Moments That Matter? On the Complexity of Using Triggers Based on Skin Conductance to Sample Arousing Events Within an Experience Sampling Framework

SJOERD VAN HALEM[1]*, EESKE VAN ROEKEL[1], LARA KROENCKE[2,3], NICLAS KUPER[2,4] and JAAP DENISSEN[1]

[1]*Department of Developmental Psychology, Tilburg University, The Netherlands*
[2]*Department of Differential Psychology and Psychological Assessment, Universität Hamburg, Germany*
[3]*Department of Psychology, University of Münster, Germany*
[4]*Department of Differential Psychology, Bielefeld University, Germany*

Abstract: To sample situations that are psychologically arousing in daily life, we implemented an experience sampling strategy in which 82 Dutch young adults (M_{age} = 20.73) were triggered based on random time intervals and based on physiological skin conductance scores across a period of 5 days. When triggered, participants had to fill in short surveys on affect, situational characteristics and event characteristics on their smartphone. We found theoretically expected relationships between the skin conductance signal on the one hand and self-reported arousal and positive energy (e.g. energetic and enthusiastic) on the other hand, although effect sizes were small. Unexpectedly, none of the negative affective scales (i.e. irritation, anxiety, and negative valence) were predicted by skin conductance levels. Despite the (partial) validity of the signal, a simple algorithm that triggered the survey based on relative increases of skin conductance levels produced counterintuitive results due to a dependence between level and slope. Additional exploratory analyses highlighted other skin conductance signal characteristics (i.e. autocorrelation, number of peaks, and change points) that might be worth examining when designing future algorithms to sample arousing moments. Overall, our experiences highlight not only the promise but also the complexity of real-time measurement of physiological processes in daily life. © 2020 The Authors. European Journal of Personality published by John Wiley & Sons Ltd on behalf of European Association of Personality Psychology

Key words: skin conductance; psychological arousal; emotions; situations; CAPTIONs

Our daily lives are filled with routines. We regularly spend our time doing familiar activities within highly repetitive and similar contexts (Wood & Neal, 2007; Wood, Quinn, & Kashy, 2002). Although the systematic patterns of our routines are informative and predictive of several psychological well-being outcomes (Csikszentmihalyi & Hunter, 2003; Quinn, Pascoe, Wood, & Neal, 2010), every now and then something exciting occurs that awakes our spirit, be it positive or negative. For instance, an argument with our partner, an important job interview, or watching the final episode of our favourite show might rattle our routines. These events tap into key goals and motivations of humans, either because they allow important rewards to be obtained or important resources to be threatened (Carver & White, 1994; Frijda, 2004; Kuppens, 2008; Sideridis, 2008). The arousal associated with these moments in turn fuels some of the most consequential human responses, such as romantic behaviour, exploration,

aggression, and flight. Identifying these arousing situations as well as people's response to them therefore has relevance for educational, clinical, social, and personality psychology. For example, in clinical psychology, identifying what situation triggers anxiety can facilitate the training of coping skills to better cope with the associated demands. Moreover, personality psychologists might investigate individual differences in the types of situations that trigger arousing responses, for example, individual differences in affective responses to situations of affiliation (cf., Dufner, Arslan, Hagemeyer, Schönbrodt, & Denissen, 2015). The current study explores a relatively new method of capturing these arousing moments in daily life and summarizes first findings on its feasibility, validity, and usefulness.

Psychological arousal is understood as a dimension denoting the level of a person's activation or alertness ranging from states of *low arousal*—described by words such as relaxed, calm, depressed, or bored—to states of *high arousal*—described by words such as fear, anger, excitement, or pure joy (Russell, 2003). Because our daily lives are mostly composed of routine activities (Wood et al., 2002; Wood & Neal, 2007), moments in which participants are psychologically aroused only constitute a small part of our daily lives and are therefore relatively rare. This makes it difficult to filter these moments of psychological

*Correspondence to: Sjoerd van Halem, Department of Developmental Psychology, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands.
E-mail: s.vanhalem@tilburguniversity.edu

arousal from a participant's day-to-day routine. Experience sampling methodology (ESM; Csikszentmihalyi & Larson, 1987; Larson & Csikszentmihalyi, 1983) might be a useful approach to study arousing moments. Studies might for example ask participants to selectively report on arousing events every time they occur, referred to as *event-contingent sampling* (Dimotakis, Ilies, & Judge, 2013; Reis & Gable, 2000; Scollon, Prieto, & Diener, 2009). However, participants are usually busy during these moments, and they might not always consciously experience that they are aroused or forget soon thereafter, resulting in missing data. Alternatively, using fixed or random time intervals in experience sampling designs might work relatively well for sampling routine activities (Neal, Wood, & Quinn, 2006) but might sample only few arousing moments. Some earlier approaches have asked participants end-of-day questions to describe a conflict or a stressful episode that took place during the day (e.g. Nezlek & Plesko, 2003), but this approach suffers from retrospective bias. A more objective assessment method is therefore needed.

As an alternative, to more selectively sample certain contextual information, studies across different psychological fields have started to use different types of sensors, such as GPS, accelerometer, and heart rate sensors (for applications, see Ebner-Priemer, Koudela, Mutz, & Kanning, 2013; Myrtek et al., 1988; Pejovic, Lathia, Mascolo, & Musolesi, 2016). Skin conductance sensors appear particularly suited for sampling moments of psychological arousal more objectively. By applying continuous currents of electricity on the skin, researchers are able to study the electric resistance or *skin conductance* of the skin, which varies in response to sweat secretion (Boucsein, 2012; Boucsein et al., 2012). Increases in sweat secretion are a common physical marker for autonomic activation and have been related to arousal, attention, and emotional responses (e.g. Boucsein, 2012; Boucsein et al., 2012; Kreibig, 2010). Kreibig (2010) reviewed 134 published studies to provide more insight in the nature of the relationship between autonomic response patterns, such as skin conductance, and emotions. Specifically, some typical low arousal emotions such as (non-crying) sadness, contentment, and relief were related to a decreased skin conductance level (SCL), whereas almost all other typical high arousal emotions, such as anger, anxiety, happiness, amusement, and joy, were accompanied by increases in SCL (Kreibig, 2010). The Kreibig (2010) review concluded that relationships between specific emotions and skin conductance overall followed a general dimension denoting a person's level of activation or tendency for action (see Russell, 2003). Consequently, measures of skin conductance could be used to objectively sample moments of psychological arousal from individuals' daily life.

Nevertheless, it is unclear how the lab findings from studies in the Kreibig (2010) review translate to measurements of emotions as reported in everyday life. Autonomic response patterns in the 134 studies included by Kreibig (2010) mostly followed experimental tasks such as looking at pictures, watching movies, by reading text, or by recalling certain events. In the lab, researchers can standardize emotional stimuli. In most cases, researchers have exact information

of when the stimulus starts and can therefore easily track the skin conductance response during and after the stimulus takes place. This is obviously different in daily life, in which there is no control over a participant's environment.

Although some studies have recorded skin conductance in daily life over longer periods of time (e.g. Doberenz, Roth, Wollburg, Breuninger, & Kim, 2010; Hoehn-Saric, McLeod, Funderburk, & Kowalski, 2004), we are aware of only one study that specifically attempted using skin conductance sensor data in daily life to sample arousing moments (i.e. Westerink et al., 2009). In a sample of 31 participants that were measured across a period of eight hours, Westerink et al. (2009) compared momentary self-reports of arousal between fixed time interval triggers (i.e. every two hours) with triggers based on continuous scores of skin conductance. To trigger questionnaires based on skin conductance, Westerink et al. (2009) took a dynamic approach. For any given moment, the algorithm generated a distribution of the last half hour of SCL scores. If the live SCL score would surpass the 95 percentile of this distribution, a questionnaire was sent to the participant's phone. This way the algorithm ensured the cut-off value was person specific and adaptive across time. Westerink et al. (2009) found that self-reported arousal was higher during triggers based on skin conductance compared with fixed time interval triggers. Although the results of Westerink et al. (2009) are promising, there are some limitations to their design. Participants were aware that they would receive triggers based on their physiological skin conductance responses. This could have created demand characteristics towards high self-reports of arousal. In addition, Westerink et al. (2009) only reported about self-reported arousal in their study, but there is a range of other variables that could be relevant in relation to these triggers. It is for instance questionable if such a triggering approach works equally well for positive and negative affective scales. Finally, in the study of Westerink et al. (2009), participants wore skin conductance wristbands with dry electrodes located at underside of the wrist. As the device is almost identical to a common watch, this makes it convenient to wear for longer periods of time. Nonetheless, the inside of the wrist is not a preferred location for measuring skin conductance (Van Dooren & Janssen, 2012). Additionally, using dry electrodes is generally not recommended because of the slow humidity buildup under the metal plate, which can result in extended periods of unstable electrodes (Boucsein et al., 2012; Fowles et al., 1981). Doberenz, Roth, Wollburg, Maslowski, and Kim (2011) have shown that it is also feasible to use skin conductance measures with gel-based electrodes for longer periods of time (i.e. 24 hours). Westerink et al. (2009) acknowledged that the quality of the data signal could influence the appropriateness of their signal. Nonetheless, despite these limitations, the study by Westerink et al. (2009) provides at least some support for the potential of this methodology.

