UNIVERSITY OF TWENTE.

Faculty Behavioral, Management and Social Sciences

The association between self-reported emotional arousal and electrodermal activity of individuals in daily life.

A seven-day longitudinal Experience Sampling Study based on EDA components and both Momentary and Retrospective Self-reports.

In Partial Fulfilment of the Requirement

for the Degree of the Master of Science

by Sascha Bödder

November 16th 2020

1st: Dr. M. L. Noordzij,

2nd: Drs. Y. P. M. J. Derks,

Department of Psychology, Health & Technology

Faculty of Behavioural, Management and Social Sciences (BMS)
Department of Psychology, Health & Technology
University of Twente

PHYSIOLOGICAL & PSYCHOLOGICAL AFFECTIVE AROUSAL IN DAILY ACTIVITY

Table of Content

PHYSIOLOGICAL & PSYCHOLOGICAL AFFECTIVE AROUSAL IN DAILY ACTIVITY

Abstract

Objective: In the past few years, an association between emotional arousal and sympathetic arousal measured by electrodermal activity (EDA) was found in laboratory studies mainly focused on inter-individual estimates. Based on earlier found differences in psychophysiological associations between laboratory and real-life, and suggestions that emotions determined to a great extent by intra- individual variety, exploring the association between individuals' emotional intensity and physiological data in the form of EDA in real life is the aim of this paper. The study explored inter- and intra-individual associations between physiological- and psychological-emotional arousal in real life based on a measurement period of seven days.

Design: Quantitative longitudinal study with experience sampling method (ESM) with both between and within-subject designs.

Method: The study was conducted among 35 students (23 male, 12 female). Physiological arousal was measured by three components of Electrodermal Activity (EDA) (SCL, SCR-frequency & SCR-amplitude) by the 'Empatica E4 wristband' (Empatica Inc., Boston, MA, USA). Emotional arousal in the form of momentary (last minute) and retrospective (last two hours) self-reports were collected by the mobile applications a) 'TIIM' (2018) and b) 'm Quest' (2017) in a fixed time-dependent-sampling method over seven days in a row. A Continuous Decomposition Analysis (CDA) was done to analyze the three single EDA components. For analyses, both within-subject and between-subject analyses were conducted, and inter-individual and intra-individual correlations were computed. Additional histograms and boxplots are given to visualize different patterns in a variety of data.

Results: The data showed a wide range of intra-individual associations between sympathetic and emotional arousal (positive and negative correlations), although most showed a weak association. On an inter-individual level, only very weak, mostly non-significant correlations were found. Besides, no significant difference could be measured between momentary and retrospective measurements in all three EDA components.

Conclusion: The association of emotional arousal found in prior laboratory studies was not found in this study. By conducting a real-life study, many variables that could be controlled within prior laboratory studies may have influenced our data. The data suggest that presumably, other factors, both at the group- and the individual level, play a much more significant role in real-life environments than prior findings suggested. Future studies should consider more external and internal variables and several different physiological measurements to provide more insights into possible variables of interest in the association of emotions and physiology in daily life.

PHYSIOLOGICAL & PSYCHOLOGICAL AFFECTIVE AROUSAL IN DAILY ACTIVITY

Introduction

Depending on which of the numerous theories of emotions (James-Lange theory (Lange & James, 1922), the Cannon-Bard theory (Cannon, 1927), Schachter and Singer's two-factor theory (Schachter & Singer, 1962)) is consulted, emotions are often associated with physiological changes in our bodies. Especially physiological arousal, which is often measured by Electrodermal activity (EDA), shows in previous studies positive connections with emotional intensity (emotional arousal) (Boucsein, 1992; Farrow, 2013; Kosogonov et al. 2017; Västfjäll & Gärling, 2007). With the rapidly increasing use of technology in our daily lives, especially wearable devices (e.g., smartphones, smartwatches), which use physiological measurements to assess individuals' psychological states, such findings in this area are of interest to a wide range of groups. In addition to the IT-industry, psychology and therapy can also benefit from this. Recognizing or clarifying emotional intensity with the help of physiological measurements can be useful. Here people with difficulties in expressing emotions (e.g., infants, deaf-dump, and paralyzed people) or borderline patients (De Visser, Bohlmeijer, Noordzij, & Derks, 2017) could benefit. Also, people with alexithymia who have problems describing their emotional states (Lumley, Neely, & Burger, 2007; Sifneos, 1973) could profit from such a connection.

The connection of physiological arousal and emotional intensity is several times studied in lab studies with between-subjects designs like Kreibig (2010) showed in his review. Nevertheless, for further application into individuals' real-life, two problems arise. First, the majority of the research that was done focuses on between-subjects designs. These designs can lead to misleading assumptions due to a phenomenon known as unverified 'group-to-individual generalizability' (Fisher, Medaglia & Jeronimus, 2018; Molenaar & Campbell, 2009). Especially in the context, emotions are determined to a great extent by intra-individual variety (Russell, 2009). Second, little research on electrodermal activity (*EDA*) and emotional arousal is done in a real-life setting. That is a significant problem when providing miss-/ information about humans' individual and subjective affective states based on only laboratory studies.

This study investigates the extent of the relation between psychological-emotional intensity and physiological arousal (in EDA) between- and within-subjects in their real-life environment. This is done through wearable sensors, momentary, and retrospective self-reports with experience sampling for a 7-day period.

Core Affect and its first Dimension of Arousal

The work on the theory on core affect can be considered a progression of the work on the dimensional representation of emotions that led to the *circumplex model of affect*. The

circumplex model of affect is a two- dimensional approach that proposes that all affective states arise from two fundamental neurophysiological systems. So core affect is a neurophysiological state that is composed to some degree of valence (pleasant-unpleasant; i.e., the degree to which an emotion is perceived as positive or negative) and some degree of arousal (high activation–low activation; i.e., how strongly an emotion is felt) in all affective states (Barrett & Russell, 1999; Bradley & Lang, 2000; Lang, 1994; Posner, Russell, & Peterson, 2005).

Following Russell and Feldman Barrett (2009), this neurophysiological state is 'consciously accessible as a simple primitive non-reflective feeling most evident in moods and emotions but always available to consciousness' (Russell & Feldman Barrett, 2009, p. 104). Indeed its 'perception in consciousness differs from being focal to peripheral to out of sight'. Following Cummins (2014), 'if core affect is intense, it can be the focus of consciousness; when it is weak, it recedes into the conscious background' (p. 1299).

With this, core affect represents a single, subjective experienced feeling that has not to be directed to anything and can be "altered by real events, but also by imaginary, remembered, and foretold events" (Russell & Feldman, 1999, p.806). That does not implicate that people always know what caused their current Core affect, like 'free-floating emotions and moods and everyday feelings' (p.1233) like Russel (2009) suggested. So Core affect also 'can exist without being labeled, interpreted, or attributed to any cause.' (Russel, 2003; p.148).

The focus of this thesis is on arousal. Following Russel and Barrett (1999), terms like arousal, activation, intensity, or energy all refer to 'a sense of mobilization or energy' (p.809) at the level of subjective experience. They stated that 'subjective feelings of activation are not illusions, but a summary of one's physiological state' (p.806). This quote is a central statement in this study. It implies two connections. First, arousal can be mentally recognized by an individual (subjective arousal), and secondly that the subjective arousal underlays in physiological changes (objective arousal) that can be measured by technology.

Measuring Subjective Arousal. The most used tools that allow insight into each individual's inner states' subjective experience are self-reports. Two different types exist. First, retrospective self-reports ask for experience in the past ('remembering self'), and momentary self-reports ('experiencing self') that asks for an experience that is 'happening in or close to real-time' (Conner & Barrett, 2012; Kahneman, 2010). Notably, following Rijs and Kahneman (2005), the remembering self is not a pure summary of the experiencing self. The remembering self offers an after-the-fact interpretation and is a sort of deformed memory of experienced moments (Conner & Barrett, 2012; Kahneman & Rijs, 2005). In the eyes of Rijs and Kahneman (2005), it 'seems to be as if the remembering self is sometimes simply wrong' (p. 286) because it

is influenced by different factors (Singer & Salovey, 2010). An example of distortions in retrospective self-reports was given by Conner and Feldman Barret (2012) in a clinical setting. They mentioned some examples where patients revealed overestimations in comparison to earlier momentary self-reports (e.g., overestimated their prior experience of distress when quitting smoking (McConnell, 2011) and premenstrual symptoms (Shiffman et al., 1997)). Another problem seems to be that this sort of experiences is judged by psychological heuristics called the 'peak-end rule' (Fredrickson & Kahneman, 1993). With this, long-term emotional experiences are mostly interpreted based on a) its most intense point (peak) and b) its end (Stone et al., 2000). Robinson and Clore (2002) summarized that currently experienced, and reported emotions seem to be more valid because they are less cognitively biased than remembered ones.

Electrodermal Activity (EDA) & Sympathetic Arousal

EDA, also called skin conductance (SC), relates to the skin's electrical conductance changes in response to sweat secretion. EDA includes two components: a) rapid phasic component (skin conductance responses (SCRs) and b) background tonic (skin conductance level (SCL)), which derive from the sympathetic nervous system (SNS). The SNS is like its antagonistic-part of the parasympathetic nervous system (PNS), part of the autonomic nervous system (ANS) that controls different physiological functions in our body (Boucsein, 2012). The PNS is associated with relaxation, the SNS with activation of the ANS (a quick response activating system of the body). The main benefit to use EDA to measure sympathetic arousal is that the skin is only influenced by the SNS and not the PNS like in other measures of heart rate, blood pressure, or cortisol levels (Boucsein, 2012; Braithwell, 2015; Critchley, 2002; Dawson Schell, & Filion, 2016; Sequeira et al., 2009).

Measuring objective sympathetic arousal. The activation of the SNS is called sympathetic arousal. Thereby, that the SNS is connected to the activity of eccrine sweat glands (Boucsein, 2012), subtle changes in sweat secretion serve as a marker for sympathetic arousal. By applying a low constant voltage onto the skin, it is possible to measure SC's changes. These changes can be detected non-invasively (Fowles et al., 1981) and are measured in microsiemens (μS) (Stern et al., 2000).

The Relation of Skin Conductance (SC) & Emotional Arousal in Core Affect

Russell sees core affect's arousal dimension partly connected to ANS-changes (Alexander et al., 2005; Russell, 2003; Russell, 2009). While ANS-changes in the form of sympathetic activity are frequently related to SC, the sympathetic arousal measured by these changes does not represent a direct link to emotional arousal in core affect. Changes in sympathetic arousal are also assumed

to be related to attentional processing and cognitive states, next to mental states like emotions (Boucsein et al., 2012; Bradley, 2009; Larkin, 2006). Nevertheless, Lang, Greenwald, Bradley, and Hamm (1993) could report a positive correlation between SC (SCR-amplitude) and self-reported arousal in almost 77% of their participants (33% of all positive correlation showed a p < .05) by using arousal-inducing pictures. Also, Västfjäll and Gärling (2007) found a significant inter-individual correlation between the composite ratings of activation and SCR (r=0.88, p<.001). Here SCRs increased monotonically and linearly with the level of arousal (F(1,60) = 53.08, r* = .81). Same inter-individual results were found earlier by Winton, Putnam and Krauss in 1984 as well as later by Kosonogov et al. in 2017 for SCR-amplitude (T = 17; N = 24; p < .001) and for SCR-frequency (T = 0; N = 24; p < .001) as well as emotional words (Manning & Melchiori, 1974; Silvert, Delplanque, Verpoort, & Sequeira, 2004) Additionally, Boucsein (1992) summarized in his eminent book 'Electrodermal activity' that next to SCL both, the mean SCR-amplitude and mean SCR-frequency are mostly related to psychophysiological arousal. While Farrow (2013) declared the same as Russell that EDA components are a reliable and objective psychophysiological indicator of psychological arousal, but with the difference that this relation would appear, especially within subjects. In studies where other physiological parameters were used for emotion recognition (e.g., heartbeat), partly contradictory results were found. Evers et al. (2014) gave an interesting explanation, which could partly explain these contrary results in the strengths and directions of the correlations between subjective emotions and physiological measurements. Their dual-response theory distinguishes between reflective emotions from automatic emotions, with the strength of the relationship between physiological changes depending on which response system is involved. They suggest two independent reaction systems, a reflective system and an automatic one. The automatic system includes, among other physiological changes and signals and the accessibility of emotions. The reflective response system includes emotional experiences. Following this theory, only automatic emotions would result in physiological changes and would be detectable.

