Psychophysiology in the digital age

van der Mee, Denise Johanna

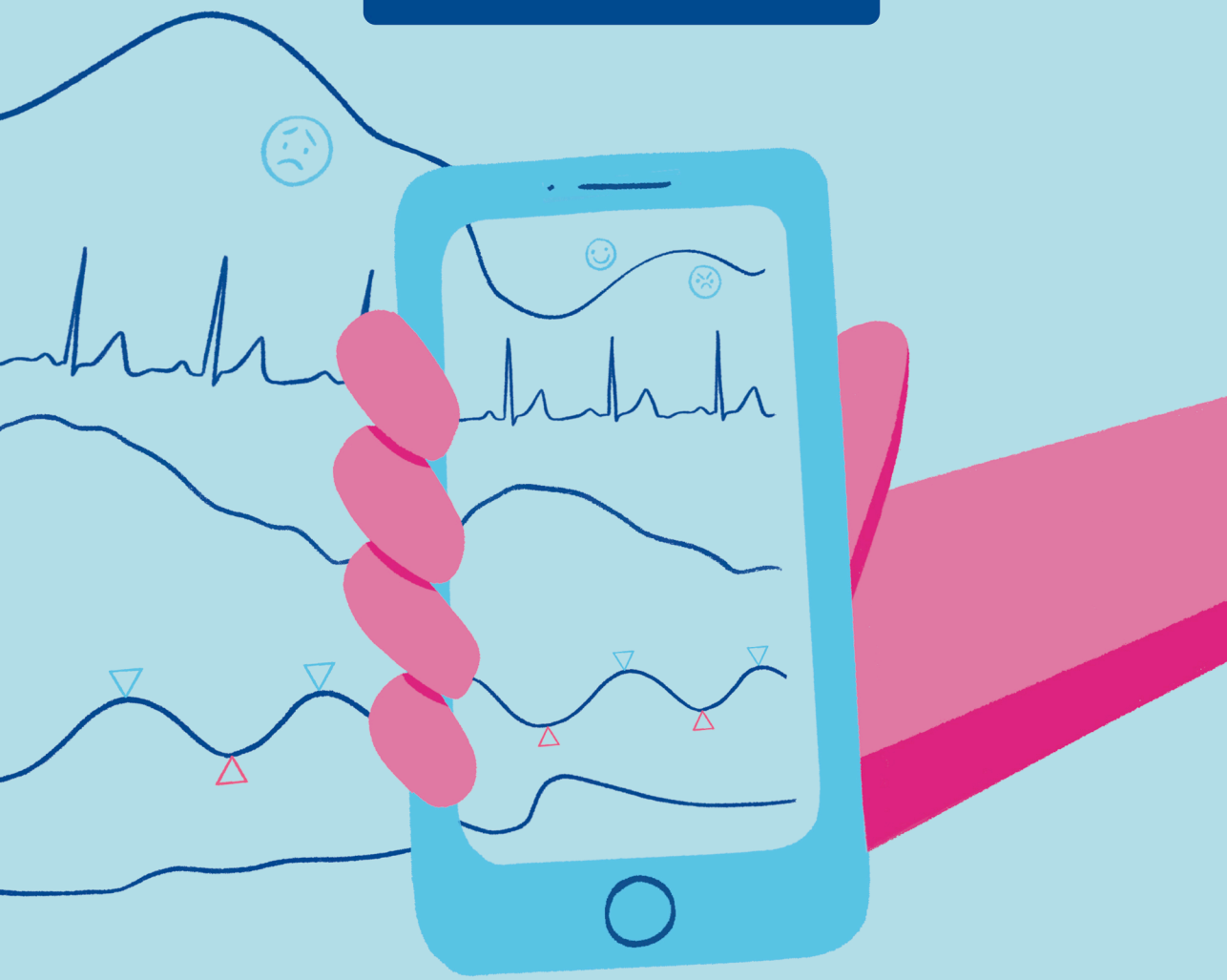
2023

DOI (link to publisher)
10.5463/thesis.258

document version
Publisher's PDF, also known as Version of record

Psychophysiology

in the Digital Age



Denise J. van der Mee - Hoek

Psychophysiology in the digital age

Denise Johanna van der Mee

Colofon

Provided by thesis specialist Ridderprint, ridderprint.nl

Printing Ridderprint

Layout & design Indah Hijmans, persoonlijkproefschrift.nl

Cover design Indah Hijmans

ISBN 978-94-6483-267-9

VRIJE UNIVERSITEIT

PSYCHOPHYSIOLOGY IN THE DIGITAL AGE

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op 12 september 2023 om 9.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Denise Johanna van der Mee

geboren te Ter Aar

promotor:	prof.dr. J.C.N. de Geus
copromotoren:	dr. M.J. Gevonden dr. J.H.D.M. Westerink
promotiecommissie:	prof.dr. A.M. Brouwer dr.ir. F. Lamers dr. L.M.C. Nauta-Jansen prof.dr. M.L. Noordzij prof.dr. A.C. Krabbendam

TABLE OF CONTENTS

CHAPTER 1

The current state of psychophysiological research

An introduction to my thesis

INTRODUCTION

Stress-related disorders are a leading cause of disability worldwide and major contributors to the overall global burden of disease (Ferrari, et al., 2013; Liu, et al., 2020; Ahola & Hakanen, 2014). Currently we are unable to predict when, why, and how people develop stress-related disorders, reducing our ability to prevent them. Decades of research have focused on disturbances in physiological and affect regulation as core processes underlying vulnerability for stress-related disorders. It is long believed by scholars that changes in physiology occurring in coherence with affective states aid the body to respond rapidly to prospected or imminent threats. This hypothesis was first postulated in the work of Charles Darwin and fueled the development of various emotion theories such as Angelo Mosso's visceral physiology of emotion (Mosso, 1881,1884), the James-Lange specificity theory of emotion (James, 1884; Lange, 1994), the Cannon–Bard thalamic theory of emotions (Bard, 1934a; Bard 1934b; Cannon, 1927), and Carroll Izard's differential emotion theory (Izard & Tomkins, 1971). These different theories have different hypotheses on how emotional states should be quantified and what markers distinguish one state from another. Nevertheless, they all come back to Darwin's original hypothesis: each affective state is associated with a differential physiological response to provide an optimal response (from a survival perspective) to the situation that gives rise to them. This viewpoint is still held by the majority of current day researchers.

Large individual differences exist in affective and physiological responsivity to comparable situations. The dominant approach to unravel how such affective and physiological responses are moderated by individual characteristics (i.e., genetics, childhood trauma) is to expose individuals to a variety of stressors in a controlled laboratory setting (Ellis et al., 2006; de Geus, et al., 2015; Matthews et al., 2000; Sapolsky, 1994). The benefits of a laboratory setting are that the influence of various confounders (such as changes in posture, physical activity, or psychosocial factors) are under experimental control. This enables researchers to study the effect of stressors on affective and autonomic nervous system (ANS) responsivity with minimal error. A drawback is that the type of stressors applied in the laboratory may translate poorly to daily life experiences, hence their ecological validity is debated (Holleman et al., 2020; Kihlstrom, 2021; Orne & Holland, 1968; Plaza, Delarue & Saulais, 2009; Schmuckler, 2001). The advent of the digital age has provided us with new tools to overcome the postulated poor ecological validity of laboratory stressors. Technological advances have opened the doors for ecological momentary assessment (EMA) with smartphones and ambulatory physiological measurement with wearable devices such as wristbands. With EMA, affective state and key

contextual factors (activity currently engaged in, bodily posture, social environment) can be assessed multiple times a day with the use of digital questionnaires. The development of wearable physiological measurement devices provides the means to measure ANS activity directly in daily life under varying conditions. When combining EMA with such wearable devices, the relationship between affect and physiology can be studied directly in daily life settings. The benefit of this approach is that stress can be studied when it occurs in real life, providing excellent ecological validity.

Yet, daily life monitoring is not a holy grail either. It comes with several drawbacks compared to a laboratory design, such as an increased measurement error due to confounders and the unpredictability of the event of interest. Furthermore, the newly developed devices apply a different technique to measure ANS activity as compared to the devices used in a laboratory setting. This complicates a direct comparison of affect-ANS dynamics observed in daily life to that of the laboratory. Due to these issues the psychophysiological research community faces a dilemma regarding the best design for studying affect-ANS dynamics. Should they go for a highly controlled but potentially less ecological valid setting, or an excellent ecological valid but noisy setting? And should they use well validated techniques that limit study duration to a couple of days or less validated techniques that enable a study duration of multiple years? In my thesis I seek to provide answers to solve this dilemma. To this end I set up the following research questions: 1) Do wearable devices have sufficient validity to capture ANS activity? 2) To what extent is the laboratory design ecologically valid to measure affect-ANS dynamics? 3) Are the affect-ANS dynamics subject to individual differences? These three questions form the common thread of the research I performed for my thesis. It is not my aim to answer these questions completely and conclusively, for this my PhD trajectory was far too short. Nevertheless, I hope that my findings will support my fellow psychophysiological researchers in selecting an optimal design for their psychophysiological studies, taking into account the relative strengths and limitations of the laboratory and daily life designs.

Measuring physiology

One of the most used methods to study physiological activity is by indexing the activity of the ANS. The ANS is considered the primary mechanism in control of the fight-or-flight response and consists of two branches, the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) (Buijs, 2013). Activation of the SNS leads to an increased neural outflow to many target organs and the secretion of epinephrine and norepinephrine from the adrenal medulla. The general effect of this SNS activation is to increase the transportation and tissue exchange capacity of oxygen and carbon dioxide

and the availability of energy substrates to enable active behavior to deal with a threat (Martini, Nath, & Bartholomew, 2012). Activation of the PNS signals the body to focus on recovery and digestion. It signals the digestive tract to increase its rate of digestion, the pancreas to release insulin, lowers the breathing rate, and relaxes the muscles involved in waste removal (Martini, Nath, & Bartholomew, 2012).

The main benefit of studying ANS activity is that this can be done non-intrusively, continuously, and with little participant burden. The oldest known index of ANS activity is heart rate (HR). Each heartbeat is initiated by the sinoatrial (SA) node, also referred to as the pacemaker of the heart. The SA node has an intrinsic rate between 60 and 100 beats per minute, depending on an individual's age and gender (Kashou, Basit, & Chhabra, 2017). Resting state HR is usually lower than this intrinsic rate due to tonic parasympathetic influences on the SA node. When PNS activity is high it acts as a brake on the intrinsic rate, while this brake almost disappears when PNS activity is low (Porges, 1995). Increased SNS activity leads to an increase in HR while increased PNS activity leads to a decrease in HR (Kashou, Basit, & Chhabra, 2020; Smith et al., 2017). Therefore, the rate of the heart at any given moment in time is determined by the interplay between the SNS and PNS on the SA node. Due to the mixed effects of both SNS and PNS activity on the resulting HR it can be used as an index of arousal, defined as the level of alertness or activation on a continuum ranging from extreme drowsiness to extreme wakefulness (Duffy, 1962). Higher levels of arousal are associated with increased SNS and decreased PNS activity. However, the relative contribution of the separate ANS branches cannot be derived from HR alone. Furthermore, the general belief that when a situation is perceived as threatening the SNS gets activated and when the stressful situation is averted the PNS gets activated (Buijs, 2013) is likely more nuanced. In their autonomic space model on cardiac physiology, Berntson, Cacioppo and Quigley (1993) discuss that the dual innervation of organs by the SNS and PNS does not happen along a single axis continuum. Rather, the SNS and PNS both independently vary along their own continuum. Activity of dually innervated target organs, such as the heart, is thus the result of a myriad of bivariate SNS and PNS activity combinations. Luckily, better knowledge of ANS anatomy has led to the discovery of other measures that can index the relative contributions of both branches, which will be discussed in the following two sections.

Measures that capture PNS activity

A frequently used measure of PNS activity is heart rate variability (HRV). HRV represents the variability in the time interval between successive heartbeats, called the inter-beat interval (IBI). The IBI, also referred to as R-R interval, is the time difference in milliseconds

(msec) between two successive electrocardiogram (ECG) R-peaks. Due to the fast temporal kinetics of the parasympathetic signaling (<1s) at the SA node, opposed to the much slower (>3s) sympathetic signaling (Chapleau & Sabharwal, 2011), changes in PNS activity affect the heart rhythm on a beat-to-beat scale while SNS changes do not. Therefore, relatively fast changes in HRV can be attributed to changes in PNS activity. HRV can be quantified in both time domain and frequency domain to assess PNS activity. The most frequently used time domain measures are respiratory- sinus-arrhythmia (RSA)(Berntson et al., 1997) and root-mean-squared successive differences (RMSSD) (Thayer, Hansen & Johnsen, 2010). RSA and RMSSD differ from one another in that the former more directly takes into account the modulation of the heartbeat in response to respiration, a process called respiratory gating (Yasuma & Hayano, 2004). The most often used frequency domain measure quantifies HRV as the spectral power in the higher (0.15 – 0.4 Hz) frequency band. This high-frequency HRV (HF-HRV) also captures respiratory gating (Berntson et al., 1997).

Measures that capture SNS activity

To assess the counterpart of the PNS, the much slower SNS, two measures are used frequently. The first measure of SNS is the cardiac pre-ejection period (PEP)(Sherwood, et al., 1990; Kelsey, 2012). PEP is an indirect measure of the strength with which the myocardium of the ventricles of the heart contracts. The strength of the contractions is under control of SNS activity solely. The PEP is the systolic time interval between the start of ventricular depolarization (Qonset) in the ECG and the time the aortic valve opens (B-point) in the impedance cardiogram (ICG), measured in msec. Shorter time periods reflect stronger contractions of the myocardium and thus more SNS activity (Newlin & Levenson, 1979; Turner, Sherwood & Light, 1991).

The second source of SNS activity indicators is the tonic and phasic activity of the sweat glands. The innervation of the sweat glands is entirely through sympathetic nerves and sweat gland activity is therefore considered one of the purest measures of SNS activity (Critchley, 2002). Activation of the SNS results in both an increase in the total number of activated sweat glands and in more secretion of sweat by the eccrine sweat glands. Both these changes in sweat gland activity in turn lead to changes in the conductance of electrical activity through the skin, also denoted as electrodermal activity (EDA). From EDA, measures of SNS activity can be derived such as skin conductance level (SCL) and the frequency of non-specific skin conductance responses (ns.SCRs). SCL is calculated in μS as the mean level during a certain time period compared to a baseline, a higher level reflects more SNS Activity (Bach, Friston & Dolan, 2010). Ns.SCRs are skin conductance responses not studied in association with a specific external stimulus. The

ns.SCR frequency is measured as the number of peaks per minute, more peaks reflect more SNS activity (Bach, Friston & Dolan, 2010).

Measuring affect

To date, the most frequently used way to measure affective state is to ask participants to rate their past or current emotions with the use of questionnaires. These affect questionnaires, such as the positive and negative affect (PANAS) scale (Watson, Clark & Tellegen, 1988) or the Maastricht scale for momentary mood state assessment (3MQ) (Schneiders, et al., 2006), are based on the circumplex model of emotions put forward by Russell in 1980 (Russell, 1980; Posner, Russell, & Peterson, 2005). In this model a distinction is made between four affective states: high arousal positive affect, low arousal positive affect, high arousal negative affect, and low arousal negative affect (Figure 1). Based on the various emotion theories these four states should be differentially related to ANS activity.

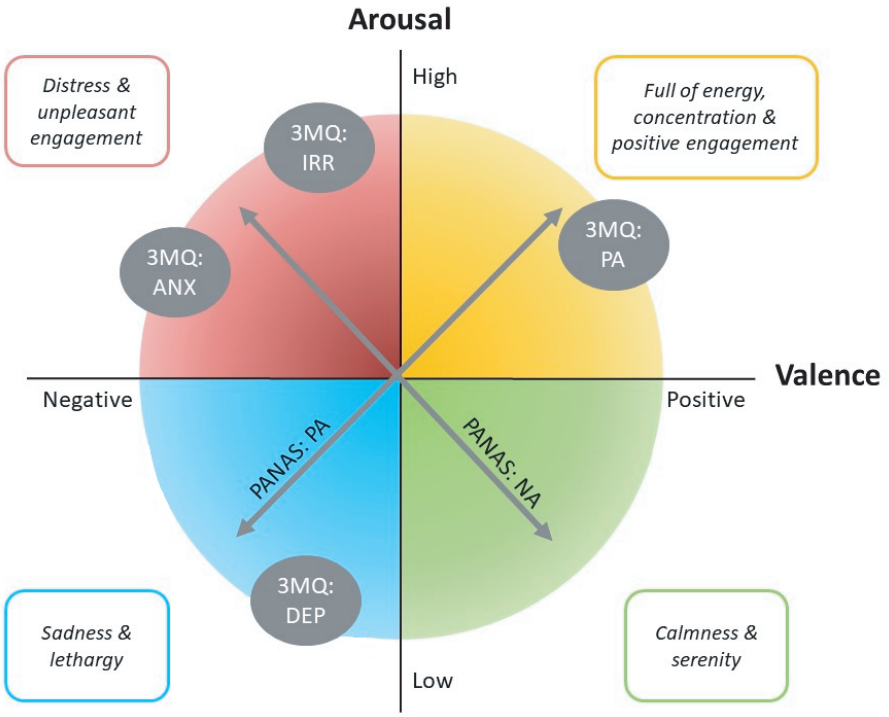


Figure 1. A visual representation of the affect quadrants.

Note: ***PANAS*** *items positive affect (PA): attentive, interested, alert, excited, enthusiastic, inspired, proud, determined, strong, and active; items negative affect (NA): distressed, upset, hostile, irritable, scared, afraid, ashamed, guilty, nervous, and jittery.* ***3MQ*** *items positive affect (PA): happy, enthusiastic, good-humored, satisfied, self-assured, and cheerful; items anxious (ANX): nervous, anxious, and tense; items depressed (DEP): sad, lonely, and listless; items irritated (IRR): irritated.*

The relationship between ANS activity and affect

The laboratory

Measuring human behavior with the use of laboratory protocols has long been the default in psychological research. In general, studying behavioral constructs in a laboratory has several benefits. First, the laboratory provides high control over measurement quality and context, thereby reducing noise due to measurement error and/or many confounding variables. This is especially important when studying a variable, or relationship between variables, that is sensitive to factors such as posture, movement, temperature, and the broader social context. Second, in the laboratory it is possible to design tasks in such a way that they selectively isolate a psychological construct of interest, for example a go/no-go task to measure inhibitory control. Third, as laboratory stressors follow a protocol, they can be easily replicated across multiple individuals. This allows us to compare relations within and between individuals in a setting that is similar across all participants. Fourth, it is possible to continuously check whether measurement devices are still working in order to avoid poor data quality or a complete loss of data. Last, a single short laboratory assessment takes little time and has a low participant burden. However, since the 1960's the ecological validity of the laboratory paradigm has been questioned, specifically whether the evoked physiological and psychological responses to stress under "artificial laboratory circumstances" reflect how one would respond to naturally occurring stressors (Orne & Holland, 1968; Schmuckler, 2001).

Ecological validity in the context of this thesis refers to whether the relationship between physiology and affect as measured in the laboratory can be directly translated to daily life. Two factors inherent to the laboratory paradigm might hamper such translations. First, laboratory tasks are often designed to measure the effect of a clear and single psychological state on ANS activity and affect, such as mental effort (Stroop test, mental arithmetic, N-back), physiological stress (cold pressor test, bicycle ergometer, treadmill), social evaluative stress (Trier Social Stress Test, Sing A Song Stress Test), or emotion (movie clips, affective pictures). Such tasks may translate poorly to the types of events individuals encounter in daily life because these events consist of multiple different psychological components that are present for longer periods of time. For example, when working towards an important work deadline people do not only experience mental effort but also social evaluative stress. Second, laboratory tasks have little relevance and/or importance to the individuals. Good or bad performance on a laboratory stress task has no consequence for their daily lives, while missing a work deadline may cost them a promotion or get them fired.

In addition, laboratory studies focus mostly on the relation of physiology with high arousal negative affect, while much less is known on the relationship with the other three affect quadrants. In daily life it is likely that periods within each affect quadrant alternate rapidly or might even co-occur, for example when watching an exciting football game. Of further relevance is that laboratory studies focus mainly on averaged differences in ANS reactivity to experimentally manipulated affect. In 2010 Kreibig performed a systematic review on 134 of such studies where affective states were manipulated with the use of emotion inducing imagery (like film clips, personal recall of events, and picture viewing). The average ANS activity during conditions in which affective states were induced were compared to each other or to a neutral condition. Such comparisons revealed a different pattern of ANS responsivity to such distinct affective states. For example ANS activity was different during induced anger states versus neutral conditions. Though most of these studies make use of a repeated measures design, the ANS response of an individual across all conditions that encompass the same affect state are aggregated. Such methods provide valuable insight for theories on the ANS response to different emotions, but they come at the cost of losing information on nuances of the within-person relationship between affect and ANS activity.

Daily life

Most of the concerns with the ecological validity of laboratory studies can be solved by measuring the construct of interest in daily life through EMA. Even before the advent of the digital age, several researchers took the plunge and studied this direct relationship in daily life with the use of paper and pencil questionnaires combined with a pager (Sloan, 1994; Hawkley 2003; Bacon, 2004; Campbell, 2006; Pollard, 2007), but a steep increase of such studies is seen since digital questionnaires became easily available (Shiffman, Stone & Hufford, 2008). In the current day EMA studies, digital devices such as smartphones are used to elicit and obtain questionnaire data multiple times a day. These so-called prompts are either random or scheduled around an event of interest. By combining EMA studies with wearables that measure ANS activity, relevant events can be captured at the moments they occur with relatively little subject burden. Furthermore, in EMA studies within-subject relationships and their variation across participants are central. By applying multilevel models, a unique relationship can be estimated for each individual encompassing the whole range of affective states that occur. Though it sounds enticing to just completely move to daily life studies only, this approach also comes with its limitations. First, the randomness of the occurrence of the event of interest in a daily life setting comes with the cost of increased risk of missing out on relevant events unless

participants are followed over very long periods of time. A second limiting factor is the lack of experimental control over various confounders. In daily life studies participants are free to lead their life as they would usually do, thereby increasing the diversity of activities engaged in that can influence ANS in different ways. Although there are ways to capture a variety of interesting covariates (social environment, location, travel, and movement pattern) with the use of passive sensing (for example, tracking their GPS signal to infer location travel pattern) some information can only be obtained by asking the participants explicitly through questionnaires, such as the perceived affiliation with the other person during social interactions. This leads to the third and most important limitation of the EMA approach, its dependency on subjective reporting (van Genugten, et al., 2020; Stone et al., 2007). To obtain a detailed description of a participant's day, whereabouts, and feelings extensive questionnaires are prompted several times a day. Though this increases the capture of context dynamics and reduces recall bias, it increases the participant burden, especially so when the aim is to measure over long periods of time (i.e., multiple weeks or even months).

Outline of my thesis

Researchers are well aware of the limitations and benefits of both the laboratory and daily life design. In the laboratory, researchers are expanding on the experimental manipulations currently applied. By testing various forms of stress reactivity (for example mental versus social stress) within- and between-individual differences can be studied better. For one specific type of stress, social-evaluative stress, there has been one dominant stress paradigm for many years, namely the Trier-Social-Stress Test (TSST). This test is quite labor intensive and requires the presence of multiple confederates, making it difficult to include in large scale experiments. In 2014, Brouwer and Hogervorst developed a simpler paradigm to measure social-evaluative stress, the Sing-a-Song Stress Test (SSST; Brouwer & Hogervorst, 2014). However, their paradigm was still quite long and still included multiple confederates. In **chapter 2** of my thesis, I validate a shortened version of this new stress paradigm: the short Sing-a-Song Stress Test ($SSST_{short}$). In a subsample of my study population, we compared the ANS and affective stress reactivity of the $SSST_{short}$ to a speeded reaction time task.

In daily life, effort is being undertaken to capture the multitude of measured and unmeasured covariates. The digital age offers various opportunities to measure a variety of such covariates through passive sensing. Commercially available smart watches (that are already in use by a substantial portion of society) include GPS sensors, accelerometers, and physiological signals. Information from these sensors can be used to

gain knowledge on an individual's day-to-day lives without increasing burden. For example, a study by Tobias and colleagues (2016) has shown that it is feasible to add diary triggers to an EMA study based on GPS derived data such as location and population density. Their method increased the diversity of the context in which affect was measured as opposed to triggering according to a time schedule. However, smart watches are not always the best solution for the recording of physiological signals. The majority of daily life studies thus far have relied on more burdensome wearable devices, such as Holter monitors and chest belt sensors, which make use of different techniques than those that are available in smart watches. With these devices, high quality cardiac ANS activity is obtained with electrocardiography (ECG) using chest electrodes, and EDA with wet electrodes on the hand or fingers. In contrast, smart watches typically make use of photoplethysmography (PPG) to assess cardiac ANS activity and dry electrodes to assess EDA. A few studies have evaluated these wristband wearable technologies compared to the laboratory golden standard (e. a. Milstein & Gordon, 2020; Schuurmans, et al., 2020; Xie, et al., 2018) and they unanimously show imperfect correspondence. The exact impact of this imperfect correspondence on the predictive validity of ANS recording and its association with affect needs to be assessed in both the laboratory and daily life. Therefore, in **chapter 3**, I tested the validity of a new technology to measure EDA on the top of the wrist. Over a hundred healthy young adults participated in a laboratory study including posture manipulation, stressors (mental, social, and physical), and daily life activities. Validity of the wristwatch EDA signal was assessed by exploring the correspondence, construct, criterion, and predictive validity against the golden standard on the palm of the hand and the cardiac PEP measures.

An additional issue that plagues the study of the physiology-affect interplay in daily life is the striking lack of studies that directly compare the affect-physiology relationship between the laboratory and daily life. Only a few laboratory studies have investigated the congruence of changes in physiology and affect in response to induced stress directly (Feldman et al., 1999). However, the approach taken by these studies are that of a classical laboratory design, which limits the direct comparison of these laboratory studies to existing daily life literature. It is of importance to know whether the affect-physiology dynamics in the laboratory differ from those in daily life. If they are similar, it can be concluded that the laboratory paradigms are ecologically valid and that their results on within-person relationships can be generalized to daily life. In **chapter 4**, I performed the first study that directly compares the within-person relationship of physiology with affect in the laboratory to that same relationship in daily life. In around a hundred individuals the

relationship of various ANS measures with valence and arousal was assessed in both the laboratory and a 24-hour daily life setting.

Lastly, it is well known from both laboratory and daily life studies that there are substantial individual differences in the strength of the ANS and affective response to stressors. According to the stress reactivity theory, such differences are considered a stable trait-like factor that is associated with various characteristics of the person such as personality (Boyce & Ellis, 2005) and lifestyle. In **chapter 5**, I explore aerobic fitness and regular physical activity as examples of individual differences that can modulate the physiological and affective response to stress exposure. Specifically, I test the so called cross-stressor adaptation hypothesis that individual differences in fitness are associated with differences in cardiovascular and/or affective response to stress. In over a hundred participants we investigated whether aerobic fitness and moderate-to-vigorous physical activity (MVPA) was associated with the strength of the ANS and affective response to stressors in both a standardized laboratory and a daily life setting.

In **chapter 6**, I summarize the findings of my studies. In **chapter 7**, I discuss how my research has contributed to answering the three research questions that are central in my thesis. I embed my findings in the work of other researchers to provide a more complete picture. I end my thesis by providing suggestions on how to improve psychophysiological research and by discussing the impact this research field can have on society.