PRESENT STUDY

With the aim to specifically sample moment that are psychologically arousing as they occur in daily life, the present study

implements an ESM sampling strategy in which participants are triggered based on their continuous physiological SCL scores. First, following the recommendations of Boucsein et al. (2012), to ensure the quality of signal, we used a skin conductance device that relies on gel-based electrodes on the palmar site. Second, by collecting continuous scores of skin conductance alongside a variety of momentary self-reports in the daily life of participants, we could validate previous lab findings on these relationships (Kreibig, 2010). Based on the results of these studies, we expected to find a relationship between skin conductance and self-reports of psychological arousal. Finally, following previous studies that have used contextual sensors to selectively sample specific moments in the daily life of participants (e.g. Ebner-Priemer et al., 2013; Myrtek et al., 1988; Pejovic et al., 2016), the present study explored if the skin conductance signal can be used to specifically sample moments that are psychologically arousing.

METHOD

Sample

For this study, we aimed to have a sample of at least 60 participants. The study of Westerink et al. (2009) found a mean self-reported arousal difference between skin conductance triggers and random triggers of $d = 1.10$. With a sample of 60 participants, the power to detect such an effect is greater than 0.99. In total, 84 participants signed up for our study. As two participants did not show up for the ESM part of the study, the final sample consisted of 82 young adults, aged between 18 and 32 years old ($M = 20.73$, $SD = 2.65$). In this sample, 52 were female (63%), and for 67 participants (82%), both parents were born in the Netherlands. All participants in the current study were first-year psychology students at Tilburg University and were recruited through the university's online participant system.

Procedure

Two weeks before the start of our study, we conducted a pilot study to test and optimize our study design. In the pilot study, we also collected surveys and continuous skin conductance data among a sample of 10 participants, $M_{age} = 19.67$, $SD_{age} = 2.55$, 90% female, and 90% both Dutch parents. On the first day, we tested whether the skin conductance set-up would work as intended for the entire day. For many participants on the first day, the electrodes fell off during the day. After adjustments on the second day, the electrodes were holding. The pilot data are not part of the current study.

After the pilot, we started the data collection of the present study, which took place across 10 weeks between March and June 2019. The data collection consisted of two parts: a baseline questionnaire followed by five consecutive days of momentary assessments. The online baseline questionnaire included demographic questions and several psychological questionnaires to measure personality, emotional clarity, depression, and so forth. For the daily assessments, participants completed several state questionnaires on their smartphones while wearing a skin conductance wristband. If the phone was compatible with the experience sampling mobile app, participants used their own smartphone. In all other cases, a phone was provided by the experimenters (MovisenXS is not compatible with iPhone or androids with a MediaTek chipset). The momentary assessments always took place between Monday and Friday (10:00 a.m.–5:00 p.m.), and measurement cycles roughly consisted of testing between 6 and 12 participants each week. We trained 13 student assistants, which were divided in groups of two or three students and were each responsible for 1 day of the week.

For the present study, participants were required to visit the lab twice a day: once in the morning between 9 and 10 a.m. and once in the afternoon between 5 and 6 p.m. Between 9 and 10 a.m., a research assistant would apply the skin conductance device and couple the ESM app to the participant's smartphone. Hereafter, participants could go about their days as usual while filling in ESM surveys when notified. On the first morning of data collection, we briefed participants to fill in the ESM questionnaires directly after being notified. We specifically told participants to report on their mood and whereabouts focused on the moment leading up to the notification. Participants were notified by means of the smartphone's alarm that lasted maximally 10 seconds, after which participants had a maximum of 20 minutes to fill in the questionnaire before becoming unavailable. Each day at 4:55 p.m., participants would also fill in a small questionnaire that provided insight about the most important event of the day as well as about their experience of wearing the skin conductance device. The day would end between 5 and 6 p.m. when participants came back to return the devices, and their data were saved.

To measure skin conductance, there are many options available, such as the ring-mounted Moodmetric (Vigofere Oy) or the wrist-worn E4 (Empatica) (Cowley et al., 2016). We chose to use the EdaMove 3 (movisens GmbH, see www.movisens.com), which is a wrist-worn skin conductance sensor that contains wired gelled electrodes. We preferred to use gelled electrodes as these came recommended over using dry electrodes in the review by Boucsein et al. (2012). In addition, movisens already had the option of sending triggers based on sensor data preprogrammed in their ESM environment, which made it convenient for us to use the EdaMove 3 compared with other sensors. The EdaMove 3 sensor has a sampling rate of 32 Hz and applies a constant voltage of 0.5 V on the sensor's silver-chloride electrodes. The two electrodes are filled with electrode gel before they are attached to the skin of the palmar surface using adhesive tape rings. To ensure that the device would hold for the period of the study, we applied kinesiology tape on the electrodes, and participants wore special gloves that were made of a breathable fabric (i.e. bamboo viscose), had good stretch, and fitted tightly. The research assistant applied the device to the participant's non-dominant hand. Please see Figure 1 for an image of this set-up.

We programmed two different triggers for this study: random triggers and triggers based on skin conductance. For the triggers, we had the following general restrictions: questionnaires would only be sent when participants wore the skin conductance device—i.e. between 10 a.m. and

Figure 1.   Movisens EdaMove 3 set-up.

5 p.m. Triggers were sent at least 15 minutes apart, and participants would receive a maximum of five random triggered questionnaires. For triggers based on skin conductance, real-time skin conductance scores were averaged across one minute (i.e. SCL) and compared with SCL scores of the minute before that. If the relative increase surpassed a threshold that was set at the start of the day (see below) a trigger was sent to the smartphone. To control for influences of physical activity, we suppressed triggers when SCL increases accompanied increases in step count, which is an established procedure (Myrtek, Aschenbrenner, & Brügner, 2005).

The triggering algorithm was provided to us by movisens. As we were limited in adjusting the algorithm, our algorithm deviated from the personalized algorithm of Westerink et al. (2009). To determine the threshold for sending a trigger, we therefore had to account for personal differences in SCL, because some people are more reactive or have higher baselines skin conductance scores (Boucsein, 2012). Using the same threshold for the entire sample might mean different things for each participant. We took the following steps to realize a more personalized approach:

1 *Day 1*. For the first day, we did not have any information on the person's skin conductance scores. We therefore used prior information of our *pilot study* to formulate an average threshold that would be the same for all participants on the first day. For each day of skin conductance data that were available for the participants of our pilot study, we created a distribution of the relative changes in skin conductance, excluding the moments in which also an increase in step count was observed. To determine the threshold for what would be a substantial increase for that particular day, we used the value that corresponded to two standard deviations above the mean in this distribution. This threshold represented the average of all the cut-off scores across measures in the pilot and equalled an increase of 19.3% relative to the preceding minute. We used this threshold on the first weekday for all participants in the study.

2 *Day 2–Day 5*. For the other days, we used the same principle to define the cut-offs but then for each person individually. Again, we defined two standard deviations above the mean in the distribution of relative changes as substantial. The cut-off value was then averaged with the

cut-offs that were set on the previous days for the participant, giving each day an equal weight in the equation. As with each day, we obtained more personalized information, we let the weight of the general sample cut-off that was used for Day 1 (i.e. 19.3%) decrease with each day of data collection. The weights for this value were as follows: 1 on Day 2, 0.75 on Day 3, 0.5 on Day 4, and 0.25 on Day 5.

Our study was approved by the Ethics Review Board (EC-2018.83) of the School of Social and Behavioral Sciences of Tilburg University, and all participants provided active consent for participation. Participants received 10 participation hours when they completed the baseline questionnaire and at least 75% of the ESM surveys. The participation hours were part of a first-year academic course in which students were required to participate in different studies for a total of 20 hours. Additionally, we provided personalized feedback reports to the participants when they indicated they were interested. These reports included several descriptive statistics on personality, emotions, and situations.

Measures

Self-reported arousal

To operationalize self-reported momentary arousal, we used two measures that were administered in the experience sampling part of the study. First, we used the item "I feel active" as a direct indication of psychological arousal without referencing to valence. Second, we relied on 19 affective adjectives, which were selected to cover both dimensions of the circumplex model of affect (e.g. Russell, 2003), that is, valence and arousal. All items were administered on a visual analogue scale (VAS) ranging from 0 to 100. Because this specific set of items had not been used previously, we first explored the underlying factor structure using multilevel exploratory factor analyses to reduce the data and possibly extract an arousal scale from scores of self-reported affect. The analyses and items are reported in the Results section.