Inter-individual vs. Intra-individual. Russel (2009) stated that emotions are determined mainly by peoples' individual variety. He also stated that contextual factors, which can vary heavily between people (especially in real-life studies), can influence the relationship between emotions and ANS activity (Russell, 2009). People can vary in their interpretation and awareness of physiological symptoms like arousal in different contexts while facing different external factors. In the light of mostly inter-individual analyses focusing on emotions and EDA that were done and presented in this thesis, Fisher, Medaglia, and Jeronimus (2018) advises analyzing data

intra-individually. They stated that psychological processes that underlie individual variability obviate the risk of unverified group-to-individual generalizability by showing two to four times larger intra-individual variability than inter-individual analyses would presume.

Real-life vs. laboratory. Concerning the expected differences, Schmidt, Reiss Dürichen, and Laerhoven (2019) pointed out that although EDA is strongly influenced by the SNS, external parameters such as temperature, humidity, or physical activity can influence physiological data outside the laboratory to a much greater extent than EDA data. Myrtek and Brügner (1996) concluded that, based on comparing results between laboratory studies and real-life studies, emotions in everyday life could have very different physiological changes (measured by the heartbeat). The study showed that the same positive correlations between heartbeat and emotional arousal could not be found. A possible explanation could be the awareness of subjects in laboratory studies of perceiving particular circumstances that may increase the positive relation between physiological- and self-reported arousal.

The current study

Current findings provide evidence for the relation of emotional arousal and sympathetic arousal measured by EDA components (D'Hondt, 2010, Kosonogov et al., 2017; Manning & Melchiori, 1974; Silvert et al., 2004; Västfjäll & Gärling, 2007). While all the above findings are based on laboratory studies, Myrtek and Brügner (1996) provided evidence that physiological changes showed different results in indicating emotional arousal between lab- and real-life ambulatory settings. However, their study focused only on associations between emotional arousal and the heartbeat instead of EDA, which showed stronger positive correlations in laboratory studies.

RQ1: *What is the range of inter-individual associations between all three EDA components with the self-reported level of arousal?*

When focusing on emotional arousal, we have evidence that momentary self-reports show higher reliability than retrospective self-reports in reporting distress (McConnell, 2011; Shiffman et al., 1997; Robinson & Clore, 2002). Besides, in general, retrospective self-reports showed to be based on memory, which can be cognitively biased and interpreted differently in contrast to momentary self-reports. Nevertheless, we have no evidence if and to which extent the two types of self-reports (a. momentary, b. retrospective) influence the relation of emotional arousal and sympathetic arousal in real-life and both intra- and inter-individual level.

RQ 2: *Is the inter-individual correlation between the three EDA components and the reported arousal level significantly higher for momentary- than for retrospective self-reports?*

Moreover, none of the studies used an intra-individual study design in an everyday context to study the relationship between emotional arousal and sympathetic arousal. Following Fisher et al. (2018), inter-individual analyses used in all studies above could be exposed to unverified group-to-individual generalizability risk. Besides that, Russel (2009) suggested that emotions are determined to a great extent by intra-individual variety. In order to gain insight into the connection between emotional and sympathetic arousal intra-individually, the third research question is as follows:

RQ 3: *What is the range of intra-individual correlations between all three EDA components with the self-reported level of arousal?*

The current study aims to get more insight into emotional and sympathetic arousal in a real-life environment. This relation is analyzed for both inter- and intra-individual levels considering the differences between momentary and retrospective self-reports. This is studied via seven-day longitudinal with the experience sampling method (ESM) to measure emotional arousal by self-reports and physiological sympathetic arousal by EDA components of SCL, SCR-amplitude, and SCR-frequency.

Method

Participants

In total, 60 participants were recruited for the study by convenience sampling. Different researchers of the 'Faculty of Behavioural, Management and Social sciences' (BMS) at the University of Twente (NL) participated in a broader study focusing on emotional- and physiological states. There were two different years where data collection took place, 2017 and 2018. For data collection of 2017, no nationality of the participants was collected. Since master and bachelor students executed this experiment in the same manner, the nationality is expected to be comparable to those of the participants of the second part of the data collection. Based on the same structure and execution of the study, we can assume equal nationality distribution within the participants of 2017 and 2018.

A number of 35 out of 60 participants were considered for the current analyses. 25 participants were examined because of insufficient data points ($\leq$ 20 measurements of self-reported level with the corresponding EDA data). Data of 24 participants of the year 2017 and 11 participants of the year 2018 were taken into account. The 35 participants had an average age of 25.83 (SD=7.28), ranging from 19 to 70; 65.71% were male, 34.29 % female. Nationalities were only recorded for the 11 participants (2018), with a distribution of 45.45% German and 54.54% Dutch participants.

The participants had to be above the age of 18, needed to have at least some pc and mobile phone knowledge (IOS or Android), access to these devices during the study, and a running internet connection in the evening to store the data on pc. All participants who fulfilled these requirements and could use their mobile for nearly all two hours by day were accepted. Participation was voluntary. No payment was provided except for 5.25 credit points for students of the University of Twente and their data insight after the study.

The Ethics Committee approved this research of the 'Faculty of Behavioural, Management and Social sciences' (BMS) at the University of Twente (NL).

Design

The study was realized through a longitudinal interval-based experience-sampling method ESM) (Csikszentmihalyi & Larson, 1987) with a between-subject and within-subject and repeated-measures design that was applied. All data were collected during a continuous seven-day period per participant. The period's starting date was freely selectable; only the starting time was fixed at 8 am.

The E4 wristband device automatically collected the data of EDA. Self-reported arousal was collected with non-verbal real-time surveys administered through Experience Sampling Methods (ESM) with the help of a smartphone app and flexible pre-defined time intervals to reduce unanticipated interruption of participant's daily life. Every two hours, self-reports occurred with an additional time reserve of 20 minutes to answer two questions regarding momentary self-reports and retrospective self-reports.

To collect sufficient data points across the waking day, Munsch et al. (2009) suggested a minimum of five measurement points per day. The applied two-hour-spacing between the measurements allows even six to eight measurements per day during the waking hours. The high quantity of measurements per individual (maximal number of 64) was chosen to deal with possible technical issues while data-acquisition and acquire high data validation.

As a consequence of evaluations during the broader study, the participants used two slightly different apps, including different rating systems, to collect psychological data (i.a. level of arousal). The psychological data of 24 participants was collected by the mobile application mQuest' within 2017 and for eleven participants by the mobile application 'TIIM' in 2018.

Materials

Basic hardware. Participants needed access to pc/laptop with a USB-connection and a smartphone with IOS or Android regardless of the used version. Direct internet access was only required to sync data, not during data acquisition.

Psychological instruments I – TIIM. The smartphone application 'The Incredible Intervention Machine' (TIIM; Version 1.3.0), created by the University of Twente in May 2017, was used to collect psychological data. The language could be chosen in Dutch or English. Also, the starting time could be adjusted. TIIM can be used on mobile devices running both the operating systems Apple iOS and Android. The mobile application can also send push messages, such as a reminder for answering the questions about the personal level of arousal. Every two hours for seven days in a row, the same two requests were displayed:

Figure 1. Illustration of an alternated version of 'Affect Grid' that was displayed on the participants' mobile devices with an adjustable cursor [yellow point] and description of four emotional area states within the TIIM application (The Incredible Intervention Machine).

1. *Mark the position on the Affect Grid that best corresponds to how you felt the last minute!'*

2. *Mark the position on the Affect Grid that best corresponds to how you felt the past two hours!*

Having regard to the circumstances that this paper is part of large-scale research that deals with emotions in the context of both dimensions of Core affect, the only well-validated and suitable tool to measure self-ratings, the Affect Grid was chosen for this study (Russell, Weiss, & Mendelsohn; Watson & Clark, 1997). Developed by Russell et al. (1989), the single-item scale *Affect Grid* was designed to assess the two dimensions of core affect. Psychometric quality was assessed as 'adequate' for reliability, convergent validity, and discriminant validity (Russell, Weiss, and Mendelsohn (1989)), later as a moderately valid measure of arousal by Killgore (1998) and as strongly to moderately reliable in experience sampling methods (ESM) studies (Müller, 2019). The participants could respond via the Affect Grid in the form of a modified version (*Figure 2*) of the original 'Affect Grid' created by Russell, Weiss, and Mendelsohn (1989). By moving a yellow cursor (the starting position was always placed in the middle/neutral position) with one finger to the desired position, *the participants could rate their arousal-level (see Figure 2)*. Intern, the questions were rated on a
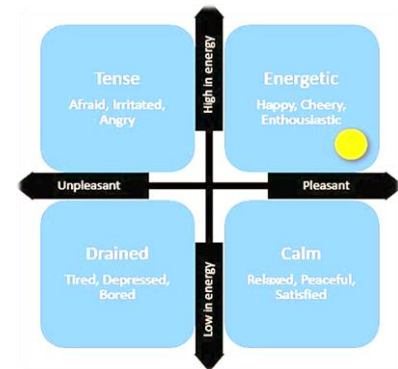
scale from -100 to +100 ('unpleasant' to 'pleasant') and -100 to +100 ('low in energy' to 'high in energy') and uploaded via CSV data-file. The Affect Grid shows adequate reliability, discriminant, and convergent validity according to emotional words, current mood, and feelings conveyed by facial expressions (Russel, Weiss, & Mendelsohn, 1989).

Psychological instruments II – mQuest. The smartphone application mQuest (cluetec GmbH) was also was used to collect psychological data. (Version 11.0) In the frequency of two hours for seven days in time, the same two requests were asked only in English:

1. *How intense were your emotions during the last minute?*

2. *How intense were your emotions during the last two hours*?

The participants could respond by using a 10 point Likert-scale ranging from zero ('very low') to ten ('very high'), where 5 represents the neutral score of arousal. The application mQuest can be used on mobile devices running both the operating systems Apple iOS and Android. Besides, the mobile application can send push messages, such as a reminder for answering questions.

Physiological instrument. Concerning the use of a wristband to measure EDA signals, a study with emotionally induced sweating (van Dooren, de Vries, & Janssen, 2012) compared with 16 different body locations found that EDA measured at the wrist was positively correlated with measurements at the fingers. Fingers are known as the best location to measure SNS-activity. However, compared with wrist-worn sensors, more obtrusive in daily activity**.** The Empatica E4 wristband (Empatica Inc., Boston, MA, USA) was chosen to collect data most efficiently (Figure 2). The E4 contains different sensors to measure different physiological signals, an internal real-time clock, the

Figure 2. The Empatica E4 wristband with the two metal electrodermal activity sensors (left) to measure sympathetic nervous system arousal

recording mode of 48+ hrs, a fast-charging time of two hours, and a robust body that qualifies it for this sort of study. The E4's EDA sensor is composed of two silver-coated (Ag) electrodes. A non-noticeable alternating current is applied to the skin through two equal dry electrodes. The sensors can measure conductance in μS-range [0.01, 100] with a default sampling rate of 4Hz. A charging cradle and a Micro USB to USB cable were used to sync data and load the E4.

For data-synchronization, the free software 'E4-Manager' (Empatica Inc., Boston, MA, USA) were used by the participants. To view and manage the acquired data on a secure cloud platform, the researchers used the 'E4 Connect' website (https://www.empatica.com). All data

could be checked roughly, sent to participants on demand, checked, and downloaded as raw data in CSV format.