CHAPTER 2

The short Sing-a-Song Stress Test: a practical and valid test of autonomic responses induced by social-evaluative stress

van der Mee, D. J., Duivestein, Q., Gevonden, M. J., Westerink, J. H. D. M., & de Geus, E. J. C. (2020). The short Sing-a-Song Stress Test: A practical and valid test of autonomic responses induced by social-evaluative stress. *Autonomic Neuroscience*, *224*, 102612

ABSTRACT

The Sing-a-Song Stress Test (SSST) was recently developed as an alternative to the Trier Social Stress Test (TSST) to investigate autonomic nervous system responses to social-evaluative stress. In the SSST, participants are suddenly cued to sing a song in the presence of confederates. However, the SSST is still quite long (~15 min) and the requirement for confederates makes it labor-intensive. The current study tested whether a shorter (~6.5 minute), single-experimenter, version of the SSST can still reliably elicit subjective and physiological stress reactivity.

Our sample consisted of 87 healthy young adult participants (age range: 18-35 years). During the short SSST and a speeded reaction time task, in which aversive loud tones were to be avoided (TA), we measured heart period (HP), sympathetic nervous system (SNS) activity using pre-ejection-period (PEP), skin conductance level (SCL), and non-specific skin conductance responses (ns.SCR), and parasympathetic nervous system (PNS) activity using respiratory-sinus-arrhythmia (RSA) and the root-mean-square of successive differences (RMSSD).

The short SSST induced significant decreases in positive affect and increases in negative affect. MANOVAs on the clusters of SNS and PNS variables showed that the short SSST elicited significant HP (- 118.46 ms), PEP (-7.76 ms), SCL (+4.85 μS), ns.SCR (+8.42 peaks/min) and RMSSD (-14.67) reactivity. Affective, SNS, and PNS reactivity to the new SSST social-evaluative stress task were of comparable magnitude to that evoked by the TA mental stressor.

We conclude that the short SSST is a valid and cost-effective task for large, scaled studies to induce social-evaluative stress to a sufficient degree to evoke measurable changes in PNS and SNS activity and affective state.

INTRODUCTION

The importance of having valid tests for different kind of stressors originates from the idea that different stressors lead to different stress responses within individuals, known as the response specificity theory (Bosch et al, 2009; Dickerson and Kemeny, 2004; Skoluda et al., 2014). This theory is grounded in the belief that different physiological response patterns have evolved to effectively cope with the variety of different stressors (Wiener, 1992). Such stressor can be broadly, but not exclusively, divided in two categories: those focusing mainly on the mental effort or challenge-appraisal component and those focusing on the social self as described by the Social Self Preservation Theory (Dickerson, Gruenewald, and Kemeny, 2004) referred to as social-evaluative stressors. Many tasks haven been developed to study the different aspects of stress induced by "mental effort" type, like the Paced Auditory Serial Addition Test (Gronwall, 1977; Tombaugh, 2006; Tanosoto *et al.*, 2012), Stroop test (Stroop, 1935; Renaud and Blondin, 1997; Van Lien *et al.*, 2013) or aversive speeded reaction time tasks (De Geus, Van Doornen and Orlebeke, 1993). However, for eliciting "social-evaluative" stress there is currently only one single default paradigm: the Trier Social Stress Test (TSST), where participants prepare to give a speech in front of a panel of judges (Kirschbaum, Pirke and Hellhammer, 1993). While the TSST reliably evokes significant social-evaluative stress, it was primarily designed to induce prolonged stress to measure the slow responding adrenocortical axis reactivity. To investigate autonomic nervous system (ANS) reactivity to social-evaluative stress (e.g., through cardiac and electrodermal activity (EDA) responses), which are much faster, a shorter test that is also easier to implement would be preferred.

To this purpose, the Sing-a-Song Stress Test (SSST) was developed by Brouwer and Hogervorst (2014). In this test, participants are suddenly cued to sing a song in the presence of confederates, which elicits social-evaluative stress comparable to TSST. However, the SSST is still quite long (~15 min), and the use of multiple confederates makes it labor-intensive. This makes it impractical and costly for studies in large samples. Therefore, we developed a shorter (~6.5 minute), single-experimenter version of the SSST. The following adaptations were made to the original SSST (Brouwer & Hogervorst, 2014). First, the read-only conditions were reduced from nine to three. Second, we added a 'practice' condition in which they received an instruction to say the word vacuum out loud twice in short succession after an anticipatory countdown period. This condition was added to counteract the risk of 'too early singing' which was found in a substantial number of participants in the original SSST (Brouwer & Hogervorst, 2014). Lastly, no physical presence of a confederate was involved, but instead the crucial instruction to

prepare to sing a song explicitly mentioned that the audio and video recordings would be shared with an audience of music professionals interested in variation in musical ability. This short version of the SSST (short SSST) can easily be implemented in large-scaled epidemiological studies on the effects of stress on health outcomes.

To investigate if the short SSST can be used as an index of social-evaluative stress, positive and negative affect were measured before and after the short SSST to gauge the subjective stress response. To study the effect of social-evaluative stress on ANS activity we studied both sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity, in line with the Brouwer and Hogervorst (2014) but expanding the set of ANS measures. SNS activity was measured as the pre-ejection period (PEP; Matyas and King, 1976), skin conductance level (SCL; Bouscein 2012) and non-specific skin conductance responses (ns.SCRs; Bouscein 2012). PNS activity was measured as respiratory sinus arrhythmia (RSA, Grossman, van Beek, Wientjes, 1990) and root mean square of successive differences (RMSSD, Goedhart *et al.*, 2007). Lastly, heart period (HP), a mixture of both SNS and PNS activity, was measured. To validate the stress component, we compared the subjective and ANS stress reactivity of participants to this new social-evaluative stressor to that of an often-employed mental stressor, the tone avoidance (TA) speeded reaction time task.

We expect the short SSST to decrease positive affect and increase negative affect. Concerning ANS reactivity we expect increased SNS activity and decreased PNS activity reflected in an increase in SCL and ns.SCRs and a decrease in HP, PEP, RSA and RMSSD. The effects sizes are expected to be at least as big as those generated by the TA mental stress task.

Participants

A total of 113 participants participated in the study (age range 18-57). Exclusion criteria were a body-mass index above 30, heart disease, high blood pressure, high cholesterol, diabetes, and thyroid or liver disease, as these can all influence the functioning of the ANS. Additional exclusion criteria were the use of antidepressants or any other medication that has been shown to influence the ANS. If applicable, female participants were measured in the first two weeks after the last day of their menstrual cycle.

Due to the lack of applicants over the age of 35 (N = 5) we decided to exclude these participants in the current study. Of the remaining 108 participants, 11 were excluded because they were obese (N = 3) or had high blood pressure (N = 11) (some participants met multiple exclusion criteria). An additional 10 participants were excluded because they did not sing a song.

Participants who were students at the VU University of Amsterdam received research credits, while the other non-student participants were compensated with a €50 gift voucher. All participants provided informed consent before the start of the experiment. The study was approved by the VUmc medical ethical committee (NL62442.029.17).

Materials

The short Sing-a-Song Test

Participants were told that they had to sit as still as possible in front of a computer while they were shown several messages, followed by a counter from 60 to 0 seconds. They were informed that some of these messages only needed to be read whereas others might contain an instruction that they had to follow when the counter had reached 0. A detailed description of the experiment is given in figure 1.

For the three read-only trials, participants were instructed on-screen to quietly read the presented messages while sitting as still as possible. It was important to select phrases that did not elicit any stress or emotions. Therefore, three phrases in big black letters from the Dutch Wikipedia site about vacuums were shown on a monitor with a white background (translated example: "A vacuum is a device that sucks dust and other small particles"), similar to the original SSST (Brouwer & Hogervorst, 2014). The neutral Wikipedia phrases and read instruction were shown for 12 seconds, followed by a counter counting down from 60s to 0s. The instruction to read out aloud twice the word "vacuum cleaner" after countdown also lasted 12s, followed by a counter from 60s to 0s and a 5s period in which the word "vacuum cleaner" was shown on the screen. In the final sing-a-song trial, an instruction was provided for 12 seconds telling them to pick a song of their own choice and prepare to sing it aloud after the counter reached zero. It was also stated that their performance would be recorded and investigated by musical professionals. The short SSST ended with the instruction to sing a song that lasted for 20 seconds.

Tone avoidance test

The tone avoidance test is a stress-inducing task of the "active coping" type. During the tone avoidance task subjects have to react to a stimulus (an "X") that flares up irregularly in one of the corners of a computer screen. Subjects have to respond to this stimulus, within a 550ms timeframe, by pressing the button opposite to this corner on their response panel using one hand only.

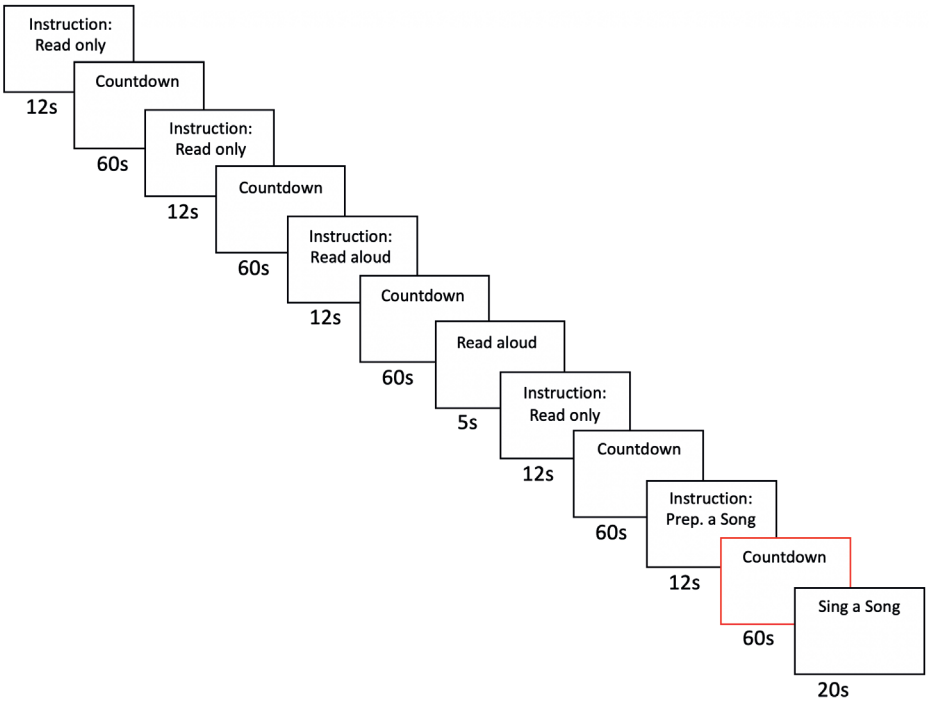


Figure 1. short Sing-a-Song stress test experimental set-up.

The task consisted of three read conditions, a speak condition and a sing condition. Each condition was followed by a countdown from 60 to 0 seconds. The first two messages were neutral text with no instruction. The third message contained the instruction to say the word "vacuum" twice when the timer reached 0. This was followed by another message with neutral text. Lastly an instruction to sing a song when the timer reached 0 was shown. In our analyses we focused on the anticipatory stress during the sing-a-song countdown, which is provided with a red outline.

Participants started with 50 points. During the task, incorrect or too slow responses were punished with a red bar, a loud noise burst (1000 Hz, 85 dB) and a loss of 1 point. Correct responses were rewarded by a green bar (Benschop & Schedlowski, 1999). When participants responded correctly for five consecutive times or more a point was added. Participants were told that they had to sit as still as possible during the test, only moving the hand they use for button pressing.

Affect Questionnaire

Positive affect scores were obtained before and after each test by asking the participants to rate on a scale of 1 (not at all) to 7 (very) whether they felt relaxed, cheerful, enthusiastic, and content. Negative affect was obtained from items rating whether they felt insecure,

lonely, anxious, irritated, and down (Myin-Germeys et al. 2001). Positive and negative affect were then defined as the mean score of the individual items.

Physiology

To measure electrocardiography (ECG), impedance cardiography (ICG), and EDA, the VU-Ambulatory Monitoring System (VU-AMS)(Vrije Universiteit, Amsterdam, the Netherlands) was used. To record ECG and ICG signals, five adhesive 55mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Heerlen, Netherlands) were placed on the subject's torso (Figure 2). A recording frequency of 1000 Hz was used.

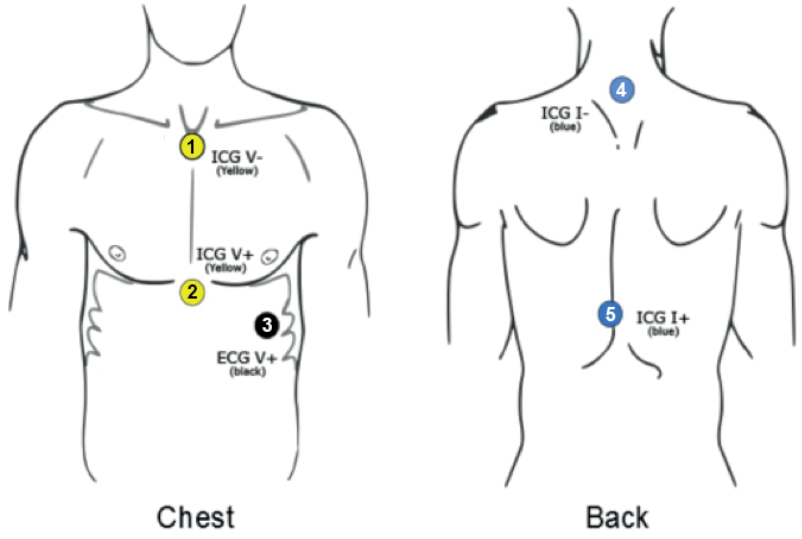


Figure 2. Electrode placement for electrocardiography and impedance cardiography recordings.

The electrodes were placed on top of the sternum at the suprasternal notch (1); at the bottom of the sternum on the processus xiphoideus (2); on the ninth left intercostal space (3); at the back, on the spine, at least 3 centimeters above electrode 1 (4); at the lower back, on the spine, at least 3 centimeters below electrode 3 (5).

EDA was recorded on the participant's non-dominant hand. No preparations were performed on the skin to preserve its electrical properties (Dawson, Schell and Filion, 2000). A 55mm Kendall H98SG hydrogel ECG electrode (Medtronic, Eindhoven, Netherlands) was placed on the inside of the non-dominant forearm approximately 15 cm below the hand electrode. Before applying this electrode, dead skin cells were removed by lightly scrubbing the skin with sandpaper. A recording frequency of 10 Hz was used.

METHODS

Procedure

This study was part of a larger study that focuses on the validation of a wristwatch-based technology, developed by Philips (Eindhoven, Netherlands), to measure EDA in a laboratory and ambulatory settings (see appendix 1 for a complete overview of the study). The larger experiment across two days was presented to the participants as a general study on the detection of stress through measurement of ANS activity using wearable technology.

When entering the lab, participants were informed that their voice, facial expressions, and posture would be recorded by video. Participants were shown the control room that the experimenter would sit in. It had a one-way mirror overlooking the experimental room where they would undergo the various tests. The control room contained multiple monitors and speakers that generate high-quality video footage and voice recordings from the camera and microphone placed in the experimental room. The participants were made aware of this intense monitoring throughout the experiment. They were *not told upfront* that the tests would involve singing nor that the recordings of their performance would be shared with an unseen audience. Throughout, no actual footage or sound was recorded and the deliberate deception about being recorded as well as its purpose was explained in the debriefing at the end of the experiment.

At the start of the experiment on day 1, resting blood pressure and body-mass index (BMI) were measured followed by a structured interview regarding the subject's demographics, medication use, perceived physical and mental health and lifestyle behaviors, to confirm that participants met inclusion criteria for the study. Next, the system for continuous monitoring of SNS and PNS activity was attached.

The experimental stress manipulations on day 2 consisted of a baseline measure, in which participants were instructed to sit as still as possible for 3 minutes, followed by the 4-minute TA task and the ~6.5-minute short SSST. In between the TA task and the short SSST, participants had a two-minute recovery period. Immediately after the baseline and after both stressors participants were instructed to fill in a short 9-item questionnaire to measure their negative and positive affect (Myin-Germeys et al. 2001).

ECG and ICG derived PEP and heart period variability measures

ECG and ICG analysis were performed using the VU-DAMS (Vrije Universiteit, Amsterdam, Netherlands) software (version 4.0). The software detects and scores all R peaks and automatically detects the start of inspiration and expiration for each breath. Possible heart period (HP) artifacts were marked by the software and visual inspection was used

to remove or correct artifacts (i.e., wrongly scored R peaks). RSA was obtained by peak-valley estimation as described elsewhere (Nederend *et al.*, 2018) combining the HP time series with the respiration signal that was extracted from the lower frequency changes in thorax impedance. RSA values were set to be zero for breaths with an invalid RSA. RMSSD was calculated by taking the root mean square of successive differences in heart period. Quality of RSA and RMSSD was checked by inspecting the respiration and heart rate signal manually, removing noisy data when necessary. For each condition, an ensemble average impedance cardiogram of all corresponding complexes of adequate quality was calculated using the VU-DAMS software. Given its sensitivity for movement artifacts, the ICG signal was filtered using a 60Hz low pass filter. Each impedance cardiogram was inspected visually, and the B, C and X points were scored automatically and manually corrected when necessary (Nederend *et al.*, 2017). PEP was obtained by calculating the time between the start of ventricular depolarization in the ECG (Q onset) and the time the aortic valve opens in the impedance cardiogram (B point). PEP has been shown to be a reliable non-intrusive way to measure SNS activity (Sherwood et al. 1990).

EDA derived SCL and ns.SCR measures

All EDA signals were cleaned with a simple automated artifact rejection algorithm (i.e., sudden drastic drops or increases in µS, flattening of the signal) in MATLAB (2016a). SCL and ns.SCRs per condition were obtained using the EDA master toolkit (Joffily, 2012) in MATLAB (2016a). The SC signal was filtered using a low-pass 0.5Hz Butterworth filter (Taylor *et al.*, 2015). SCL was calculated as the average over the artifact-free, filtered signals. A ns.SCR was identified when the peak amplitude exceeded 0.01µS but was not larger than 2.5µS and the rise time was between 0.1 and 5 msec. Overlapping responses (a ns.SCR that occurs during the rise time of a preceding ns.SCR) were counted to detect stacking of responses. The total number of identified responses was divided by the artefact free time to obtain ns.SCR frequency per minute.

Statistical Analysis

Data were analyzed using SPSS (ver. 25.0, 2017). For the analyses, the mean of all ANS measures during the 3-minute baseline, 4-minute TA task and 60 seconds short SSST sing anticipation was used. All ANS variables were checked for normal distribution and outliers. If a variable was not normally distributed, it was log-transformed. A value was considered an outlier if it deviated from the mean with more than three standard deviations. All outlier values were removed. Concerning to SCL, values over 35 µS were deemed implausible and therefore censored at 35 µS.

We expected all neutral anticipatory conditions to be different from the SSST anticipation but had no reason to expect differences between the neutral anticipatory conditions and the baseline or between the neutral anticipatory conditions themselves. This was borne out by preliminary comparisons. Therefore, our analyses were simplified to physiological reactivity of the short SSST by focusing on the contrast between the sitting baseline condition and a) the Sing anticipation condition and b) the TA task.

To investigate the effect of the two stressors on ANS activity, a repeated measures MANOVA on the clusters of SNS and PNS variables was performed with type of stressor (short SSST vs TA) and condition (baseline vs stress exposure) as the repeated measures. The multivariate cluster of SNS variables included HP, PEP, SCL and ns.SCR. The multivariate cluster of PNS measures included HP, RMSSD, and RSA. Significant main effects of the repeated measures MANOVA were followed by post-hoc testing on each stressor and each ANS measure separately. To obtain the effect size of each ANS measure cohen's d was calculated. We notice that HP reactivity was used twice in this approach. Although HP reactivity does truly reflect both SNS and PNS reactivity, we did risk the results being dominated by HP effects. To examine whether this was the case, we repeated the analyses without HP in both clusters. As this did not noticeably alter the pattern of results, we report only on the MANOVA on clusters with HP left in for brevity.

All analyses were performed with age and sex as covariates. Age was transformed into a binary variable with 0 for participants under the age of 25 and 1 for participants of 25 years and older. Since respiration rate (RR) has been associated with RSA (de Geus *et al.*, 1995; Grossman, Karemaker and Wieling, 1991) and RMSSD (Schipke, Arnold, and Pelzer, 1999) RR was added to the analyses of these variables. If the assumption of sphericity was violated the Greenhouse-Geisser results were reported. The large number of tests performed (N = 26) required a correction of our experiment-wise p-value to reduce type I errors. Since the variables in the PNS and SNS clusters are highly interrelated, a Bonferroni correction would be overly conservative. Instead, we used Matrix Spectral Decomposition (matSpD) to estimate the equivalent number of independent variables in the full correlation matrix of all SNS and PNS variables tested, and we adjusted the p-value accordingly (QIMR Genetic Epidemiology Laboratory, Dale's homepage https://gump.qimr.edu.au/general/daleN/matSpD/). This led to a p-value threshold of .002 for a result to be declared significant.

RESULTS

Descriptive statistics of the study population (N = 87) can be found in Table 1. RSA and RMSSD were not normally distributed and therefore log transformed. Regarding data quality: two outlier values (> 3SD) were removed for log-transformed RSA, eight for log-transformed RMSSD and four for PEP. For two participants the data quality of the EDA recording was considered too low for reliable peak detection.

Affect

To test whether the short SSST affected subjective reporting of positive and negative affect and whether this effect was similar to the TA task, Wilcoxon signed rank tests were performed. Figure 3 shows that participants felt both significantly less positive (short SSST: N = 87, Z = -3.65, $p < .001$; TA: N = 87, Z = -5.54, $p < .001$) and more negative (SSST: N = 87 , Z = -4.69 $p < .001$; TA: N = 87, Z = -6.44, $p < .001$) after these tests. There was no significant difference in affect scores between the two tests (N = 87, Z = -2.29, $p = .022$).

Table 1. Population descriptive statistics.

	Excluded (N = 26)	Included (N = 87)	Male (N = 35)	Female (N = 52)
Smoking (%)	30.0	34.5	31.4	36.5
Exercise (%)	42.7	39.1	31.4	44.2
Student (%)	62.2	82.8	80.0	84.6
Age (M ± SD)	27.01 ± 9.36	22.37 ± 3.52	22.37 ± 3.62	22.37 ± 3.48
BMI (M ± SD)	26.29 ± 3.42	23.58 ± 2.78	23.53 ± 2.76	23.63 ± 2.81
SBP (M ± SD)	124.34 ± 17.75	114.28 ± 9.15	117.50 ± 7.57	112.15 ± 9.54
DBP (M ± SD)	78.35 ± 9.16	69.68 ± 8.34	67.28 ± 8.59	71.30 ± 7.84
Education (years)	10.5 ± 2.61	10.21 ± 2.15	9.97 ± 2.03	10.37 ± 2.24

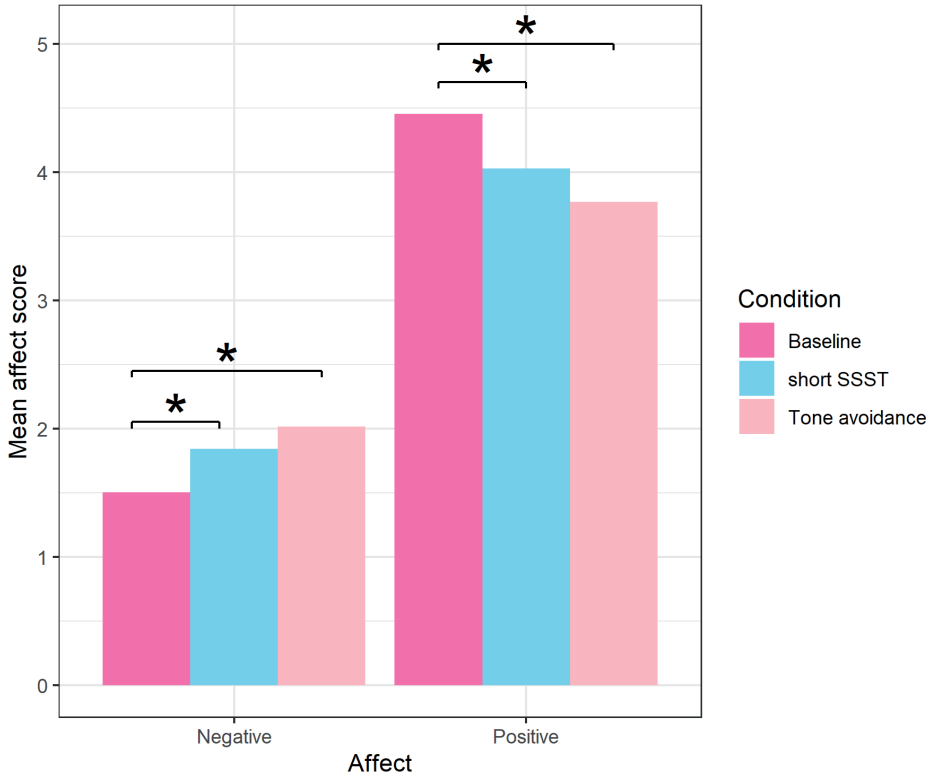


Figure 3. Affective response to the stress tasks.

ANS reactivity

Table 2 shows the means and standard deviations of the ANS measures and RR during the different conditions. There was a significant difference in SNS activity between the three conditions (Greenhouse-Geisser: $F(1.86,140) = 24.86$, $p < .001$). Contrasts analyses indicated a significant difference between baseline and SSST ($F(1,70) = 690.08$, $p < .001$), baseline and TA task ($F(1,70) = 1582.84$, $p < .001$) and SSST and TA task ($F(1,70) = 1657.38$, $p < .001$). The difference in SNS activity between baseline and the short SSST was driven by all individual SNS measures, with medium to large effect sizes (Table 3).

There was also a significant difference in PNS activity between the three conditions (Greenhouse-Geisser: $F(1.80, 162) = 24.70$, $p < .001$). Contrasts analyses indicated a significant difference between baseline and SSST ($F(1,81) = 1775.77$, $p < .001$), baseline and TA task ($F(1,81) = 3094.80$, $p < .001$) and SSST and TA task ($F(1,81) = 2279.60$, $p < .001$). The difference in PNS activity between baseline and the short SSST was driven by HP with a large effect size, with a trend for RMSSD with a medium effect size (Table 3).

The observed difference between the two stress tasks was entirely driven by the larger HP reactivity to the short SSST compared to the TA task, the individual SNS and PNS variables all showed comparable reactivity.

Response stereotypy across short SSST and TA tasks

To assess response stereotypy, Pearson correlations were computed on the reactivity scores (stress test – baseline) for the TA and short SSST tasks across all 6 variables. There was a significant positive correlation ($P < .001$) between the short SSST and TA reactivity for all SNS and PNS variables (Figure 4) showing autonomic stress reactivity to be a stable individual characteristic across the mental and social-evaluative domains.

Table 2. Means and SD of the ANS measures at baseline and during stress.