Situational characteristics

To measure situational characteristics, we used 14 items that in pairs represented the seven CAPTIONs dimensions. The CAPTIONs represents a situational taxonomy that was

Table 1. Inter-item correlations between item pairs of the dimensions the CAPTIONs-SF

Caption dimension	Item pair	Between-person r	Within-person r
Complexity	Academic – Scholarly	.75	.65
Adversity	Tiresome – Stressful	.70	.49
Positive valence	Cherished – Heartwarming	.81	.67
Typicality	Regular – Standard	.94	.61
Importance	Useful – Productive	.82	.73
Humour	Wacky – Goofy	.91	.63
Negative valence	Repulsive – Malicious	.73	.42

Note: Situational items were administered as "The current situation is ...." $N = 82$, surveys = 3868.

developed by Parrigon, Woo, Tay, and Wang (2017) using a lexical approach. They used an extensive sample of adjectives and reduced them to seven dimensions representing different psychological situation characteristics: Complexity, Adversity, Positivity, Typicality, Importance, humOr, and Negative valence. The items were phrased as "the current situation is …" and were administered on a VAS (0–100). The specific situational adjectives of the CAPTIONs as well as the inter-item correlations on both the between-person level and within-person level are reported in Table 1.

Event characteristics
We used three VAS (0–100) items to ask participants about the event that took place. Participants were asked to rate whether the event that took place on the moment of the trigger was expected, important, and pleasant.

Analytical strategy

We used multilevel exploratory factor analyses for an indication of how our affect items clustered together and to determine whether it was justified to formulate different scale scores from the 19 items that were included in our questionnaire. For our factor structure, we focused only on the within-person variance–covariance matrix. The between-person variance covariance matrix was estimated without any restrictions. Additionally, to avoid that our factor structure was influenced by demand characteristics, we conducted the factor analyses on the input from the random triggers only. After determining the best model fit using absolute fit indices (i.e. comparative fit index, Tucker–Lewis index, root mean square error of approximation, and standardized root mean square residual) and model tests based on the deviance statistics, we computed the scale scores by aggregating the items that were indicative for each factor in the analysis. For each scale, we selected the items that had a loading of at least 0.40 and did not cross-load more than 0.10 on any of the other factors. We estimated these models using Mplus version 6 (Muthén & Muthén, 2010).

To examine how skin conductance relates to affective and situational self-reports in everyday life, we examined how SCL related to our affect dimensions, situational characteristics, and event characteristics, accounting for variance between individuals. We used the lme4 package

version 1.1-21 (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.6.0 (R Development Core Team, 2016) to estimate multiple separate multilevel models with skin conductance as independent variable. Both the intercept of the model and the slope of skin conductance with the outcome variable (i.e. affect, situation, and event) were allowed to vary across individuals. We centred our predictor variables in these models within each person before running the analyses.

In the final part of our analyses, we directly examined differences between the triggering conditions—that is, random triggers versus triggers based on skin conductance—in self-reported affect, situational characteristics, and event characteristics. The intercepts of the models and slopes of the triggering condition on the self-reported outcome variables (i.e. affect, situation, and event) were estimated freely across individuals. The predictor variable (i.e. type of trigger) was not centered. Instead, we upheld its dichotomous coding to facilitate a condition-relative interpretation. We pre-registered these final analyses before collecting the data[1].

For all our multilevel regressions, explained variances were estimated using the guidelines of Nakagawa and Schielzeth (2013), who provided an R function for calculating both the explained variance of a model's fixed effects, as well as the explained variance of the random effects. To obtain this statistic, the function divides the variance of the fixed effects by the total variance.

Exploratory analyses
Guided by suggestions that were made during the review process, we also ran additional exploratory analyses. These analyses were conducted after having obtained the outcomes

---

[1]For our pre-registration, see https://osf.io/prs7j/?view_only= a6b1bfebe7fb49f1970b8372fc334bbb. Some things were not included in the preregistration but added or changed during the study. First, initially we assumed that we could use only the first day of data collection to estimate the personalized thresholds. We reconsidered this strategy after having collected and examined the pilot data. Additionally, we pre-registered the multilevel analyses in which we examined the differences between both triggers (skin conductance triggers and random trigger) on our outcome variables. The analyses with the raw SCL data were not included before collecting the data but nonetheless planned before analysing the data. Finally, the exploratory analyses were added during the review process. Our data and scripts are openly available. See https://osf.io/v4qh9/?view_only= 054234b95e454a7db553e38d0490f686.

of the planned analyses presented above and were not pre-registered. For these analyses, we set out to examine the skin conductance features of arousing moments more in detail. To do so, we extracted people's most and least arousing moments, to then further compare and characterize the skin conductance signatures of each category.

To obtain a large enough sample for 'high' and 'low' arousing moments, we selected the two most and two least arousing and emotional moments of each individual. For this purpose, we used both self-reports of arousal as well as each of our four affective scales: positive energy, irritation, anxiety, and negative valence. This resulted in a total of five categorizations. Ties for the top two highest or lowest scores were handled randomly. With 82 participants for each split off, this resulted in a sample of 164 high arousing moments and 164 low arousing moments. After making these selections, we extracted the skin conductance signal of each of these moments across three different timeframes: 1 minute, 10 minutes, and 30 minutes before the trigger.

We used multilevel regression analyses to compare both categories on five different outcomes measures related to the skin conductance signal: (i) mean, (ii) standard deviation, (iii) autocorrelation, (iv) number of peaks, and (v) number of change points. Going beyond simply extracting mean levels of the signal, these parameters together better capture the complex and dynamic characteristics of physiological time series data. For the autocorrelation measure, to avoid ceiling effects (e.g. autoregressive correlations approaching 1), we used relatively broad lags of four seconds to calculate the autocorrelation. To estimate the number of peaks in the signal, we first centred the data within timeframes of four seconds. This way the tonic signal was largely removed from the data. Next, to smoothen the signal and remove background noise, we used a 500-ms frame Savitzky–Golay filter. This signal was used to detect the peaks using the *findpeak* function in the *pracma* package in R. We specified a peak when a consistent increase of 0.5 seconds would follow a consistent decrease of 0.5 seconds. Our method of identifying peaks is comparable with standard methods of extracting frequencies of non-specific skin conductance responses, although in contrast these methods include absolute cut-offs values, such as minimum amplitudes of 0.05 micro Siemens. Finally, we calculated the number of change points using the *change point* package in R. A change point is defined by a substantial change in the distributional properties of the skin conductance data in terms of mean and variance. We used the *cpt.meanvar* function with a penalty value of 10 000 (Killick & Eckley, 2014).

RESULTS

In total, 3868 questionnaires were sent, which on average is 47.17 questionnaires per participant ($SD = 8.62$). Of these questionnaires, 1516 (39%) were randomly triggered and 2356 questionnaires (61%) were triggered by our skin conductance algorithm. The response rate for this study was 89%, and overall, 3432 questionnaires were completed. On average, it took participants 77 seconds to start the form after being notified and two minutes and 27 seconds to finish the entire survey of 40 items. Overall, participants did not feel bothered by wearing the skin conductance devices. When we asked participants about their experiences wearing the skin conductance device, on a VAS (1–100), they reported an average of $M = 30.42$ ($SD = 19.24$) on whether they were annoyed by the device. We found an average of $M = 29.34$ ($SD = 23.28$) on whether the device had impeded them in their daily routines and an average of $M = 19.32$ ($SD = 18.15$) on whether they thought the device influenced their behaviour. The means of all three items seem to suggest participants experienced the EdaMove as relatively unobtrusive. We visually inspected the physiological skin conductance data each day and made a note when we suspected that electrodes might have shifted or became loose during the day. On 17 (4.21%) occasions, we listed the data to contain probable artefacts from loose electrodes. Most of the time this occurred during the final hours of the day. As this was only a small proportion of the total data, we kept these data in the sample. Nevertheless, we conducted the analyses of this manuscript with and without these cases included. Overall, removing the data with the potential artefacts resulted in almost identical results.

Affective structure

We used multilevel exploratory factor analyses to explore the underlying within-person structure of the 19 affect items. The models were estimated on the data that were collected with random triggers, using an unrestricted between-person variance covariance matrix. Fit indices of acceptable models ranged from a one-factor solution to a five-factor solution, which are depicted in Table 2. Based on these indices, a four-factor solution fitted the data best. Both the root mean square error of approximation and standardized root mean square residual for this solution were below 0.05 (i.e. both equalled 0.03), and the Tucker–Lewis index and comparative fit index were 0.96 and 0.85, respectively.

