., Hamo, A., Lord, C., 2017a. Placebo-like response in absence of treatment in children with autism. *Autism Res.* 10 (9), 1567–1572. <https://doi.org/10.1002/aur.1798>.
- Jones, R.M., Southerland, A., Hamo, A., Carberry, C., Bridges, C., Nay, S., Rozga, A., 2017b. Increased eye contact during conversation compared to play in children with autism. *J. Autism Dev. Disord.* 47 (3), 607–614. <https://doi.org/10.1007/s10803-016-2981-4>.
- Jones, R.M., Tarpey, T., Hamo, A., Carberry, C., Lord, C., 2018. Smartphone measures of day-to-day behavior changes in children with autism. *Nat. Digit. Med.* 1 (34).
- Jones, R.M., Plesa Skwerer, D., Pawar, R., Hamo, A., Carberry, C., Ajodan, E.L., Meyer, S., 2019. How effective is LENA in detecting speech vocalizations and language produced by children and adolescents with ASD in different contexts? *Autism Res.* 12 (4), 628–635.
- Khalaf, A., Nabian, M., Fan, M., Yin, Y., Wormwood, J., Siegel, E., Chou, C.-A., 2020. Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat. *Expert Syst. Appl.* 140, 112890.
- Kinnaird, E., Stewart, C., Tchanturia, K., 2019. Investigating alexithymia in autism: a systematic review and meta-analysis. *Eur. Psychiatry* 55, 80–89.
- Kleckner, I.R., Wormwood, J.B., Simmons, W.K., Barrett, L.F., Quigley, K.S., 2015. Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology* 52 (11), 1432–1440.
- Kleckner, I.R., Zhang, J., Tourotoglou, A., Chanes, L., Xia, C., Simmons, W.K., Barrett, L.F., 2017. Evidence for a large-scale brain system supporting allostatic and interoception in humans. *Nat. Hum. Behav.* 1 <https://doi.org/10.1038/s41562-017-0069>.
- Kleckner, I.R., Jones, R.M., Wilder-Smith, O., Wormwood, J.B., Akcakaya, M., Quigley, K.S., Kleckner, I.R., 2018. Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. *IEEE Trans. Biomed. Eng.* 65 (7), 1460–1467. <https://doi.org/10.1109/TBME.2017.2758643>.
- Kleckner, I.R., Feldman, M.J., Goodwin, M.S., Quigley, K.S., 2021. Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behav. Res. Methods* 53 (2), 518–535.
- Kristjansson, S.D., Kircher, J.C., Webb, A.K., 2007. Multilevel models for repeated measures research designs in psychophysiology: an introduction to growth curve modeling. *Psychophysiology* 44 (5), 728–736. <https://doi.org/10.1111/j.1469-8986.2007.00544.x>.
- Kurinec, C.A., Stenson, A.R., Hinson, J.M., Whitney, P., Van Dongen, H., 2022. Electrodermal activity is sensitive to sleep deprivation but does not moderate the effect of total sleep deprivation on affect. *Front. Behav. Neurosci.* 16, 885302.
- Kylliainen, A., Hietanen, J.K., 2006. Skin conductance responses to another person's gaze in children with autism. *J. Autism Dev. Disord.* 36 (4), 517–525. <https://doi.org/10.1007/s10803-006-0091-4>.
- Lakens, D., 2022. Sample size justification. *Collabra: Psychology* 8 (1), 33267.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2008. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-8. University of Florida, Gainesville, FL.
- Lonsdorf, T.B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Richter, J., 2019. Navigating the garden of forking paths for data exclusions in fear conditioning research. *Elife* 8, e52465.
- Lord, C., Rutter, M., Dilavore, P.C., Risi, S., Gotham, K., Bishop, S., 2012. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4. Western Psychological Services, Torrance, CA, USA.
- Lykken, D.T., Rose, R., Luther, B., Maley, M., 1966. Correcting psychophysiological measures for individual differences in range. *Psychol. Bull.* 66 (6), 481–484. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/5974620>.
- McVeigh, K., Kleckner, I.R., Quigley, K.S., Satpute, A.B., 2023. Fear-related psychophysiological patterns are situation and individual dependent: A Bayesian model comparison approach. *Emotion*. <https://doi.org/10.1037/emo0001265>.
- Schell, A.M., Dawson, M.E., Rissling, A., Ventura, J., Subotnik, K.L., Gitlin, M.J., Nuechterlein, K.H., 2005. Electrodermal predictors of functional outcome and negative symptoms in schizophrenia. *Psychophysiology* 42 (4), 483–492.
- Schulter, G., Papousek, I., 1992. Bilateral electrodermal activity: reliability, laterality and individual differences. *Int. J. Psychophysiol.* 13 (3), 199–213.
- Tolin, D.F., O'Bryan, E.M., Davies, C.D., Diefenbach, G.J., Johannessen, J., 2023. Central and peripheral nervous system responses to chronic and paced hyperventilation in anxious and healthy subjects. *Biol. Psychol.* 176, 108472.
- Tronstad, C., Amini, M., Bach, D.R., Martinsen, Ø.G., 2022. Current trends and opportunities in the methodology of electrodermal activity measurement. *Physiol. Meas.* 43 (2), 02TR01.
- Troy, A.S., Shallcross, A.J., Mauss, I.B., 2013. A person-by-situation approach to emotion regulation: cognitive reappraisal can either help or hurt, depending on the context. *Psychol. Sci.* 24 (12), 2505–2514.
- Vaidyanathan, U., Isen, J.D., Malone, S.M., Miller, M.B., McGuire, M., Iacono, W.G., 2014. Heritability and molecular genetic basis of electrodermal activity: a genome-wide association study. *Psychophysiology* 51 (12), 1259–1271.
- Venables, P.H., Christie, M.J., 1980. Electrodermal activity. In: Venables, I.M.P.H. (Ed.), *Techniques in Psychophysiology*. Wiley, New York, pp. 3–67.
- Webb, E.K., Etter, J.A., Kwasa, J.A., 2022. Addressing racial and phenotypic bias in human neuroscience methods. *Nat. Neurosci.* 25 (4), 410–414.
- Wilhelm, F.H., Grossman, P., 2010. Emotions beyond the laboratory: theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biol. Psychol.* 84 (3), 552–569. <https://doi.org/10.1016/j.biopsych.2010.01.017>.
- Wormwood, J.B., Khan, Z., Siegel, E., Lynn, S.K., Dy, J., Barrett, L.F., Quigley, K.S., 2019. Physiological indices of challenge and threat: a data-driven investigation of autonomic nervous system reactivity during an active coping stressor task. *Psychophysiology* 56 (12), e13454.
- Zimmer, H., 2000. Frequency and mean amplitude of spontaneous electrodermal fluctuations are not interchangeable indicators of psychological processes. *Zeitschrift für Experimentelle Psychologie: Organ der Deutschen Gesellschaft für Psychologie* 47 (2), 129–143.