ANS measure	Condition	N	Mean	SD
PEP (msec)	Baseline	84	111.58	17.84
	short SSST	84	103.70	19.18
	TA	83	104.57	17.66
SCL (µS)	Baseline	81	11.85	7.46
	short SSST	85	16.26	8.61
	TA	82	14.61	7.90
ns.SCR (pm)	Baseline	81	5.21	3.24
	short SSST	85	13.71	4.74
	TA	82	15.50	4.86
HP (msec)	Baseline	87	842.25	133.07
	short SSST	87	722.79	132.10
	TA	87	784.63	131.82
RSA (msec)[1]	Baseline	88	87.84	46.37
	short SSST	87	67.24	34.90
	TA	89	63.22	30.44
RMSSD (msec)[1]	Baseline	86	55.54	28.03
	short SSST	86	40.87	21.05
	TA	87	49.50	24.74
RR (bpm)	Baseline	86	15.28	2.72
	short SSST	86	16.40	3.04
	TA	87	19.36	3.51

1. Raw data is reported. In further analyses log transform is used.

Table 3. ANS stress reactivity

ANS measure	Condition	N	Difference	Cohen's	df	F	p
			M ± SD	*d*			
PEP (msec)	short SSST – Baseline	83	-7.76 ± 14.03	-0.55	1, 79	18.05	<.001
	TA – Baseline	83	-7.49 ± 9.55	-0.78	1, 79	34.29	<.001
	short SSST – TA	82	-0.34 ± 12.35	-0.03	1, 78	0.39	.53
SCL (µS)	short SSST – Baseline	81	4.85 ± 4.17	1.16	1, 77	53.46	<.001
	TA – Baseline	78	3.46 ± 3.37	1.02	1, 74	46.67	<.001
	short SSST – TA	82	1.43 ± 3.28	0.43	1, 78	6.50	.013
ns.SCR (pm)	short SSST – Baseline	81	8.42 ± 4.60	1.83	1, 77	123.76	<.001
	TA – Baseline	78	10.52 ± 4.71	2.23	1, 74	183.89	<.001
	short SSST – TA	82	-1.78 ± 5.26	-0.33	1, 78	4.56	.036
HP (msec)	short SSST – Baseline	87	-118.46 ± 110.59	-1.07	1, 83	56.42	<.001
	TA – Baseline	87	-57.61 ± 94.74	-0.61	1, 83	22.03	<.001
	short SSST – TA	87	-60.84 ± 97.09	-0.62	1, 83	16.32	<.001
RSA0 (msec)	short SSST – Baseline	85	-19.09 ± 48.74	-0.39	1, 79	2.05[1,2]	.15
	TA – Baseline	86	-25.02 ± 44.51	-0.56	1, 81	0.28[1,2]	.59
	short SSST – TA	86	43.46 ± 30.77	1.41	1, 80	4.40[1,2]	.039
RMSSD (msec)	short SSST – Baseline	87	-14.67 ± 26.65	-0.55	1, 77	8.69[1,2]	.004
	TA – Baseline	87	-6.04 ± 22.55	-0.27	1, 79	0.55[1,2]	.46
	short SSST – TA	87	-8.63 ± 19.88	-0.43	1, 78	19.12[-2]	<.001

1. Log transform was used in the analysis
2. Adjusted for respiration rate

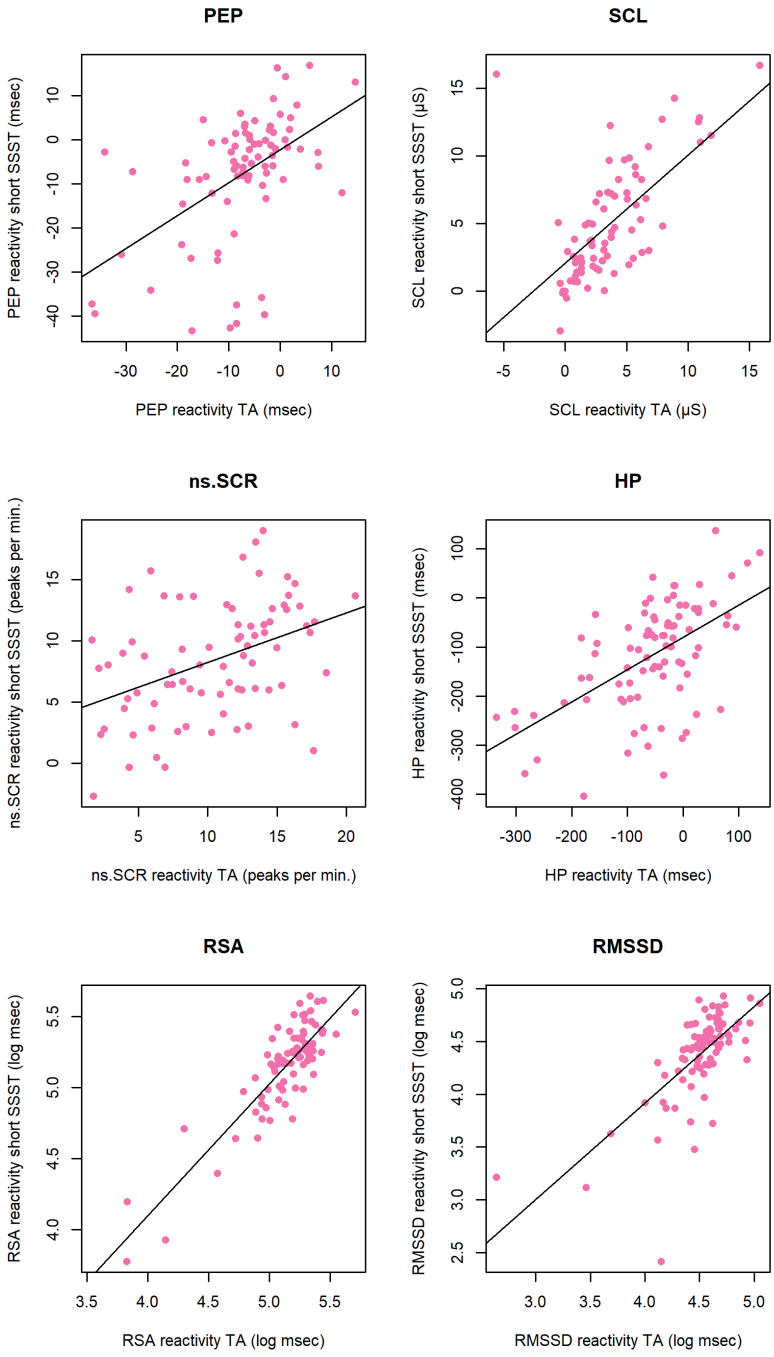


Figure 4. Response Stereotypy.

Note: A scatter plot with regression line of short SSST reactivity with tone avoidance reaction time task (TA) reactivity is shown for each ANS measure. A) PEP ($r = .51$, $p < .001$), B) SCL ($r = .63$, $p < .001$), C) ns.SCR ($r = .41$, $p < .001$), D) HP ($r = .56$, $p < .001$), E) RSA ($r = .88$, $p < .001$) and F) RMSSD ($r = .71$, $p < .001$).

DISCUSSION

The "sing-a-song stress test" (SSST) is shown to be a valid shorter alternative for the longer and labor-intensive Trier Social Stress test (TSST) to evoke social-evaluative stress (Brouwer & Hogervorst, 2014). The current study shows that a shorter and more practical version of the SSST still effectively induces social-evaluative stress reflected by both affective responses and physiological reactivity.

In the current study several improvements have been made to the original SSST. First, the short SSST contains fewer trials, thus decreasing the overall duration from ~15 (SSST) to ~6.5 (short SSST) minutes. Second, confederates are no longer required. Third, by adding a training condition (read aloud), we were able to eliminate the problem of participants starting to sing too early as none of the participants started singing before they were instructed to do so. Last, the short SSST was validated using a more diverse range of ANS measures providing broader insight into the ANS reactivity caused by this stressor.

In accordance with our expectations, the short SSST significantly decreased positive affect, increased negative affect, and shifted the ANS to a state of increased SNS activity and decreased PNS activity, with medium to large effect sizes. With regard to the cardiac ANS measures, our HP results (converted to HR for comparison) are consistent with those of the original longer SSST (Brouwer and Hogervorst (2014) but of slightly smaller magnitude ($HR_{\text{short SSST}}$: 11.8 bpm vs. $HR_{\text{original SSST}}$ 15.3bpm). When compared to an often-employed mental stress test, a speeded reaction time task in which incorrect and slow responses are punished by aversive loud tones, the short SSST evoked reactivity of similar direction and magnitude for all cardiac ANS measures. These findings are consistent with that of previous studies investigating cardiac ANS reactivity to a wide array of other stress tasks in both direction and effect size (Brindle et al. 2014). Consistent with the findings of Bosch et al. (2009), the short SSST led to higher HP and RMSSD reactivity compared to the TA task. However, such an effect was not observed for PEP.

With regard to our EDA measures, the reactivity to the short SSST and TA test were also of similar direction and magnitude. The SCL results, however, showed larger differences between the short and the original SSST ($SCL_{\text{short SSST}}$: 4.4μS vs. $SCL_{\text{original SSST}}$ 10.9μS). This may be partly explained by the strong sensitivity of absolute SCL levels to the type and placement of the electrodes as well as the room temperature. Therefore, we measured ns.SCRs as an alternative read-out of skin SNS. Measuring ns.SCRs has two advantages. First, it is less sensitive to temperature and type and placement of the electrodes. Second, as thoroughly discussed by Boucsein in his book on electrodermal activity (2012), several

studies in the 1970s focusing on the anticipatory stress preceding an electrical shock have shown that ns.SCR frequency is a potent indicator of this type of stress. These studies even suggest that this type of stress is captured better by ns.SCR frequency than SCL (Boucsein, 2012). Taking these findings, a step further, Erdmann, Janke, and Bisping (1984) studied the EDA response to the anticipation of public speaking. They compared public speaking to 1) white noise (95 dB) presented discontinuously, 2) anticipation of a painful electric shock and 3) a Charlie Chaplin film (as a "eustress" condition) and found that ns.SCR frequency was higher during speech anticipation compared to all other conditions. Interestingly, in our study ns.SCR frequency also showed the largest effect size among all ANS measures. This provides support for ns.SCR frequency as a potent measure of anticipatory social-evaluative stress.

The increase in ns.SCR frequency tended to be even higher in the TA task, although not formally significant. This could be due to the physical activity component of this task (rapid button pressing). Note that movement artefacts per se are unlikely, button pressing was allowed only with the dominant hand, i.e., contralateral to the hand containing the EDA electrodes which rested on the table. Support for this notion is given by the study of Novak, Mihelj and Munich (2010) who showed that ns.SCRs frequency increased substantially when physical workload is increased, independent of mental workload. This could also explain the relatively modest correlation for ns.SCR frequency reactivity between the short SSST and TA test.

Using a variety of ANS and affective measures our results show that the short SSST is a potent stress-inducing task. This is further supported by the substantial correlation of our ANS measures between the short SSST and the TA task, suggesting that the short SSST captures the general trait of being a low or high 'stress-reactor' rather well. This suggests that social-evaluative stress can be effectively induced even without the need for confederates. However, we do note several limiting factors to the use of the short SSST. First, several participants refused to sing entirely, causing their data to be unusable for this study. The nature of their incompliance is unknown. It might be that they just did not feel like singing or that they were to stressed to even start singing. It would be interesting to investigate this in future studies. Second, during debriefing we informed the participants that none of their singing was actually recorded. We noticed that some of the participants indicated that they already suspected this because they had not signed formal informed consents that their performance would be shared, but that, even so, they were not entirely sure. Unfortunately, we did not document this, therefore we could not investigate a possible effect on task outcome. Third, it is unlikely that the short SSST can be used repeatedly within the same subject to the same effect. The task, like the

original TSST and the SSST, requires a form of deception that demands full debriefing from an ethical point of view, which may greatly reduce its impact on repeated exposure. For the same reason, a direct comparison of the original longer SSST with the new short SSST in the same participants was not feasible. Last, the observed ANS effects were a little smaller than the those found by Brouwer and Hogervorst. Though this could be due to differences in study population, we cannot exclude that the difference might be due to the lack of physically present confederates.

In conclusion, the short SSST is a more time-efficient and less labor-intensive alternative to the SSST and TSST. It induces social-evaluative stress to a sufficient degree and evokes measurable changes in affective state, PNS and SNS activity. We believe that this test can be successfully used in large scale studies on the causes and consequences of individual differences in autonomic responding to stress.

CHAPTER 3

Validity of electrodermal activity-based measures of sympathetic nervous system activity from a wrist-worn device.

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. C. (2021). Validity of electrodermal activity-based measures of sympathetic nervous system activity from a wrist-worn device. *International Journal of Psychophysiology*, *168*, 52-64.

ABSTRACT

Measuring electrodermal activity (EDA) on the wrist with the use of dry electrodes is a promising method to help identify person-specific stressors during prolonged recordings in daily life. While the feasibility of this method has been demonstrated, detailed testing of validity of such ambulatory EDA is scarce. In a controlled laboratory study, we examine SCL and ns.SCR derived from wrist-based dry electrodes (Philips DTI) and palm-based wet electrodes (VU-AMS) in 112 healthy adults (57% females, mean age = 22.3, SD = 3.4) across 26 different conditions involving mental stressors or physical activities. Changes in these EDA measures were compared to changes in the pre-ejection period (PEP) and stressor-induced changes in affect. Absolute SCL and ns.SCR frequency was lower at the wrist compared to the palm. Wrist-based ns.SCR and palm-based ns.SCR and SCL responded directionally consistent with our experimental manipulation of sympathetic nervous system (SNS) activity. Average within-subject correlations between palm-based and wrist-based EDA were significant but modest (*r* SCL = 0.31; *r* ns.SCR = 0.42). Changes in ns.SCR frequency at the palm (*r* = -.48) and the wrist (*r* = -.47) were correlated with changes in PEP. Both palm-based and wrist-based EDA predicted changes in affect (6% - 14%). Our data suggest that wrist-based ns.SCR frequency is a useful addition to the psychophysiologist's toolkit, at least for epidemiology-sized ambulatory studies of changes in sympathetic activity during daily life.

INTRODUCTION

The European parliament recognizes mental health as a fundamental human right and launched the EU Action Plan on mental health for 2021-2027, which is a continuation of the World Health Organization's Mental Health Action Plan 2013-2020 (World Health Organization, 2013). A core element is the development of effective strategies for stress detection and management. In the past decade, a lot of effort has been put in the development of biosensors that help identify person-specific stressors by inspection of their body's physiological responses to daily life settings (Wu, et al., 2012; Carbonaro, et al., 2013; Jung & Yoon, 2017; Jebelli, 2019). Sweat gland activity on the wrist is one of these physiological signals. It builds on a rich psychophysiological research tradition and recording is feasible for prolonged periods of time in daily life. The innervation of the sweat glands is entirely through sympathetic nerves and sweat gland activity is considered one of the purest measures of sympathetic nervous system activity (Critchley, 2002). The sympathetic nervous system (SNS) is rapidly activated when an individual is faced with a situation that is perceived as threatening or challenging, eliciting the so-called "fight or flight response" (Brindle, et al., 2014; Jansen, et al., 1995), in parallel to subjective feelings of arousal and negative affectivity often denoted as 'stress'. Activation of the SNS results in both an increase in the total number of activated sweat glands and in more secretion by the sweat ducts. These changes in sweat gland activity in turn lead to changes in the conductance of electrical activity through the skin, also denoted as electrodermal activity (EDA).

EDA is relatively easy to measure and has been used in a wide variety of research fields, notably attention, information processing, and emotion (Dawson et al., 2000). Decades of research have shown that various laboratory stressors increase Skin Conductance Level (SCL) compared to conditions of low arousal during pre- or post-stress baselines. These same stressors also systematically increase the frequency of non-specific skin conductance responses (ns.SCRs), by some referred to as spontaneous fluctuations (SF) (Bach et al., 2010). These skin conductance responses are not studied as a directly evoked response to a specific experimenter-controlled external stimulus. Rather we define ns.SCRs, consistent with Posada-Quintero & Chon (2012), to reflect "fluctuations in EDA in the presence of an ongoing sustained stimulus over a period of time" which differs slightly from Bouscein et al. (2012), who state that ns.SCR "occur in the absence of external stimuli and in the absence of artifacts such as movements and sighs". Frequency of ns.SCRs is measured in peaks per minute over longer time periods. Both SCL and ns.SCRs are considered indicators of SNS activity that show sensitivity to stress. Both resting levels

and responses to stress of these EDA measures show relatively stable inter-individual differences (Boucsein. et al., 2012) which are substantially heritable (Crider, et al., 2004; Schell, et al., 1988; Tuvblad, et al., 2010; Wang, et al., 2015). Since these EDA measures can be measured independent of knowledge on the content or timing of specific stimuli, they are in principle very suitable as indicators of SNS activity outside of the controlled laboratory environment.

The classical approach records EDA by passing a small electrical current through a pair of active/reference electrodes placed on the hand, either the middle phalanges of two adjacent fingers or the palm of the hand (Boucsein, et al., 2012). These locations are preferred because the hand contains the highest density of eccrine sweat glands (Posada-Quintero & Chon, 2020). However, a practical problem facing ambulatory measurement of EDA is that the typical location for electrode placement on the fingers or the palm of the hand is quite obtrusive and interferes with daily activities. This introduces bias in the behavioral repertoires assessed and increases risk for noisy or lost signals. Another practical problem for ambulatory measurement of EDA with the classical approach is the use of wet electrodes. Wet electrodes make contact to the skin through the use of electrolyte paste (Boucsein, 2012). When measuring over longer periods of time the electrolyte gel may gradually spread out on the skin and hydrate the corneum (Boucsein, et al., 2012). This can lead to both an increase in the recording area of the electrode (and thus observed EDA) and danger of electrode loosening. Especially the latter has a large influence on data quality and limits the length of the recording. Moreover, electrolyte paste might need to me reapplied when considering measuring over multiple days of even weeks, making it impracticable for these types of recordings.

A solution to both of these limitations is using electrodes without electrolyte paste on the wrist. Dry electrodes are generally reusable and easier to apply than wet electrodes making it a promising method to measure EDA in daily life over longer periods of time (Posada-Quintero & Chon, 2020). The wrist is a good alternative location as many smart watches already make contact with the skin on the wrist, and these are readily tolerated for prolonged wear time. Van Dooren et al. (2012) showed that measuring EDA on the wrist is indeed a good alternative to the hands and Westerink et al. (2009) and Poh et al. (2010) have shown that measuring EDA on the wrist with dry electrodes is feasible. However, while the ambulatory assessment of EDA on the wrist with dry electrodes is attractive and feasible it should be noted that these electrodes come with their own set of problems including the dependency on sufficient amounts sweat to detect EDA (Boucsein, et al., 2012). Even though the study by Poh et al. 2010 showed high correlations between SCL on the wrist and fingers, the results of other studies have not been encouraging (Konstantinou, et al., 2020;

Milstein & Gordon, 2020; Menghini, et al., 2019; van Lier, et al., 2019; Kleckner, et al., 2020). The evidence for ns.SCR responses rather than absolute SCL levels is more encouraging. A study by van Lier et al. (2019) showed that the mean amplitude of the ns.SCRs of dry wrist electrodes increases in a similar fashion to wet palm electrodes in response to a social stressor (sing-a-song stress test). In addition, Kleckner and colleagues (2020) have shown that exposure to a mental arithmetic stressor and physical activity led to an increase in the detection of ns.SCR of dry electrodes on the wrist.

At present, detailed testing of the validity of ambulatory EDA remains scarce. We therefore set up a controlled laboratory study and examined construct, criterion, and predictive validity for wrist-based dry electrode EDA monitoring in response to various mental and physical stressors. We employed an existing wrist-based dry-electrode device that evolved from the Emotion Measurement platform and monitors SCL and ns.SCR frequency (DTI5, Philips Ltd, The Netherlands) and compared the wrist-based EDA measures to parallel recorded EDA measures using an active electrode on the palm of the hand (thenar eminence). Because the DTI5 uses a proprietary algorithm to extract ns.SCR frequency, we added a second scoring of ns.SCR frequency from the raw wrist-based signal that was identical to the scoring of the palm-based signal (Jofily, 2012). First, construct validity was assessed by exposing participants to known experimental manipulations of SNS activity and testing whether the wrist-based EDA measures display the expected response pattern. We hypothesized that the mean SCL and ns.SCR frequency would increase from pre-task baseline level during exposure to mental and physical stressors, and then decrease again during recovery from those stressors. Second, to test criterion validity, we compared the within-subject changes across 25 experimental conditions in wrist-based EDA to changes in the Pre-Ejection Period (PEP: the time interval between the start of left ventricular depolarization and the opening of the aortic valve). We hypothesized that mental or physical stress-induced decreases in PEP, a proven and validated measure of cardiac SNS activity (Berntson, et al., 1994a; Berntson et al. 1994b), would be associated with increases in wrist-based EDA, indexing SNS activity on the skin. Finally, to assess predictive validity, we tested whether the changes in EDA measures predicted parallel changes in self-reported positive and negative affect induced by the mental stress tasks. We hypothesized that changes in the wrist-based EDA measures are predictive of the changes in affect induced by mental stress.

MATERIAL AND METHODS

Study population

Participants were required to be between the age of 18 and 48, Dutch speakers, and currently employed, or in a schooling trajectory. Exclusion criteria were a body-mass index above 30, heart disease, high blood pressure, high cholesterol, diabetes, thyroid or liver disease, and use of antidepressants, anticholinergics, or any other medication that has been shown to influence the SNS. Female participants were measured within the first two weeks following the last day of their menstrual cycle to account for hormonal changes.

Recruitment of potential participants was done through several routes. First, advertisements were placed on the Vrije Universiteit (VU) campus and the VU participant recruitment system SONA (a cloud-based participant pool software) to recruit students and VU employees. Second, participants were recruited from the local community through social media, by advertising on a Dutch Facebook page dedicated to participant recruitment (Proefbunny) and the investigators' personal social media pages. Finally, co-workers, friends, and family of the investigators, who themselves were excluded from participating, were asked to widely share the advertisement for this experiment in their social networks.

Interested participants could contact the research team through the contact information in the advertisement. During an ensuing telephone call, it was established whether the potential participant met the study criteria and was interested to receive the full information on the study. In case of a positive response, participants received the study information letter by e-mail. After a period of two weeks the research team contacted the participants and gauged their interest for actual participation in this study. After the volunteers were given complete, adequate written and oral information regarding the nature, aims, possible risks and benefits of the study, they were scheduled for the study visit at the Vrije Universiteit in Amsterdam.

Participants who were students received research credits, while other participants were compensated with a €50 gift voucher. All participants provided written informed consent before the start of the experiment. The study was approved in institutional review by the VUmc medical ethical committee (METc VUmc #2017.374, ABR #NL62442.029.17).

Physiological measurements

Electrodermal activity

Data acquisition

- *Palm-based EDA*

As the "ground truth", exosomatic palmar EDA was obtained with the Vrije Universiteit Ambulatory Monitoring System (VU-AMS) on the thenar eminence with direct current. However, the classical placement of the electrodes was adjusted to fit better with the daily life character of the experimental procedures. By limiting the number of electrodes placed on the hand to one, participants had more freedom to move their hand. We considered the thenar eminence of the non-dominant hand as the least interfering placement at the hand. The reference electrode was placed on a less obtrusive location, at the ventromedial forearm approximately 15 cm below the hand electrode (Figure 1A). This is considered a relatively inactive reference site (Venables & Christie, 1980) which reduces signal amplitude but greatly adds to participant comfort. On the thenar eminence adhesive tape was used to reduce movement and improve fixation to the skin and the skin curvature. In addition, the wire was fixed by means of tape to the skin 10–15 cm from the electrode, so the participants were able to move their hand in all directions without exerting pull on the electrode.

We chose to use different electrodes for the active and reference sites to optimize signal quality. On the thenar eminence disposable Biopac Systems EL507 EDA isotonic gel electrodes (Biopac systems Inc, Goleta, US) were used. These electrodes are designed for electrodermal activity measurement and are pre-gelled with isotonic gel (Ag/AgCl contact, wet liquid gel (0.5% chloride salt) electrolyte, 11 mm diameter contact area). Following guidelines, no preparations were performed on the skin to preserve its electrical properties (Dawson, Schell, & Filion, 2000) and electrodes were placed at least 5-10 minutes before the start of the experimental procedure to avoid decreased conductance due to electrolyte penetration of the stratum corneum from the isotonic gel (Boucsein, et al., 2012). On the ventromedial forearm 55mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Eindhoven, Netherlands) were used. ECG electrodes are designed to detect the electrical currents of the heart. For ECG recording EDA is considered an artefact, therefore ECG electrodes contain a layer of electrically conductive gel between the skin and the electrodes to reduce resistance. By lightly scrubbing the skin with abrasive paper part of the stratum corneum was removed to further lowering resistance (Boucsein, et al., 2012). Placing the inactive electrode on an electrodermal inactive site with very little resistance provides a higher and cleaner EDA signal compared to placing the inactive electrode on an electrodermal active site which resistance fluctuates with ongoing EDA.

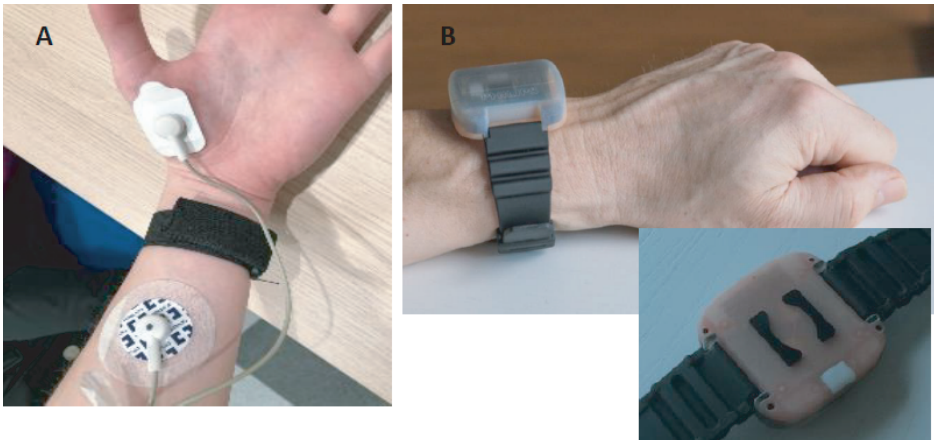


Figure 1. EDA electrode placement.

Note: A) VU-AMS: an active wet electrode is placed on the thenar eminence and a non-active electrode on the volar forearm. Medical tape used to secure the palmar electrode is not shown for clarity reasons. B) DTI5: the band is placed directly behind the head of the ulna. The dry electrodes contact the skin on the dorsal wrist.

EDA was recorded with a direct voltage of 0.5V, a sampling frequency of 10 Hz, and 16-bit (A/D converter) precision in the 0-100 microSiemens (μS) range. EDA signal quality assessment was performed after completion of the recording with a simple automated artefact rejection algorithm (i.e., sudden drastic drops or increases in μS based on the first derivative, and flattening of the signal, verified by visual inspection) in MATLAB. Segments flagged as artefact were removed from further analysis. The EDA signal was filtered using a low-pass 0.5 Hz Butterworth filter to deal with noise and motion artefacts (Doberenz, et al., 2011).