Table 3 shows the four-factor solution including the factor loadings of the items on each of the four factors. By selecting the items with a primary loading of at least

Table 2. Model comparisons of the different within-person exploratory factor solutions of the affect scale using a subsample of random-beeps only

Model[†]	χ²	df	SRMR	RMSEA	CFI	TLI	Δχ²	Δdf	p
1-factor	3777.40	152	0.10	0.09	0.69	0.30			
2-factor	2116.73	134	0.07	0.06	0.83	0.57	1660.67	18	<.001
3-factor	1341.86	117	0.05	0.05	0.90	0.69	774.87	17	<.001
4-factor	701.66	101	0.03	0.03	0.95	0.83	639.34	16	<.001
5-factor	380.59	86	0.02	0.02	0.98	0.90	321.07	15	<.001

Note: $N = 82$, surveys = 1333.
[†]All exploratory factor models were based on the within-person covariance matrix and estimated using unrestricted between-person variance.

Table 3. Exploratory factor analysis of affect with four within-person factors and unrestricted between-person covariance

Affect indicator items	Factor loadings			
	Positive energy	Irritation	Anxiety	Negative valence
Energetic	.76			
Enthusiastic	.71			
Cheerful	.66			
Satisfied	.49	.41		
Happy	.45			
Relaxed		.44		
Angry				.64
Nervous			.75	
Calm		.48		
Worried			.63	
Irritated		−.67		
Fearful			.70	
Stressed			.58	
Irritable		−.63		
Sad				.75
Insecure			.67	
Gloomy				.69
Bored				
Lifeless	−.54			

Note. $N = 82$, surveys = 1333. Maximum likelihood estimation with Geomin oblique rotation. Only factor loadings (= pattern loadings) above |0.40| are reported in the table

0.40 and without any substantial cross-loadings (i.e. secondary loadings higher than 0.40 or a difference with the primary loading of less than 0.10), the first factor was marked by five items: energetic, enthusiastic, cheerful, happy, and lifeless—the latter having a negative loading. We labelled this factor (1) *positive energy*. The second factor had four indicators: irritated, agitated, calm, and relaxed, which we labelled as (2) *irritation*. The third factor was marked by the items nervous, worried, fearful, stressed, and insecure and was labelled (3) *anxiety*. The final factor was marked by the items angry, gloomy, and sad and was labelled as (4) *negative valence*.

After reverse coding the negative items and constructing the scale scores of the four constructs by aggregating the scores of the indicators of each scale, we used McDonald's omega total to examine the reliability of the scale including the data that were administered by random triggers and skin conductance triggers. To estimate the between-person and within-person omega coefficients, for each scale, we fitted a multilevel confirmatory factor analysis in Mplus (see Geldhof, Preacher, & Zyphur, 2014). The person omega's for the scales were positive energy ($\Omega_b = 0.84$ and $\Omega_w = 0.74$), irritation ($\Omega_b = 0.82$ and $\Omega_w = 0.77$), anxiety ($\Omega_b = 0.95$ and $\Omega_w = 0.83$), and negative valence ($\Omega_b = 0.94$ and $\Omega_w = 0.81$).

Descriptive analyses

We calculated the intercorrelations of all continuous variables that were included in the present study. Both the within-person and between-person correlations are depicted in Table 4 together with the means and standard deviations of the variables. Although correlations at the within-person level were somewhat weaker than the correlations at the between-person level, the direction of the relationships was consistent across both levels. All four affective scales were correlated with another, with positive energy having negative relationships with all three other affective scales. Concerning the situational characteristics, there were relationships between importance, complexity, and, to a lesser extent, adversity on both levels. Additionally, at both levels, adversity was

Table 4. Within-person and between-person correlation matrix of the continuous measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	*M*	*SD*	ICC
Affect																	
1. Positive energy	—	−.47	−.34	−.41	−.09	−.29	.37	.06	.08	−.06	−.28	.01	.05	.41	61.64	10.57	.43
2. Irritation	−.69	—	.52	.48	.09	.39	−.25	−.16	<.01	.17	.31	−.07	.04	−.38	26.82	9.96	.38
3. Anxiety	−.40	.53	—	.51	.14	.41	−.13	−.12	.07	.23	.30	−.02	.11	−.26	16.19	12.10	.53
4. Negative valence	−.43	.57	.69	—	−.01	.22	−.09	−.16	−.07	.21	.37	−.08	.00	−.22	9.62	7.80	.37
Situations																	
5. Complexity	−.03	.22	.20	.14	—	.54	−.30	.16	.69	−.09	.06	.22	.49	−.32	38.05	14.65	.22
6. Adversity	−.32	.52	.74	.44	.57	—	−.34	−.07	.41	.12	.31	.08	.35	−.45	26.43	12.16	.30
7. Positive valence	.09	.11	.30	.19	.07	.15	—	−.05	−.19	.07	−.13	−.05	−.06	.50	26.21	12.80	.27
8. Typicality	.16	−.16	.11	−.03	.13	.05	−.03	—	.21	−.36	−.22	.28	.03	.07	66.08	11.10	.28
9. Importance	.00	.18	.24	.13	.78	.53	.00	.31	—	−.13	−.05	.27	.63	−.23	51.44	10.84	.13
10. Humour	−.10	.35	.43	.54	.19	.35	.51	−.23	.06	—	.35	−.20	−.03	−.10	9.38	6.78	.22
11. Negative valence	−.16	.39	.53	.65	.23	.41	.38	−.01	.09	.64	—	−.10	.02	−.30	7.46	6.69	.34
Events																	
12. Expected	.13	−.12	.04	−.02	.05	−.01	−.10	.50	.15	−.21	−.09	—	.25	.06	65.47	12.80	.22
13. Important	.20	−.01	.07	−.11	.47	.31	<.01	.07	.60	−.03	−.05	.28	—	−.20	55.34	11.72	.15
14. Pleasant	.52	−.48	.15	−.19	.32	−.35	.30	.16	−.19	−.04	−.09	−.32	−.08	—	60.82	9.36	.16

Note: Correlations above the diagonal represent within-person correlations; correlations below the diagonal represent between-person correlations. *M*, mean; *SD*, standard deviation; ICC, intraclass correlation.

Table 5. Multilevel models examining the within-person relationships between skin conductance level (SCL) and affect, situation, and event characteristics

Outcome variables	B	95% CI		p	R^2	σ^2_{SCL}	p
Affect							
Self-reported arousal	0.42	0.08	0.77	**.017**	.008	1.31	<.001
Positive energy	0.32	0.07	0.56	**.011**	.007	0.61	<.001
Irritation	0.04	−0.22	0.29	.782	<.001	0.67	<.001
Anxiety	0.09	−0.14	0.31	.453	<.001	0.52	<.001
Negative valence	−0.08	−0.26	0.11	.410	<.001	0.32	<.001
CAPTIONs							
Complexity	0.51	0.09	0.95	**.019**	.005	1.01	.052
Adversity	0.23	−0.09	0.55	.165	.002	0.84	<.001
Positive valence	−0.02	−0.31	0.27	.882	<.001	0.30	.221
Typicality	−0.16	−0.48	0.16	.325	.001	0.95	<.001
Importance	0.18	−0.26	0.61	.406	<.001	1.51	<.001
Humour	0.05	−0.13	0.24	.529	<.001	0.14	.151
Negative valence	0.13	−0.07	0.33	.219	.002	0.46	<.001
Event							
Expected	−0.42	−0.76	−0.08	**.018**	.005	0.64	.058
Importance	0.30	0.06	0.67	.169	.002	1.29	.018
Pleasant	−0.17	−0.40	0.06	.143	.001	0.02[a]	.457

Note: Each row represents a separate multilevel model with a random intercept and random slope. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect; $\sigma^2_{trigger}$, the variance estimate for the random slope of the trigger type. We have displayed significant p values in bold.
[a]Model produced singular fit; fixed effect only model was interpreted.

related to negative valence. Medium correlations were found between typicality of the situation and the expectedness of the event that took place. Surprisingly, humour was mainly related to negatively valenced items. Instead of referring to funny situations, this construct could refer more to weird or odd situations. Despite this surprising relation, within this nomological network, all other variables mostly behaved as expected.

Skin conductance signal and triggering algorithm

Skin conductance level

The average SCL values across participants of the minute preceding the trigger was $M = 6.89$, $SD = 3.42$ expressed in micro Siemens (mS). As can be expected, this is considerably higher than the averages (between 3 and 4 mS) reported by Westerink et al. (2009) who used dry electrodes. Table 5 provides insight into how these SCL values relate to our other study variables in daily life. Relationships were mostly in the expected direction. Higher SCL values in the minute preceding the trigger were related to moments in which participant reported being more active. Also, higher levels of positive energy were reported during moments of higher SCL. We did not find a relationship between SCL and irritation, anxiety, or negative valence, however. Finally, looking at the situations and events, situations of high SCL were rated as more complex and generally occurred during events that were rated as less expected. The effect sizes of the significant fixed relationships as reported in terms of explained variance (R^2) ranged between 0.005 and 0.008. This means that SCL explained between 0.5% and 0.8% of the variance of the outcome variable. We found no relationships for any of the other situation and event variables. Nonetheless, for all affect variables and most of the other situational and event characteristic variables, there was a significant variance around the slope, indicating that the effects differed among individuals.