Procedure

Short briefing and check for interest. Before data gathering could start, all possible participants were personally briefed about the study's content, procedure, and duration to check for their interest. To search for participants that would like to participate and collect as much data as possible was important because it was asked for endurance for all daily steps for one whole week. The informed consent was read and signed after participants complied with all conditions and their right to stop the study anytime. The contact was in person at no particular place, just orientated on the needs of the participants.

Quick software installation, information & usage. Once the participation was formal, the researcher or the participant on participants' devices installed the software (TIIM & E4-Manager or mobile app mQuest).

Second, the participant had to sign in (with their e-mail address and password) for the study using TIIM or mQuest. Afterward, the researcher sent an activation link to the participant to activate the study. Here the researcher asked for the desired starting-day. The researcher also showed a screenshot of the two questions and explained how to use the cursor to answer them correctly. Participants with an IOS-mobile device were also informed about a software bug while using the cursor and how they could work around it.

Third, the participants were shown how to use the E4-Manager to synchronize their data. Therefore the researcher used a previously closed session that was saved on the E4-wristband to show a test synchronization. Further, the participant received a printed instruction-guideline, where also, these steps were described (Appendix II).

E4-Wristband instructions. At the start, the participants were asked to watch a short instruction video to the E4-wristband (Appendix I). After that, the E4-wristband was handed out and placed on the participant's non-dominant wrist to demonstrate the electrodes' correct placement onto the skin. Then the participants were asked to start their test session and read the instruction guide. Afterward, the participants were asked to stop their first session on their own to check for possible problems. Following a successful tryout, the researcher handed out the E4-Wristband inclusive of a USB charging device.

Additional instructions. After the first own test trial, every participant was personally informed to wear the E4 wristband when possible the whole day-time on their non-dominant

hand (because of fewer movement disruptions). Furthermore, they were asked to sync the E4-wristband before going to bed and charge it by night while sleeping. The daily data storage was necessary because of some problems with the E4-Manager software and or the E4 wristband. In cases like that, the researcher could search for help in the morning so that that data could be saved and new devices/software-updates could be provided the same day.

Further, the participants were informed about the TIIM application questions (or the mQuest application) that could be answered 24 hours a day every two hours. The two questions provided by TIIM could be answered with a maximum delay of 20 mins. When a time-window was missed, the next two questions could be answered two hours later, counted from the moment the last question was provided. At least the participants were informed about an automatic reminder to answer the questions that are displayed at the mobile notification-window to reduce missing data collection.

Combined datasets. To increase the validity of both the inter-individual calculations (by a decrease in the number of participants) and the intra-individual calculations (excluding participants with < 20 measurements), we combined three datasets of the same broader study at the University of Twente (Department of Psychology, Health & Technology). Two datasets were collected in 2017 (survey periods: March-May and October–December). The third dataset was collected in 2018 (survey period: October – December). The difference between these two years of the same study is the application used to collect the self-reported arousal level. In 2017 the mQuest application was used; in 2018, the TIIM application was used.

Data Analysis

All statistical data analyses were performed using SPSS version 24.0.0.0 (IBM Corp, 2016). Regarding the two datasets collected in 2017 (via mQuest application) and one dataset of 2018 (via TIIM application), all self-reported data was entered into one dataset. Data of self-reports was utilized from the CSV files provided by both the TIIM- and mQuest- mobile application. The next step was converting these two different scales used to measure the self-reported level of arousal. The TIIM application that used the Affect Grid had a scale from -100 (low arousal) to +100 (high arousal).

In contrast to that, the mQuest application used a 10 point Likert scale with a range of 0-10 in which 5 represents a neutral arousal score. Therefore, we chose to convert the TIIM application

scale to values from 0 to 10 instead of -100 to +100. The following formula with 'x' as the data

from TIIM and 'y' the new recoded data was used for the conversion: $\frac{\left(\frac{x}{2}\right)+10}{2} = y.$

Next, all variables (a. momentary arousal, b. retrospective arousal, EDA components of c. SCL, d. SCR-amplitude, and e. SCR-frequency) were provided with timestamps and joined together into one SPSS dataset based on the timestamps per participant. Before, deconvolution of the raw skin conductance (SC) signal, the tonic component (SCL), and the phasic components (SCR-amplitude & SCR-frequency) were calculated by the Continuous Decomposition Analysis (CDA; Benedek & Kaernbach, 2010) with the MATLAB toolbox Ledalab (Version 3.46). Next, the dataset was screened for participants with less than 20 complete measurement points to exclude them from the dataset. Subsequently, the participants were renamed from 1 to 35.

We replaced single negative data points of the EDA readings with missing values. These negative values were shown as negative due to the applied CDA analysis. Missing data replaced only five negative SCL measurement values out of 2160 SCL measurement values.

First, all variables' descriptive statistics were calculated, all separated by the two types of self-reports (a. momentary, b. retrospective). All variables were controlled for normality by a Shapiro-Wilk test. Here a $p > 0.05$ indicates normally distributed data, a $p < 0.05$ indicates that the data has not been distributed normally. Depending on these results either the Pearson correlation (for p-values > 0.05) or the Spearman's correlation (for p-values < 0.05) was used. Furthermore, a Mann-Whitney U test for the two different datasets (mQuest in 2017; TIIM in 2018) was carried out. This test was done for the two types of self-reports (a. momentary, b. retrospective) separately.

The first research question was explored by Spearman's rank correlation coefficient analyses one-sided (inter-individual) for all three EDA- components with the self-rated level of arousal, also separated by the two types of self-reports (a. momentary, and b. retrospective).

The second research question was tested by the inter-individual correlations, which were transformed with a Fisher's z- and tested by an Independent T-test one-sided to compare the two types of self-reports (a. momentary versus b. retrospective).

The third research question was analyzed by Spearman's rank correlation coefficient analyses (intra-individual) for all three EDA components with the self-rated arousal level per individual. The correlations were separated by the two types of self-reports (a. momentary and b. retrospective) (see Appendix V-VI). Here, the distribution of correlations' strengths was visualized by histograms (see figure 3) to give a broader overview of the range and possible differences and similarities in all six conditions given by histograms. All strengths of correlation

were predefined by Pearson as negligible (.01 < r < .19), weak (.20 < r < .39), moderate (.40 < r < .59), strong (.60 < r < .79) or very strong (r ≥ .80). During all analyses, two-tailed tests were utilized with a p < 0.05.

Results

Descriptive statistics are given in table 1 for all variables of interest, separated per time interval/ type of self-report (momentary/ retrospective) to make both comparable. By comparing the means of self-reported arousal level, the data shows a higher arousal level for retrospective self-reports (M = 3.63) than for momentary self-reports (M = 3.10). Both means of self-reported arousal level are below the expected average value of '5' that represent the neutral middle (between 'low arousal' and 'high arousal') of both scales (mQuest vs. TIIM) that were used to report the level of arousal and ranged between 0 – 10.

Table 1

Descriptive Statistics of Self-reported arousal, Mean SCL, Mean SCR-frequency and Mean SCR-amplitude all separated by the two types of self-reports.

Variables	Last minute (momentary)			Past 2 hours (retrospective)		
	M (SD)	Range	N	M (SD)	Range	N
Self-reported arousal	3.10 (2.07)	0 - 10	1108	3.63 (1.93)	0 - 10	1108
SCL (μS)	.67 (2.05)	.29 - 38.58	1058	.80 (1.99)	.01 - 26.30	1102
SCR-frequency	12.10 (6.74)	0 - 29	1058	11.45 (5.35)[+]	.46 - 24.98[+]	1102
SCR-amplitude (μS)	.05 (.17)	0 - 4.39	1056	.05 (.10)	.01 - 1.33	1102

Note. SCR: skin conductance response, SCL: skin conductance level; [+]calculated to mean per minute

Before exploring the three research questions, all variables were controlled for normality by a Shapiro-Wilk test that reveals no normal distribution in all variables (see Appendix IV). Concerning the two different scales (2017: mQuest vs. 2018: TIIM) that were used to self-report the level of arousal, a Mann-Whitney U Test for momentary self-reported arousal and retrospective self-reported arousal was done (see Appendix IX). The classification of Cohen (1992) was used to assess the power of this effect. The results show (*U* = 92920.50, *Z* = -9.837, *p* < .000) a moderate effect size for momentary self-reports (d= 0.2955), for retrospective self-reports (*U* = 127074.00, *Z* = -3.178, *p* < .001.) a weak effect size (r = 0.0955). These results

mean that the two different scales used to measure the arousal level influence both momentary self-reports and retrospective self-reports. In comparison, the two different scales show stronger influence in momentary self-reports. For this, figure 1 visualize the distribution in the form of boxplots for the momentary arousal level per participant. In contrast, figure 4 does the same for the *retrospective arousal level*. Note that participants 1-23 self-reported their arousal level in 2017 with mQuest; participants 24-35 self-reported their arousal level in 2018 with TIIM.

In Figures 1 and 2, a higher level of self-reported arousal can be found for participants of 2018 (TIIM) than for 2017 (mQuest) participants. 8 out of 11 medians of participants of 2018 (TIIM) show higher values than the mean (M =3.096) of all participants. Figure 2 shows that 6 out of 11 medians of participants of 2018 (TIIM) show higher values than the mean (M =3.625) of all participants. For a comparison between both measurements (mQuest/ TIIM), the distribution of self-reported level of arousal is presented in two boxplots separated for both time-intervals (momentary/ retrospective) (see Appendix VIII).

Figure 1. Momentary self-reported level of arousal per participant. The dotted blue line represents the mean of all self-reported scores of the level of arousal; the black line represents the expected middle score (value '5') of both scales (mQuest vs. TIIM) that were used to report the level of arousal. The color-coded boxplots show the number of significant correlations (green = positive / red = negative) with the three EDA components per participant as follows: light green/ light red = 1, green = 2.