- *Wrist-based EDA*

Wrist-based exosomatic EDA was obtained with a CE approved wearable skin conductance sensor type DTI5 (Discreet Tension Indicator version 5, Philips) (Figure 1B), under development as a smartwatch for commercially availability to the consumer market. The DTI5 has a 47.1 * 15.5 * 47.8 mm casing and weighs 40 grams. It contains two 'banana' shape electrodes made of black hydrophilic silicone rubber that are placed at a distance of approximately 1 cm (see Figure 1). The band is placed directly behind the head of the ulna. Upon arrival at the laboratory participants had been wearing the device for already ~24h. This allowed moisture under the silicone rubber to build up, which, from past experience in prototype testing, yields a better conductive contact between the skin and the electrodes.

The DTI5 applies a direct voltage of 1V between both electrodes to measure skin conductance with a frequency of 160 Hz within a range of 0 to 24 µS and a precision of 22 bits. The DTI5 has an internal, on-line signal quality rating. The maximal quality rating of 3 is proportionally lowered based on the presence of certain features, for instance a change rate that exceeds plus 10 percent or minus 1 percent per second. The lowered quality value not only holds for the moment of the actual change rate disturbance but starts 0.5 seconds prior to the detected disturbance and ends 5 seconds after. Data of quality 1 was considered to reflect an artefact and segments with quality rating lower than 1 were removed from further analysis. The 160 Hz data subsequently is low-pass filtered (cross-over 5 Hz) to remove repetitive distortions in the skin conductance signal that coincide with motion.

EDA measures

The measures of interest that can be derived from both EDA signals are skin conductance level (SCL) and frequency of non-specific skin conductance responses (ns.SCR). Both these measures typically increase with increased SNS activity (Boucsein, et al., 2012; Posada-Quintero & Chon, 2020). For both devices SCL is calculated as the mean EDA level in µSiemens on the filtered artefact-free portion of every experimental condition. Peaks from both palm and wrist recordings were detected using the EDA master toolkit (Joffily, 2012) in MATLAB on the filtered artefact-free fragments of the EDA signal. As suggested by Braithwaite et al. (2013) ns.SCRs were counted if they had a peak amplitude threshold of 0.01 µS and rise time range of 0.1 - 5 msec (Braithwaite, et al., 2013). The parameter for detecting responses in rapid succession (overlapping responses) was set to ON. The resulting total number of ns.SCRs_mat during an experimental condition were counted and divided by the artifact-free minutes of the corresponding condition to obtain ns.SCR frequency in peaks per minute. The DTI contains an internal method of peak detection that makes use of a curve fit method, yielding ns.SCR_cf (for details you can contact Luc Vosters (luc.vosters@philips.com). The correlation within each participant between the peaks detected on the wrist signal by the two methods (internal algorithm ns.SCR_cf vs. MATLAB algorithm ns.SCR_mat) was high (*r* mean = .80, IQR = .71 – 94). Even so, we present the results from both the device-internal and toolkit scoring algorithms jointly throughout. This allows comparison of palm and wrist using the same method of peak detection across, as well as comparison of palm and wrist that additionally uses a different method of peak detection for the wrist location.

Pre-ejection period

The PEP has been shown to be a reliable non-intrusive cardiac measure of SNS activity (Sherwood, et al., 1990; Kelsey, 2012). PEP was obtained by calculating the time between the start of ventricular depolarization (Q onset) in the electrocardiogram (ECG) and the time the aortic valve opens (B point) in the impedance cardiogram (ICG) collected by the VU-AMS device. ECG and ICG were recorded from five adhesive 55 mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Eindhoven, Netherlands) placed on the chest and back of the participants (Figure 2) with a recording frequency of 1000 Hz. The locations of the Q onset and B point are automatically placed by the Vrije Universiteit Data Acquisition and Management Software (VUDAMS, available at: http://www.vu-ams.nl/support/downloads/software/) and manually corrected after visual inspection when necessary.

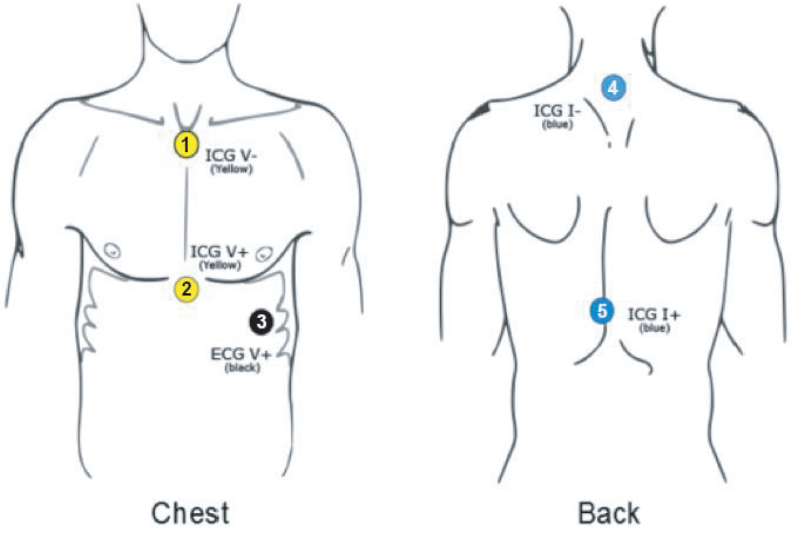


Figure 2. Electrode placement for ECG and ICG recordings.

Note: the electrodes were placed on top of the sternum at the suprasternal notch (1); at the bottom of the sternum on the processus xiphodius (2); at the apex of the heart on the ninth left intercostal space (3); at the back, on the spine, at least 3 centimeters above electrode 1 (4); at the lower back, on the spine, at least 3 centimeters below electrode 2.

Anthropometrics

The participant's body weight (kg) and body mass index (kg/m^2) were measured to reflect adiposity. After removal of shoes and coats, height was measured to the nearest millimeter using a stadiometer and weight was assessed to the nearest 0.1 kg using a digital scale. Body mass index (BMI) was calculated as body weight in kilograms divided by height in meters squared. Second, body fat distribution was measured using waist circumference (cm) and waist-to-hip ratio (W/H).

Interview and Questionnaires

A structured interview regarding the participant's demographics, medication use, perceived physical and mental health and lifestyle behaviors was performed to reconfirm participants met the inclusion criteria and to obtain a series of potential confounders/ explanatory variables. Two additional questionnaires were supplied: 1) the Edinburgh handedness inventory (Oldfield, 1971) to determine to participants hand preference and 2) the Profile of Mood States – short form (POMS), a psychological rating scale used to assess current overall mood state (McNair, et al., 1971).

Affect was repeatedly rated during the experiment directly following certain tasks (see table 1) by the Maastricht Questionnaire (Myin-Germeys, et al., 2001). Positive affect scores were obtained by asking the participants to rate on a scale of 1 (not at all) to 7 (very) whether they felt relaxed, cheerful, enthusiastic, and content and averaging the score over the 4 items. Negative affect was obtained by averaging the scores for 5 items: insecure, lonely, anxious, irritated, and down.

Experimental tasks

Posture

Changes from a supine to a sitting to a standing position are well-known to generate a stepwise increase in SNS activity. Impact of postural manipulation on our SNS activity measures was obtained by having the participants lie down on a stretcher bed, sit upright on a comfortable chair with both feet on the ground and stand upright, each for 3 minutes.

Mental stressors

To measure SNS responsivity to artificially induced emotionally engaging mental stress, participants performed a set of often used stress tasks. These include: Tone Avoidance reaction time task (TA) (2x), Paced Auditory Serial Addition Task (PASAT) (2x), short Sing-a-Song Stress Test ($SSST_{short}$) (1x), and the Raven's progressive matrices (RPM) IQ test (1x).

The TA aims to induce "effortful active coping" (de Geus, et al., 1990; van der Mee, et al., 2020). During the TA participants have to react to a stimulus (an "X") that flares up irregularly in one of the corners of a computer screen. Participants have to respond as fast as possible to this stimulus by pressing the button opposite to this corner on their response panel. During the tone avoidance task incorrect or too slow responses are punished with a red bar and a loud noise burst. Correct responses are rewarded by a green bar.

The PASAT is a measure of cognitive function that assesses capacity and rate of information processing and sustained and divided attention (Tombaugh, 2006). The PASAT is presented using prerecorded audio to ensure standardization in the rate of stimulus presentation. Single digits are presented at short intervals, traditionally every 3 seconds,

and the respondent must add each new digit to the one immediately prior to it. Responses are made by clicking the corresponding answer (0-18) using a mouse and must be given before the next stimulus is presented. Feedback is given by a green checkmark in case of a correct and timely answer, or a red x when the answer is wrong or too late. Shorter inter-stimulus intervals are known to increase the difficulty and perceived stressfulness of the task.

In the current implementation of the TA and PASAT tasks a staircase algorithm was used that adapted the criterion reaction time to the participant's average reaction time. This ensures that the level of difficulty is tailored to the skills of the participants which may vary due to e.g., age or educational attainment. In addition, the application of such a staircase maintains task difficulty during repeated exposure: both the TA and PASAT tasks were repeated twice which might induce habituation. To further ensure sufficient effort and engagement of the participants with these tasks a competition was set up in which the three best performing participants would gain an additional monetary reward of 50 Euros. A large and visible score board was used to keep the score, identifying participants by their participant ID code.

The $SSST_{short}$ is a recently developed adaptation of the Sing-a-Song Stress Test aimed at measuring social-evaluative stress in a quick and easy manner (van der Mee, et al., 2020). In this test participants are told that they had to sit as still as possible in front of a computer (surrounded by cameras and voice recording equipment) while they are shown several messages, followed by a clock counting down from 60 to 0 seconds. They are informed that some of these messages only need to be read whereas others will contain instructions they have to follow when the counter reaches 0. One of these instructions is to sing a song of their choice out loud. The instructions additionally mentions that their performance is recorded and will later be studied by conservatory students. The anticipatory interval of 60 seconds before the participant started singing was the stressor of interest, unaffected by the movement involved in the act of singing itself.

Raven's progressive matrices test is a nonverbal IQ test typically used in educational settings. It is a 60-item test, listed in order of difficulty, used in measuring abstract reasoning and regarded as an estimate of non-verbal fluid intelligence (Raven, 2003). In each test item, the participant is asked to identify the missing element that completes a pattern. We used the original test items; however, we only gave the participants 4 minutes to complete the test, which is far too short to complete all items. The test was administered on a tablet computer and the remaining time was shown in bright red in the right corner of the screen. Beneath the timer their progress and number of errors were presented, further increasing the ego-threatening aspect of IQ testing.

Table 1. Experimental timeline.

	Experimental condition	Duration (minutes)	Mood measurement
Postural	Lying down	3	
	Standing	3	
	Sitting	3	V
Mental stressors	Tone Avoidance	4	V
	Recovery (sitting)	2	
	short Sing-a-Song Stress Test	6.5	V
	Recovery (sitting)	2	
	Paced Auditory Serial Addition Test	4	V
	Recovery (sitting)	2	
	Raven's progressive matrices	4	V
	Short break	*5-10*	
Physical stressors	Walking at natural pace	2	
	Fast walking	2	
	Biking	4	
	Stair climbing	4	
	Recovery (standing)	2	
	Dish washing	2	
	Vacuum cleaning	2	
	Recovery (sitting)	2	
Mental stressors	Tone Avoidance (repeat)	4	V
	Recovery (sitting)	2	
	Paced Auditory Serial Addition Test (repeat)	4	V
Physical stressors	Treadmill intensity 1 (4.5 – 5 km/h)	4	
	Treadmill intensity 2 (6 - 6.5 km/h)	4	
	Treadmill intensity 3 (7.5 - 8 km/h)	4	
	Treadmill cooling down (3.7 – 4 km/h)	3	V
	Recovery (sitting)	3	V

Physical stressors

To examine how the EDA measures captured the effects of general everyday life activities on SNS activity, several typical everyday life activities were conducted during the laboratory session (see Table 1, experimental timeline). Mild to moderate physical activity

was induced by self-paced walking (at the pace they normally walk), fast walking (the pace they walk when they are in a hurry), bicycling, stair climbing and descending, mock dishwashing (without actual water and soap) and vacuum cleaning. To examine how the EDA measures captured standardized physical activity, participants had to jog/run on a treadmill at 3 incremental stages of speed (males: 5, 6.5, 8 km/h; females: 4.5, 6, and 7.5 km/h), each lasting 4 minutes. After a 3-minute cooling-down on the treadmill (males: 4 km/h, females: 3.7 km/h) participants sat down for a 3-minute recovery stage.

Procedure

The full research project included an initial data collection phase in a real-life ambulatory (~24 h, including the night) setting, but here we focus on the second phase, the standardized laboratory validation (~2,5 h of experimental manipulations) of wrist-based EDA obtained from the DTI5 device. During their initial visit to the laboratory (~1 h) at the start of ambulatory recording, participants provided informed consent, anthropometrics were measured, and the structured interview and questionnaires were administered. Subsequently, equipment for monitoring SNS activity was applied to the participant, with the EDA electrodes of the VU-AMS device and the DTI5 device on the non-dominant hand and wrist. Once equipped with the measuring devices, participants left the laboratory for a day of ambulatory monitoring. They returned the next day for participation in the laboratory protocol. Upon their return, it was verified that all the measurement equipment was still in working order.

Next, participants were informed that footage of their facial expressions, posture and voice would be recorded during the experiment. Furthermore, the participants were informed that during the tasks, including the $SSST_{short}$, the experimenter would monitor their performance through a one- way mirror to ensure good compliance and quality of the recordings. Then all experimental manipulations were presented in a fixed order (see Table 1).

After the experimental session, all devices were removed, and participants were provided the option to use a nearby shower. The experiment ended with a debriefing in which they were informed that the TA, PASAT and Raven tasks were purposefully made so difficult so that they would be impossible to perform without errors. They were explicitly told that the test score rankings were only added to increase the stressfulness of the task and did not reflect their actual ability, and their performance on the RPM test is no meaningful reflection of their intelligence. Furthermore, they were informed that their singing during the $SSST_{short}$ was not actually recorded and is not going to be studied by

conservatory students. Nevertheless, the best performers on the TA and PASAT tasks were rewarded with an extra 50 euros, as promised.

Analytic strategy

Data inclusion and quality

To assess data quality the average percentage of artifact free signal per participant was calculated for both EDA signals: a condition of a participant was considered useable for analysis when the duration of valid data in the condition was at least 30 seconds, and when at least 20% of the signal of the entire experimental condition was artifact free for both DTI5 and VU-AMS signals. Otherwise, the data for the whole condition was rejected. We decided to only include participants that had at least 3 useable conditions.

Under classical signal detection theory, we expect a lower EDA level in the wrist signal compared to the palmar signal since the density of sweat glands is ~5 times larger on the palm than on the wrist. The amount of detected peaks is therefore also expected to be lower on the wrist. A study by Payne et al. (2016) showed that only in 30% of the cases when an SCR occurred at the fingers there was a simultaneous SCR at the wrist. However, during a stress task this percentage rose to 72% (Payne, et al., 2016).

To assess the extent in which EDA levels were very low, making it difficult to filter signal from noise, the percentage of participants in which the average SCL was below 0.5 µS was calculated for both EDA signals (Milstein & Gordon, 2020). Due to the lower EDA on the wrist, we also expect less ns.SCRs to be detected. The percentage of conditions where the number of ns.SCRs detected by either internal or the matlab method was zero, i.e. no detected peaks at all, was calculated and compared for the EDA signals of each participant.

Data alignment and reduction

For accurate device-to-device comparisons, we synchronized the DTI5 and VU-AMS recordings by temporally aligning the EDA signals to the maximal cross correlation between the tri-axial accelerometer signals of both devices. Next, we retained only data from the artefact free segments that fell within one of the experimental conditions.

In the present study the parameters of interest were defined as responses to short-term stressors and physical activities. Therefore for all wrist-based and palm-based SCL measures and the PEP, a mean value was generated across the same start and stop times for all conditions for each participant up to a total of 26 conditions, consisting of 3 posture conditions (lying, sitting, and standing), 4 first-exposure mental stressors, 2 repeated mental stressors, 6 daily life activities, a physical stressor consisting of 4

levels, and 7 recovery periods separating the stressors. For the ns.SCR measures, the frequency of the peaks for each of the conditions was retained. Outlier detection and removal was performed on these measures using a 3.5 SD criterion together with careful visual inspection of the histograms.

Multilevel analyses

Across participants, EDA data were available for analyses from at least 13 conditions, with an average of 25.8 within-subject observations. To take into account that these observations are nested within participants we performed multilevel (ML) analyses, also referred to as linear mixed models or hierarchical linear models. Although Bland-Altman plots have been suggested as the appropriate method for device comparisons (van Lier, et al., 2019), they are less suitable here, as we anticipate large between-subject differences in e.g. absolute SCL values at the palm and the wrist and are primarily interested in the correspondence of within-subject changes in e.g. palm-based and wrist-based EDA, wrist-based EDA and PEP, and wrist-based EDA and affect.

A basic two-level ML model can be represented by the following formula in which the outcome variable Y is a function of the intercept β_{0j}, a predictor variable X and a random error term (Blackwell, et al., 2006):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij.}$$

The lower level (level 1) is indexed by the subscript i and the higher level (level 2) by the subscript j. In this study the individual participants are treated as the level 2 unit, and the repeated measures across the various conditions within a participant as the level 1 unit.

ML analysis possesses a number of favorable characteristics suited well to our design. For instance, it does not require the number of repeated measures to be equal for all subjects and therefore is robust to missing data (assuming missingness at random). ML analysis can also explicitly test for the need to model inter-individual variation of the intercept and slope of the relationship between predictor and outcome. This means that each individual participant has its own intercept and slope coefficient value. Therefore, the β_{0j} and β_{1j} coefficients that are predicted by the model can be further broken down into a mean intercept Υ_{00} and mean slope Υ_{10} with deviations from that mean U_{0j} and U_{1j} :

$$\beta_{0j} = \Upsilon_{00} + U_{0j}$$

$$\beta_{1j} = \Upsilon_{10} + U_{1j}$$

For each model specified it can be tested whether allowing the intercepts and slopes to vary improves the fit of the model. However, to allow for direct comparison of the different analyses, all models were run including a random intercept and random slope even if this did not improve model fit. Cross-level interaction effects between our level 1 predictor variable with age, biological sex and BMI were tested by adding all interactions to a single model. However, none of the analyses showed an interaction effect of $p < .05$ with any of the level 1 variables rendering them obsolete.

Predictor variables were participant-mean centered. To do so the mean predictor value over all experimental conditions was calculated for every participant. This participant-specific mean was then subtracted from this participant's observed values during all experimental conditions. By centering, the intercept of each individual participant can be interpreted as the expected value of the outcome when the predictor values equal their own mean score.

For all outcome variables the total variance across conditions and participants was calculated, as well as two intra-class correlations (ICC) representing the amount of total variance that could be explained by inter-individual differences, and the amount of variance that could be explained by the experimental manipulations.

Our main validation analyses revolve around the prediction of a criterion outcome by the wrist-based EDA measures. In these analyses we are primarily interested in the proportion of variance in our outcome variable explained by the predictor variable. Because no standard solution is available for calculating this explained variance in a full ML model, we applied two strategies. First, the explained variance of the outcome by the predictor was calculated by the formula:

$$R^2 = \beta_{1j}^{\ 2} * \varepsilon_{ij}(predictor) / (\beta_{1j}^{\ 2} * \varepsilon_{ij}(predictor)) + \varepsilon_{ij}(model))$$

In which β_{1j} is the estimated slope of the full ML model, ε_{ij}(predictor) the residual variance of a ML model using a random intercept only, and ε_{ij}(model) the residual variance of the full ML model. This formula is an adjustment of the standard coefficient of determination, the proportion of the variance in the dependent variable that is predicted by the independent variable, used in linear regression. Secondly, we calculated the more intuitive within-subject correlation between the outcome and predictor for each individual participant separately and report the mean correlation and interquartile range, as well as the squared mean correlation as an approximation of the average proportion of variance in our outcomes that could be explained by the predictors across all participants.

All analyses were performed in R version 3.5.2. All ML analyses were performed using the packages lme4 and lmertest. Models were estimated under restricted maximum likelihood, with random intercepts and random slopes set as correlated and using the optimizer "nlminbwrap" to aid convergence problems. To test for autocorrelation effects the ML models were rerun with lme, of the nlme package, setting correlation to corAR(). The results with and without specified autocorrelation were almost identical. Therefore, the results presented in this paper are limited to models without autocorrelation. The threshold for significance was set to $p = .001$.

Correspondence between palm-based and wrist-based EDA measures

To test the correspondence between within-subject changes in classic VU-AMS palm-based EDA and the new DTI5 wrist-based EDA, ML regression analyses including all 26 experimental manipulations were performed for our EDA measures SCL and ns.SCR_mat and ns.SCR_cf. The VU-AMS EDA measures were added as outcome variables and DTI5 measures as the predictor variables.

Construct validity of palm-based and wrist-based EDA measures

In testing the effects of our experimental manipulations on SNS activity we focus on the classical reactivity contrasts of 'stress level compared to baseline level'. For the four EDA measures and the PEP we performed a ML model in which the experimental conditions are entered as a categorical variable. The baseline was specified as the contrast. The intercept estimate of this model represents the predicted value for the baseline category, and the estimate of all the conditions represent the deviations of these conditions from the baseline with the p-value specifying whether the difference is significant, very much like a repeated measures ANOVA.

Because posture itself has an effect on SNS activity, two separate analyses were performed. One for the mental stress tasks, which were all performed while sitting and therefore have the sitting quietly posture condition as baseline, with ten contrasts: four first exposure mental stressors, two repeated mental stressors and four recoveries. And one for the physical stress tasks, which were all performed upright and therefore have the standing quietly posture condition as baseline, with eleven contrasts: six daily life activities, one recovery and four-levels of physical stress on a treadmill).

Criterion validity of palm-based and wrist-based EDA measures

In order to test whether changes in the skin-based measures of SNS activity show the same pattern as cardiac measures of SNS activity, we performed ML regression analyses

using the cardiac measure PEP as the outcome variable and the EDA measures of both the VU-AMS and DTI5 as the explanatory variables. We excluded the lying down condition because PEP is known to be sensitive to the large preload effects in this posture (Houtveen, et al., 2005), leaving 25 conditions for the PEP-EDA comparisons. Earlier work has shown that PEP was significantly correlated with the EDA measures SCL and ns.SCR frequency, particularly when both showed large variation due to the inclusion of physical stressors (Goedhart, et al., 2008). Because the current study used multiple mental and physical stressors, a sensitivity analysis was performed by repeating the analyses separately for the mental stressors, including the four first exposure mental stressors, two repeated mental stressors and four recoveries, and the physical stressors, including the six daily life activities, one recovery and four-levels of physical stress on a treadmill.

Predictive validity

To test whether changes in EDA measures could predict concurrent changes in affect induced by our experimental manipulations we performed ML regression analyses using positive and negative affect as the outcome variable and the five EDA (SCL - palm & wrist, ns.SCR - palm, wrist ns.SCR_mat, and wrist ns.SCR_cf) measures as the predictor variables. Due to the potential effects of the long physical activity session on affect, mixed with potential effects of task habituation, we limited the scope of our analysis to the baseline mood report and the 4 reports taken after the *first exposure* to the mental stress tasks, which all took place before any of the physical stressors. For comparison, the predictive validity of the PEP is also given.

Power calculation

To determine the power to detect an effect we use a regression relationship between X (predictor) and Y (dependent) using T = ~ 26 repeated measures (level 1), that are nested in individuals (level 2). We tested the power to reject the hypothesis that there is no relationship between X and Y at all. We determined power with a sample size of N=120 individuals, R^2 of 1 % and a p-value of $p = .01$. The power calculations to reject H-null = .997. Further specifications of the power calculation can be found in Appendix 1: Full power calculations.

RESULTS

Study population

A total of 121 healthy young adults (56% females, age range = 18 – 32, mean age = 22.3, SD = 3.3) with a mean BMI of 23.6 (SD = 2.9) participated in the study. The majority were in a schooling trajectory (81%) and right-handed (84.3%, 5.8% ambidex), 76.9% were non-smokers and 64.4% reported to exercise on at least a weekly basis. The mean POMS total mood disturbance score was 22.9 (SD = 12.8), the mean of the subscales tension = 8.9 (SD = 3.1), vigor = 14.9 (SD = 3.4), fatigue = 10.5 (SD = 3.8), anger (8.8, SD = 2.8), and depression = 9.5 (SD = 3.3).

Four participants did not have VU-AMS recordings for EDA or PEP because of a failure of the memory card. Six participants had a too low ICG quality to be included in the criterion validity analyses. Eight participants did not have DTI5 wrist-based data for the following reasons: 1) DTI5 battery was insufficiently charged (N = 2); 2) DTI5 removed because of participant discomfort (N = 1); 3) DTI5 recording error (N = 5). Nine of the above participants overlapped, in that they suffered from multiple sources of data loss (e.g., low ICG quality and DTI5 removed). This resulted in a population of 112 participants that could be included in the analyses (57% females, age range = 18 – 32, mean age = 22.3, SD = 3.4).

Data quality

On average the length of the active experimental conditions added together was 76 minutes. This excluded down-time between conditions. Because some conditions were not performed or shortened for certain individuals (e.g., speed of the treadmill was too high for their fitness level) gross recording lengths ranged from 67 minutes to 79 minutes. Data quality of both devices was good. On average 86.5% of the recorded palm EDA signal was artefact free, while 88.8% of the recorded wrist EDA signal was considered artefact free. When assessing the occurrence of low absolute levels of skin conductance, we observed that in 14.9% of the participants the average wrist SCL was below 0.5 µS. This is considerably less problematic than suggested in a previous study that found that 73% of all wrist EDA data was below 0.5 µS (Milstein & Gordon, 2020). Structural low SCL did not occur on the palm.

For the wrist, 23.1% of the participants had an absence of ns.SCR_mat in more than half of the experimental conditions (>= 13), this was only 9.9% for ns.SCR_cf. This is more than at the palm where all participants had at least one ns.SCR in the majority of experimental conditions, only 1.6% of the participants had an absence of ns.SCRs in less than 3 out of 26 conditions. In 28.4% of all observations the wrist did not detect a single ns.SCR_mat,

while there was at least one ns.SCR detected at the palm. For ns.SCR_cf this was 23.1%. The better performance of the curve fit method (ns.SCR_cf) with respect to peak detection can be explained by the optimization of the curve fit method for the detection of peaks at the wrist specifically, taking into account the lower absolute level of conductance and the morphology of motion artefacts at that location.

Figure 3 shows that most of the experimental conditions in which peaks were only detected at the palm and not at the wrist (using the matlab method, the results for the curve fit method where highly similar and are shown in Supplementary Figure 1) are during conditions with no or low physical activity. We believe that this effect is driven by the build-up of moisture seen at the wrist in the more physically demanding conditions. During physically non-engaging activities there is usually low sweat production at the wrist. When sweat levels are low, no moisture build-up has taken place between the sensors and the skin. This makes it very difficult for the sensors to detect the ns.SCRs despite them being present. This dependency on sufficient amounts of sweat to detect EDA is a known disadvantage of dry electrodes. Figure 4 shows an example of a wrist and a palm EDA signal for a single participant during three different conditions.