Triggering algorithm

To have an indication of stability of the personalized threshold that we used, we estimated the intraclass correlations across days. The intraclass correlation was 0.67, which indicated that the personalized thresholds were relatively consistent across days.

To examine whether our algorithm successfully selected moments of relative increase, we extracted skin conductance

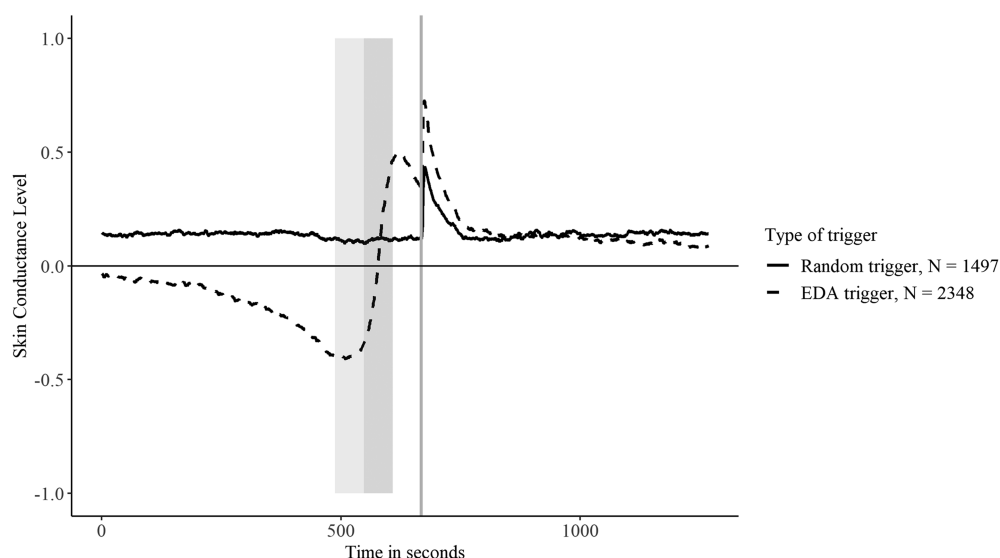Figure 2. Aggregated skin conductance scores of the 10 minutes preceding the trigger and the 10 minutes following the trigger split by random triggers and triggers based on skin conductance. The vertical line represents the trigger. The two grey surfaces represent the two minutes that were used by the skin conductance algorithm. We standardized the skin conductance scores within-person, within the day.

values of the 668 seconds preceding a trigger (i.e. 10 minutes + 68 seconds) and 600 seconds following a trigger (i.e. 10 minutes). Figure 2 shows the skin conductance values leading up to either a random trigger or a trigger based on skin conductance. The two minutes in which the algorithm determines an increase took place between 480 and 600 seconds, denoted by the grey areas. The other 68 seconds before the trigger represent the delay between administering the increase and sending a trigger. This delay was not intentional but was necessary for the algorithm to process the data and send out the trigger. There are three things to note based on this graph. First, it shows that the random triggers on average had a relatively flat approach towards the trigger, whereas for the trigger based on skin conductance, a clear positive slope can be observed. In line with this observation, multilevel regressions showed that the relative increase in SCL across the two minutes preceding a trigger was significantly higher for triggers based on skin conductance ($B = 0.31$, 95% CI [0.29, 0.34], $p < .001$, $R^2 = .134$), which is consistent with the way our algorithm was programmed. A second point that stands out from the graph is that the skin conductance values of the triggers sent by our algorithm generally had a lower baseline value than the random trigger. Although there was no significant difference across the SCL in the minute before the trigger (i.e. the dark grey area; $B = 0.24$, 95% CI [−0.02, 0.49], $p = .067$, $R^2 < .001$), the overall SCL value across the entire 10 minutes of Figure 2 preceding the trigger turned out to be significantly lower for the trigger based on skin conductance ($B = −0.95$, 95% CI [−1.24, −0.67], $p < .001$, $R^2 = .013$). It therefore seems that there is a dependency between level and slope. Moments during which SCLs showed a steep increase occurred mostly when participants had low SCLs to begin with. Finally, both types of triggers sparked a spike in skin conductance that overall led up to the highest value of skin conductance in the graph.

Differences between random triggers and triggers based on skin conductance

We examined if both types of triggers differed in step count, the time of the day that the survey was administered, and whether there were differences on how much time had passed since the last trigger was sent. We found that in the minute preceding the trigger (i.e. the dark grey area in Figure 2), participants reported to have taken more steps during the random trigger ($B = −1.79$, 95% CI [−3.54, −0.05], $p = .046$, $R^2 = .001$). This corresponds with the fact that we controlled for step count increases in our algorithm. Additionally, no differences were found between both types of triggers for the time of the day the trigger was sent ($M_{random} = 13.53$, $M_{SCL} = 13.43$, $t = 1.47$, $p = .142$), but the time that passed (in hours) since the last trigger was significantly shorter for skin conductance-based trigger ($M_{random} = 0.90$, $M_{SCL} = 0.57$, $t = 20.76$, $p < .001$).

Multilevel regression as reported in Table 6 indicated that participants reported to be less aroused before skin conductance-based triggers and felt less positively energized. There were no differences on the other negative affective scales irritation, anxiety, or negative valence. Concerning

Table 6. Multilevel models examining the within-person differences on affect, situation, and event characteristics between random triggers (0) and triggers based on skin conductance values (1)

Outcome variables	B	95% CI		p	R^2	$\sigma^2_{trigger}$	p
Affect							
Self-reported arousal	−1.92	−3.12	−0.75	**.001**	.002	0.04[a]	.949
Positive energy	−1.34	−2.18	−0.49	**.002**	.002	<0.01[a]	.994
Irritation	0.44	−0.54	1.42	.379	<.001	3.59	.352
Anxiety	−0.18	−0.98	0.61	.654	<.001	<0.01[a]	.999
Negative valence	0.68	−0.02	1.38	.056	<.001	0.65[a]	.090
CAPTIONs							
Complexity	−0.68	−2.56	1.19	.478	<.001	15.46	.290
Adversity	−0.12	−1.40	1.17	.858	<.001	<0.01[a]	.995
Positive valence	−0.32	−1.89	1.23	.683	<.001	6.46	.700
Typicality	0.07	−1.36	1.48	.923	<.001	11.75	.080
Importance	−2.44	−4.30	−0.60	**.011**	.002	2.42	.497
Humour	0.02	−0.95	1.01	.966	<.001	4.23	.116
Negative valence	0.28	−0.50	1.07	.486	<.001	4.56	<.001
Event							
Expected	−1.12	−2.73	0.48	.173	<.001	0.02	.988
Importance	−2.01	−3.87	−0.14	**.035**	.001	0.72	.705
Pleasant	0.18	−1.23	1.59	.803	<.001	0.01	.998

Note: Each row represents a separate multilevel model with a random intercept and random slope. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect; $\sigma^2_{trigger}$, the variance estimate for the random slope of the trigger type. We have displayed significant p values in bold.
[a]Model produced singular fit; fixed effect only model was interpreted.

the contextual variables about the situational and event characteristics, we found that participants generally reported the situation and the event to be *less* important during triggers that were based on increased skin conductance. The differences we found were of small effect size with explained variances ranging from 0.001 to 0.002. In exception of negative valence for which there was significant variance around the slope, none of the other random variance estimates indicated that the effects were different across individuals.

Exploratory analyses

To examine the skin conductance features of arousing moments more in depth, we used multilevel regressions to compare the skin conductance signatures of participants' two most and two least arousing and emotional moments. We compared both categories (i.e. high and low) on five different outcomes measures related to the skin conductance signal: (i) mean, (ii) standard deviation, (iii) autocorrelation, (iv) number of peaks, and (v) number of change points. The results of these additional analyses are presented across five tables, each using a different self-report measure: arousal, positive energy, irritation, anxiety, and negative valence.