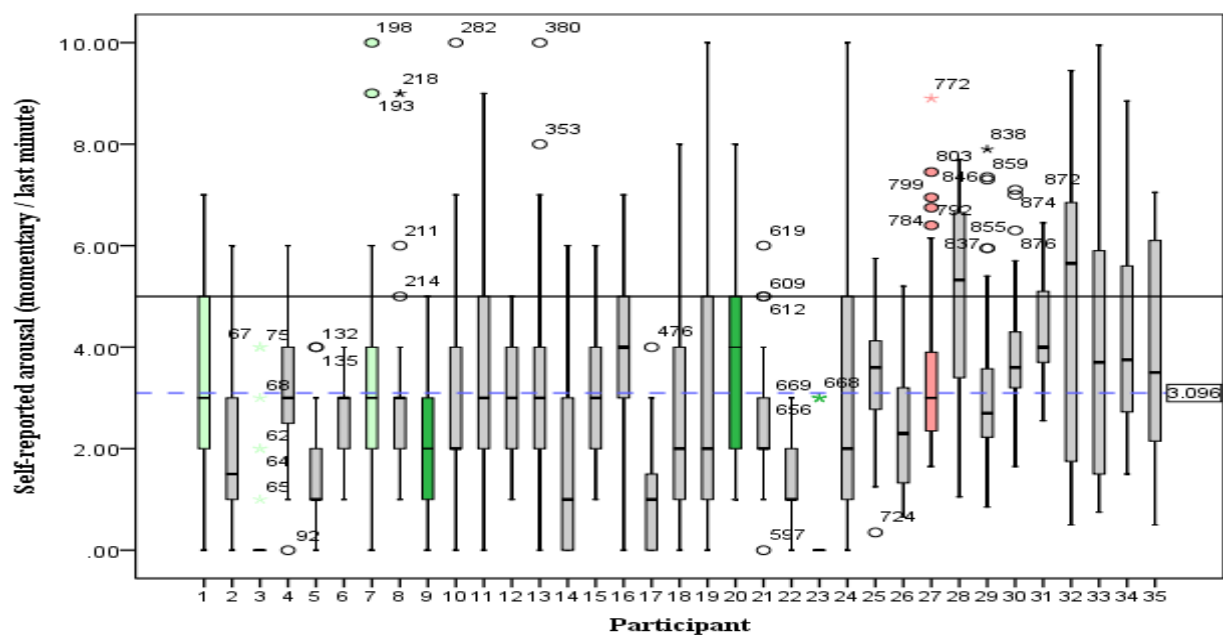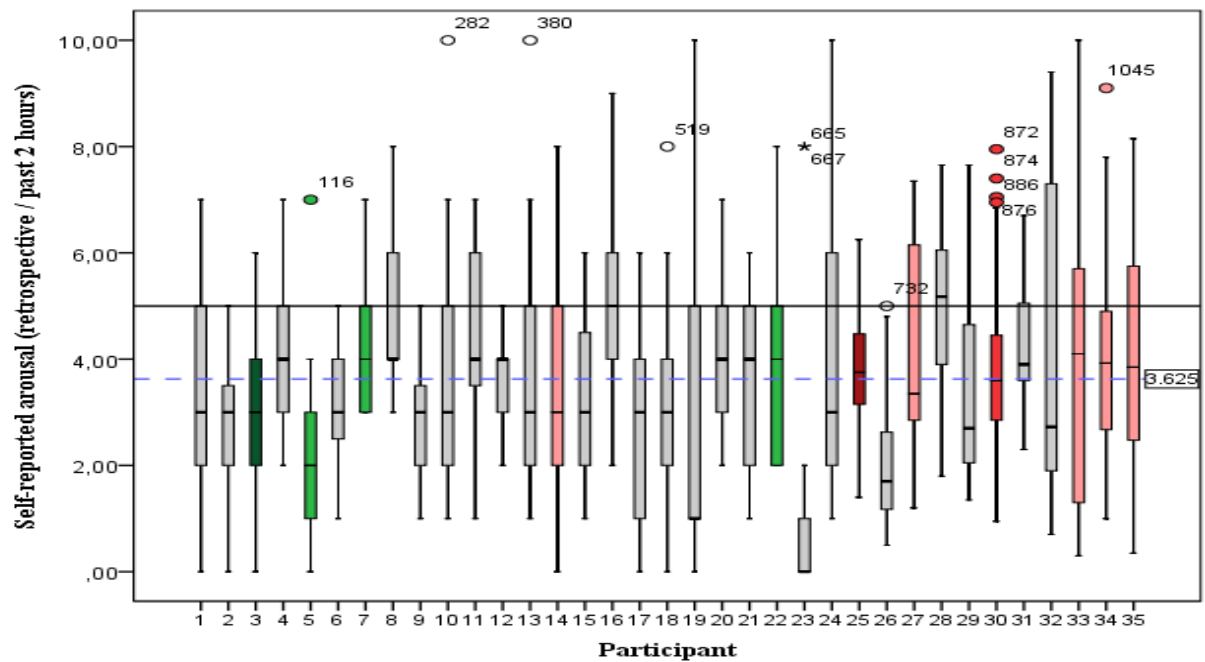Figure 2. Retrospective self-reported level of arousal per participant. The dotted blue line represents the mean of all self-reported scores of the level of arousal; the black line represents the expected middle score (value '5') of both scales (mQuest vs. TIIM) that were used to report the level of arousal. The color-coded boxplots show the number of significant correlations (green = positive / red = negative) with the three EDA components per participant as follows: light red = 1, green/ red = 2, dark green / dark red = 3

Inter-individual correlations between the three EDA components and self-reported arousal level

The data shows two significant yet negligible correlations. First, SCL with momentary self-reports and second, SCR-amplitude with retrospective self-reports. All other correlations were non-significant and ranged around zero between -.044 to .010. In total, we found two positive (range: .010 to .076) and four negative correlations (range: -.012 to -.09). No moderate to strong (positive) inter-individual correlation could be found.

Table 2

Inter-individual Spearman correlation of all used EDA components with the self-reported level of arousal all separated by the two types of self-reports/ time intervals

Variables		Self-reported arousal	
		Momentary (last minute)	Retrospective (past 2 hours)
SCL	Correlation coefficient	.076**	-.029
	Sig. (2-tailed)	.007	.165
SCR-frequency	Correlation coefficient	-.012	-.044
	Sig. (2-tailed)	.344	.072
SCR-amplitude	Correlation coefficient	.010	-.09**
	Sig. (2-tailed)	.371	.001

Note.; **Correlation is significant at the level of 0.01 (one-sided)*; SCR: skin conductance response; SCL: skin conductance level

The range of Inter-individual correlations for momentary and retrospective self-reports.

Table 2 shows more positive values of momentary inter-individual correlations than retrospective inter-individual correlations in all three EDA components. All momentary correlations show more positive values than retrospective correlations. We tested with Fisher's Z comparison calculation if these differences were statistically significant (see Appendix X). The differences between momentary and retrospective results of the SCL were not statistically significant (Z = .433, ns). The same was found for differences of SCR-frequency correlations that were not statistically significant (Z = .144, ns). Additionally, the SCR-amplitude correlations were also not statistically significant (Z = .408, ns).

The range of intra-individual correlations.

The histograms in figure 3 show the intra-individual correlations for the three components of EDA and the two conditions of the self-reported arousal level. All six histograms display that the intra-individual correlations are spread around zero, which indicates the presence of mostly weak correlations within individuals. Moderate correlations to strong correlations are less likely to occur but are also present in all histograms; only very strong correlations could not be found. Figure 3 shows that the correlations do not give a clear direction by indicating both positive and negative correlations in an even frequency. It is noticeable that the normal curve and the direction of the correlations show differences in the two conditions of self-reported arousal (A. momentary vs. B. retrospective). In contrast to the retrospective self-reported intra-individual correlations (B) the momentary self-reported intra-individual correlations (A) only show a positive means

between M=.06 to M=.14 and standard deviations between SD=.21 and SD=.25. Furthermore, retrospective self-reports (B) show a more flat distribution than momentary self-reports (A) and also only negative values of skewness (B1: -.585, B2: -.263, B3: -.015) (see Appendix XI).
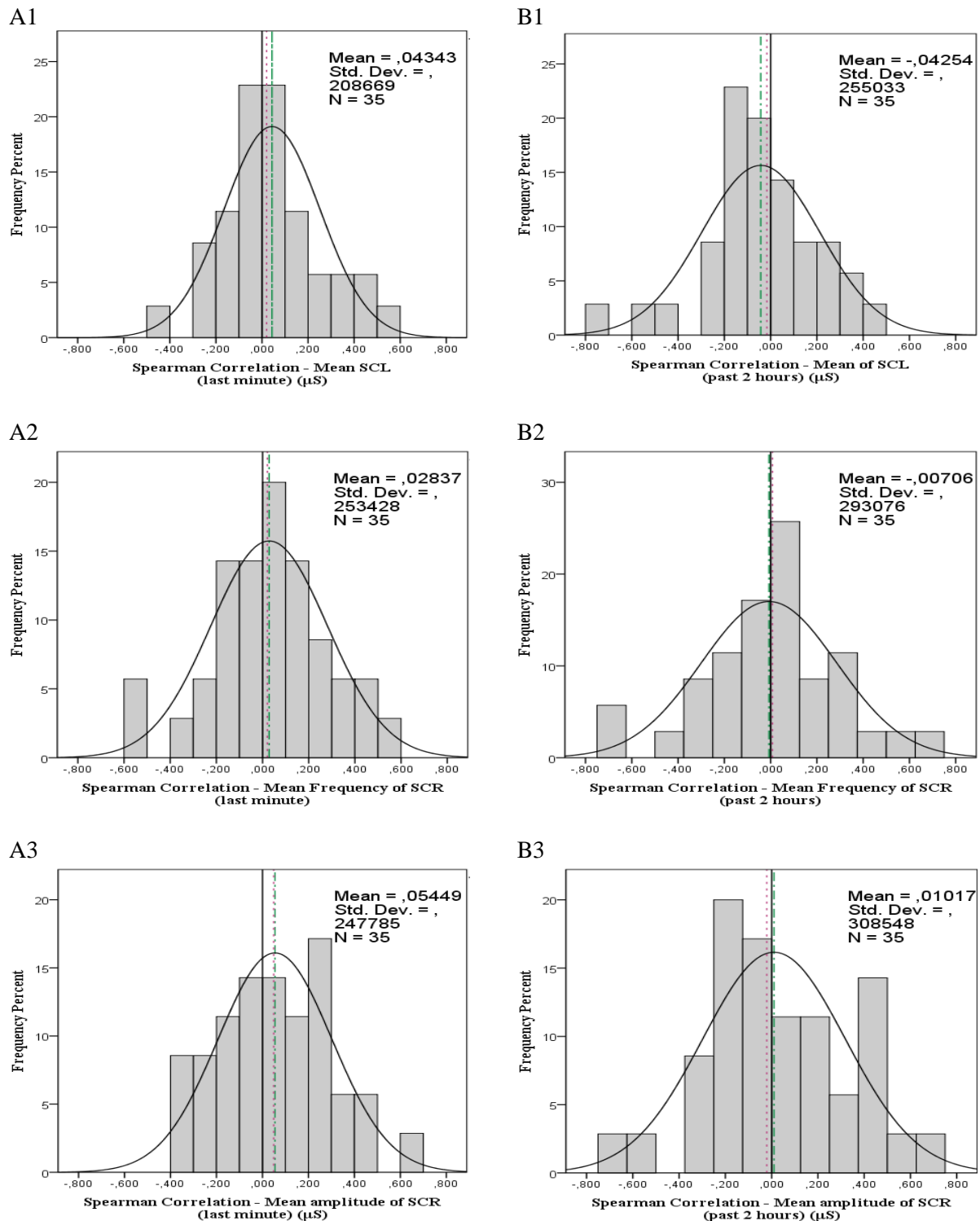

Figure 3. Histograms of the frequency distribution (in percent) of intra-individual correlations between the self-reported level of arousal and all three components of EDA ((1) SCL, (2) SCR-frequency, (3)

SCR-amplitude) separated by the two types of self-reports ((A) momentary, (B) retrospective). The green striped line represents the mean; the purple punctured line represents the median.

In total we found more positive (51.9%; 110 correlations) and 48.1% negative (102 correlations) intra-individual correlations. In this 19 significant positive (Range: r=.373 – r= .635) and ten significant negative correlations (Range: r=-.294 – r=-.764) were found (see Appendix V and VI). Momentary intra-individual correlations show 56% positive and 44% negative values. Significant were ten positive and one negative correlation. Retrospective data showed 49% positive and 51% negative correlations, with nine significant positive and nine significant negative correlations. No significant positive correlation could be found in the data-sample of 2018 (TIIM), but nine out of ten significant negative correlations

Looking at individuals, some participants (3, 15, 19, 20, 25, 31, and 33) showed little some (see participants 7,8,14, 17, 21, 27, 29, and 32) greater differences in the strength of variation between the two time-intervals. When looking at the participants who showed the significant correlations in their data, there is no specific distribution pattern between participants with negative and participants with positive correlations, participants with significant correlations, and those who had no significant correlations or more than one (see Figures 1 and 2).

Discussion

This research focused on the association between emotional- and sympathetic arousal in daily life measured by the three components of Electrodermal Activity (SCL, SCR-frequency & SCR-amplitude). Our results could not support the prior findings that were found in laboratory studies (D'Hondt, 2010; Kosogonov et al. 2017; Melchiori, 1974; Silvert, Delplanque, Verpoort, & Sequeira, 2004; Västfjäll & Gärling, 2007) that could demonstrate the relationship between emotional arousal and sympathetic arousal measured via EDA. Mostly non-significant or very weak significant positive but also negative correlations were found inter-individually. Intra-individually correlations were also mostly non-significant, but significant weak to moderate positive and negative correlations could be found in the data. Intra-individually, no patterns could be found between the participants with significant positive and the participants with significant negative correlations or participants who showed both (significant positive and negative correlations). Furthermore, no significant differences were found between the two time intervals (momentary / retrospective) regarding the strength of the association between emotional and sympathetic arousal. The inter-individually correlations found between SCL, SCR-frequency,

and SCR-amplitude with the self-reported arousal level also reflect the direction of the distribution found in the intra-individual correlations. Only occasionally significant positive but also negative intra-individual correlations were found in all measurements of the EDA components. However, a precise centering of these correlations around the value zero indicates no clear association between the two variables observable in the data. The results indicate no association between sympathetic arousal measured by EDA and the level of emotional arousal in a real-life environment. Besides, no significant difference between momentary and retrospective correlations could be found in this study either.