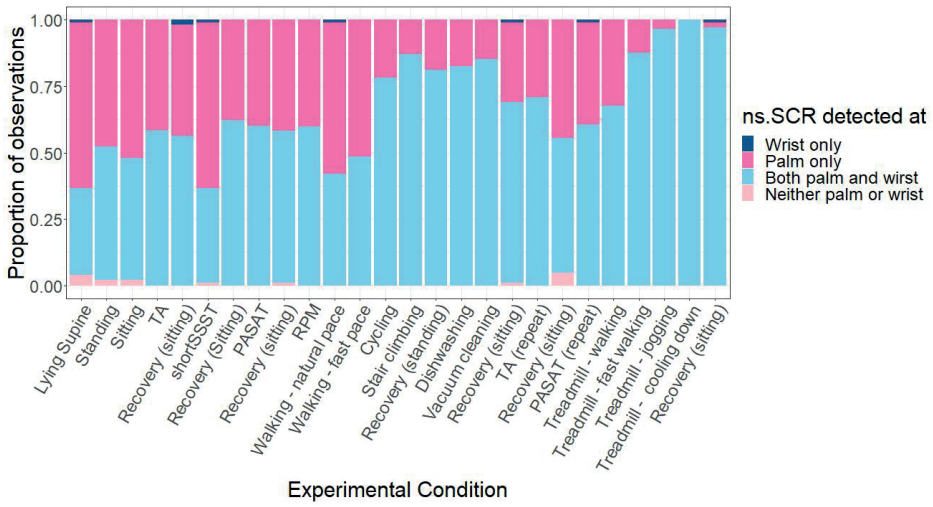


Figure 3. Percentage of participants that had at least one ns.SCR_mat detected at wrist, palm, both, or neither, separately per experimental condition.

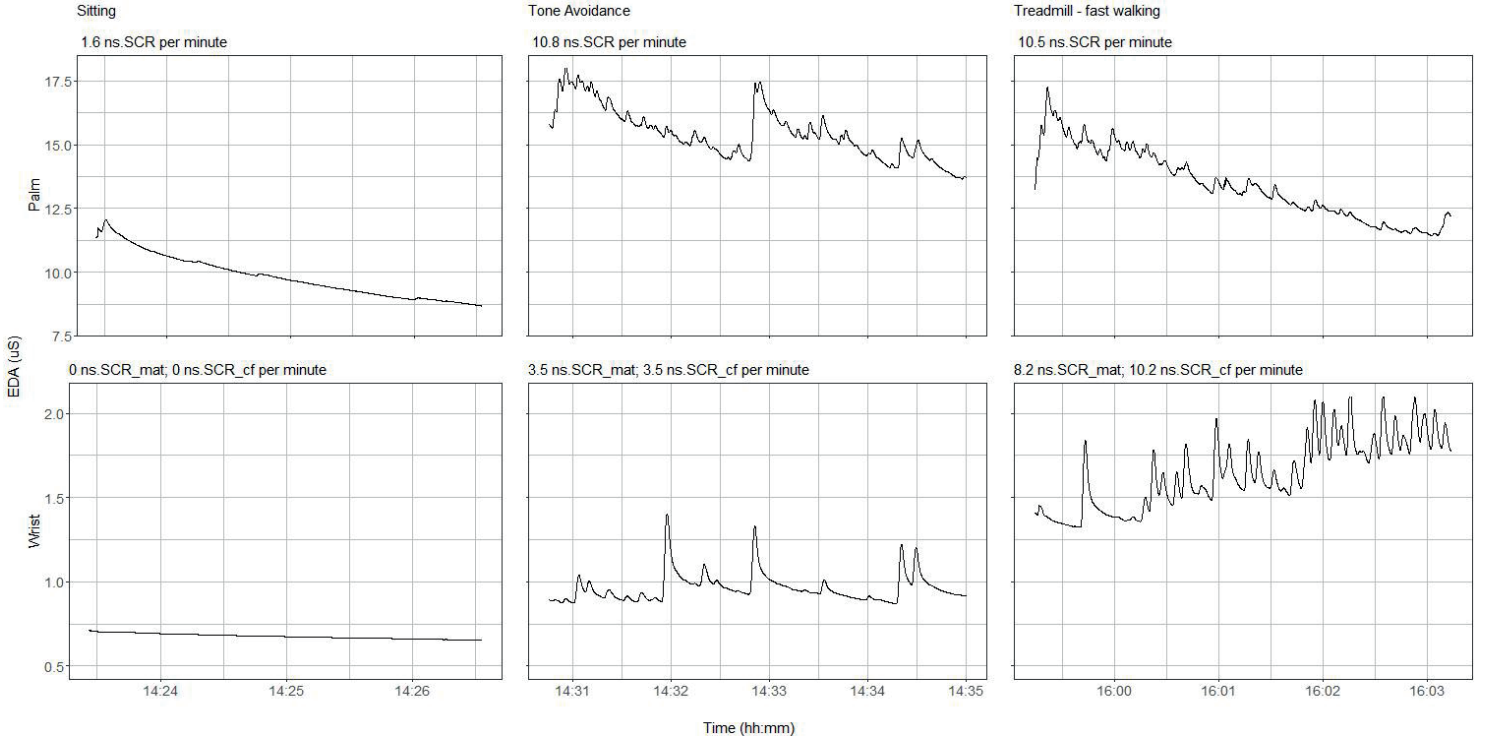


Figure 4. Direct comparison of EDA on the palm and the wrist of a single participant during rest (sitting), a mental stressor (Tone Avoidance task) and a physical stressor (walking on a treadmill at a fast pace.

Note: The top panels show the EDA recording of the palm, the bottom panels the recording on the wrist.
Above each plot the ns.SCR frequency in peaks per minute is reported. The Y-axis of the left panels applies also to the middle and right panels.

Table 2. Correspondence between palm and wrist-based EDA.

Palm	Wrist	N	β_{0j}	U_{0j}	B_{1j}	U_{1j}	SE	*p*	Model R^2	Correlation		
										R^2	M	IQR
SCL	SCL	108	14.32	7.35	0.83	0.99	0.12	**<.001**	14%	10%	.31	.05 – .59
ns.SCR	ns.SCR_mat	109	11.53	2.26	0.44	0.11	0.02	**<.001**	20%	18%	.42	.34 – .58
ns.SCR	ns.SCR_cf	107	11.50	2.24	0.32	0.06	0.01	**<.001**	19%	19%	.43	.33 – .56

β_{0j} and B_{1j} are the average intercept and slope of the regression of wrist-based on palm-based measures, SE, and p values of the ML model with random intercept and slope show that individual variation around both (U0j and (U1j) are significant. Model R^2 is the percentage of within-subject variance explained in palm EDA by wrist EDA. This value is also approximated by the correlation R^2 which derives from squaring the mean (M) within-subject correlations, the range in which is indicated by the IQR. Significant results are depicted in bold.

Correspondence

The ICC analysis showed that inter-individual differences explained 80.7% of the variance in palm SCL, while only 5.8% was explained by the experimental manipulations. For wrist SCL, 58.4% and 18.7% of the variance was explained by inter-individual differences and experimental manipulations, respectively.

For palm ns.SCR frequency, 15.0% of the variance was explained by inter-individual differences and 46.3% by the experimental manipulations. For wrist ns.SCR frequency, 16.1% of the variance in ns.SCR_mat and 3.9% of ns.SCR_cf was explained by inter-individual differences and 52.0% of variance in ns.SCR_mat and 63.9% in ns.SCR_cf by the experimental manipulations, respectively. In general, we find that inter-individual differences explain the largest part of variation in SCL, whether at palm or wrist, whereas the experimental manipulations are the major source of variance for the ns.SCR frequency at both locations.

To test the correspondence between the EDA measures from different locations, we predicted the palm-based EDA measures by their wrist counterparts. There is a significant correlation between wrist and palm EDA (Table 2). Of the variance in palm SCL 14% could be explained by wrist SCL, while 20% of the variation in palm ns.SCR frequency could be explained by wrist ns.SCR_mat and 19% by ns.SCR_cf.

Table 3. Experimental manipulation of SNS activity.

	SCL (µS)		ns.SCR (pm)			PEP (msec)
	Palm	Wrist	Palm	Wrist *mat*	Wrist *cf*	Heart
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Exp. Condition						
Baseline Sitting	**9.92 (5.74)**	**2.02 (2.01)**	**5.11 (3.49)**	**0.93 (2.03)**	**0.59 (1.46)**	**112.48 (15.35)**
TA	13.39 (7.18) ▲	2.00 (2.02) =	15.58 (4.67) ▲	1.95 (3.07) ▲	1.86 (2.97) ▲	105.94 (16.26) ▲
Recovery	13.12 (7.93) –	1.97 (1.99) =	7.34 (4.24) –	1.43 (2.04) – –	0.94 (1.59) –	112.66 (15.55) –
$SSST_{short}$	14.75 (7.91) ▲	1.84 (1.88) =	13.55 (4.68) ▲	2.12 (4.22) ▲	2.06 (4.49) ▲	102.93 (17.91) ▲
Recovery	13.00 (7.19) –	1.88 (1.84) =	7.65 (4.29) –	1.80 (2.68) –	1.35 (2.44) –	110.35 (15.19) –
Pasat	14.56 (7.84) ▲	1.78 (1.79) =	13.85 (3.80) ▲	1.74 (2.70) ▲	1.43 (2.13) ▲	104.20 (18.05) ▲
Recovery	13.46 (7.73) –	1.74 (1.75) =	7.72 (4.29) –	1.71 (2.56) –	1.31 (2.59) –	111.69 (15.71) –
Raven	13.68 (7.46) ▲	1.71 (1.72) =	9.74 (3.55) ▲	1.27 (2.30) =	1.02 (2.36) =	108.70 (16.96) ▲
TA repeat	15.91 (8.46) ▲	2.60 (2.50) ▲	12.04 (5.59) ▲	2.57 (3.38) ▲	2.96 (3.61) ▲	105.60 (16.50) ▲
Recovery	14.00 (7.77) –	2.47 (2.37) ▲	5.85 (4.07) –	1.70 (2.67) –	2.68 (3.51) –	111.00 (15.89) –
Pasat repeat	15.97 (9.11) ▲	2.33 (2.18) =	11.55 (5.21) ▲	1.90 (2.87) ▲	2.86 (3.56) ▲	107.26 (15.68) ▲
Baseline Standing	**10.54 (5.87)**	**2.09 (2.18)**	**8.08 (3.89)**	**1.25 (2.47)**	**1.29 (2.64)**	**115.19 (17.98)**
Walking own pace	13.76 (7.84) ▲	1.61 (1.62) ▼	11.52 (3.83) ▲	1.07 (2.69) =	4.19 (4.93) ▲	93.45 (13.98) ▲
Walking fast pace	13.92 (7.54) ▲	1.59 (1.61) ▼	13.08 (3.79) ▲	1.60 (3.30) =	5.30 (5.47) ▲	83.93 (12.34) ▲
Cycling	12.13 (6.65) ▲	1.56 (1.54) ▼	14.32 (3.28) ▲	5.34 (5.54) ▲	6.47 (5.64) ▲	80.52 (12.89) ▲
Stair climbing	13.29 (6.93) ▲	1.90 (1.70) =	14.00 (3.39) ▲	7.14 (6.29) ▲	8.97 (7.70) ▲	75.04 (10.17) ▲
Standing recovery	13.38 (7.43) =	2.65 (2.31) +	10.19 (4.83) –	6.74 (5.31) –	9.40 (6.37) +	103.04 (20.43) –

Table 3. Experimental manipulation of SNS activity. (continued)

	SCL (µS)		ns.SCR (pm)			PEP (msec)
	Palm	Wrist	Palm	Wrist *mat*	Wrist *cf*	Heart
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Dish washing	12.92 (6.90) ▲	2.85 (2.62) ▲	16.85 (3.66) ▲	9.66 (7.06) ▲	5.98 (5.12) ▲	104.80 (21.12) ▲
Vacuuming	13.68 (7.22) ▲	2.77 (2.48) =	15.92 (4.60) ▲	8.26 (7.13) ▲	4.59 (4.19) ▲	93.57 (16.94) ▲
Treadmill walking	15.48 (8.14) ▲	2.25 (2.08) =	13.58 (3.80) ▲	2.69 (4.34) ▲	6.90 (6.87) ▲	89.38 (15.86) ▲
Treadmill fast walking	15.59 (7.96) ▲	2.39 (2.15) =	15.54 (2.89) ▲	5.31 (5.06) ▲	11.00 (6.35) ▲	78.40 (12.10) ▲
Treadmill jogging	17.63 (8.03) ▲	3.39 (2.43) ▲	18.52 (2.26) ▲	15.12 (5.52) ▲	20.74 (7.31) ▲	71.65 (10.22) ▲
Treadmill cool down	18.10 (8.19) ▲	5.28 (2.84) ▲	16.94 (4.74) ▲	14.71 (4.78) ▲	22.70 (6.02) ▲	75.56 (10.72) ▲

Note: for PEP higher levels indicate less SNS activity. So levels that are lower compared to baseline are marked as an increase in SNS activity

▲ *significant increase in SNS compared to posture-specific baseline*

▼ *significant decrease in SNS compared posture-specific to baseline = no significant change from stress level*

\+ *unexpected significant increase in SNS compared to stress level*

– *expected significant decrease in SNS compared to stress level, but incomplete recovery compared to baseline*

Construct validity

The results of the construct validity analyses are shown in Table 3. Clear confirmation of successful experimental manipulation of SNS activity comes from the changes in PEP across conditions. Both mental and physical stressors lead to a decrease in the PEP over the appropriate sitting and standing baselines. During recoveries, the PEP values bounce back to the baseline values, with the exception of standing recovery after stair climbing where an increase in SNS activity remains evident.

All experimental effects on the palm EDA measures were in the expected direction. Both mental and physical stressors increased SCL and ns.SCR over the appropriate sitting and standing baselines. During recoveries, values decrease compared to the previous stress level, although they remain elevated compared to the baseline. When comparing the beginning and ending of the experiment there is no sign of a clear drift in the palm SCL signal that keeps a steady average of around 13 µS with reliable increases during experimental manipulations.

Results for wrist SCL do not show the expected pattern, with SCL showing (non-significant) decreases rather than increases during exposure to the mental stressors, and SCL even decreased in response to some of the physical stressors indicating a clear drift in the signal over time. Furthermore, absolute SCL levels on the wrist were considerably lower compared to the palm. In contrast, results for wrist ns.SCR are again very consistent with the expectations in both mental and physical stressors for both methods: Generally, ns.SCR frequency increases over the appropriate sitting and standing baselines and during recoveries the ns.SCR frequency is seen to drop compared to the previous stress level. This indicates that ns.SCR on the wrist can track SNS activity independent of thermoregulatory need. The exception was the cool-down phase of the treadmill protocol were SNS activity should have abided but ns.SCR frequency was seen to remain high at both palm and wrist. This exception is likely to reflect ongoing thermoregulatory sweating to restore core body temperature, which is known to be prolonged after exercise cessation (Kenny & McGinn, 2017).

Figure 5 presents a direct visual comparison between the mean ns.SCRs of the palm and wrist over the course of the experiment.

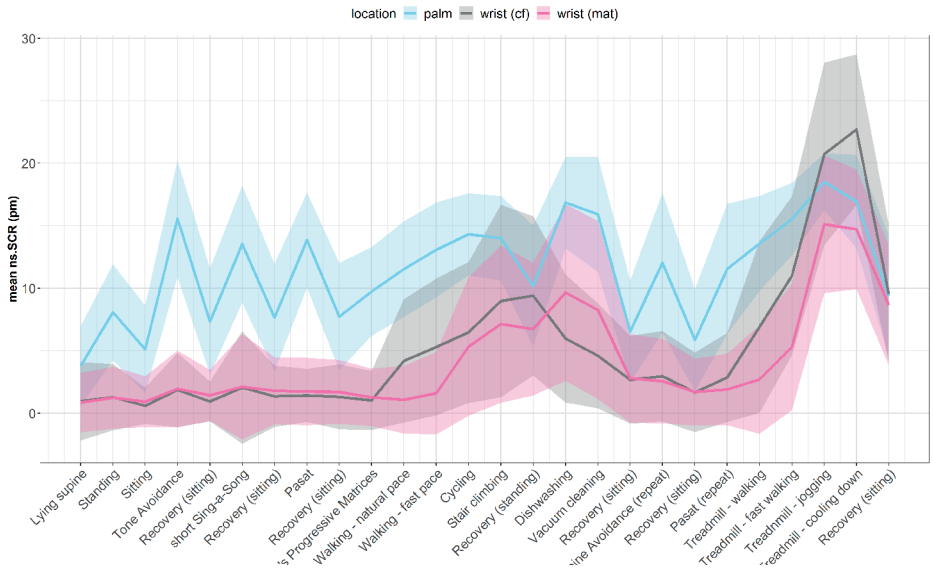


Figure 5. Mean palm and wrist ns.SCR reactivity to the experimental manipulations.

Note: Shaded areas around the lines represent the standard deviations around the means.

Criterion validity

The ICC showed that 32.1% of the variance in the PEP could be explained by inter-individual differences and 42.9% by the experimental manipulations. To test whether electrodermal measures of SNS activity show the same pattern as cardiac measures of SNS activity, we compared the EDA measures of the palm and wrist to the PEP. As expected, all EDA measures showed a significant negative relationship with the PEP (Table 4), with a lower PEP being associated with higher EDA. When comparing the available EDA measures, for both palm and wrist EDA variability in PEP was best explained by ns.SCR frequency, in which palm ns.SCR frequency explained 21% of the variability in PEP and wrist ns.SCR_mat frequency explained 14% and ns.SCR_cf explained 25% of the variability. SCL correlated more poorly to PEP, particularly SCL at the wrist. The sensitivity analyses in Supplementary Table 1 showed that overall these relationships are attenuated but still present when computed across the mental and physical stressors separately.

Table 4. Correspondence between PEP and EDA.

EDA	Location	N	β_{0j}	U_{0j}	β_{1j}	U_{1j}	SE	*P*	R^2	Correlation		
										R^2	M	IQR
SCL	Palm	108	100.15	11.74	-1.50	1.83	0.22	**< .001**	10%	6%	-.24	-.48 – .00
	Wrist	103	99.00	11.34	-1.19	1.02	0.28	**< .001**	1%	1%	-.08	-.26 – .06
ns.SCR	Palm	109	100.19	11.59	-1.52	0.74	0.09	**< .001**	21%	20%	-.44	-.60 – -.32
	Wrist *mat*	103	98.92	11.34	-1.26	0.61	0.09	**< .001**	14%	13%	-.36	-.54 – -.27
	Wrist *cf*	102	98.58	11.46	-1.23	0.61	0.07	**< .001**	25%	22%	-.47	-.65 – -.33

β_{0j} and B_{1j} are the average intercept and slope of the regression of the PEP on the EDA measures, SE, and p values of the ML model with random intercept and slope show that individual variation around both (U_{0j} and (U_{1j}) are significant. Model R^2 is the percentage of the within-subject variance in the PEP explained by the EDA measures. This is also approximated by the Correlation R^2 which derives from squaring the mean (M) within-subject correlations, the range in which is indicated by the IQR. Significant results are depicted in bold.

Predictive validity

Figure 6 shows that all first exposures to the mental stressors significantly increased negative affect and decreased positive affect. We found that 43.0% of the variance in positive affect could be explained by inter-individual differences and 15.3% by the experimental manipulations, while 68.4% of the variance in negative affect could be explained by inter-individual differences and 6.8% by the experimental manipulations.

As shown in Table 5, positive and negative affect were significantly related to palm SCL, palm ns.SCR frequency, and wrist ns.SCR frequency (with an exception for ns_SCR_mat which only showed a trend for negative affect), but not to wrist SCL.

The relationships between EDA and affect were in the expected direction with a higher SCL and ns.SCR being associated with lower positive affect and higher negative affect (Cacioppo, et al., 1993; Nikula, 1991; Boucsein, 2012). However, the explained variance was very low, in part because the variation in this Likert type scale was very low, and R^2 of the ML model converged to zero (not shown). When we approximate explained variance by the squared within-subject correlations only palm SCL and ns.SCR frequency explained a meaningful part of the variance in affect (6% - 14%).

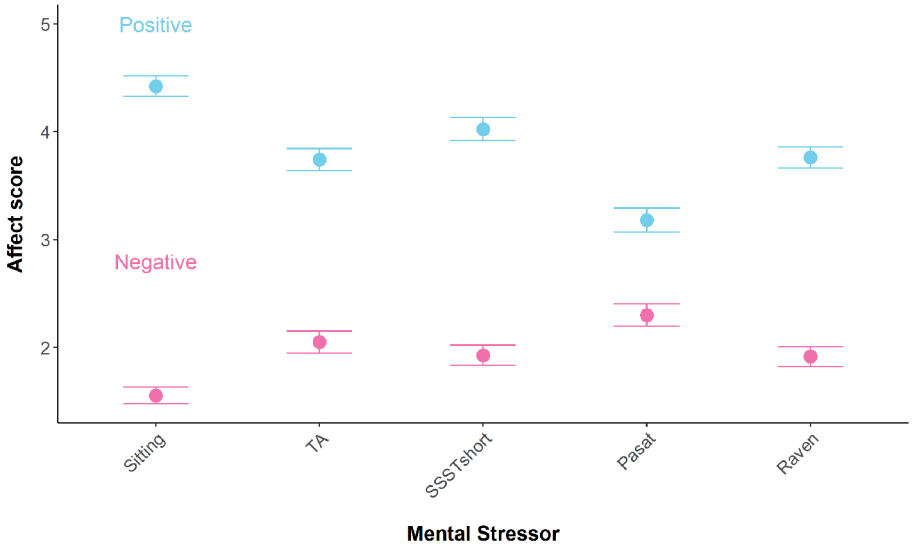


Figure 6. Experimental manipulation of affect.

Note: all first exposure stress tasks significantly decreased positive affect ($p < .001$) and increased negative affect ($p < .001$). *Dots represent the mean; error bars represent the standard error of the mean.*

Table 5. EDA predicting affect.

EDA	Affect	Location	N	β_{0j}	U_{0j} SD	B_{1j}	U_{1j} SD	B_{1j} SE	*p*	Correlation		
										R^2	M	IQR (r)
PEP	**Positive**	Heart	107	3.76	0.75	0.00	0.00	0.00	.10	3%	.16	-.23 – .63
	Negative	Heart	107	2.01	0.79	-0.01	0.00	0.00	**<.001**	5%	-.30	-.60 – .09
SCL	**Positive**	Palm	112	3.76	0.70	-0.05	0.01	0.01	**<.001**	6%	-.24	-.68 – .07
		Wrist	109	3.86	0.74	0.11	0.21	0.08	.19	3%	.17	-.21 – .56
	Negative	Palm	112	2.01	0.85	0.04	0.03	0.01	**<.001**	9%	.29	-.04 – .65
		Wrist	109	1.95	0.85	-0.03	0.03	0.06	.64	1%	-.08	-.50 – .23
ns.SCR	**Positive**	Palm	112	3.81	0.73	-0.05	0.02	0.00	**<.001**	10%	-.31	-.72 – .03
		Wrist *mat*	109	3.66	0.73	-0.05	0.00	0.01	**.001**	2%	-.13	-.45 – .19
		Wrist *cf*	106	3.56	0.72	-0.06	0.00	0.01	**<.001**	3%	-.17	-.58 – .19
	Negative	Palm	112	1.98	0.87	0.04	0.03	0.00	**<.001**	14%	.37	.12 – .70
		Wrist *mat*	109	2.09	0.85	0.01	0.05	0.01	**.005**	3%	.16	-.27 – .53
		Wrist *cf*	106	2.19	0.95	0.05	0.06	0.01	**.001**	3%	.29	-.25 – .61

β_{0j} and B_{1j} are the average intercept and slope of the regression of the EDA measures and affect, SE and p values of the ML model with random intercept and slope show that individual variation around both (U_{0j} and (U_{1j}) are significant. Because the variance in the Likert-scale based Affect outcomes was low, Model R^2 did not produce interpretable values, so we just report the approximation of the percentage of within-subject variance in affect explained by the EDA measures by the Correlation R^2. This R^2 derives from squaring the mean (M) within-subject correlations, the range in which is indicated by the IQR. Significant results are depicted in bold.

DISCUSSION

In an extensive controlled laboratory study in 112 participants we recorded two measures of skin SNS activity, SCL and ns.SCR frequency, using both wrist-based dry electrodes and classical palm-based wet electrodes. Throughout we find that the variance in absolute SCL at both palm and wrist is mainly determined by between-subject differences but only weakly by experimental manipulations even if these induced a large range of SNS activity. In contrast, variance in ns.SCR frequency at the palm and the wrist was predominantly governed by experimental conditions. The ns.SCR frequency therefore seems a superior measure than SCL to detect within-subject changes in SNS activity across conditions. In addition, whereas both SCL and ns.SCR frequency may tackle relevant individual specific factors like chronic stress or personality traits, SCL may also entail anatomical features like the number of sweat glands per mm^2 skin, exact electrode positioning, hydration status at the time of recording, and other factors that are usually not germane to psychophysiological research. Those factors seem to plague the non-specific SCRs to a lesser degree.

The analysis of the correspondence between palm and wrist measures also favored ns.SCR frequency, as the explained variance in palm EDA by wrist EDA was larger for ns.SCR than for SCL (20% vs. 14%), although both were modest in keeping with findings for SCL from previous studies (Milstein & Gordon, 2020; Menghini, et al., 2019). The preferred use of ns.SCR frequency over SCL is further supported by a comparison of the performance of ns.SCR vs. SCL in the validity tests. Construct validity was higher for ns.SCR frequency than for SCL most notably at the wrist, and the criterion validity, using the PEP as the criterion for SNS activation, was also better for ns.SCR frequency at both palm and wrist. Furthermore, we found stronger predictive validity for changes in positive and negative affect using ns.SCR frequency than using SCL, and for the wrist only ns.SCR frequency was a significant affect predictor, although the explained variance was low for all signals (<14%). Finally, for wrist EDA the dependency of absolute SCL on thermoregulation was observed to a much lesser extent for the wrist-based ns.SCR frequency. Although thermoregulation remains a powerful co-determinant, our results for wrist ns.SCR frequency bolster the accumulating evidence (Machado-Moreira & Taylor, 2012; van Dooren et al., 2012) that refutes the traditional idea that detection of emotional sweat gland responding is confined to the palmar and plantar skin surface whereas only thermal sweating evokes responses of the sweat glands across other parts of the body (Dawson et al., 2000; Edelberg, 1967; Ogawa, 1975). Evidence of mental stress effects on the sweat glands at the wrist can, however, be detected only by ns.SCR frequency, and is indeed not seen in the SCL.

Taken together, our results lead us to conclude that ns.SCR frequency is a more suitable measure than SCL for prolonged ambulatory recording of SNS activity. We now turn to the question of whether the less invasive wrist-based recordings of ns.SCR can sufficiently capture SNS activity or whether the more obtrusive palm-based recordings are needed. In wet electrodes, the use of electrolyte cream on the palm increases conductance, while the dry electrodes are dependent on the presence of sweat to act as an electrolyte between the electrodes and the skin. This leads to much lower levels of absolute SCL, which in turn could make it more difficult to detect ns.SCRs. Indeed, in around 15% of the participants the average SCL on the wrist was below 0.5 µS, while this low level did not occur on the palm. While substantial, this percentage is considerably less problematic than suggested in a previous study that found that 73% of all wrist EDA data was below 0.5 µS (Milstein & Gordon, 2020). Even so, we did find that the absolute number of ns.SCR peaks detected was considerably higher at the palm than at the wrist. The issue of a lower number of EDA responses at non-palmar sites is a longstanding one (Rickles & Day, 1968) and was investigated by Payne et al. (2016). In two small samples of students they noted that only 16% to 31% of the SCRs to emotionally salient pictures at the palm were simultaneously

detected at the wrist, i.e. 69% to 84% of the orienting response induced palm SCRs were not detected at the wrist (Payne, Schell, & Dawson, 2016). This is likely related to the lower amount of eccrine sweat glands on the wrist (Harker, 2013) but also the use of wet vs. dry electrodes.