In general, we found most of our effects for self-reported arousal (Table 7: $M_{high\ arousal} = 79.19$; $M_{low\ arousal} = 19.51$) and positive energy (Table 8: $M_{high\ positive\ energy} = 80.77$; $M_{low\ positive\ energy} = 37.33$). During moments in which individuals reported relatively high levels of arousal, the skin conductance signal generally also had a higher mean, a lower

Table 7. A comparison of different skin conductance parameters across different timeframes (i.e. 1 minute, 10 minutes, and 30 minutes before the trigger) between a person's two highest ($N = 164$, $M = 79.19$) and two lowest ($N = 164$, $M = 19.51$) self-reported arousal scores

	High	Low	B	95% CI		p	R^2
Mean							
1 minute	7.09	6.32	0.77	0.19	1.34	**.009**	.010
10 minutes	6.71	5.90	0.81	0.25	1.38	**.005**	.011
30 minutes	6.76	6.08	0.68	0.14	1.22	**.014**	.008
SD							
1 minute	0.87	0.88	−0.01	−0.17	0.15	.930	<.001
10 minutes	0.89	0.89	0.01	−0.11	0.12	.911	<.001
30 minutes	1.04	1.10	−0.07	−0.20	0.06	.306	.002
Autocorrelation							
1 minute	0.76	0.79	−0.03	−0.06	0.01	.101	.008
10 minutes	0.86	0.89	−0.03	−0.05	−0.01	**.002**	.021
30 minutes	0.89	0.91	−0.02	−0.04	−0.01	**.007**	.016
No. of change points							
1 minute	0.10	0.18	−0.08	−0.15	−0.01	**.034**	.013
10 minutes	1.27	1.65	−0.38	−0.62	−0.13	**.003**	.024
30 minutes	3.92	4.95	−1.02	−1.54	−0.51	**<.001**	.035
No. of peaks							
1 minute	4.56	4.61	−0.06	−0.80	0.69	.885	<.001
10 minutes	17.42	15.04	2.24	−0.27	4.76	.082	.008
30 minutes	49.37	43.63	5.79	−0.32	11.90	.064	.007

Note: Each row represents a separate multilevel model with a random intercept. The high category was coded as 0, and the low category was coded as 1. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect. We have displayed significant p values in bold.

Table 8. A comparison of different skin conductance parameters across different timeframes (i.e. 1 minute, 10 minutes, and 30 minutes before the trigger) between a person's two highest ($N = 164$, $M = 80.77$) and two lowest ($N = 164$, $M = 37.33$) self-reported positive energy scores

	High	Low	B	95% CI		p	R^2
Mean							
1 minute	7.36	6.45	0.91	0.35	1.46	**.001**	.013
10 minutes	6.93	6.06	0.88	0.36	1.39	**.001**	.013
30 minutes	7.06	6.24	0.81	0.30	1.32	**.002**	.011
SD							
1 minute	0.91	0.88	0.03	−0.11	0.17	.691	<.001
10 minutes	0.97	0.90	0.07	−0.04	0.19	.221	.002
30 minutes	1.08	1.05	0.03	−0.10	0.16	.641	<.001
Autocorrelation							
1 minute	0.77	0.80	−0.02	−0.05	0.01	.165	.005
10 minutes	0.87	0.89	−0.02	−0.04	0.00	.058	.008
30 minutes	0.89	0.91	−0.02	−0.04	−0.01	**.010**	.014
No. change points							
1 minute	0.09	0.18	−0.09	−0.16	−0.02	**.013**	.018
10 minutes	1.52	1.66	−0.14	−0.38	0.10	.251	.004
30 minutes	3.95	4.64	−0.69	−1.20	−0.18	**.009**	.017
No. of peaks							
1 minute	4.79	4.46	0.31	−0.51	1.13	.461	.002
10 minutes	18.59	15.72	2.73	0.08	5.39	**.045**	.010
30 minutes	53.09	45.31	7.56	1.11	14.01	**.022**	.011

Note: Each row represents a separate multilevel model with a random intercept. The high category was coded as 0, and the low category was coded as 1. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect. We have displayed significant p values in bold.

autocorrelation, and less change points than in the low condition. For positive energy, we found similar results. During moments in which individuals were relatively positively energized, their skin conductance was overall higher, and the signal contained fewer change points. Across 30 minutes, we also found that the autocorrelation was lower. Finally, in contrast to arousal, for positive energy, the number of peaks were more numerous during moments of high positive energy. It is notable that most of the effects that we found in Tables 7 and 8 were present across all the three different timeframes. For the negative affective scales, irritation (Table 9: $M_{\text{high irritation}} = 54.51$; $M_{\text{low irritation}} = 11.49$), anxiety (Table 10: $M_{\text{high anxiety}} = 41.50$; $M_{\text{low anxiety}} = 4.65$), and negative valence (Table 11: $M_{\text{high negative valence}} = 33.16$; $M_{\text{low negative valence}} = 1.66$), in line with our previous findings (see Table 5), relationships were mostly absent for these three negative affect scales.

DISCUSSION

To capture moments that were psychologically arousing, the present study adopted a sampling technique in which 5 days of experience sampling self-reports were linked to physiological measures of skin conductance. Our study indicates that although it is feasible to obtain continuous measurements of skin conductance in a naturalistic setting, an integration of skin conductance signals with momentary self-reports in daily life appears challenging. Specifically, sampling situations based on relative increases of skin conductance in this study did not actually result in a sample of more psychologically arousing situations. In fact, because of an unexpected dependency between level and slope, an inverse pattern was found. Our additional exploratory analyses show that there might be additional skin conductance features related to arousal in daily life. In what follows, we first discuss measuring skin conductance outside the lab in real life as well as the relationships between skin conductance measures and self-reports. Second, we turn our attention to our findings concerning the triggering algorithm and inform future studies that aim to use physiological signals to sample arousing situations.

Skin conductance and momentary self-reports in daily life

Our study is among the few to have linked physiological measures of skin conductance to momentary self-reports as they occurred in everyday life. Based on our experiences and the physiological output that was administered by the skin conductance devices, measuring skin conductance for longer periods of time throughout the day outside the lab

Table 9. A comparison of different skin conductance parameters across different time-frames (i.e. 1 minute, 10 minutes, and 30 minutes before the trigger) between a person's two highest ($N = 164$, $M = 54.51$) and two lowest ($N = 164$, $M = 11.49$) self-reported irritation scores

	High	Low	B	95% CI		p	R^2
Mean							
1 minute	7.02	7.00	0.02	−0.58	0.61	.957	<.001
10 minutes	6.64	6.61	0.03	−0.54	0.61	.910	<.001
30 minutes	6.68	6.69	−0.01	−0.57	0.55	.974	<.001
SD							
1 minute	0.90	0.80	0.10	−0.04	0.25	.166	.004
10 minutes	0.95	0.87	0.08	−0.04	0.20	.174	.003
30 minutes	1.10	1.03	0.07	−0.06	0.19	.280	.002
Autocorrelation							
1 minute	0.75	0.76	−0.01	−0.04	0.02	.559	.001
10 minutes	0.86	0.89	−0.03	−0.05	−0.01	**.007**	.016
30 minutes	0.89	0.91	−0.02	−0.04	−0.01	**.009**	.015
No. of change points							
1 minute	0.07	0.13	−0.06	−0.13	0.00	.070	.010
10 minutes	1.46	1.49	−0.03	−0.28	0.22	.811	<.001
30 minutes	4.40	4.62	−0.21	−0.73	0.30	.417	.002
No. of peaks							
1 minute	4.36	4.66	−0.28	−1.08	0.52	.492	.002
10 minutes	15.66	17.01	−1.22	−3.95	1.51	.382	.002
30 minutes	42.91	50.85	−7.56	−14.53	−0.61	.034	.011

Note: Each row represents a separate multilevel model with a random intercept. The high category was coded as 0, and the low category was coded as 1. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect. We have displayed significant p values in bold.

Table 10. A comparison of different skin conductance parameters across different timeframes (i.e. 1 minute, 10 minutes, and 30 minutes before the trigger) between a person's two highest ($N = 164$, $M = 41.50$) and two lowest ($N = 164$, $M = 4.65$) self-reported anxiety scores

	High	Low	B	95% CI		p	R^2
Mean							
1 minute	6.71	6.42	0.28	−0.29	0.86	.336	.001
10 minutes	6.43	6.09	0.34	−0.20	0.87	.221	.002
30 minutes	6.56	6.32	0.24	−0.30	0.77	.386	.001
SD							
1 minute	0.80	0.79	0.01	−0.12	0.14	.928	<.001
10 minutes	0.85	0.89	−0.03	−0.16	0.09	.599	<.001
30 minutes	1.02	1.06	−0.04	−0.17	0.09	.577	<.001
Autocorrelation							
1 minute	0.75	0.78	−0.03	−0.06	0.01	.115	.006
10 minutes	0.87	0.88	−0.01	−0.03	0.01	.155	.004
30 minutes	0.90	0.91	−0.01	−0.02	0.01	.342	.002
No. of change points							
1 minute	0.09	0.15	−0.07	−0.13	0.00	**.048**	.011
10 minutes	1.39	1.55	−0.16	−0.41	0.08	.196	.005
30 minutes	4.20	4.64	−0.45	−0.92	0.03	.070	.008
No. of peaks							
1 minute	4.30	4.65	−0.31	−1.08	0.45	.422	.002
10 minutes	15.00	15.87	−0.79	−3.22	1.65	.526	.001
30 minutes	44.78	43.96	0.88	−5.10	6.86	.773	<.001

Note: Each row represents a separate multilevel model with a random intercept. The high category was coded as 0, and the low category was coded as 1. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect.

appeared feasible. Participants were not too bothered by the device and only in a small proportion our measurements we had to deal with loose electrodes. This gives some rise to optimism regarding future use of continuous measures of skin conductance in daily life. Particularly in the case of the EdaMove that uses gel-based electrodes that are fixed to the skin and that are recommended as opposed to rings or bracelets that use dry electrodes (Boucsein et al., 2012; Fowles et al., 1981).