Theoretical reflection and implications

Contrary to the assumed association between emotional arousal and the three components of EDA, we could find no evidence of such an association in a real-life environment neither between-subjects nor within-subjects like Farrow (2013) suggested. Our results do not show a linear association of the presumed association of emotional arousal with EDA data. Thus, the statement of Russell and Barret that 'subjective feelings of activation are not illusions, but a summary of one's physiological state' (p.806) and their theory of the core affect and its second dimension of arousal, which is believed to be related to physiological arousal, cannot be supported. This only refers to the association investigated here in the time intervals of 1 minute and 2 hours.

Existing laboratory findings suggested a linear regression between our two variables of interest. First, the study of Västfjäll and Gärling (2007) found a highly significant inter-individual correlation of SCR with ratings of emotional activation (r=.88, p<.001). They used 16 pictures of the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2005) to induce emotional arousal. Their used sample was comparable to the one in this study (20 students, ten men, and ten women; mean age was 26.1 years (SD=4.0), and the construction of their dimensional scales used for self-reports. Kosogonov et al. (2017) reported the same positive association of SCR and emotional arousal also induced by the International Affective Picture System. They used 20 pleasant, 20 unpleasant, and 20 neutral pictures, whereby all neutral pictures were indicated to induce lower arousal values than the chosen pleasant and unpleasant pictures. They also studied EDA components separately like we did but found significant associations between neutral and pleasant as well as unpleasant pictures for both SCR-amplitude (T = 17; N = 24; p < .001) and SCR-frequency (T = 0; N = 24; p < .001).

In contrast to our results, the studies mentioned above' measured self-reported arousal level also seem to be generally higher. In the present study, generally weaker arousal values were

indicated below the mean value of '5' of the used scale. This is true for both momentary self-ratings (M=3.10; SD=2.07) and retrospective self-ratings (M=3.63; SD=1.93). In contrast, Västfjäll and Gärling (2007) reported a mean of 1.25 (SD=2.14; the range of scale: -4 to +4). Kosonogov et al., 2017, who reported no direct means or arousal ratings, but based on a figure that was given, the 20 pleasant and 20 unpleasant pictures of the IAPS were averaged rated with M=7.5, and the 20 neutral pictures were averaged rated with M=1.0 (range of scale: 1 to 10). Following Cummins (2014), who pointed out that the individual may more readily perceive subjective experiences that are perceived as more intense, the difference in arousal level between the above studies may influence the relationship between emotional and sympathetic arousal.

Comparing the studies by hand that showed positive correlations between emotional and sympathetic arousal (Västfjäll & Gärling, 2007; Kosonogov et al., 2017; Winton, Putnam, and Krauss, 1984; D'Hondt, 2010), Evers and colleagues' (2014) dual-response framework refers to an explanation for these differences at both the inter-individual and intra-individual level. As already briefly mentioned above, according to Evers et al. (2014), emotions can occur both reflectively and automatically. An automatic emotion is activated by a quick and unconscious reaction to clues from the environment. This automatic activation is associated with physiological signals such as EDA and heartbeat. Reflective is an emotion when it becomes conscious and can be reported. According to Evers et al. (2014), there is no direct correlation between reflective and automatic emotions. According to previous laboratory studies and this paper, the dual-response framework in which these two systems seem to operate can probably provide a rationale for the different results. The self-reported emotional arousal level represents conscious emotions processed by the reflective system following Evers and colleagues' (2014) idea of the dual-response framework. In contrast to that following Evers et al. (2014), the three EDA components' measurements, which did not show any differences in strength of correlations in between, would be influenced by automatic, unconscious emotions rather than reflective, conscious emotions. Conversely, this would indicate that the self-reported core affect (reflective) may not be part of the same system as emotional arousal (automatic), measured physiologically.

We could not find a significant difference in the between-subject correlations between the retrospective and current self-ratings in our data. Robinson and Clore (2002) summarized that currently experienced and reported emotions are more valid because they are less cognitively distorted (peak-end rule, Stone et al., 2000; overestimated stress, McConnell, 2011; overestimated pain, Shiffman et al., 1997) than currently experienced emotions. Our results' classification seems to be limited due to the majority of non-significant inter- and intra-individual correlations between the two different time intervals. However, one could at least assume that

momentary estimates could be more reliable than retrospective in real life. The data shows that all momentary inter-individual correlations are more positive than retrospective ones. Also, on the intra-individual level, this seems to be visible in the correlations. However, this trend seems only marginal here and cannot be statistically proven within our data.

The high variability of intra-individual correlations might indicate that more personal variables and external variables influence the relationship of emotions that is the focus of this study. Here Myrtek, Aschenbrenner, and Brügner (2005) referred to other factors like cognitive schemata and personality dimensions that could be linked more importantly to emotions and emotional arousal. Furthermore, the different directions and strengths of the intra- and inter-individual correlations (mostly not significant) could be found in individuals' mental awareness (Myrtek et al., 2005; Sze, Gyurak, Yuan & Levenson, 2010). In contrast to the laboratory setting, real-life challenges the individual capacity for the individual's interoceptive awareness far more. Because it needs cognitive processing to become aware of inner sensory information, it can be assumed that individuals differ in their ability and the situations in which they have to report their actual level of arousal during a real-life study.

Additionally, the possible reason is that incoming information of all kinds could take more cognitive capacity than participants experienced in prior laboratory studies. These factors limit the chances of enough cognition or awareness needed to become aware of the inner sensory information related to emotions resulting in different results between laboratory and real-life studies in this field. Here, the positive intra-individual correlations found in some of the participants throughout the six conditions in our data might be explained. These participants could show higher awareness of physiological changes in their bodies than others were able at the time of data collection.

Strong and weak points of the study

This study has some strengths that add value to the existing and non-existing body of literature and limitations that need to be considered.

Strengths. Firstly, the actual study is one of the first that studied the relation between emotional arousal and sympathetic arousal measured by EDA and emotional arousal in real-life outside the laboratory. Additionally, it includes next to an inter-individual, an intra-individual study design to provide more specific information about this specific part of arousal's physiological and psychological relation. As a result, more individuals' differences become visible, leading to better generalizability and understanding of this topic and more specific future

research. Besides that, this study obviates misleading assumptions like supposing an unverified 'group-to-individual generalizability' (Fisher et al., 2018; Molenaar & Campbell, 2009).

Secondly, combining the ESM (experience sampling method) with a longitudinal study-design of seven days gave several benefits. On one side, the individuals had time to familiarize themselves with the equipment and self-reports. On the other side, the seven-day period offered the opportunity to minimize the participants' daily interruptions in combination with the choice of the used equipment. On the other hand, we got enough data points throughout the study so that the chances to get the full range of individuals' emotions were increased. The ESM with a fixed interval-based sampling method gives more insight into the actual world people live in, increases the ecological validity by measuring in individuals' natural environment, and further allows greater generalizability of the resulting data.

Thirdly, the three used EDA components' observations combined with two different time intervals of self-reports and inter- and intra-individual analyses offer a wide range of different insights into the topic. It can serve as a basic framework for future research investigating the physiological and psychological relation of arousal in individuals' daily and natural environment. Furthermore, it gives the first points of reference regarding expected correlations, variation, and other interest estimates like the different ranges of real-life emotional arousal within individuals.

Limitations. At first, some limitations that influence the measurement of EDA could be found in the last years. In 2002, Wilson found out that EDA's collected data was also influenced by cognitive activity, measured by pilots during take-offs and landings that required a higher amount of attention. Furthermore, reports of Kallinen and Ravaja (2005) and Stepanski (2003) indicate that EDA is also sensitive to body movements, changes in external electromagnetic fields, humidity, and shifting temperatures that are more common and mostly uncontrollable in non-laboratory studies. In short, EDA is a direct measure of sympathetic nervous system activity and can still be regarded as a convenient measure for cognitive workload and the dynamics of emotional arousal (Stepanski, 2003). It is frequently used and the best choice to measure changes in the SNS induced by physiological arousal, but uncontrollable variables influence variables in open field studies where it is impossible to control external factors that could affect the data. Second, another limitation that is present in this study is the use of two different scales ((a) mQuest, (b) TIIM) that were used in self-reports to measures the subjective level of arousal, which could also influence the data. Here the effect sizes showed different values for the two types of self-reports (a. momentary, b. retrospective). For momentary self-reports a moderate effect of (U($n1$=696, $n2$=412) = 92920.5, p=.000, r=0.30) and for retrospective self-reports a

weak effect of (U(n1=696, n2=412) =127074.0, p=.001, r=0.10) were found (see Appendix IX). The difference in the scales that were found could influence the data we found in this study.

Besides, the EDA data measurement by the E4 sensor does not seem to capture subtle stimuli due to its placement at the wrist. A study by van Lier et al. (2019) suggests that wrist EDA measurements can mainly reliably detect more severe social stressors. As a result, the study participants' subtle changes could not be fully detected and mapped.

Suggestions for further research

By finding more insights into the relationship between emotional arousal and sympathetic arousal in a more realistic environment, it seems essential to include more (contextual) factors. Based on the high variability of intra-individual correlations, external factors that can influence physiological measurements (humidity, physiological activity, and temperature; Reiss, Dürichen & Laerhoven, 2019) and more personal factors like interoceptive capacity of awareness should be considered when studying emotions in real-life (Csikszentmihalyi, 2014; Myrtek et al., 2005). These factors should receive attention to get a more detailed picture of variables that could play a role in the association between emotional and sympathetic arousal found in prior laboratory studies. The ESM research design is such a tool that we recommend it for further research. It is assumed to be suitable to capture the variable nature of individuals' emotional experiences over time in real life between- and within-subjects. Besides, it is also viable to identify situational and personal factors (Myin-Germeys et al., 2018).

Compared to laboratory studies, which usually show more retrospective recall bias by which the answers are often less valid, ESM studies offer the possibility to minimize the time components between the different measurements and measure 'at the moment'. ESM reduces the influence of heuristic biases in retrospective judgments (e.g., the peak-end rule; Fredrickson & Kahneman, 1993). By definition, these measurements' ecological and external validity exceeds conventional investigations performed in artificial environments such as laboratories.

Concerning the exploratory study that took place here, several considerations should be made in subsequent studies. First, the self-reported arousal level shows a distribution towards the middle with very few high and low self-rated arousal values in percentage. An option for event-based sampling could be added to collect data of more extreme arousal levels. Besides, Cummins (2014) pointed out the possibility that subjective experiences that are perceived as more intense become more easily aware by individuals. Here possible stronger associations could be expected and studied in real life, especially when using the E4 that seems more suitable for more intense stimuli (Lier et al., 2019).

Second, based on the high number of non-significant results in this study, especially at the intra-individual level, we recommend a more extended sampling period instead of increasing the number of measurements per day. Extended periods allow the participants to adjust to the sampling devices and processes by the same effort per day and give more insight into the variability in measurements over a more extended period. Based on the review study of ESM studies by Trull and Ebner-Priemer (2020), an average of 5.65 (SD=3.01) measurements per day and an average sampling duration of 12.30 days (SD=10.78) can be used as a benchmark.

Third, the datasets' low usability is caused by problems with the soft- and hardware and a low compliance rate. About 50% of all participants fulfilled the condition of at least 20 complete measurements, whereby 63 measurements were possible within the seven days of sampling. Next to well-functioning and studied technology, preliminary statistical power analyses, best possible prior training of participants, and provided help during the sampling period to reduce data losses, ESM studies should be use percentage rewards. This could be done in the form of pro-rating payments or other resources that keep participants' motivation for compliance high for the period of sampling.