In spite of the lower absolute number of ns.SCRs, good construct, and criterion validity for wrist-based ns.SCR was found. Mental and physical stressors, followed by recovery periods induce the expected changes in ns.SCR and stressor-induced decreases in PEP are significantly associated with increases in the ns.SCR frequency for wrist-based measurements. About 14-25% of the variance in PEP was recaptured by the wrist-based ns.SCR measures. Also for palm-based ns.SCR this value is low at 21%. This may appear modest if we presume the SNS to always act as a completely unitary system with tightly parallel changes in outflow to all organs at once. However, such a unitary SNS response is unlikely to occur. Direct sympathetic nerve activity recordings and noradrenaline spillover studies have shown substantial regional specificity of SNS activity (Wallin, 2004) that would allow the SNS activity to skin and heart to be less than perfectly correlated. Moreover, various other factors will act to reduce the PEP – EDA correlation even if SNS activity to all organs was perfectly aligned: PEP is sensitive to preload and afterload effects (Lewis, et al., 1977) and the indirect action of circulating levels of catecholamines on the ventricular β1 and β2 receptors. With these caveats on 'unitary SNS activity to all organs' in mind, the correlation found here between the PEP and the wrist- and palm-based ns.SCR measure is in the expected range.

A previous study had suggested that the ns.SCR-PEP relationship might be seen only when using a wide range of SNS activity, i.e. by adding intense physical exercise (Goedhart, et al., 2008). For stress researchers, the more relevant question is whether changes in skin and cardiac SNS activity are also correlated during exposure to mental stressors. We therefore repeated the analysis separately for the mental stress and recovery periods (all sitting), excluding all physical active conditions. These analyses showed that changes in ns.SCR were still related to changes in the PEP althaugh the effect was strongly attenuated.

Regarding predictive validity, we found significant within-subject correlations between changes in affect induced by the mental stressors and changes in both palm and wrist ns.SCR frequency. Consistent with the literature, higher ns.SCR was associated with decreased positive affect and increased negative affect (Cacioppo, et al., 1993; Nikula, 1991; Boucsein, 2012). The mean correlation for wrist-based ns.SCR frequency with affect was very modest, at -.13 for positive affect and .16 for negative affect. Such low correlations are in keeping with a long history of modest relationships being reported

between affective states and physiology (Cacioppo, et al., 1993). This may in part reflect the restricted range of variance in affect. Although our mental stressors significantly lowered positive affect and increased negative affect, on average these changes amounted to about 1 point on a 7-point Likert scale suggesting that the induced stress was relatively mild, which was further corroborated by the average PEP reactivity of -6.5 ms across all mental stress tasks. This is a limitation of artificial laboratory stressors in general; they cannot fully reconstitute the more profound stress experienced in real life daily situations.

Taken together, the results of our validity analyses suggest that wrist-based ns.SCR frequency is a useful addition to the ambulatory psychophysiologist's toolkit. It responds to our experimental manipulations of SNS activity shows a decent overlap with parallel recorded cardiac SNS effects as measured by the PEP. Performance of wrist-based ns.SCR was in many aspects comparable to the more obtrusive palm-based ns.SCR, but the latter shows higher absolute levels of ns.SCR frequencies and remains superior in higher predictive validity for changes in affect. In controlled laboratory studies, palmar based EDA recording, therefore, remains the preferred method. The inherent limitations of wrist-based EDA recording should, however, be properly weighed against its huge advantages. Wrist-based ns.SCR frequency detection could feasibly be scaled up to epidemiology-sized studies including thousands of participants. In addition, prolonged recording for days to weeks and even months is possible, allowing the monitoring of daily SNS activity in relation to sleep and sleep quality (Sano & Picard, 2011; Sano, et al., 2014), academic performance (Zhang et al., 2018), weekly fluctuations in work stress exposure, and longer-term mood regulation in naturalistic social settings (Sano et al., 2018; Weise, et al., 2013). In contrast to all other known 'pure' SNS measures wrist-based EDA has the potential to be employed as a biofeedback tool for just-in-time adaptive interventions (Heron, et al., 2017). Such interventions use early signs of stress from the physiological state of a client to time the provision of (smartphone-based) alerting and/or coaching to prevent the client from cascading into a chronic stress response. We therefore see our study as a strong justification for the further technical and methodological development of wristwatch-based recording of skin conductance as way to measure SNS responses to perturbations by mental, emotional, and physical stressors.

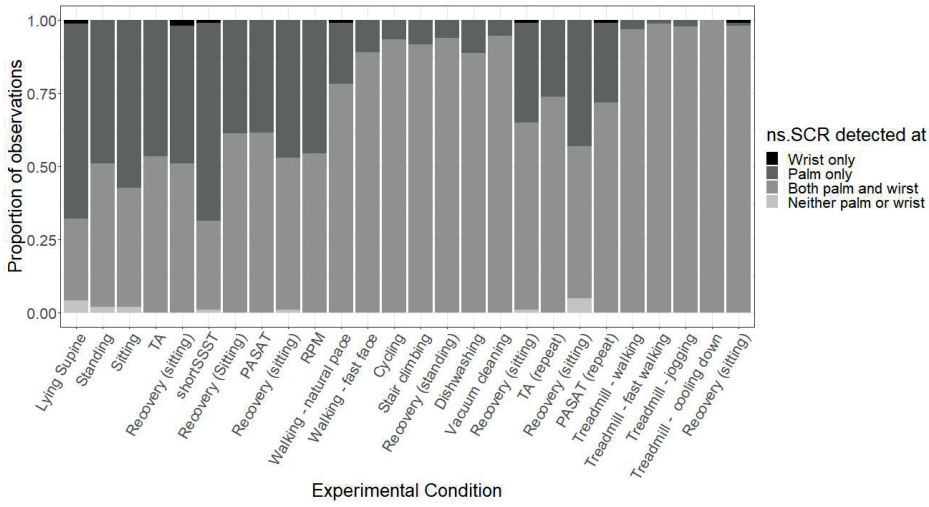
APPENDIX 1. FULL POWER CALCULATIONS

The a priori power calculations for ML models necessarily involve many choices concerning peripheral (nuisance) parameters (variance of the intercept, auto correlation of predictor X, autocorrelation of residual Y, amount of missing data), and concerning the focal parameters of interest (mean and variance of the slope). The most critical parameter is the within-person correlation. which we conservatively set to 0.1 for the fluctuations in EDA and fluctuations in 3MQ mood, i.e. an R2 of 1%. All other correlations, i.e. within the physiological domain across different instruments measuring the same variable (e.g., wrist-based or hand palm-based EDA) and across different variables measuring the same construct (e.g. EDA and PEP) are expected to be (much) higher. So, powering our study for the within-person EDA – mood correlation was expected to give a lower boundary for power. To determine the power to detect this effect we use a regression relationship between X (predictor) and Y (dependent) using T = ~ 26 repeated measures. In this regression model, we accommodate individual differences in the regression relationship (slope and intercepts) by adopting a two-level model, in which individual observations (X and Y; level 1 variables) that are nested in individuals (level 2).

We tested the power to reject two hypotheses. H-null 1, which states that on average there is no relationship between X and Y (mean(slope)=0), but X and Y may be related at the level of the individual (variance(slope)>0). H-null 2, which states that there is no relationship between X and Y at all. We determined power with a sample size of N=120 individuals, with 26 X and Y measures per individual. We set the intercept mean to equal 10 and the intercept standard deviation to equal 2.5. As X and Y are timeseries, we allowed for autocorrelation. The autocorrelation of the predictor was set to 0.5 and autocorrelation of the residuals to 0.7. In evaluating the power to reject H-null 1 and 2, the intercept is a random parameter not subject to testing, as it has no bearing on the relationship between X and Y. As missing data are often observed in time series models, we considered the power afforded by the complete data (no missing) and by the data with 40% missing. The power was established empirically by means of the analysis of simulated data. Each power calculation was based on 1000 replications.

Given an effect size of R^2 = 1% on average, the power to reject H-null 1 is > .999 (α = .05) and > .999 (α = .01), and the power to reject H-null 2 is >.999 and .999 (α = .05 and α = .01, respectively). Given 40% missingness the power remained similar: H null 1: .994 (α = .05) and .975 (α = .01); and H null 2: .999 (α =.05) and .997 (α = .01).

SUPPLEMENTARY TABLES AND FIGURES



Supp. Figure 1. Percentage of participants that had at least one ns.SCR detected at wrist, palm, or both, separately per experimental condition (curve fit method).

Supp. Table 1. Correspondence between PEP and EDA during mental and physical stressors.

EDA	Location	N	β_{0j}	U_{0j}	β_{1j}	U_{1j}	SE	*P*	R^2	Correlation		
										R^2	M	IQR
SCL	**Palm**											
	Mental	108	107.95	14.16	-0.48	0.78	0.15	**.003**	2%	3%	-.17	-.46 – .08
	Physical	105	88.42	10.67	-1.49	1.43	0.24	**< .001**	8%	7%	-.26	-.57 – ..06
	Wrist											
	Mental	103	107.62	14.41	0.49	1.55	.48	.32	0%	0%	.06	-.25 – .35
	Physical	101	86.82	9.97	-0.68	0.79	0.37	.086	0%	1%	-.12	-.33 – .09
ns.SCR	**Palm**											
	Mental	109	107.49	14.01	-0.57	0.35	0.07	**< .001**	9%	12%	-.34	-.61 – -.12
	Physical	107	90.47	10.41	-1.08	0.43	0.13	**< .001**	8%	8%	-.28	-.54 – -.09
	Wrist mat											
	Mental	103	106.64	14.30	-0.34	0.23	0.14	.03	0%	0%	-.08	-.32 – .20
	Physical	101	87.76	9.72	-0.39	0.10	0.08	**< .001**	2%	2%	-.15	-.45 – .13.
	Wrist cf											
	Mental	101	106.09	14.08	-0.37	0.19	0.12	**.007**	1%	0%	-.08	-.31 – .18.
	Physical	100	88.64	10.84	-0.59	0.21	0.06	**< .001**	10%	7%	-.37	-.57 – -.11.

β_{0j} and B_{1j} are the average intercept and slope of the regression of the EDA measures and the PEP, SE, and p values of the ML model with random intercept and slope show that individual variation around both (U_{0j} and (U_{1j}) are significant. Model R^2 is the within-subject explained variance in the PEP by the EDA measures. Correlation R^2 derives from squaring the mean (M) within-subject correlations, the range in which is indicated by the IQR. Significant results are depicted in bold.

CHAPTER 4

Comparing the relationship of physiology with affect across laboratory and real-life settings.

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. C. (2023). Comparing the relationship of physiology with affect across laboratory and real-life settings. Psychosomatic medicine *submitted February 2023.*

ABSTRACT

The debate on the ecological validity of laboratory-induced stress as a representation of real-life stress reactivity is ongoing. While there have been many studies on the generalizability of individuals' tendency to respond with low or high stress reactivity in laboratory and real-life settings, the extent to which the affect-ANS coupling in the laboratory generalize to real life settings is understudied. In the current study this cross-domain relationship is directly compared between a laboratory and daily life setting. Data was collected from the same individuals in both settings using measures of ANS activity (inter-beat-interval (IBI), respiratory sinus arrythmia (RSA), pre-ejection-period (PEP), and non-specific skin conductance responses (ns.SCR)) and affect (nine questionnaire items differing in valence and arousal). Multilevel modeling was used to analyze the relationship between ANS and affective valence and arousal in both the laboratory and daily life. The average fixed regression coefficients from these models were compared between the two contexts using a Z-test. The results of this study show that it is possible to validly measure the affect-ANS dynamics in a laboratory setting. This allows for the development of emotion prediction algorithms in a controlled and low-burden laboratory setting with a large number of individuals, which offers researchers the opportunity to optimize their methods before applying them in a real-life setting.

INTRODUCTION

The ecological validity of laboratory-induced stress as a proxy for real-life stress is hotly debated. This debate was already brought up in the mid-20th century by Donald T. Campbell (Campbell, 1957). Rapid developments in wearable technology over the past two decades greatly enabled empirical testing of the generalizability of laboratory stress responses to responses seen during stressful events in daily life. Most of the lab-to-real life studies test the generalizability of individual differences in the amplitude of the reactivity of physiological stress systems. The basic question is whether hyper- or hyporeactors to laboratory stressors are also characterized by parallel hyper- or hyporeactivity to real life stressors. The results generally show low correspondence between the amplitude of reactivity to laboratory and real-life stressors (Zanstra & Johnston, 2011). Correspondence increases to moderate when the laboratory and daily life stress conditions are matched better with regard to environmental and psycho-social factors. For example, when the laboratory and daily life are matched for demand and situational control (Kamarck et al., 2005) or by comparing a Trier Social Stress Test (TSST) to real public speaking (Henze et al., 2017).

While there is an abundance of studies on the lab-real life generalizability of the tendency to respond with low or high stress reactivity, the question that has not been addressed is whether the affect-ANS coupling in response to laboratory stress generalize to real life settings. An important cross-domain correlation for the theory of stress is the psycho-physiological link between stress-induced affective states and stress-induced changes in the activity of the autonomic nervous system (ANS). Such a correlation takes into account individual differences in the appraisal of the stressors during different contexts. In the classical laboratory design, the ANS and affective response to stressors are computed separately by comparing them to the baseline ANS and affective state using repeated measures AN(C)OVA. The correlation between the affective and ANS responses is rarely a focus of interest. Often the affective response is used as a manipulation check, whereas the amplitude of ANS stress reactivity is the main parameter of interest. Only a few studies directly investigated the strength of the relationship between co-occurring changes in ANS activity and affect in response to induced stress in the laboratory. Feldman and colleagues (1999) conducted a meta-analysis across nine studies and 16 tasks that induced negative affect by acute laboratory stressors. Despite the stressors increasing both the average negative emotions and ANS activity the across-participant correlation between these two responses was modest with only 2% to 12% of the variance in negative emotions being accounted for by physiology (Feldman et al., 1999). Similarly, significant

but very modest correlations to reactivity in various ANS measures were found in studies inducing a wider range of affective states, as reviewed by Kreibig and colleagues (2010).

In contrast to classic laboratory testing, studies using ambulatory assessment have more often taken the within-participant relationship between affect and physiology as their main focus. With the combined use of wearable devices to measure ANS activity, and ecological momentary assessment (EMA) to assess mental states the knowledge on the relationship between ANS and affect in daily life has been rapidly growing. To establish the strength of the relationship EMA studies apply time-series analyses such as multilevel modeling (de Vries et al., 2021). The interest in the relationship between ANS and affect in daily life is fueled by a desire to predict affective states by ANS states. Such ANS based affect predictions can be used to provide individuals more insight into their stress triggers and be used as input for Ecological Momentary Interventions (EMI) also called just-in-time adaptive interventions (JITAI) (Versluis et al., 2016; Nahum-Shani et al., 2018; Balaskas et al., 2021; Koch et al., 2021). Such interventions require the development of (machine learning) algorithms to predict emotional states from physiological signals to be able to use them as a proxy for these emotional states (Wang et al., 2022).

Thus far, the results of daily life studies have shown that even "in the wild" it remains difficult to relate physiology to specific affective state, particularly when affective arousal and valence are considered separately. Studies show that HR was positively associated with negative affect when arousal is high (Pieper et al., 2007; Pollard et al., 2007; Lumley et al., 2014; Kennedy et al., 2015; Ensari et al., 2020; Gordon & Mendes, 2021; Kim et al., 2021; Simon et al., 2021) or low (Kim et al., 2021; Simon et al., 2021; Liu et al., 2022). For general (i.e., not taking arousal levels into account) negative affect mixed results are found, with both positive (Pieper et al., 2007; Ilies et al., 2010), negative (Sloan et al., 1994), and non-significant relationships (Kamarck et al., 1998; Kimhy et al., 2010; Dennis et al., 2018; Määttänen et al., 2021) identified. Furthermore, HR is also positively associated with general positive affect (Ensari et al., 2020; Määttänen et al., 2021; Simon et al., 2021) and with positive affective states characterized by high arousal (Gordon & Mendes, 2021). Low arousal positive affect on the other hand showed a negative association with HR (Kennedy et al., 2015). These latter results are in line with findings that arousal is positively associated with HR, independent of valence (Kamarck et al., 1998).

A similar pattern is observed for HRV, but with more variability: it is negatively associated with general negative affect (Bacon et al., 2004; Pieper et al., 2007; Conrad et al., 2008; Kimhy et al., 2010), and with negative affect with high arousal (Bacon et al., 2004; Pieper et al., 2007). When arousal is low, HRV has been negatively (Simon et al., 2021, Liu et al., 2022), positively (Kim et al., 2021), or not (Dennis et al., 2018; Määttänen et

al., 2021) associated with negative affect. The overall reduction in HRV is also seen during general positive affect (Bacon et al., 2004; Conrad et al., 2008; Kennedy et al., 2015; Simon et al., 2021) and positive affect paired to high arousal (Gerteis & Schwerdtfeger, 2016). Low arousal positive affect on the other hand showed a positive relationship with HRV (Bacon et al., 2004; Gordon & Mendes, 2021; Gerteis & Schwerdtfeger, 2016; Zenker et al., 2021). In line with the findings for HR, lower HRV was seen at higher arousal independent of affective valence (Bacon et al., 2004; Zenker et al., 2021).

The difficulties to identify specific physiological effects for different combinations of valence and arousal in daily life are partly due to sparse occurrence of some states, and the potential co-occurrence of positive and negative affect in real life (Zelendski & Larsen, 2000; Vansteelandt et al., 2005; Larsen et al., 2017). Using short laboratory protocols to assess the affect – physiology relationship could provide a number of clear advantages over daily life recording. In the laboratory one can induce various different affective states which might need days of daily life assessment to capture. Added to this are the 'standard' advantages of laboratory assessment, in that it takes place in a controlled environment, excluding the influence of homeostatic (posture, physical activity, temperature) and contextual behavioral and social factors. This is of specific importance to detect small to moderate correlations such as those between physiology and affect. Finally, a short laboratory protocol also causes lower researcher and participant burden than long-term subjective monitoring in daily life, allowing for larger sample sizes for the same research budget.

However, to justify the use of laboratory tests to estimate cross-domain correlations between affect and ANS activity in real life, these cross-domain correlations should be directly comparable between the laboratory and daily life using the same methodology. The aim of the current study is to provide such a direct comparison between the relationship of ANS and affect in the laboratory and in daily life. To this end we collected data from the same healthy individuals in both a laboratory setting (in which they perform different stress tasks) and during a 24h daily life EMA period (in which natural stressors occur), while measuring 4 indicators of ANS activity with wearable devices as well as affect via a diary. We believe that the observed differences in physiological responding between the laboratory and daily life result from differences in affective appraisal: a stronger physiological response in daily life is coupled to a corresponding stronger affective response. Therefore, we expect the estimated relationship between the ANS measures and affect to be similar across measurement context. If this holds, reasonable expectations can be drawn with respect to the ANS – affect dynamics in daily life based on laboratory studies.

MATERIAL AND METHODS

Study population

Recruitment of participants and the laboratory protocols are described in detail elsewhere (van der Mee, Gevonden, Westerink, & de Geus, 2021). Briefly, participants were required to be between the age of 18 and 48, Dutch speakers, and currently employed, or in a schooling trajectory. Exclusion criteria were a body-mass index above 30, heart disease, high blood pressure, high cholesterol, diabetes, thyroid or liver disease, and use of antidepressants, anticholinergics, or any other medication that has been shown to influence the SNS. Female participants were measured within the first two weeks following the last day of their menstrual cycle to minimize the impact of hormonal changes.

Participants who were students received research credits, while other participants were compensated with a €50 gift voucher. All participants provided written informed consent before the start of the experiment. The study was approved in institutional review by the VUmc medical ethical committee (METc VUmc #2017.374, ABR #NL62442.029.17).

Usable data were obtained in 115 participants out of 121 participants originally recruited in the study for the laboratory section and 108 from the daily life section. Due to the following reason participants were not included: two participants were not included because after data collection was completed, they deviated too much from the mean age (they were > 45, while all other participants were ≤ 30), one participant was excluded because it was an outlier on all ANS measures, one participant had too poor data quality, one participant had incomplete laboratory data, and one participant requested the data to be removed from the study. Seven additional participants could not be included in the ambulatory data section due to faulty SD cards.

We decided only to retain participants that had both laboratory and daily life data. The resulting study population consisted of 108 participants with a mean age of 22.55 years (SD = 3.60, range = 18-35) and an average educational attainment of 9.14 years (SD = 2.17). It consisted of 56% females and 80% were students. Of the participants 77% were of Western-European ancestry. Of the remaining participants at least one of the parents was of Asian (11%), African (7%), or South American (5%) ancestry. With regard to health behaviors 63 % reported regular leisure time exercise, 25% were regular smokers, and 70% drank alcohol on a weekly basis. Furthermore, 83% reported good subjective physical health, 82% good subjective mental health, and 75% low experienced daily life stress. The overall good mental health of the study population was verified by the results of the POMS questionnaire with a mean total mood disturbance score of 23.49 (SD = 13.13, range = 6 – 78), mean anger score of 8.95 (SD = 2.87, range = 7 – 24), mean depression score of 9.60

(SD = 3.50, range = 8 – 26), mean fatigue score of 10.67 (SD = 3.84, range = 6 – 24), mean tension score of 8.94 (SD = 3.25, range = 6 – 21), and mean vigor score of 14.68 (SD = 3.30, range = 7 -22).

Physiological measures

Cardiac Autonomic Nervous System Activity

The cardiac physiological measures inter-beat interval (IBI), respiratory sinus arrythmia (RSA), and pre-ejection period (PEP) were obtained with the VU-AMS device (version 5-wire 5fs). The VU-AMS is a lightweight portable device that has been used to measure ANS activity in over 300 scientific studies (for an overview see http:/www.vu-ams.nl/research/publications). It records electrocardiogram (ECG) and impedance cardiogram (ICG) from five adhesive 55 mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Eindhoven, Netherlands) placed on the chest and back of the participants with a recording frequency of 1000 Hz (de Geus et al., 1995; Willemsen et al., 1996). VU-AMS data was analyzed with the Vrije Universiteit Data Acquisition and Management Software (VUDAMS version 4.6, available at: https://vu-ams.nl/software-solutions/). The IBI is calculated as the time difference between two successive R peaks and reflects the combined influence of SNS and PNS activity on the heart. RSA is calculated by means of a peak valley method, which combines the R-peaks time series with the impedance-derived respiration cycle. In this method the shortest IBI during each inspiration period (prolonged by a 750 ms delay) and the longest IBI during each expiration period (prolonged by a 750 ms delay) are detected. Then the former is subtracted from the latter. When the calculation of the RSA results in zero or negative values they are coded as zero (de Geus et al., 1995; Goedhart et al., 2007). RSA is well validated measure of PNS activity (Katona & Jih, 1975; Berntson et al., 1993; Migliaro, 2020) and has been frequently studied in relationship to stress (ea. Beauchaine, 2015; Beauchaine et al., 2019; Campbell et al., 2019; Lane et al., 1992). The PEP is obtained by calculating the time difference between the start of ventricular depolarization (Q onset) in the ECG and the time the aortic valve opens (B point) in the ICG (Nederend et al., 2018; Willemsen et al., 1996). For each time segment of interest, a single averaged ICG complex was derived by means of ensemble averaging of the ICG signal over all R-peaks in the condition, as explained by Riese and colleagues (2003). Various studies have shown that the PEP shortens in response to mental and social stressors, reflecting higher contractility due to increases in SNS activity (Brindle et al., 2014; van der Mee et al., 2021; Rahman et al., 2018). The VU-DAMS software automatically detects and scores the various attributes necessary to calculate the IBI (R-peaks), RSA (R-peaks and respiration) and PEP (Q-onset and B-point). All data scoring of the VU-DAMS was manually checked and if necessary

corrected. In the laboratory a mean IBI, RSA, and PEP score was calculated for each stress task for each participant. During daily life, a mean IBI, RSA, and PEP score was calculated during the five minutes preceding each diary entry for each participant.

Electrodermal activity

Electrodermal activity was measured with a CE approved wearable skin conductance wrist sensor, type DTI5 (Discreet Tension Indicator version 5, Philips). This wristwatch has been shown to sufficiently capture SNS activity (van der Mee, Gevonden, Westerink, & de Geus, 2021). The ns.SCRs frequency is defined as the number of peaks per minute obtained by an internal method of peak detection that makes use of a curve fit method. Ns.SCR frequency has been shown to relate to negative emotions (Nikula, 1991; van der Mee et al., 2021), arousal (Nikula, 1991), and stress (Miller & Shmavonian, 1965; Kelsey, 1991). We recently showed that this measure performs even better than the widely used skin conductance level to index changes in SNS activity across a wide variety of stressors (van der Mee et al., 2021). Similar to the cardiac ANS measures a mean ns.SCR score was calculated for each stress task for each participant in the laboratory and during the five minutes preceding each diary entry for each participant in daily life.

EMA and Affect

An iPad containing an in-house built electronic diary application was provided to participants to report their affect at set times in the laboratory. For the daily life section participants received an iPod containing the same electronic diary application. In both settings, affect was rated with nine items derived from the Maastricht Questionnaire (Myin-Germeys, 2001). Affect scores were obtained by asking the participants to rate whether they felt anxious, cheerful, content, down, enthusiastic, insecure, irritated, lonely, and relaxed on a scale of 1 (not at all) to 7 (very).

A single valence score and a single arousal score were calculated from these measurements, combining emotions when co-occurring. To calculate valence the negative affect items anxious, down, insecure, irritated, and lonely were multiplied by -1 to give them a negative value. Following, the average valence score for each time point was calculated by taking the average of all positive and recoded negative affect items. A positive valence score indicates that the dominant valence was positive, while a negative valence score indicates that the dominant valence was negative. Likewise, the arousal score was calculated by multiplying the low arousing items content, down, lonely, and relaxed were multiplied by -1, followed by calculating an average over all high arousing and

recoded low arousing items. A positive arousal score indicates that the dominant arousal was high, while a negative arousal score indicates that the dominant arousal was low.

In both the laboratory and during daily life the far majority of the observations had a positive valence, indicating an overrepresentation of positive affect in the current sample (see Supp. Figure 1). In the laboratory, 33 participants had more than one observation with a dominant negative valence, while in daily life this was only the case for 16 participants. Because we expect ANS activity to differ as a factor of both valence and arousal, this limited our ability to study the relationship of ANS with a strong dominant negative affect state. Therefore, we decided to omit negative valence scores below -1 (26 observations in the laboratory and 20 in daily life) and focus on valence as a representation of positive valence.