Nonetheless, there are some practical challenges to consider when using skin conductance devices in daily life. To make sure the electrodes would stick for an entire day, we used additional tape. Although this worked well, it can also produce possible artefacts in the data. It might for instance further induce sweating, the pressure on the palmar surface could elicit changes in blood circulation, and changes in pressure could produce local changes in skin resistance (i.e. Ebbecke waves; Boucsein, 2012). Also, participants in our study had to come in twice a day to get a device administered, which not only intensified the design for participants but also made it challenging to test many participants concurrently or to measure during evenings and weekends. Students are in this regard a convenient population as in many cases it was easy for them to adjust their schedule. This is something to bear in mind when testing among other, possibly less flexible populations. With the help of 13 student assistants, it

took us a total of 10 weeks to gather the data for 82 participants.

In addition to these practical considerations, it is important to consider the complexity of skin conductance data. While skin conductance data are generally sensitive to arousing events, the signal lacks specificity as it depends on many non-psychological factors such as digestion, movement, and temperature. It is therefore at least promising that when we examined the raw skin conductance signal for the minute preceding a trigger (i.e. both random triggers and triggers based on skin conductance), we found several relationships with self-reported affect, situations, and events that were in the expected direction. In line with skin conductance being indicative for psychological arousal (Boucsein, 2012), we found that self-reported arousal was positively related to SCL in the minute preceding a questionnaire. Additionally, higher levels of positive energy were related to higher levels of skin conductance, which aligns with findings of previous lab studies that found that experimentally induced positive emotions such as happiness, joy, and amusement generally coincide with elevated skin conductance responses (Kreibig, 2010). Also, we found relationships with the complexity of the situation, as well as the expectedness of the event, which is again consistent with levels of skin conductance denoting moments of psychological significance. Accordingly, we collected some evidence that skin

Table 11. A comparison of different skin conductance parameters across different timeframes (i.e. 1 minute, 10 minutes, and 30 minutes before the trigger) between a person's two highest ($N = 164$, $M = 33.16$) and two lowest ($N = 164$, $M = 1.66$) self-reported negative valence scores

	High	Low	B	95% CI		p	R^2
Mean							
1 minute	6.63	6.88	−0.26	−0.82	0.31	.378	.001
10 minutes	6.14	6.58	−0.44	−0.98	0.10	.109	.003
30 minutes	6.31	6.74	−0.43	−0.95	0.09	.106	.003
SD							
1 minute	0.93	0.72	0.20	0.07	0.33	**.003**	.016
10 minutes	0.99	0.82	0.17	0.06	0.29	**.004**	.012
30 minutes	1.10	0.99	0.12	0.00	0.24	.060	.005
Autocorrelation							
1 minute	0.75	0.76	−0.01	−0.04	0.02	.592	.001
10 minutes	0.87	0.87	0.00	−0.02	0.02	.723	<.001
30 minutes	0.90	0.90	0.00	−0.02	0.01	.874	<.001
No. change points							
1 minute	0.12	0.12	0.01	−0.06	0.07	.858	<.001
10 minutes	1.54	1.59	−0.05	−0.30	0.20	.702	<.001
30 minutes	4.34	4.26	0.08	−0.41	0.57	.753	<.001
No. of peaks							
1 minute	4.44	4.54	−0.12	−0.85	0.62	.758	<.001
10 minutes	14.59	16.52	−2.09	−4.71	0.55	.120	.007
30 minutes	42.35	46.36	−4.68	−11.25	1.92	.164	.004

Note: Each row represents a separate multilevel model with a random intercept. The high category was coded as 0, and the low category was coded as 1. B, unstandardized regression coefficient of the fixed effect; CI, confidence interval; R^2, explained variance for the fixed effect. We have displayed significant p values in bold.

conductance in daily life is meaningfully associated with experience sampling measures of emotions.

It was surprising that we did not find any relationships between skin conductance and the negative affective dimensions: anxiety, irritation, and negative valence. We expected anxiety and irritation to be accompanied by states of high physiological activation. Anxiety, for example, is a common emotion that has been related to elevated skin conductance responses in previous lab studies (Kreibig, 2010). The negative valence scale comprised sadness, gloominess, and anger. Previous lab studies found sadness to be one of the emotions that was accompanied by decreases in skin conductance but found the opposite for anger (Kreibig, 2010). It could be that for negative valence, the low arousal of sadness and gloominess and the high arousal of anger balanced each other out. Post hoc, we have looked at each item separately, but for all three items separately, no relationship was found. Looking at the significant random variance around the slopes of anxiety, irritation, and negative valence, we conclude that the absence of these effects might not hold for the entire sample. Also, based on the low means that we found of the three negative affect dimensions, moments of severe anxiety,

negative valence, or irritation were relatively rare. As a result, it could be that for many participants, we did not have enough variance on the higher end of the scale to find a potential effect.

Of note, the effect sizes of the relationships that we found were relatively small (between 0.5% and 0.8% explained variance), and they were only significant because of the large sample size. The small effect size is not completely unexpected due to common discrepancies between using two converging operationalizations of emotional experiences. First, for the skin conductance signal, this might be due to the complex nature of everyday physiological information. In the lab, participants are generally requested to remain as still as possible while working through an experimental task in a controlled environment. In daily life, we cannot, and do not desire, to constrain participants similarly because participants must be able to go about their day undisturbed. Because of this, there are many factors that can cause artificial jumps in the data. Participants could for instance squeeze their hands or grab something. To deal with this, researchers could consider triangulating the skin conductance data with other relevant sensor data, such as heart rate or respiratory rate. If only one of the sensors shows an increase due to noise in the data, researchers are less likely to misinterpret this increase as potentially relevant. In addition to these artificial jumps, skin conductance is not an exclusive indicator of emotional experiences. Skin conductance is a complex physiological signal that relates to non-psychological processes as well. It, for instance, also varies as a function of digestion, homeostasis, medicine intake, and/or temperature (Boucsein, 2012; Mauss & Robinson, 2009). As for our study design, the signal had to be processed instantly, we were limited in options for data filtering. In this particular design, we only filtered the data on movement, but future research should consider incorporating additional features to ensure the algorithm processes a clean signal.

For self-reports of emotions, although a common, fast, and relatively effective method for assessing subjective emotional experiences, researchers are dependent on subjective judgements of participants. In the moment, participants might not always consciously experience their feelings. They might also lack the knowledge to cluster their feelings into emotional adjectives, over-rationalize their feelings, or even report about their feelings more desirably (Mauss & Robinson, 2009). Previous research indicates that people tend to respond to items of affective scales more as continuous valence dimensions, instead of pertaining to concrete situations that elicit certain discrete emotions (Barrett, 1998). As these general dimensions are not necessarily bound to specific situations, it might be harder to relate this to the specific timeframe of the skin conductance signal (i.e. a minute before the trigger). Our exploratory analyses further underline this as we found relationships between self-reports and skin conductance which were generally present across three different time spans: 1 minute, 10 minutes, and 30 minutes. Finally, what further complicates relating skin conductance signal dynamics with self-reported measures is that in ESM, participants usually get some time to fill in the questionnaire. Where our participants were generally fast responders, to ensure participants report in the

moment, we recommend future studies to use stricter timeframes to fill in the questionnaire.

Using the skin conductance signal to sample arousing moments

Our study indicates that sampling arousing situations based on physiological skin conductance data is more complex than simply administering the relative increase of the current minute with the preceding minute. We found that selectively sampling situations based on relative increases in skin conductance actually resulted in a sample of *less* psychologically arousing situations as experienced by participants—the opposite of what we tried to achieve. In our case, the algorithm successfully selected moments of skin conductance increases, but the overall SCL in the 10 minutes approaching the trigger was lower in our triggering condition. This general level-slope dependency was not anticipated but of course makes sense if relative increases are the focus. After all, as the algorithm selected moments on the basis of relative increase, lower values required a less substantial increase to be selected. As a result, these moments had an increased chance to trigger a questionnaire, yet participants in these moments were in a relatively low arousal mood. The negative relation between level and slope has earlier been described as the law of initial values (Wilder, 2014). This law predicts that low initial values are generally related to a higher difference scores.