However, due to the growing technical possibilities and the increasing spread of the ESM design in research, methodological and reporting problems often occur. Meanwhile, various studies by van Roekel, Keijsers, and Chung (2019) for adolescents and Trull and Ebner-Priemer (2020) for mental health research offer guidelines and checklists for planning, conducting, and reporting ESM studies. Their suggestions allow for more effective design, replication, comparison, and use in ESM studies' developing field.

Since we cannot control many external influences in real-life studies, we should try to use the possibilities of the increasingly far-reaching and unobtrusive technology in future studies. Factors such as humidity, ambient temperature (Sequeira, 1999), and physical activity (Boucsein, 2012) should also be included in evaluating study design and including sensors such as GPS, thermometers, and humidity sensors. Concerning EDA measurements, the use of several different body positions (palm, wrist, fingers, right and left side of the body) should be considered, if possible, since different measurements were found in some cases.

Conclusion

Finally, this study's results suggest that the connection between emotional arousal and sympathetic arousal in individuals' daily and natural environments does not show the same positive connection in the form of correlations found in different prior laboratory studies. The same applies to the group level. Except for the SCL that showed a very weak significant positive

correlation, prior findings could not be replicated in this study. Furthermore, no statistically significant difference between momentary and retrospective self-reports in the correlations (inter- and intra) between emotional arousal and sympathetic arousal could be found. In summary, if the relation of physiological and psychological of emotional arousal exists outside the lab, this study found no clear evidence for it. At least, concerning the intra-individual findings, we suggest more real-life research, especially with those focusing on various, and in this study, unobserved external and internal variables. These could play a more prominent role in investigating the relation of physiological and psychological emotional arousal in real-life.

References

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, *190*(1), 80-91. doi.org/10.1016/j.jneumeth.2010.04.028

Boucsein, W. (1992). *Electrodermal activity*. doi.org/10.1007/978-1-4757-5093-5

Boucsein, W. (2012). *Electrodermal activity*. doi.org/10.1007/978-1-4614-1126-0

Bradley M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology*, *46*(1), 1–11. doi.org/10.1111/j.1469-8986.2008.00702.x

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, 2, 235-245 Retrieved from https://books.google.de/books?hl=en&lr=&id=A2s963AzymYC&oi=fnd&pg=PA242&dq=Measuring+emotion:#v=onepage&q=Measuring%20emotion

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, *39*(1/4), 106-124. doi.org/10.2307/1415404

Cohen J., (1988). *Statistical power analysis for the behavioral sciences*, 2nd edn, L. Erbaum Press, Hillsdale, NJ, USA.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic medicine*, *74*(4), 327–337. doi.org/10.1097/PSY.0b013e3182546f18

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132-142. doi.org/10.1177/107385840200800209

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience-Sampling Method. *The Journal of nervous and mental disease*, *175*(9), 526–536. doi.org/10.1097/00005053-198709000-00004

Dawson, M. E., Schell, A. M., & Filion, D. L. (2016). The Electrodermal System. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. Cambridge university press. doi:10.1017/9781107415782.010

De Visser, T., Bohlmeijer, E. T., Noordzij, M. L., & Derks, Y. P. M. J. (2017). mHealth in Mental Health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. *International Journal of Human Factors and Ergonomics,* 5(1), 61. doi:10.1504/ijhfe.2017.10009438

D'Hondt, F., Lassonde, M., Collignon, O., Dubarry, A. S., Robert, M., Rigoulot, S., Honoré, J., Lepore, F., & Sequeira, H. (2010). Early brain-body impact of emotional arousal. *Frontiers in human neuroscience*, *4*, 33. doi.org/10.3389/fnhum.2010.00033

Farrow, T. F. D., Johnson, N. K., Hunter, M. D., Barker, A. T., Wilkinson, I. D., & Woodruff, P. W. R. (2013). *Neural correlates of the behavioral-autonomic interaction response to potentially threatening stimuli. Frontiers in Human Neuroscience, 6.* doi:10.3389/fnhum.2012.00349

Farrow, T. F., Johnson, N. K., Hunter, M. D., Barker, A. T., Wilkinson, I. D., & Woodruff, P. W. (2013). Neural correlates of the behavioral-autonomic interaction response to potentially threatening stimuli. *Frontiers in human neuroscience*, *6*, 349. doi.org/10.3389/fnhum.2012.00349

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. doi.org/10.1073/pnas.1711978115

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, *65*(1), 45.

Kahneman, D. (2010, March) Daniel Kahneman: The riddle of experience vs. memory [Video file]. Retrieved from https://www.ted.com/talks/daniel_kahneman_the_riddle_of_experience_vs_memory

Kahneman, D., & Riis, J. (2005). Living, and thinking about it: Two perspectives on life. *The science of well-being*, *1*. doi.org/10.1093/acprof:oso/9780198567523.003.0011

Kahneman, D., & Riis, J. (2005). Living, and thinking about it: two perspectives on life. The Science of Well-Being, 284–305. doi:10.1093/acprof:oso/9780198567523.003.0011

Kallinen, K., & Ravaja, N. (2005). Effects of the rate of computer-mediated speech on emotion-related subjective and physiological responses. *Behaviour & Information Technology*, *24*(5), 365-373. doi.org/10.1080/01449290512331335609

Killgore, W. D. S. (1998). The Affect Grid: a moderately valid, nonspecific measure of pleasure and arousal. *Psychological reports*, *83*(2), 639-642. doi.org/10.2466/pr0.83.6.639-642

Kosonogov, V., De Zorzi, L., Honoré, J., Martínez-Velázquez, E. S., Nandrino, J. L., Martinez-Selva, J. M., & Sequeira, H. (2017). Facial thermal variations: A new marker of emotional arousal. *PloS one*, *12*(9), e0183592. doi.org/10.1371/journal.pone.0183592

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394-421. doi.org/10.1016/j.biopsycho.2010.03.010

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). International Affective Picture System (IAPS): Digitized photographs, instruction manual, and affective ratings (Tech. Rep. No. A-6). Gainesville: University of Florida, Center for Research in Psychophysiology.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261-273. doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lange, C. G., & James, W. (Eds.). (1922). *A series of reprints and translations.The emotions, Vol. 1*. Williams & Wilkins Co. doi.org/10.1037/10735-000

Larkin, K. T. (2006). Psychophysiological assessment. In *Clinician's handbook of adult behavioral assessment*, 165-185. Academic Press. doi.org/10.1016/b978-012343013-7/50009-4

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: implications for understanding and treating health problems. *Journal of personality assessment*, *89*(3), 230–246. doi.org/10.1080/00223890701629698

McConnell, A. R. (2011). The Multiple Self-Aspects Framework: Self-Concept Representation and Its Implications. *Personality and Social Psychology Review*, *15*(1), 3–27. doi.org/10.1177/1088868310371101

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current directions in psychological science*, *18*(2), 112-117. doi.org/10.1111/j.1467-8721.2009.01619.x

Myrtek, M., & Brügner, G. (1996). Perception of emotions in everyday life: studies with patients and normals. *Biological psychology*, *42*(1-2), 147–164. doi.org/10.1016/0301-0511(95)05152-x

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, *17*(3), 715-734. doi.org/10.1017/s0954579405050340

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. Journal of personality and social psychology, 83(1), 198–215. doi.org/10.1037/0022-3514.83.1.198

Russell J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145–172. doi.org/10.1037/0033-295x.110.1.145

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161. doi.org/10.1037/h0077714

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and emotion*, *23*(7), 1259-1283. doi.org/10.1080/02699930902809375

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, *76*(5), 805. doi.org/10.1037/0022-3514.76.5.805

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, *57*(3), 493. doi.org/10.1037/0022-3514.57.3.493

Schachter, S., & Singer, J. E.. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review*, *69*, 379–399. doi.org/10.1037/h0046234

Schmidt, P., Reiss, A., Dürichen, R., & Laerhoven, K. V. (2019). Wearable-Based Affect Recognition-A Review. *Sensors (Basel, Switzerland)*, *19*(19), 4079. doi.org/10.3390/s19194079

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *71*(1), 50–56. doi.org/10.1016/j.ijpsycho.2008.07.009

Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of consulting and clinical psychology*, *65*(2), 292–300. doi.org/10.1037/0022-006x.65.2.292.a

Sifneos P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and psychosomatics*, 22(2), 255–262. doi.org/10.1159/000286529

Silvert, L., Delplanque, S., Bouwalerh, H., Verpoort, C., & Sequeira, H. (2004). Autonomic responding to aversive words without conscious valence discrimination. *International Journal of Psychophysiology*, *53*(2), 135-145. doi.org/10.1016/j.ijpsycho.2004.03.005

Singer, J. A., & Salovey, P. (2010). *Remembered self: emotion and memory in personality*, 10 (6) Retrieved from https://books.google.de/books?id=vzbIVja8_NAC&printsec=frontcover&dq=the+Remembered+self:+emotion+and+memory+in+personality+singer+salovey+++2010sa

Stepanski, E., Lampere, J., Roehrs, T., Zorick, F., Roth, T., Sterman, M. B., ... & Multiband Topographic, E. E. G. 4.1. 5 Electrodermal Activity. *Operator Functional State Assessment*, *81*, 27.

Retrieved from https://pdfs.semanticscholar.org/2f2d/264aedb72aacd5035134d74542ccc9fca2f0.pdf#page =139

Stone, A. A., Broderick, J. E., Kaell, A. T., DelesPaul, P. A., & Porter, L. E. (2000). Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics?. *The journal of pain: official journal of the American Pain Society*, 1(3), 212–217. doi.org/10.1054/jpai.2000.7568

Sze, J. A., Gyurak, A., Yuan, J. W., & Levenson, R. W. (2010). Coherence between emotional experience and physiology: does body awareness training have an impact?. *Emotion (Washington, D.C.)*, *10*(6), 803–814. doi.org/10.1037/a0020146

van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., de Haan, H. A., Vollenbroek-Hutten, M., Schraagen, J. M., & Noordzij, M. L. (2020). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behavior research methods*, *52*(2), 607–629. doi.org/10.3758/s13428-019-01263-9

Västfjäll, D., & Gärling, T. (2007). Validation of a Swedish short self-report measure of core affect. *Scandinavian journal of psychology*, *48*(3), 233–238. doi.org/10.1111/j.1467-9450.2007.00595.x

Winton, W. M., Putnam, L. E., & Krauss, R. M. (1984). *Facial and autonomic manifestations of the dimensional structure of emotion. Journal of Experimental Social Psychology, 20(3), 195–216.* doi:10.1016/0022-1031(84)90047-7

PHYSIOLOGICAL & PSYCHOLOGICAL AFFECTIVE AROUSAL IN DAILY ACTIVITY

Appendix

Appendix I. Instruction E4 Wristband & E4 Manager (SYNC)

Overview

The E4 WRISTBAND enables you to record pulse, activity level, skin conductance and temperature while you go about your daily routine and while you sleep (not in this study).

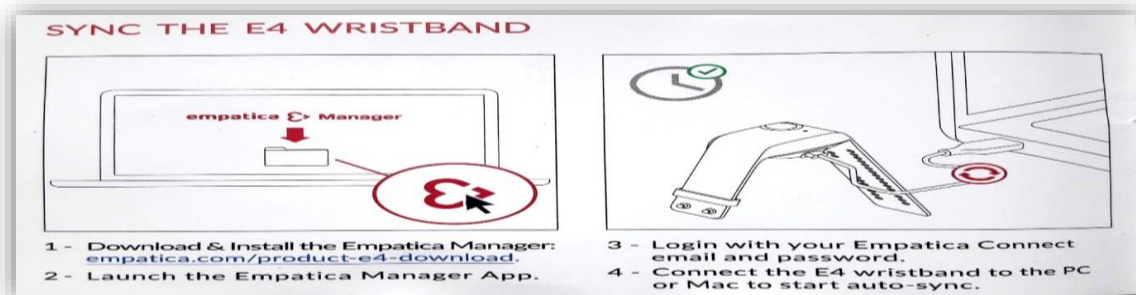It allows you to record sessions and sync those sessions to the web using E4 Manager on PC or Mac.