Contextual EMA variables

In addition to question on affect the iPod diary application collected contextual variables at each beep. These contextual variables included in this study consisted of a multiple-choice question regarding the activity they were engaged in at that moment in time with the answer options: work/study, leisure, household chores, transportation, relaxing, sleeping, and other. Participants were informed that the iPod would go off hourly between the hours of 07:30 am and 11:00 pm. Each participant received 15 diary prompts but was allowed to manually fill in extra diaries by opening the app if they went to bed after 11 pm or woke up earlier than 7:30 pm. They were also informed a random jitter of 15 minutes was added around each diary prompt to reduce expectation. Compliance for e-diary entries during the daily life part of the experiment was good. On average participants completed 12 out of the 15 prompted diaries (SD = 3.00).

Interview

A structured interview was performed during the laboratory session to gain information on the participants' physical activity behavior, subjective mental and physical health score (1 – very bad to 5 – excellent), ancestry (country of birth of themselves and both their parents). In addition, the participants were asked to fill in the Profile of Mood States questionnaire to obtain an indication of overall daily mood.

Procedure

Participants visited the laboratory on two consecutive days. Participants were free to choose at which time of day they would like to start their measurement as long as it was between office hours (08:00 – 18:00). During their initial visit to the laboratory (~1 h)

participants provided informed consent and the structured interview was conducted. Subsequently, the ANS measuring devices were applied to the participant to continuously measure ANS activity. The participants were provided with an iPod containing the questionnaire application. The experimenter practiced all items of the questionnaire with the participant to make sure they understood each item.

Once equipped with the measuring devices and the iPod, participants left the laboratory for a day of ambulatory monitoring. During the 24-hour recording only a few restrictions of normal activities were applied. Participants were requested not to take a bath or engage in water sports. They were asked to remove the devices (but not the electrodes) during showering and during prolonged heavy physical activity and reattach the devices afterwards.

Participants returned the next day for participation in the laboratory protocol. Upon their return, it was verified that all measurement equipment was still in working order. Next, to increase stress, participants were informed that footage of their facial expressions, posture and voice would be recorded during the experiment. Then all experimental manipulations were presented in a fixed order (see Supplementary Table 1), including a short intelligence test. Participants were informed that their scores on this task were tracked on a scoreboard containing other participants' scores for comparison on their performance." After each stressor, the participants were asked to fill out a similar affect questionnaire as in the real-life part on an iPad.

After the experimental session, all devices were removed, and participants were provided the opportunity to use a nearby shower. The experiment ended with a debriefing in which they were informed that the tasks were purposefully made so difficult that they would be impossible to perform without errors. They were explicitly told that the test score rankings were only added to increase the stressfulness of the task and did not reflect their actual ability. Furthermore, they were informed that no actual voice or video recording had been made.

Laboratory stressors

The laboratory section started with a 3-minute sitting resting conditions. The stressors used in the experimental set-up can be divided in to mental-stressors and social stressors. The mental stressors used in this study are the Tone avoidance (TA) task and the Paced Auditory Serial Subtraction (PASAT) task. The TA task aims to induce effortful active coping in which participants have to avoid a loud tone by pressing a button on the opposite site of an "X" presented on one of the four corners a computer screen (de Geus, et al., 1990; van der Mee, et al., 2020). The PASAT is a calculus task with a staircase algorithm to

measure capacity and rate of information processing and sustained and divided attention. Single digits are presented every 3 seconds and the respondent must add each new digit to the one immediately prior to it before the next digit is presented (Tombaugh, 2006; Sampson & MacNeilage, 1960).

The social evaluative stressors used in this study are the short Sing-a-Song-Stress-Test (SSSTshort) and the Raven Progressive Matrices IQ (RPM) test. In the SSSTshort participants unexpectedly have to sing a song out loud in front of a camera and the experimenter (van der Mee, et al., 2020). The RPM test (Raven progressive matrices; Raven, 2003) was timed, participants had to solve as much matrices as possible in a 4-minute time window. They were informed that the more correct answers they gave the higher their IQ score would be. To induce social evaluative stress their scores were directly compared to the scores of the other participants by the use of a score board placed in sight of the participant.

Analyses

Due to the large influence of major body movements on ANS activity (Fu & Levine, 2013) during the awake period, data with accelerometer values >= 50 milliG acceleration were excluded. Furthermore, daytime data in which the participants reported to be resting/sleeping were excluded. Participants with < 3 complete observations in either measurement context (laboratory vs. daily life) were excluded.

In both the laboratory and daily life sections, the mean and standard deviation was calculated for valence, arousal, IBI, RSA, PEP and ns.SCR scores for each individual separately. Differences in these individual mean and standard deviation scores were assessed with a repeated measures t-test on complete cases.

To investigate the relation of ANS with valence and arousal in the laboratory and in daily life, we applied multilevel models in R using the packages "lme4" and "lmerTest". The models were rune with the "lmer" function and the optimizer "nlminbwrap". Due to the clustered data structure multilevel analyses were performed with participant id as cluster indicator (for a more detailed explanation of this approach see van der Mee et al., 2021).

Intra-class-correlations (ICC) were computed for each all four ANS measures, valence, and arousal with the use of an empty multilevel model in which the respective parameter was predicted by only clustering for individuals or contextual parameters. For the laboratory, the contextual parameters are the experimental conditions. For daily life, the contextual parameters consist of beep number, time of day, and type of activity performed. Time of day was subdivided in 2–3-hour segments to reflect different parts of the day: 08:00 – 10:00, 10:01 – 12:00, 12:01 – 15:00, 15:01 – 18:00, 18:01 – 20:00, and 20:01

– 22:00. The type of activity performed was categorized as: resting, relaxing, working/studying, commuting, basic needs (household activities, grocery shopping and food consumption), and other (reflecting a specific category in the questionnaire if the activity did not fit with any of the options).

To account for individual differences in trait valence and arousal, we subtracted the mean value of the respective measurement setting from the valence and arousal values at the laboratory assessment and daily life prompts, respectively. These mean-centered values are referred to as state valence and arousal, while the mean values are referred to as trait valence and arousal. Furthermore, in view of its skewed distribution, RSA was first transformed to normality using the natural logarithm. Multilevel regression models were run for each of the ANS measures (IBI, RSA, PEP and ns.SCR), using state valence, state arousal and their interaction as level 1 predictors of the within-participant changes in ANS activity. These analyses were performed for the laboratory and daily life datasets separately.

Trait valence and arousal were included as level 2 covariates. Since there is evidence that males and females respond differently to stressors with regard to their ANS response, with males being "vascular" reactors and females "cardiac" reactors (Huang et al., 2013) and it has been shown that ANS activity changes with age (Peters et al., 2020), therefore age, and biological sex were included as level 2 covariates. Since changes in respiration rate can cause changes in RSA that are not caused by changes in PNS activity (de Geus et al., 1995; Grossman et al., 1991) repeated observations on RR (derived from the ICG signal as described by Houtveen, Groot & de Geus, 2006) were added as a level 1 covariate to the RSA analyses. Previous work has shown that the ns.SCR frequency measure is not sensitive to thermodynamic effects as compared to skin conductance levels in both the laboratory (van der Mee, et al., 2021) and in daily life (van der Mee, et al., 2022). Therefore, the analyses regarding ns.SCR were performed without inclusion of skin and ambient temperature covariates.

We first fitted models using random intercepts but fixed effects for the slopes of valence and arousal. Next, we determined per model whether allowing the slopes to vary across individuals improved the model by comparing the model parameters with an ANOVA procedure, at a nominal significance of $p = 0.05$. Explained variance in ANS activity by state valence and state arousal was determined with the r2mlm package. To obtain pure explained variance by valence and arousal and not the model as a whole the function was run on the model including only state valence or state arousal and the participants id as cluster factor.

To test our main hypothesis, that the relationships between affect and ANS activity are comparable in the laboratory and daily life, the average regression coefficients of the selected fixed or random slope models are compared across the two settings with a Z-test:

$$Z = (\beta 1 - \beta 2) / \sqrt{(\beta 1 SE^2 + \beta 2 SE^2)}$$

β1 = average (fixed/random) regression coefficient in the laboratory
β2 = average (fixed/random) regression coefficient in daily life
β1SE = standard error of the average (fixed/random) regression coefficient in the laboratory
β2SE = standard error of the average (fixed/random) regression coefficient in daily life

This was repeated for the β's reflecting arousal, valence, and interaction effects for each of the 4 ANS parameters (12 tests). To account for multiple testing significance levels were adjusted from nominal 0.05 to p = .05/12 = .004.

The data that support the findings of this study are available from the corresponding author, DJ, upon reasonable request.

Power

Based on previous laboratory studies we expected to find, and average explained variance in physiology by affect of ~5% (Feldman, et al., 1999). With regard to the laboratory dataset the average correlation among the predictor variables valence and arousal was .58 and among the ANS outcome variables was .63, therefore the predictor autocorrelation (phiX) was set to .58 and outcome autocorrelation (phiE) was set to .63. The power to detect ~5% explained variance in an outcome by a predictor in a sample of N = 100 individuals using T = ~5 repeated measures (baseline plus four stressors) at a p-value of p = .005 is .79**.** With regard to the daily life dataset the average correlation among both the predictor variables was .51 and outcome variables was .54, therefore the predictor autocorrelation (phiX) was set to .51 and outcome autocorrelation (phiE) was set to .54. The power to detect ~5% explained variance in an outcome by a predictor in a sample of N = 100 individuals using T = ~12 (average number of diary entries) repeated measures at a p-value of p = .005 is .99.

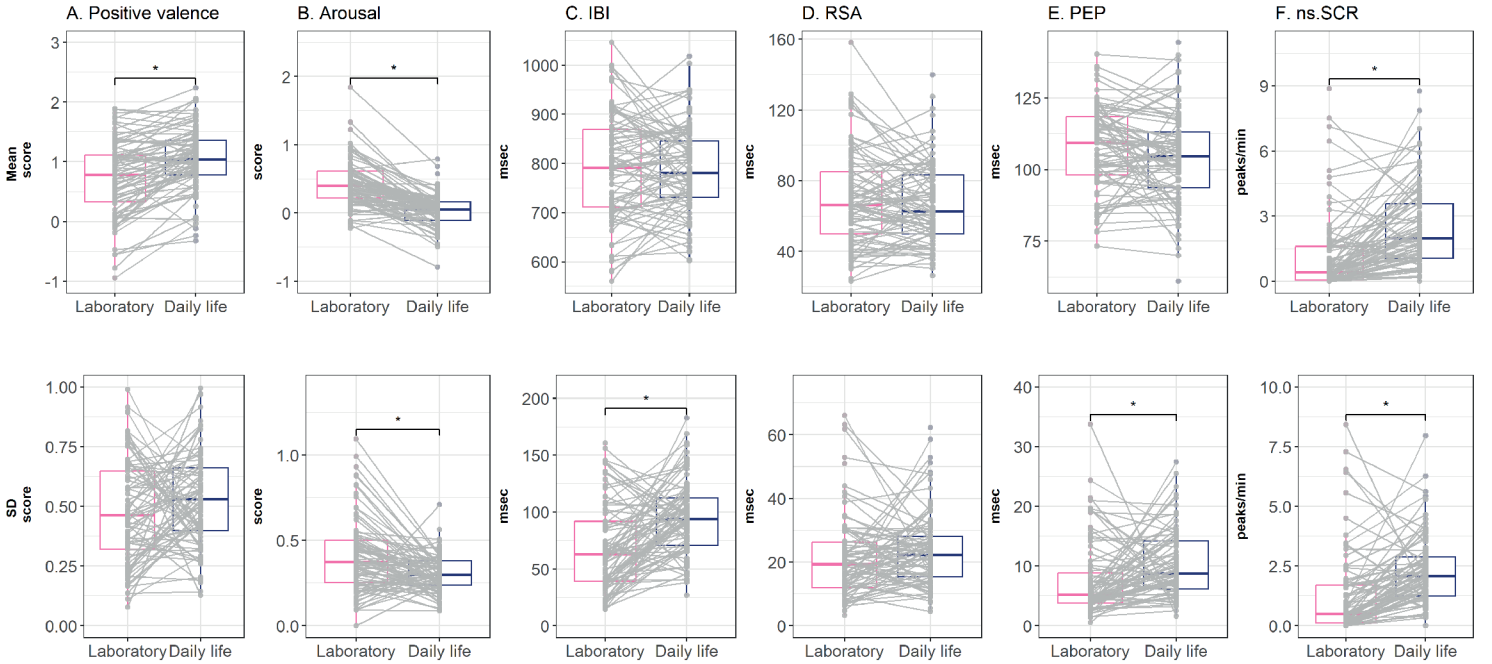


Figure 1. Differences in mean and standard deviation of trait valence, arousal, and ANS activity between the laboratory and daily life.

** p < .001*

Note: The top row depicts the box plot and individual data for the means of valence, arousal, and ANS scores across the laboratory and daily life. The bottom row depicts the box plot and individual data for the standard deviations of valence, arousal and the ANS measures across the laboratory and daily life. Paired samples t-test showed that there was a significant difference between the laboratory and daily life in mean trait valence (ΔM = 0.34, T(80) = 7.14, p < .001), arousal (ΔM = -0.42, T(80) = -10.84, p < .001), and ns.SCR (ΔM = 1.25, T(80) = 6.07, p < .001), but not mean trait IBI (ΔM = 3.27, T(80) = 0.34, p = .73), RSA (ΔM = -2.39, T(80) = -1.05, p = .29), or PEP (ΔM = -2.52, T(80) = -1.56, p = .12). For the variance in traits there was the paired samples t-test showed a significant difference between the laboratory and daily life for the SD of arousal (ΔSD = -0.12, T(80) = -5.20, p < .001), IBI (ΔSD = 29.49, T(80) = 5.91, p < .001), PEP (ΔSD = 2.83, T(80) = 2.97, p = .004), and ns.SCR (ΔSD = 0.96, T(80) = 4.63, p < .001), but not for valence (ΔSD = 0.03, T(80) = 0.86, p = .39) or RSA (ΔSD = 2.11, T(80) = 1.29, p = .20).

RESULTS

Means and Variances

Figure 1 depicts the individual trait valence, trait arousal, and the mean ANS measure with their SDs in the laboratory and daily life. In daily life, participants had significantly higher mean valence (more positive affect), lower mean arousal, and higher mean ns.SCR frequency compared to the laboratory (all $p < .001$). With regard to variance in affective arousal and valence: higher SD for arousal was seen during the laboratory compared to daily life, but the SD for valence was comparable. ANS activity was more variable in daily life than in the laboratory, with IBI, PEP and ns.SCR showing a significantly higher SD (all $p < .004$).

Between versus within participant variance

In the laboratory a large part of the variance in state valence (49%), state arousal (35%), and the cardiac ANS measures could be contributed to individual differences (IBI: 66%, log RSA: 57%, PEP: 70%, ns.SCR: 24%). The contextual parameters (experimental conditions) explained 23% of the variance in state valence and 35% of the variance in state arousal but explained relatively little variance in IBI (12%), log RSA (6%), PEP (3%) or ns.SCR (2%). This left around a third of the laboratory variance unexplained: state valence (28%), state arousal (30%), IBI (22%), log RSA (37%), and PEP (27%). For ns.SCR most of the variance was unexplained (74%).

In daily life the variance explained by individual differences dropped slightly for affect (Valence 41%; Arousal = 31%). Explained variance also dropped for all cardiac ANS measures (IBI: 40%; log RSA: 39%; PEP: 63%) save ns.SCR (ns.SCR: 28%). Of the contextual parameters in daily life, beep number and time of day explain almost none of the variance in either physiology or affect (beep number 0 – 1%, time of day 0 – 4%). Activity type explained 8% of the variance in state valence, 5% in state arousal, a substantial part of the PNS sensitive measures IBI (20%) and log RSA (10%), but almost none of the variance in the SNS sensitive measures PEP (1%) and ns.SCR (<1%). This left around a third to over half of the variance in the daily life data unexplained: state valence (46%), state arousal (65%), IBI (40%), log RSA (42%) and PEP (33%). For ns.SCR, most of the variance was again unexplained (71%).

ANS by valence and arousal

The results of the multilevel model on the fixed main effects of state valence and state arousal and their interaction on ANS activity are shown in Table 1. In the laboratory, state

arousal was negatively associated with IBI, log RSA and PEP. The explained variance was modest, with around 1-3% of the variance in ANS was explained by state arousal. In the laboratory there were no significant relationships between state valence and ANS activity.

In daily life, state arousal was also negatively associated with IBI, but there were no significant relationships between state arousal and log RSA or PEP. In further contrast to the lab, daily life state valence was negatively associated with IBI and PEP, although with low explained variance (< 1%). There were no significant interactions of state valence and arousal on ANS activity, either in the lab or in daily life.

All models in Table 1 were retested for improved fit after allowing the slopes of state valence and state arousal to vary between individuals. Model fit only improved for the daily life cardiac ANS models, but even there allowing random slopes had little impact on the size, direction and explained variance of the average estimated coefficients (see Supplementary Table 2). We therefore decided to compare to models of the laboratory and daily life based on the fixed slopes.

Figure 2 shows the results of the Z-test comparison of the estimated average regression coefficients in the laboratory and in daily life for these fixed affect-ANS relationships. As shown in Fig 2, the relationship of state valence with ANS was in the same direction in the laboratory as in daily life for all ANS measures, and effect sizes were comparable". In addition, the strengths of the valence-ANS relationships were comparable, as shown by the overlap of 95% CIs of the lab and daily life estimates. For state arousal, all relationships with ANS were also in the same direction for the laboratory and daily life, with effect sizes for log RSA and PEP appearing higher in the laboratory, and even reaching significance for PEP.

The interaction effects between state valence and state arousal on ANS activity were never significant yet tended to be in opposite directions between the laboratory and daily life (see Supplementary Figure 2). However, none of these effects reached the preset significance level correcting for multiple testing.

Table 1. ANS activity by valence and arousal in the laboratory and daily life - Fixed effects

ANS	N	Obs	β_{0j}	U_{0j}	Affect	β_{1j}	β_{1j}SE	T	*p*	R^2
Laboratory										
IBI msec	102	492	791.86	105.26	Valence	-19.58	9.41	-2.08	.038	.005
					Arousal	-62.21	10.95	-5.68	**<.001**	.027
					Valence*Arousal	-25.85	13.50	-1.91	.056	
RSA log	101	474	4.12	0.40	Valence	-0.05	0.05	-1.14	.25	.008
					Arousal	-0.23	0.05	-4.32	**<.001**	.027
					Valence*Arousal	-0.10	0.07	-1.60	.11	
PEP msec	97	459	82.39	14.26	Valence	-1.28	1.25	-1.02	.30	.006
					Arousal	-6.63	1.45	-4.57	**<.001**	.020
					Valence*Arousal	-0.91	1.75	-0.52	.60	
ns.SCR p/m	92	438	1.15	1.55	Valence	-0.22	0.29	-0.77	.44	.009
					Arousal	0.53	0.34	1.58	.11	.011
					Valence*Arousal	-0.08	0.42	-0.18	.85	
Daily life										
IBI msec	102	1150	722.22	83.88	Valence	-26.74	5.53	-4.84	**<.001**	.007
					Arousal	-43.42	9.98	-4.35	**<.001**	.009
					Valence*Arousal	18.79	16.95	1.11	.26	
RSA log	101	1119	4.24	0.33	Valence	0.00	0.02	0.16	.87	.000
					Arousal	-0.08	0.04	-1.84	.065	.003
					Valence*Arousal	0.07	0.07	0.95	.34	
PEP msec	100	1132	82.21	15.48	Valence	-3.33	0.67	-4.98	**<.001**	.006
					Arousal	-1.16	1.20	-0.97	.33	.000
					Valence*Arousal	4.97	2.05	2.34	.019	
SCR p/m	100	1091	2.17	1.41	Valence	0.06	0.15	0.40	.69	.000
					Arousal	0.26	0.27	0.96	.33	.000
					Valence*Arousal	0.39	0.45	0.86	.39	

β_{0j} and B_{ij} are the average intercept and slope of the regression models. The between participant variation around the intercept (U_{0j}) is given. R^2 gives the proportion explained variance in the outcome by the fixed effect slope estimate β_{1j}. P-values below the threshold for multiple testing are depicted in bold.

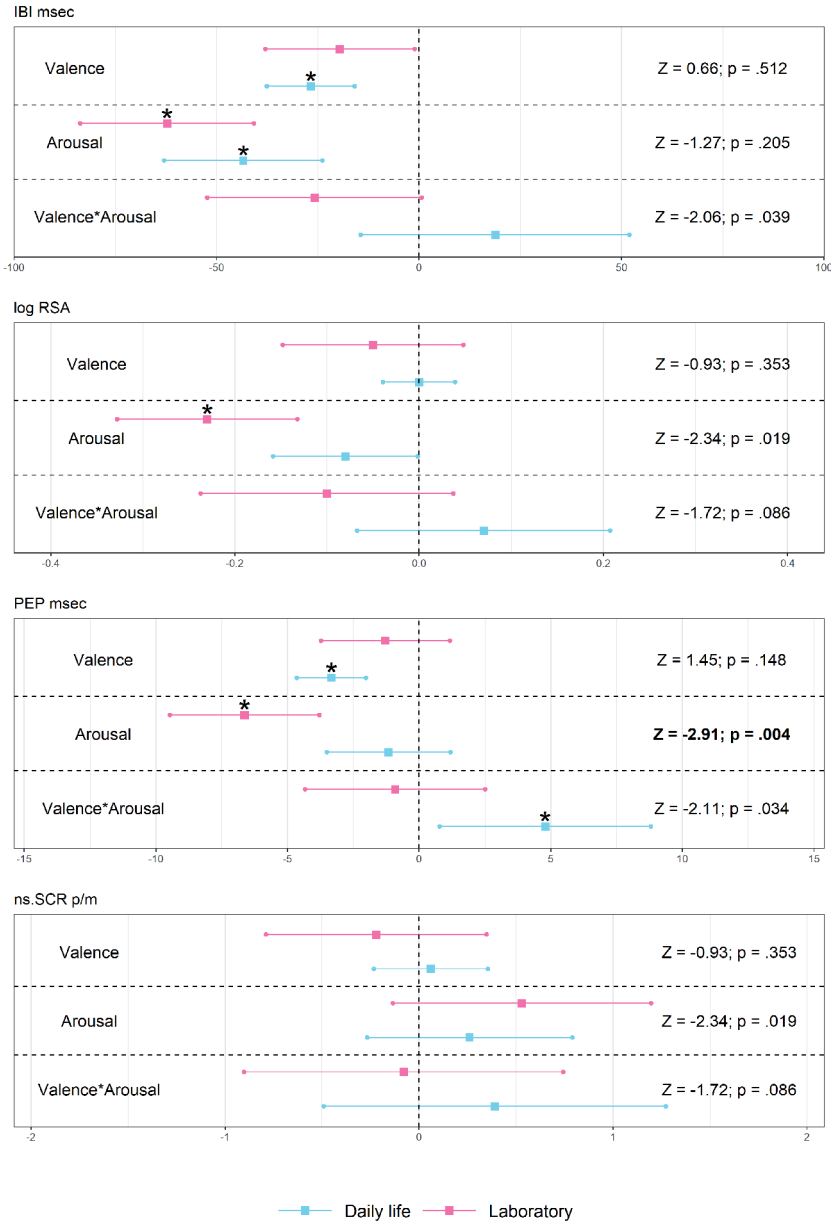


Figure 2. Comparison of fixed effect average estimated coefficients between the laboratory and daily life.

Note: This figure shows the average estimated coefficient (with 95% CI) of the fixed effects models as depicted in Table 1. Significant relationships from these models are indicated with an *.
For each Z-test comparison the Z and p value are given on the right side of the figure. When the estimated coefficients differ significantly between the laboratory and daily life the Z and p value are depicted in bold. A dotted line is presented at value 0, this provide an easy way to compare the direction of the average estimated coefficients across the laboratory and daily life (negative: left of the 0-line; positive: right of the 0-line).

DISCUSSION

The ecological validity of laboratory induced stress is hotly debated, with various researchers making the change to study their construct of interest "in the wild" rather than in the laboratory. Measuring in the wild comes with several limitations that are inherent to the design. The occurrence of the affective state of interest is left up to chance, different affective states co-occur, and there is a plethora of measured and unmeasured confounders. Such issues are not present in a controlled laboratory setting, which offers the opportunity to study the affect-physiology dynamic more clearly. However, the similarity in the relationship between physiology and affect in response to stress is understudied and complicated by the lack of studies adopting the same methodology across contexts. In the current study we provide the first direct comparison of the relationship of affect with ANS activity in the laboratory to that in daily. The overarching conclusion from our study is that the relationship of affect with ANS activity is remarkably similar in the laboratory and daily life, giving green light to researcher on the ecological validity of their laboratory designs.

Previous work in the same sample showed that our laboratory experimental manipulations were successful in decreasing positive affect and PNS activity and increasing negative affect and SNS activity (van der Mee et al., 2022) with affect sizes of expected magnitude (Brindle et al., 2014). As expected, this resulted in a lower mean valence and higher mean arousal in the laboratory compared to daily life. This difference was not seen in the mean cardiac ANS measures, only the skin measure ns.SCR frequency differed between the contexts. The overall variance in ANS activity, however, did differ between contexts. In daily life variance was higher for IBI, PEP and ns.SCR. Furthermore, arousal also showed more variance in daily life compared to the laboratory.

In both contexts a substantial amount of variance in affect and ANS activity was explained by inter-individual differences. This observation is consistent with the "stress reactivity" theory, in which stress reactivity is viewed as a stable trait-like factor that is associated with various personality characteristics such as extraversion and neuroticism (Boyce & Ellis, 2005). In this theory "hyper-reactors" are more sensitive to stressors and have a predisposition to respond to these stressors with greater affective and ANS activation. Interestingly, only little variance in daily life was explained by contextual factors, such as time of day or the activity type engaged in. Despite our efforts to exclude periods of high physical activity or supine posture, a large part of the variance in daily life likely reflects confounding by posture and physical activity.

Most significant affect-ANS relationships were in the expected direction: higher valence was associated with higher SNS and lower PNS activity. Likewise, higher arousal was also linked to higher SNS activity. While it appears counter-intuitive that more intense positive affect is linked to an increase in SNS and decrease in PNS activity, this can be explained by the distribution of observations on the valence and arousal axes in the current study. Both in the laboratory and in daily life, a considerable number of observations fell within the high arousal positive affect quadrant, implying that most positive emotions were also arousing (e.g., cheerful, enthusiastic). These findings align with prior studies in daily life (Ensari et al., 2020; Gerteis & Schwerdtfeger, 2016; Gordon & Mendes, 2021; Kamarck et al., 1998; Kennedy et al., 2015; Simon et al., 2021; Zenker et al., 2021).

All the affect-ANS relationships were in the same direction in the laboratory and daily life, with comparable effect sizes. For state arousal effect sizes for log RSA and PEP are higher in the laboratory, although this reached significance for PEP only. In both contexts allowing the slopes for valence and arousal to vary across individuals had little to no effect on the explained variance by the model. This suggest that individual differences have a minimal effect on the psychophysiological coupling. The improved fit in daily life likely reflects differences between individuals in the occurrence of affective states. This might also explain the tendency for an opposite direction of the interaction effect of valence and arousal in the two contexts. In the laboratory we induce a specific type of stress that aims to increase negative affect and arousal. This leads to a strong negative relationship between valence and arousal, with high arousal scores being associated with lower valence scores (Supp. Fig 1). In daily life, this association is much weaker. There, because of more diverse and complex individual experiences, arousal is coupled to both low and high valence (Supp. Fig. 1).