Our additional analyses indicated that there might be other skin conductance characteristics worth looking into when designing algorithms to sample arousing moments besides relative increases. Particularly in the case of positive arousal, we found that mean levels, autocorrelations, number of peaks, and number of change points were all in some way related to high arousing moments. Our results indicate that people's skin conductance signature may change in many different and complex ways when an arousing situation is encountered. For future studies, it could be interesting to create an algorithm that draws upon an even more extensive list of time series parameters. Ideally, such an algorithm also contains an extensive learning period. Participants should first wear the skin conductance device for some time and regularly provide feedback when they are in an arousing situation. The algorithm could then adjust the parameters each time the participant provides feedback. With each round of feedback, the predictive capabilities of the algorithm should then become more precise.

CONCLUSION

Overall, in this study we found that relationships between skin conductance and momentary self-report measures in daily life were not straightforward. We found that we were able to use a skin conductance device (EdaMove 3) that makes use of gel-based electrodes which were stuck to the skin for longer periods of time outside of the lab. Participants overall did not experience much discomfort, and electrodes overall remained relatively well attached to the skin. Nonetheless, bearing the complexity of the skin conductance signals in mind, an

algorithm that was concerned with selecting relative increases of SCL across intervals of a minute appeared oversimplified. Before we can confidently answer if and how it will be possible to use the skin conductance measures to select situations of psychological arousal, we need to better understand the discrepancies between skin conductance measures and self-reports. Overall, we found some indication of relationships between skin conductance and self-reported arousal in daily life, but effect sizes were small, and for many of our negatively valenced constructs, we did not find a relationship at all. Additional analyses showed that it is important for future studies to also account for other relevant parameters of skin conductance (i.e. autocorrelation, number of peaks, and number of change points) and also consider multiple and broader timeframes. Having exact information on where both measures deviate can further inform future studies that intend to use both physiological skin conductance and self-reports in their design. For additional validations, future studies might also investigate other psychological correlates of physiological arousal in daily life (e.g. personality; Eysenck, 1967; Fowles, 1980) or use multiple indicators of arousal, such as heart rate or respiratory rate.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

REFERENCES

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition and Emotion*, *12*, 579–599. https://doi.org/10.1080/026999398379574.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Boucsein, W. (2012). *Electrodermal activity*. Dordrecht, the Netherlands: Springer Science and Business Media.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*, 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*, 319–333. https://doi.org/10.1037//0022-3514.67.2.319.

Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., … Ravaja, N. (2016). The psychophysiology primer: A guide to methods and a broad review with a focus on human–computer interaction. *Foundations and Trends in Human–Computer Interaction*, *9*, 151–308. https://doi.org/10.1561/1100000065.

Csikszentmihalyi, M., & Hunter, J. (2003). Happiness in everyday life: The uses of experience sampling. *Journal of Happiness Studies*, *4*, 185–199. https://doi.org/10.1007/978-94-017-9088-8_6.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method: Mental disorders in their natural settings. *Journal of Nervous and Mental Disease*, *175*, 526–536 doi:cbo9780511663246.006.

Dimotakis, N., Ilies, R., & Judge, T. A. (2013). Experience sampling methodology. In J. M. Cortina, & R. S. Landis (Eds.),

Modern research methods for the study of behavior in organizations (pp. 345–374). New York, NY: Routledge.

Doberenz, S., Roth, W. T., Wollburg, E., Breuninger, C., & Kim, S. (2010). Twenty-four hour skin conductance in panic disorder. *Journal of Psychiatric Research*, *44*, 1137–1147. https://doi.org/10.1016/j.jpsychires.2010.04.012.

Doberenz, S., Roth, W. T., Wollburg, E., Maslowski, N. I., & Kim, S. (2011). Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology*, *80*, 87–95. https://doi.org/10.1016/j.ijpsycho.2011.02.002.

Dufner, M., Arslan, R. C., Hagemeyer, B., Schönbrodt, F. D., & Denissen, J. J. (2015). Affective contingencies in the affiliative domain: Physiological assessment, associations with the affiliation motive, and prediction of behavior. *Journal of Personality and Social Psychology*, *109*, 662–667. https://doi.org/10.1037/pspp0000025.

Ebner-Priemer, U. W., Koudela, S., Mutz, G., & Kanning, M. K. (2013). Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. *Frontiers in Psychology*, *3*, 1–8. https://doi.org/10.3389/fpsyg.2012.00596.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, Illinois: Thomas.

Fowles, D. C. (1980). The three-arousal model: Implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology*, *17*, 87–104 doi: j.1469-8986.1980.tb00117.X.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology*, *18*, 232–239 j.1469-8986.1981.tb03024.

Frijda, N. H. (2004). Emotions and action. In A. S. R. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium* (pp. 158–173). Cambridge, England: Cambridge University Press.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72–91. https://doi.org/10.1037/a0032138.

Hoehn-Saric, R., McLeod, D. R., Funderburk, F., & Kowalski, P. (2004). Somatic symptoms and physiologic responses in generalized anxiety disorderand panic disorder: An ambulatory monitor study. *Archives of General Psychiatry*, *61*, 913–921. https://doi.org/10.1001/archpsyc.61.9.913.

Killick, R., & Eckley, I. (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58*, 1–19. https://doi.org/10.18637/jss.v058.i033.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*, 394–421 Doi:J.biopsycho.2010.03.010.

Kuppens, P. (2008). Individual differences in the relationship between pleasure and arousal. *Journal of Research in Personality*, *42*, 1053–1059.

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. *New directions for methodology of social & behavioral science.*, *42*, 1053–1059. https://doi.org/10.1016/j.jrp.2007.10.007.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*, 209–237. https://doi.org/10.1080/02699930802204677.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (Sixth ed.). Los Angeles, CA: Muthén & Muthén.

Myrtek, M., Aschenbrenner, E., & Brügner, G. (2005). Emotions in everyday life: An ambulatory monitoring study with female students. *Biological Psychology*, *68*, 237–255. https://doi.org/10.1016/j.biopsycho.2004.06.001.

Myrtek, M., Brügner, G., Fichtler, A., König, K., Müller, W., Foerster, F., & Höppner, V. (1988). Detection of emotionally induced ECG changes and their behavioural correlates: A new method for ambulatory monitoring. *European Heart Journal*, *9*, 55–60. https://doi.org/10.1093/eurheartj/9.suppl_n.55.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits—A repeat performance. *Current Directions in Psychological Science*, *15*, 198–202. https://doi.org/10.1111/j.1467-8721.2006.00435.x.

Nezlek, J. B., & Plesko, R. M. (2003). Affect-and self-based models of relationships between daily events and daily well-being. *Personality and Social Psychology Bulletin*, *29*, 584–596. https://doi.org/10.1177/0146167203029005004.

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*, 642–681. https://doi.org/10.1037/pspp0000111.

Pejovic, V., Lathia, N., Mascolo, C., & Musolesi, M. (2016). Mobile-based experience sampling for behaviour research. In M. Tkalčič, B. De Carolis, M. De Gemmis, A. Odić, & A. Košir (Eds.), *Emotions and personality in personalized services* (pp. 141–161). Cham, Germany: Springer International Publishing.

Quinn, J. M., Pascoe, A., Wood, W., & Neal, D. T. (2010). Can't control yourself? Monitor those bad habits. *Personality and Social Psychology Bulletin*, *36*, 499–511. https://doi.org/10.1177/0146167209360665.

R Development Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reis, H. T., & Gable, S. L. (2000). Event-sampling methods. In H. T. Reis, & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 190–222). New York: Cambridge University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172. https://doi.org/10.1037//0033-295x.110.1.145.

Scollon, C. N., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In E. Diener (Ed.), *Assessing well-being* (pp. 157–180). Dordrecht, the Netherlands: Springer International Publishing.

Sideridis, G. D. (2008). The regulation of affect, anxiety, and stressful arousal from adopting mastery-avoidance goal orientations. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *24*, 55–69. https://doi.org/10.1002/smi.1160.

Van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology and Behavior*, *106*, 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020.

Westerink, J., Ouwerkerk, M., de Vries, G.-J., de Waele, S., van den Eerenbeemd, J., & van Boven, M. (2009). *Emotion measurement platform for daily life situations. Paper presented at the affective computing and intelligent interaction and workshops, 2009. ACII 2009.*

Wilder, J. F. (2014). *Stimulus and response: The law of initial values*. Baltimore: Williams & Wilkins.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*, 843–863. https://doi.org/10.1037/0033-295x.114.4.843.

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*, 1281–1297F. https://doi.org/10.1037//0022-3514.83.6.1281.