In combination with one of the researchers this guide will explain how to set up the E4 WRISTBAND.

It includes wearing, syncing for daily recordings, and interacting with the device in case you need to.

What do I have to do?

1. Read and follow instructions **A) –D)**
2. Wear the E4 WRISTBAND 7 whole days (excluding while sleeping)
3. Charge it every night to make sure to have enough power for the following day
4. Try to SYNC your E4 WRISTBAND with the E4 Manager every night after you stopped recording (push the E4 WRISTBAND Button for 2 seconds)
5. Is it recording?
6. At the beginning you should see a red light for 40 seconds (see **C)**)
7. If you want to control it after 40 seconds, you should see at the back of the E4 WRISTBAND a red and green light - if not, push the E4 WRISTBAND button for 2 seconds ('PRESS') (see **D)**)

A)   How to install E4 MANAGER to sync your data to the protected cloud?

SYNC THE E4 WRISTBAND

1 - Download & Install the Empatica Manager: empatica.com/product-e4-download.
2 - Launch the Empatica Manager App.
3 - Login with your Empatica Connect email and password.
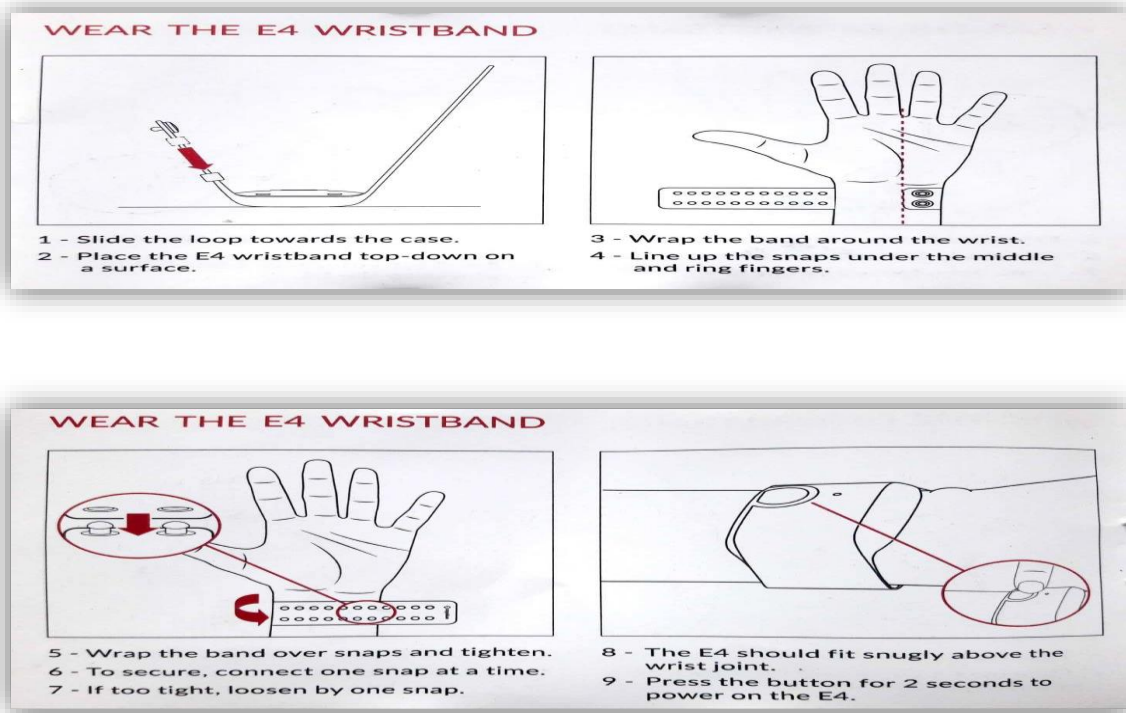4 - Connect the E4 wristband to the PC or Mac to start auto-sync.

NOTICE:

If you have a Windows 10 version – you have to install a new driver before steps 1-4 can be done.
You don't have to push the button to connect the E4 WRISTBAND to the E4 Manager you have installed non-dominant arm (e.g. if you are right-hander – try to wear it on the left).

B)   How to wear the E4 WRISTBAND

WEAR THE E4 WRISTBAND

1 - Slide the loop towards the case.
2 - Place the E4 wristband top-down on a surface.

3 - Wrap the band around the wrist.
4 - Line up the snaps under the middle and ring fingers.

WEAR THE E4 WRISTBAND

5 - Wrap the band over snaps and tighten.
6 - To secure, connect one snap at a time.
7 - If too tight, loosen by one snap.

8 - The E4 should fit snugly above the wrist joint.
9 - Press the button for 2 seconds to power on the E4.

Explanations:

5. Try to wear it on your non-dominant arm (e.g. if you're right-hander – try to wear it on the left)

7. Try to wear it on a way that the 2 metal contact surfaces always have contact with your skin

C) What do the 4 different colors mean to me?

E4 WRISTBAND LED STATUS

🟡 Yellow light indicates the battery is charging (when the E4 wristband is connected to power) or the battery is low (when E4 is worn).

🟢 Green light indicates device start-up.

🔴 Red light indicates recording has started*.

🔵 Blue light indicates data streaming has started*.

*Note: During a session the light powers off after 40 seconds to save power.

Red (recording) and Blue (streaming) lights are displayed for 1 second after a button click when an "event mark" is tagged.

D) What can I do with the button on the E4-Wristband?

Notice:

You will not need the function 'HOLD'!

When you want to start recording, you have to switch the E4 WRISTBAND on (see **D)**; 'PRESS')

You can tag a special event or something that you think is special ('CLICK' - 1s), but you don't have to do this in this study.

HAVE FUN!

Appendix II. Video-link to 'Instruction Video for E4':

https://www.youtube.com/watch?v=UEgIqRyX8Yc&feature=youtu.be

PHYSIOLOGICAL & PSYCHOLOGICAL AFFECTIVE AROUSAL IN DAILY ACTIVITY

Appendix III. Informed Consent

4.11.2018   University of Twente, The Netherlands

Informed Consent

Research: **"Does your smartwatch know how you feel?"**

I hereby declare that I have been informed in a manner that is clear to me about the nature and method of the research as described in the introduction of the research. My questions have been answered to my satisfaction. I agree with my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my given data is to be used in scientific publications or made public in any other manner, then the data will be made completely anonymous. My personal data will not be disclosed to third parties without my consent. If I request further information or help for the research, now or in the future, I may contact the researchers:

Contacts:

If you have any complaints about this research, please direct them to the researchers or the secretary of the Ethics Committee of the Faculty of Behavioural Sciences at the University of Twente, Drs. L. Kamphuis-Blikman P.O. Box 217, 7500 AE Enschede (NL), telephone: +31 (0)53 489 3399; email: l.j.m.blikman@utwente.nl).

By signing this informed consent, I agree to participate in this study as well as give the permission to the researchers to use the gained data. I have been informed, that I can withdraw from the experiment at any time.

Date: _____   **Signature**: _____

Appendix IV. Shapiro-Wilk

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Self-reported arousal for (momentary)	,950	1051	**,000**
Self-reported arousal (retrospective)	,972	1051	**,000**
Mean of SCL (momentary)	,254	1051	**,000**
Mean of SCL (retrospective)	,316	1051	**,000**
Mean amplitude of SCR (momentary)	,206	1051	**,000**
Mean amplitude of of SCR (retrospective)	,463	1051	**,000**
Mean Frequency of SCR (momentary)	,976	1051	**,000**
Mean Frequency of SCR (retrospective)	,961	1051	**,000**

Appendix V. Within-subjects Spearman correlation (momentary)

Within-subjects Spearman correlation of all used EDA components with momentary self-reports of arousal

| Partici-pant | Self-reported level of arousal (momentary) | | | | | | N |
| | SCL | | SCR-amplitude | | SCR-frequency | | |
	Correlation Coefficient	p Sig. (2-tailed)	Correlation Coefficient	p Sig. (2-tailed)	Correlation Coefficient	p Sig. (2-tailed)	
1	**,424***	,031	,244	,230	,300	,136	26
2	,121	,509	-,020	,914	,256	,158	32
3	,096	,622	**,385***	,039	,120	,535	29
4	-,019	,924	,074	,714	-,039	,846	27
5	-,067	,752	-,295	,153	-,393	,052	25
6	-,062	,723	-,046	,795	-,093	,597	35
7	**,529****	,003	,136	,481	,021	,915	29
8	-,011	,955	-,321	,078	-,187	,313	31
9	,029	,885	**,454***	,017	**,554****	,003	27
10	,041	,847	-,311	,130	,003	,989	25
11	,298	,082	,240	,166	**,402***	,017	35
12	-,007	,975	,147	,525	-,009	,967	21
13	,200	,204	,219	,164	,301	,053	42
14	-,092	,690	,049	,833	-,269	,239	21
15	,158	,357	,206	,228	,099	,565	36
16	-,144	,424	-,114	,526	,296	,094	33
17	-,210	,325	-,163	,447	-,106	,623	24
18	,304	,064	,198	,233	,085	,612	38
19	-,009	,961	,281	,126	,124	,508	31
20	**,411***	,027	,361	,054	**,465***	,011	29
21	,385*	,027	,083	,646	-,115	,523	33
22	,092	,685	,275	,216	,169	,451	22
23	-,017	,938	**,605****	,002	**,573****	,004	23
24	,012	,958	,469*	,028	,163	,468	22
25	-,245	,177	-,080	,662	,076	,680	32
26	,057	,761	,021	,910	-,120	,521	31
27	,047	,781	-,249	,137	**-,521****	,001	37
28	-,101	,662	,124	,592	,154	,504	21
29	,123	,539	-,072	,725	,277	,162	27
30	-,422	,064	-,328	,157	-,146	,538	20
31	,020	,898	-,047	,757	-,093	,545	45
32	,157	,355	,001	,995	,010	,951	37
33	-,214	,179	-,199	,212	,027	,869	41
34	-,180	,255	-,144	,365	-,204	,194	42
35	-,184	,338	-,276	,148	-,041	,832	29

Note. *$p < .05$. **$p < .01$; red marked numbers for negative correlations, green marked numbers for positive correlations.

Appendix VI. Within-subjects Spearman correlation (retrospective)

Within-subjects Spearman correlation of all used EDA components with retrospective self-reports of arousal
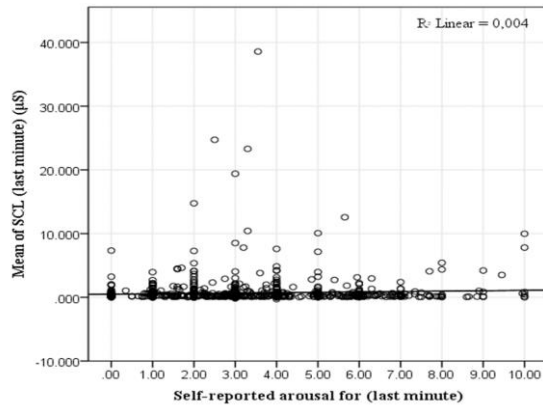
| Partici-pant | Self-reported level of arousal (retrospective) | | | | | | N |
| | SCL | | SCR-amplitude | | SCR-frequency | | |
	Correlation Coefficient	p Sig. (2-tailed)	Correlation Coefficient	p Sig. (2-tailed)	Correlation Coefficient	p Sig. (2-tailed)	
1	,133	,518	-,021	,918	,264	,192	26
2	-,153	,405	-,098	,594	,141	,442	32
3	,390*	,037	,441*	,017	,392*	,035	29
4	,081	,687	-,027	,896	,057	,776	27
5	,309	,133	,547**	,005	,632**	,001	25
6	-,016	,929	,131	,454	,356*	,036	35
7	,211	,272	,373*	,046	,511**	,005	29
8	-,283	,123	-,320	,080	,007	,970	31
9	-,167	,404	,379	,051	,301	,127	27
10	-,221	,289	-,170	,418	,008	,969	25
11	,236	,172	,088	,614	,055	,754	35
12	,112	,629	,142	,539	,181	,432	21
13	,015	,924	,007	,967	,049	,759	42
14	-,488*	,025	-,328	,147	-,168	,466	21
15	-,141	,414	,076	,659	,106	,538	36
16	-,198	,270	-,211	,239	-,031	,864	33
17	-,076	,725	,403	,051	-,069	,750	24
18	,195	,240	,138	,408	-,098	,558	38
19	,049	,792	,259	,159	,358*	,048	31
20	,231	,227	,294	,122	,106	,585	29
21	,028	,875	-,072	,691	-,066	,716	33
22	,481*	,023	,498*	,018	,255	,252	22
23	-,204	,351	-,209	,338	-,362	,089	23
24	-,002	,994	,384	,078	-,035	,877	22
25	-,586**	,000	-,542**	,001	-,666**	,000	35
26	-,001	,994	-,115	,509	,033	,852	35
27	-,054	,740	-,291	,068	-,477**	,002	40
28	-,126	,541	,143	,487	,171	,403	26
29	-,106	,571	-,171	,357	-,216	,244	31
30	-,764**	,000	-,731**	,000	-,196	,382	22
31	-,195	,179	-,245	,090	-,225	,120	49
32	-,028	,858	,007	,966	,095	,549	42
33	-,198	,182	-,105	,484	-,294*	,045	47
34	,060	,700	-,231	,131	-,334*	,027	44
35	-,013	,941	-,241	,162	-,064	,714	35

Note. *p < .05. **p < .01; red marked numbers for negative correlations, green marked numbers for positive correlations.
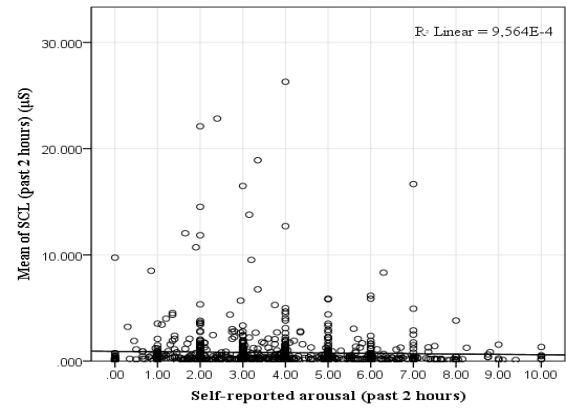
Appendix VII. Scatter plots of intra-individual correlations per condition
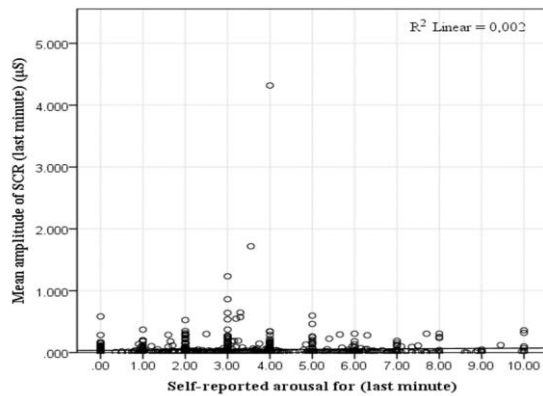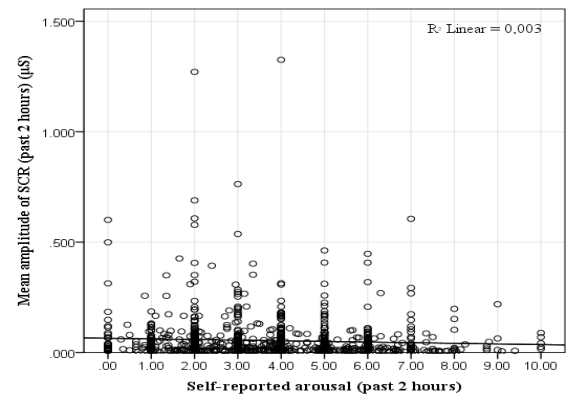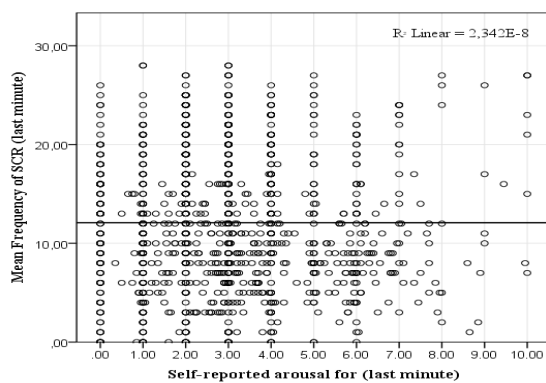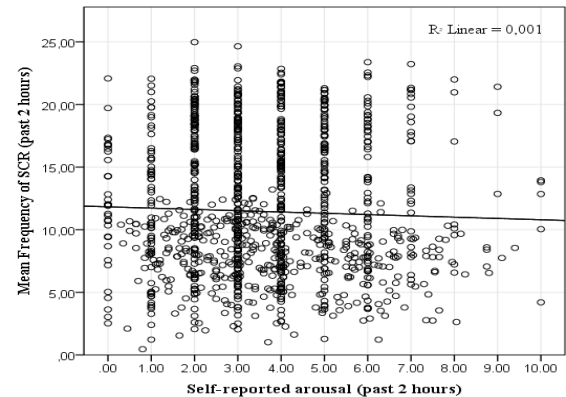
A1

B1

A2

B2

A3

B3

Scatter plots of all intra-individual correlations between all three components of EDA (1. SCL, 2. SCR-amplitude, 3. SCR-frequency) and the self-reported level of arousal - separated by the two different self-reports (A.. momentary, B. retrospective)

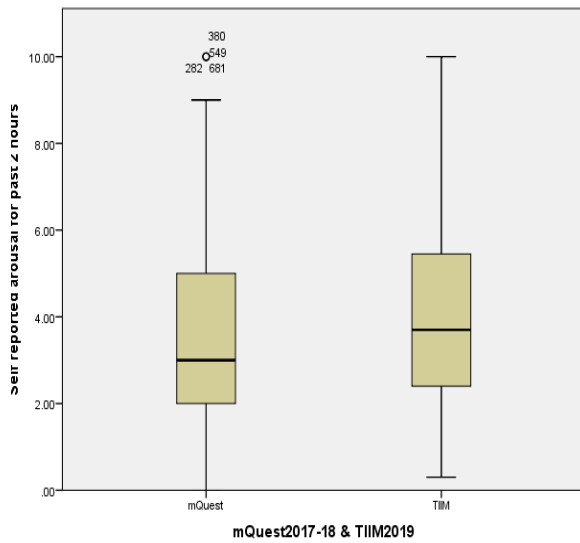*Appendix VIII. Boxplots of two-time-intervals (mQuest; TIIM)*

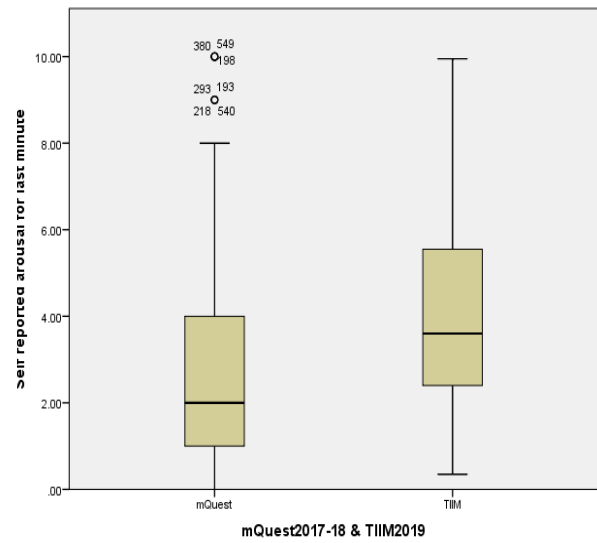*Figure 7.* Boxplots showing retrospective self-reports of arousal separated per group (mQuest ;TIIM)

Figure 8. Boxplots showing momentary self-reports of arousal separated per group (mQuest ;TIIM)

Appendix IX. Mann Whitney-U test (mQuest ;TIIM)

Mann Whitney test

Table X

Result of the Mann-Whitney U Test for momentary self-reported arousal and retrospective self-reported arousal

Variable	Mean rank		U	Z	*p*	r
	mQuest	**TIIM**				
Momentary self-reports	482,01	676,96	92920,5	-9,837	.000	.30
Retrospective self-reports	531,08	594,07	127074	-3,178	.001	.10

Note. p > 0.05

Result (further calculated by hand):

Momentary self-reports: r= 0.2955 (moderate effect size)

Retrospective self-reports: r= 0.0955 (weak effect size)

Appendix X. Fisher's Z comparison calculation (Excel)

For inter-individual correlations of self-reported level of arousal and SCL in momentary self-reports vs. retrospective self- reports:

H0: $\rho1 = \rho2$

Ha: $\rho1 > \rho2$

r'1 = FISHER(r1) = FISHER(.077) = . 077

r'2 = FISHER(r2) = FISHER(-.029) = -.029

s = SQRT(1/(n1 − 3) + 1/(n2 − 3)) = SQRT(1/32 + 1/32) = .245

z = (r'1 − r'2)/s = (.077− -.029) / .245 = .433

p-value = 1 − NORMSDIST(z) = 1 − NORM.DIST(0.433; 0; 1; TRUE) = 0.323

p-value = .32 > .05 = α

For inter-individual correlations of self-reported level of arousal and SCR-Frequency in momentary self-reports vs. retrospective self- reports:

H0: $\rho3 = \rho4$

Ha: $\rho3 > \rho4$

r'3 = FISHER(r3) = FISHER(-.016) = -.016

r'4 = FISHER(r4) = FISHER(-.044) = -.044

s = SQRT(1/(n1 − 3) + 1/(n2 − 3)) = SQRT(1/32 + 1/32) = .245

z = (r'3 − r'4)/s = (-.016− -.044) / .245 = .144

p-value = 1 − NORMSDIST(z) = 1 − NORM.DIST(0.144; 0; 1; TRUE) = 0.454

p-value = .45 > .05 = α

For inter-individual correlations of self-reported level of arousal and SCR-Amplitude in momentary self-reports vs. retrospective self- reports:

H0: $\rho5 = \rho6$

Ha: $\rho5 > \rho6$

r'5 = FISHER(r5) = FISHER(-.51) = -.563

r'6 = FISHER(r6) = FISHER(-.39) = -.412

s = SQRT(1/(n1 − 3) + 1/(n2 − 3)) = SQRT(1/32 + 1/32) = .245

z = (r'5 − r'6)/s = (-.563 − -.412) / .245 = .408

p-value = 1 − NORMSDIST(z) = 1 − NORM.DIST(0.408; 0; 1; TRUE) = 0.342

p-value = .34 > .05 = α

Appendix XI. Descriptive statistics of intra-individual Spearman's correlation

Descriptive statistics of intra-individual Spearman's correlation coefficients of all three EDA components with both momentary and retrospective self-reported larousal level

Variables correlated with self-reported level of arousal	M	SD	Range		Skewness		N
			Min	Max	Statistic	Std. Error	
SCL (momentary) (μS)	,043	,208	-,422	,529	,360	,398	35
SCR-Amplitude (momentary) (μS)	,055	,247	-,328	,605	,213	,398	35
SCR-Frequency (momentary)	,028	,253	-,573	,554	-,284	,398	35
SCL (retrospective) (μS)	-,043	,255	-,764	,481	-,585	,398	35
SCR-Amplitude (retrospective) (μS)	,010	,308	-,731	,632	-,015	,398	35
SCR-Frequency (retrospective)	-,007	,293	-,666	,632	-,263	,398	35