The current study is the first to compare the affect-ANS coupling in the laboratory to daily life in a single population using the same methodology. It thereby provides an excellent sample to address the question of the ecological validity of laboratory-induced stress. Consistent with previous laboratory studies (Feldman et al., 1999) only a small amount of the variance in ANS activity was explained by affect (0.5% – 3.0%). The results of the current study were limited by an over-representation of high arousal positive affect. During daily life participants experienced little distress and even the laboratory experimental manipulations were not strong enough to induce a dominant negative affective state in the majority of participants. This prohibited us form studying the relationship of other affective states (such as: high arousal negative affect/low arousal positive/low arousal negative affect) with physiology. Future research could address this issue by capturing more diverse affective states. The laboratory protocol should be

expanded to include a broader range of stressors, including stressors that not only aim to increase negative valence and arousal but also increase positive valence and lower arousal. For daily life studies, such data could be obtained by measuring during a period including a known stressful day and a known relaxed day, or measure over longer periods of time to capture these events naturally. With regard to ecological validity the latter has the preference. However, it comes with a major drawback. The current golden standard ANS technology is not suitable for long term measurement (up to weeks) that is needed to capture substantial variance in affective states. While wearable wrist-worn ANS devices exist, they show only modest validity with their golden-standard counterparts (van der Mee et al., 2021; Milstein & Gordon, 2020; Schuurmans, et al., 2020; Xie, et al., 2018). It is currently unknown whether these wearable devices can accurately capture the inter-relationship of affect and ANS activity.

From the current study can be concluded that we can ecologically validly measure the affect-ANS dynamics in a laboratory setting. When considering the individual differences in affective appraisal of the stressor the psychophysiological coupling appears to be a universal process. This is good news for the translatability of laboratory-based findings to daily life. Furthermore, it validates the study of the affect-physiology dynamics in a controlled and low participant-burden laboratory setting in a large number of individuals, offering researchers the opportunity to optimize their EMI/JITAI methods or validate their wearable sensors before applying them in a daily life setting.

SUPPLEMENTARY MATERIALS

Supplementary Table 1. Experimental timeline*

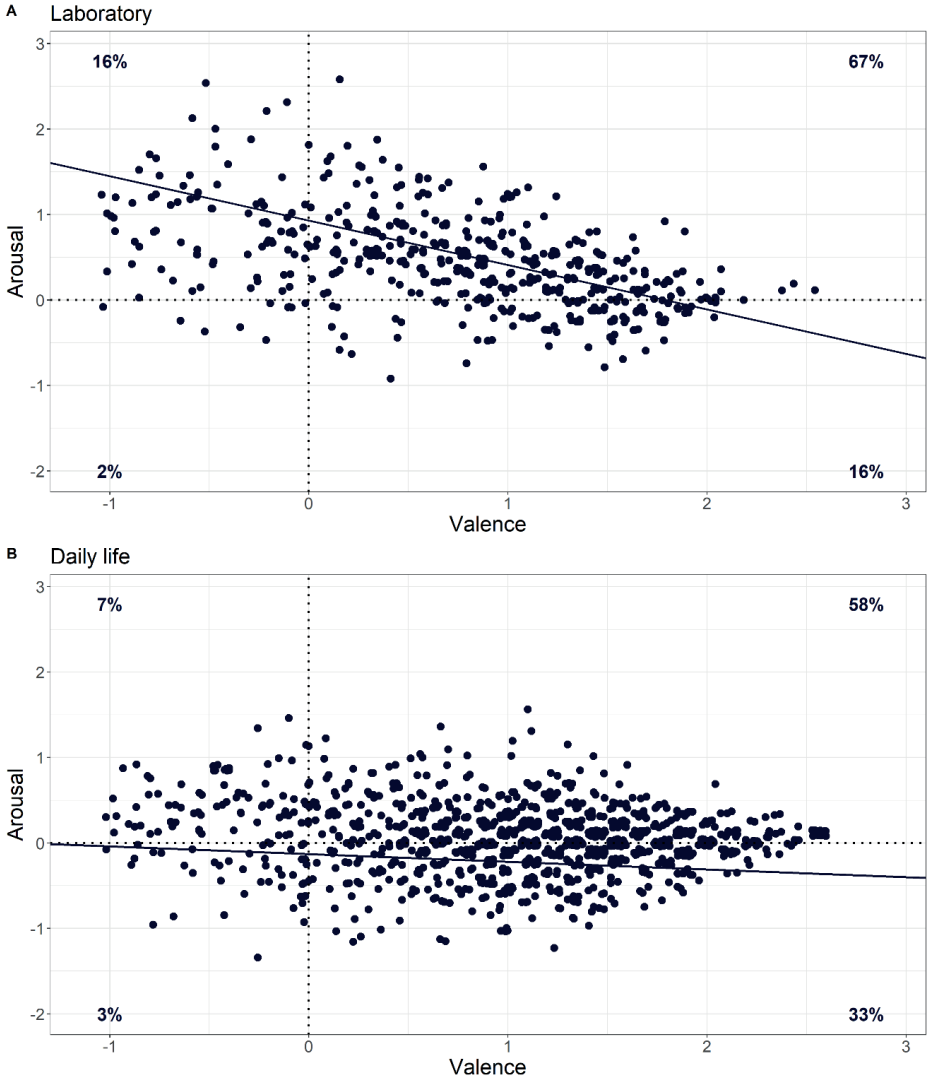
Experimental condition	Duration (minutes)	Affect measurement
Baseline	3	Yes
Tone Avoidance (TA)	4	Yes
Rest	2	No
short Sing-a-Song Stress Test ($_{short}$SSST)	6.5	Yes
Rest	2	No
Paced Auditory Serial Addition Test (PASAT)	4	Yes
Rest	2	No
Raven's progressive matrices (RPM)	4	Yes

**The timeline only presents tasks relevant for the present paper in their presentation order. The full timeline of all experimental conditions can be found in van der Mee, et al., 2021.*

Supplementary Table 2. ANS activity by valence and arousal in daily life - Random effects

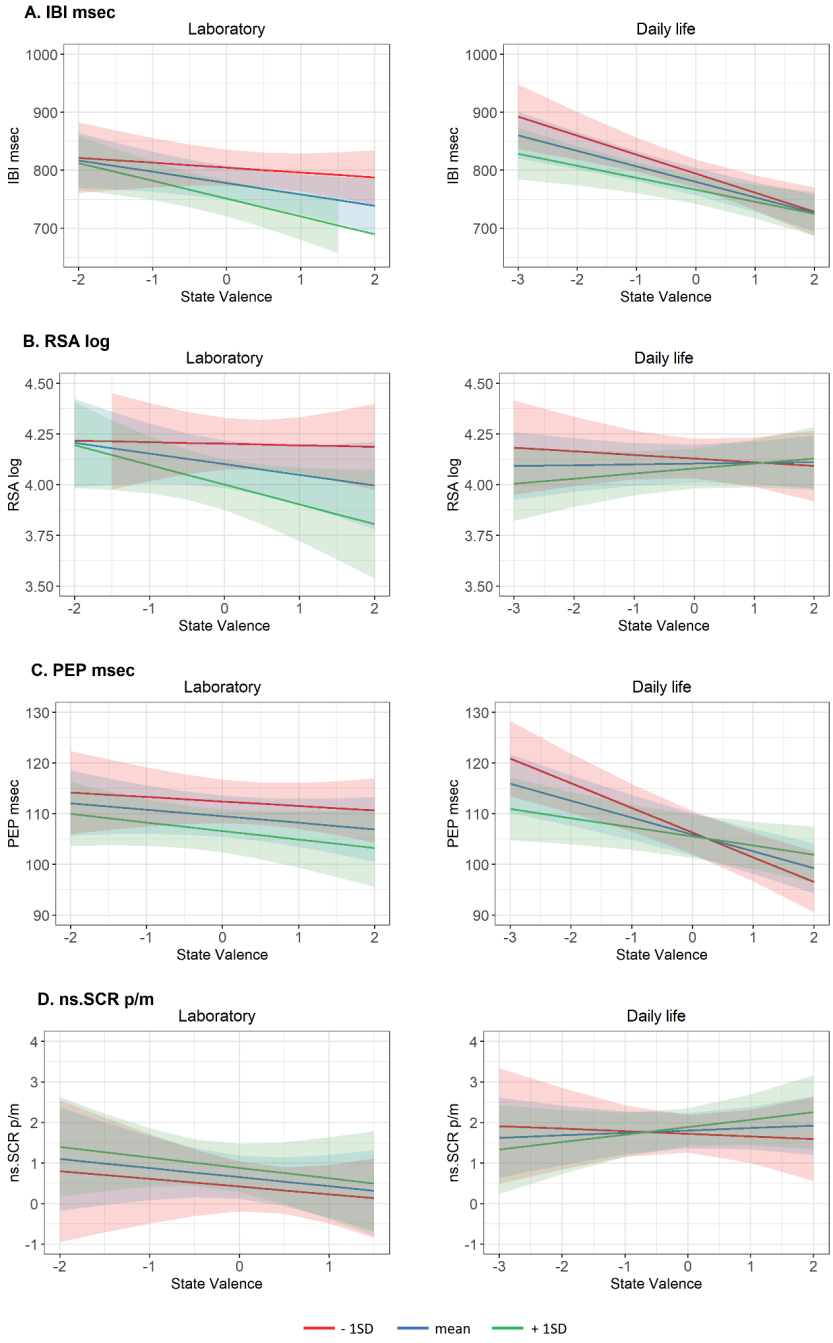
ANS	N	Obs	Affect	β_{0j}	U_{0j}	β_{1j}	β_{1j}SE	U_{1j}	T	p	R^2 fixed	R^2 random
IBI msec	102	1150	Valence	713.44	84.30	-27.36	7.29	43.01	-3.76	**< .001**	.008	.033
			Arousal			-44.10	10.00		-4.41	**< .001**	.009	-
			Valence*Arousal			21.89	17.18		1.27	.20		
RSA log	101	1119	Valence	4.26	0.33	0.00	0.03	0.16	01.12	.90	.000	.025
			Arousal			-0.09	0.05	0.16	-1.89	.061	.003	.007
			Valence*Arousal			0.06	0.07		0.81	.42		
PEP msec	100	1116	Valence	82.11	15.51	-3.07	0.93	5.76	-3.31	**.001**	.006	.022
			Arousal			-1.19	1.50	7.87	-0.79	.43	.000	.012
			Valence*Arousal			4.75	2.10		2.26	.024		

β_{0j} and B_{1j} are the average intercept and slope of the regression models. The between participant variation around the intercept (U_{0j}) is given. R^2 gives the proportion explained variance in the outcome by the fixed effect slope estimate β_{1j}. P-values below the threshold for multiple testing are depicted in bold.



Supplementary Figure 1. Scatterplot of the observations on the valence and arousal scales.

Note: In both contexts there is a negative relationship between valence and arousal (laboratory N = 102, obs = 492, $\beta 0j = 0.93$, $\beta 1j = -0.52$, SE = 0.03, $p < .001$, daily life N = 102, obs = 1157, $\beta 0j = -0.13$, $\beta 1j = -0.09$, SE = 0.016, $p < .001$, meaning that higher valence scores are related to lower arousal scores. The relationship was significantly stronger in the laboratory ($Z = -9.38$, $p < .001$). The percentage of observations within each quadrant is given in the corner of each quadrant.



Supplementary Figure 2. Interaction effects of valence and arousal in the laboratory and daily life.

CHAPTER 5

Cardiorespiratory fitness, regular physical activity, and autonomic nervous system reactivity to laboratory and daily life stress.

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. (2022). Cardiorespiratory fitness, regular physical activity, and autonomic nervous system reactivity to laboratory and daily life stress. *Psychophysiology*, e14212.

ABSTRACT

The cross-stressor adaptation hypothesis – which posits that adjustment to physical stress as a result of regular physical activity and its effects on fitness crosses over to psychological stress reactivity – has been around for over four decades. However, the literature has been plagued by heterogeneities preventing definitive conclusions. We address these heterogeneity issues in a combined laboratory and daily life study of 116 young adults (M = 22.48 SD = 3.56, 57.76% female). The exposure, i.e. the potential driver of adaptation, was defined in three ways. First, a submaximal test was performed to obtain aerobic fitness measured as the VO_2max (kg/ml/min). Second, leisure time exercise behavior, and third, overall moderate-to-vigorous physical activity (MVPA), were obtained from a structured interview. Outcomes were autonomic nervous system (ANS) reactivity and affective responsiveness to stressors. ANS activity was measured continuously and expressed as inter-beat-interval (IBI), pre-ejection-period (PEP), respiratory sinus arrythmia (RSA), and non-specific skin conductance responses (ns.SCR). Negative and positive affect were recorded after each experimental condition in the laboratory and hourly in daily life with a nine-item digital questionnaire. Linear regressions were performed between the three exposure measures as predictors and the various laboratory and daily life stress measurements as outcomes. Our results support the resting heart rate reducing effect of aerobic fitness and total MVPA in both the laboratory and daily life. We did not find evidence for the cross-stressor adaptation hypothesis, irrespective of ANS or affective outcome measure or whether the exposure was defined as exercise/MVPA or aerobic fitness.

INTRODUCTION

The hypothesis that regularly physically active individuals are not only more resilient to acute exercise but also to acute psychological stress has been around for over four decades (see Sothmann et al., 1996). The basis of this so-called cross-stressor adaptation hypothesis lies in the similarity between the physiological response to exercise and psychological stressors. One of these physiological responses is the activation of the autonomic nervous system (ANS). The basis of the idea is that the ANS response to a fixed dose of exercise becomes lower after repeated exposure to (intense) physical activity, with additional faster recovery (as reviewed by Micheal Jr., 1957). This so-called 'training' effect is a combination of increased organ responsiveness (stroke volume, muscle capillarization), changed feedback from exercising muscles, and central nervous system adaptations, including changes in the 'central command' or the feed forward engagement of ANS by the brain. These adaptations, especially the scaling down of the anticipation of the required ANS activity, may then be inherited by any other type of stressor that engages anticipatory ANS responding, like challenging cognitive tasks and social-evaluative stressors (Sothmann et al., 1996; Sothmann, 2006). Cross-stressor adaptation could be an important contributor to the well-established health benefits of regular physical activity on many major diseases by countering the detrimental effects of repeated and prolonged cardiovascular stress reactivity (Gerber & Pühse, 2009).

A large number of studies have sought to provide empirical support for the cross-stressor adaptation hypothesis, but results have been mixed and even systematic reviews and meta-analyses don't come to unequivocal conclusions (Forcier et al., 2006, Huang et al., 2013; Jackson & Dishman, 2006; Mücke et al., 2018). A first potential source of heterogeneity in findings is the mixture of studies using exercise intervention ('training') and studies using cross-sectional comparisons of regular exercisers versus less regular or non-exercisers. Both designs have strengths and weaknesses but mixing them is a strong source of heterogeneity. If duration of the exposure to regular exercise is the main determinant of cross-stressor adaptation, then many intervention studies may not have trained the participants long enough to induce the adaptation. Cross-sectional studies can be at a substantial advantage in this respect. However, if co-occurring confounders such as socioeconomic position and genetics are the main source of reduced stress-reactivity seen in regular exercisers, then the outcome of comparisons between exercisers and non-exercisers would depend on the variance of such confounders in the study population. Studies using randomization to assign participants to exercise versus control manipulations do not suffer from this bias.

A second potential source of heterogeneity in findings is the mixed use of regular physical activity versus measures of cardiorespiratory fitness as the independent variable explaining differences in cardiovascular stress reactivity. These concepts are often treated as interchangeable, whereas empirical observations of correlations between regular physical activity and fitness measures typically do not exceed .40 (Siconolfi et al., 1985; Morrow & Freedman, 1994; Aadahl et al., 2007; Emaus et al., 2010; Minder et al., 2014).Therefore, mixing cardiorespiratory fitness and physical activity effects on stress reactivity is likely to induce heterogeneity.

In addition, whereas cardiorespiratory fitness has a well circumscribed definition, regular physical activity is a complex construct which can be defined and assessed in different ways. A frequently used measure is total daily physical activity derived from a self-report questionnaire, which is subject to recall and response bias, and therefore frequently underestimated as well as overestimated (Prince et al., 2020). Fortunately, self-reporting becomes more reliable for moderate-to-vigorous activities, particularly when they are voluntary and salient like sports and exercise activities in leisure time (van der Zee et al., 2019, van der Zee, Schutte & de Geus, 2019, van der Zee et al., 2020). Reliable self-reports may be feasible when activities have a relatively fixed intensity and duration, like minutes spent on cycling to work or taking a well-defined walk but become difficult for activities which are more variable and lack clear boundaries. For the latter, accelerometer assessment is a far more reliable alternative (Slootmaker et al., 2009). The above makes clear that different definitions result in different physical activity measures, which induce heterogeneity that could distort possible cross-stressor adaptation effects on stress reactivity (Forcier et al., 2006).

Even when studies restrict themselves to measuring the uniformly defined construct of aerobic fitness, findings on cross-stressor adaptation remain confusing. This is illustrated by two meta-analyses performed in 2006 on the specific relationship between aerobic fitness and cardiovascular reactivity to acute laboratory stress. The meta-analysis by Jackson and Dishman (2006), using VO_2max as the indicator of aerobic fitness, found an overall *higher* cardiovascular stress reactivity in more fit participants, particularly for heart rate (HR) and heart rate variability (HRV) reactivity. This higher reactivity was, however, paired to a better recovery after the stressor, which could be a relevant advantage when dealing with repeated stress exposure. The meta-analysis by Forcier and colleagues (2006) with baseline HR as their aerobic fitness indicator showed partially contrasting results. They report an overall *lower* HR and systolic blood pressure (SBP) reactivity to stress, although they did fully corroborate the faster HR recovery after stress. Both studies also illustrate that the meta-analytic effect sizes are very small and strongly

heterogeneous across included studies. Moderator analyses showed this heterogeneity to be partly caused by the previously mentioned differences in study design (cross sectional or intervention studies) and the population included (healthy or at risk, general population, or high stress occupation, young or older, males and/or females) but another important determinant was the type of stressor used (physiological, mental, or social-evaluative). This issue of heterogeneity as a source of mixed results was recently addressed again in a systematic review by Chauntry and colleagues (2022). They focused solely on self-reported physical activity as their fitness measure. Two out of the six studies that measures ANS stress reactivity identified a significantly lower HR response to stress in more active individuals, and one study reported higher HR recovery in more active individuals. None of the studies observed an effect for HRV.

To specifically reduce the heterogeneity caused by the use of different types of stressors, Mücke and colleagues (2018) performed a systematic review including only a single stress paradigm, the Trier Social Stress Test (TSST). They also only included studies with a cross-sectional design, and explicitly tested for differences in the definition of the exposure variable, e.g., using dichotomies based on measured aerobic fitness level versus dichotomies based on the amount of regular physical activity. Furthermore, they took into account the assessment methods for physical activity. In spite of homogeneous cross-sectional design and the use of a single stressor, results were again mixed. This could be largely attributed to the definition and assessment of the exposure variable. From the studies included by Mücke and colleagues (2018) that included ANS reactivity as their outcome measure, a relationship between questionnaire derived physical activity and lower HR reactivity combined with faster recovery can be observed. Studies using VO_2max as the aerobic fitness measure, however, found a higher HR reactivity to the stressor (albeit only significant in women) again paired to a faster recovery. While the single study using accelerometers to obtain total physical activity did not find an effect of physical activity levels on stress reactivity or recovery. In addition, Mücke and colleagues (2018) also addressed the effect of physical activity on psychological stress reactivity. They found that overall participants who engaged more in physical activity (measured by either a questionnaire or accelerometry, but not VO_2max) showed a lower negative affective response to the stressor. The review by Mücke and colleagues (2018) again illustrates the importance of the exact construct used as the exposure variable.

While the current literature provides some support for the association of fitness with higher reactivity, and of physical activity with lower reactivity, and for both fitness and physical activity to yield faster recovery from stress, laboratory studies have not yet provided an unequivocal answer to the validity of the cross-stressor adaptation

hypothesis. Importantly, most laboratory tasks used in the studies reviewed so far typically elicit weaker physiological and psychological responses of much shorter duration than those found for naturalistic stressors (Peronnet & Szabo, 1993; Sothmann, Hart & Horn, 1991). To increase the ecological validity of cross-stressor adaptation effects, and with that their clinical relevance, Tonello and colleagues (2014) reviewed studies done between 2007 and 2013 that examined the association between questionnaire-based physical activity, daily HRV levels and the subjective experience of work stress. Overall, they found that higher levels of work stress were associated with lower HRV, but the evidence for a stress-buffering effect of physical activity on HRV remained inconclusive (Tonello et al., 2014). Four more recent studies directly linked questionnaire-based physical activity and VO_2max with the effects of daily stress on cardiovascular and affective measures (Chovanec & Gröpel et al., 2020; Gnam et al., von Haaren et al., 2016; Schilling et al., 2020). The first of these studies was performed in 61 inactive male engineering students of which half engaged in a 20-week aerobic training program. They found reduced heart rate variability reactivity during an examination period in students who participated in the aerobic exercise training program in comparison to sedentary controls (von Haaren et al., 2016). The second study was conducted in firefighters during the final exam of their vocational training program. No beneficial effect of either physical activity or VO_2max on the HR and HRV response to the exam was found. Instead, more physically active firefighters showed higher cognitive stress appraisal levels compared to the less physically active firefighters (Gnam et al., 2019). The third study was performed in a population of 173 police officers (66.5% male, mean age 37). Higher VO_2max was associated with reduced HRV reactivity to perceived acute work stress and increased HRV recovery at night. However, no relationship between VO_2max and positive or negative affect was observed (Schilling et al., 2020). The fourth study was performed in 52 female college students who engaged in either an eight-week endurance exercise training program (N = 18), an eight-week resistance training group (N = 21) or were placed on a waiting list (the control group; N = 13). Both training programs led to significantly increased VO_2max and reduced the subjective experience of daily life stress and HR recovery time from audiovisual stress stimuli, as compared to the control group (Chovanec & Gröpel, 2020). While informative, the expanding body of daily life studies does not yet elucidate the validity of the cross-stressor adaptation hypothesis. They exhibit a similar heterogeneity as the larger body of work using laboratory stressors.

The aim of the current study was to re-examine the association of both aerobic fitness and regular physical activity with stress reactivity and recovery, with specific attention to methodological aspects that could moderate these associations. We include three different exposure measures, (1) aerobic fitness, operationalized as the VO_2max derived

from a submaximal test, (2) self-reported weekly minutes spent on sports and exercise activities in leisure time, and (3) an index of the total amount of moderate-to-vigorous physical activity which includes the above sports and exercise activities, but also self-reported weekly minutes spent on walking and cycling. Whereas most previous studies have used heart rate and blood pressure reactivity as the main outcomes, we focus on the activity of the sympathetic (SNS) and parasympathetic (PNS) branches of the ANS separately. These are the main effectors causing the feed forward changes in heart rate (Robinson et al., 1966) and blood pressure (Yang et al., 2017), and should therefore more directly reflect adaptations in the central ANS control over the cardiovascular system. In addition, different patterns of SNS and PNS co-activation, co-inhibition, or reciprocal activation/inhibition can lead to similar end-organ responses (Berntson et al., 1994) and cross-stress adaptation may well depend on a change in such patterns. SNS reactivity is measured using changes in the cardiac pre-ejection period (PEP) and non-specific skin conductance response (ns.SCR) frequency. PNS reactivity is measured using changes in respiratory sinus arrhythmia (RSA), taking into account parallel changes in respiration rate. Lastly, inter-beat-interval (IBI) is included as measure that reflects both SNS and PNS activity.

ANS reactivity and recovery are measured in a controlled laboratory setting using both cognitive and social-evaluative stressors. To specifically address the cross-stressor hypothesis in a naturalistic daily life setting, we further use a continuous 24-hour measurement of ANS activity combined with an hourly digital diary to obtain information on work or leisure setting, level of like or dislike of their current activity, and positive and negative affect state.

METHODS

Study population

The main focus of the parent project of this study was the validation of a wristwatch-based technology to assess the relationship between ANS activity and stress in daily life. Recruitment of participants and the laboratory protocols are described in detail elsewhere (van der Mee et al., 2021). Briefly, participants were required to be between the age of 18 and 48, Dutch speakers, and currently employed, or in a schooling trajectory. Exclusion criteria were a body-mass index above 30, heart disease, high blood pressure, high cholesterol, diabetes, thyroid or liver disease, and use of antidepressants, anticholinergics, or any other medication that has been shown to influence the SNS. Female participants were

measured within the first two weeks following the last day of their menstrual cycle to account for hormonal changes.

Participants who were students received research credits, while other participants were compensated with a €50 gift voucher. All participants provided written informed consent before the start of the experiment. The study was approved in institutional review by the VUmc medical ethical committee (METc VUmc #2017.374, ABR #NL62442.029.17).

Procedure

Participants visited the laboratory on two consecutive days. During their initial visit to the laboratory (~1 h) participants provided informed consent and were interviewed about their physical activity behaviors. During this structured interview, the participants' systolic and diastolic blood pressure (SBP, DBP) were measured twice. Subsequently, the ANS measuring devices were applied to the participant to continuously measure ANS activity. The participants were provided with an iPod containing the questionnaire application. The experimenter practiced all items of the questionnaire with the participant to make sure they understood each item. They were informed that the iPod would go off hourly between the hours of 07:30 am and 11:00 pm. Each participant received 15 diary prompts but was allowed to manually fill in extra diaries by opening the app if they went to bed after 11 pm or woke up earlier than 7:30 pm. They were also informed that a random jitter of 15 minutes was added around each diary prompt, to reduce expectation effects.

Once equipped with the measuring devices and the iPod, participants left the laboratory for a day of daily life monitoring. During the 24-hour recording only a few restrictions of normal activities were applied. Participants were requested not to take a bath or engage in water sports. They were asked to remove the devices (but not the electrodes) during and reattach the devices afterwards.

Participants returned the next day for participation in the laboratory protocol. Upon their return, it was verified that all measurement equipment was still in working order. Next, to increase stress, participants were informed that footage of their facial expressions, posture and voice would be recorded during the experiment. Furthermore, their scores on the task were tracked on a score board containing other participants scores for comparison on their performance. Then all experimental manipulations were presented in a fixed order (see Table 1). After each stressor, the participants were asked to fill out a short affect questionnaire (see Table 1).

After the experimental session, all devices were removed, and participants were provided the opportunity to use a nearby shower. The experiment ended with a debriefing in which they were informed that the tasks were purposefully made so difficult so that

they would be impossible to perform without errors. They were explicitly told that the test score rankings were only added to increase the stressfulness of the task and did not reflect their actual ability. Furthermore, they were informed that no actual voice or video recording had been made.

Table 1. Experimental timeline*

Experimental condition	Duration (minutes)	Affect measurement
Baseline	3	Yes
Tone Avoidance (TA)	4	Yes
Rest	2	No
short Sing-a-Song Stress Test ($_{short}$SSST)	6.5	Yes
Rest	2	No
Paced Auditory Serial Addition Test (PASAT)	4	Yes
Rest	2	No
Raven's progressive matrices (RPM)	4	Yes
Break (application of CosMed)		
Treadmill intensity 1 (4.5 – 5 km/h)	4	No
Treadmill intensity 2 (6 - 6.5 km/h)	4	No
Treadmill intensity 3 (7.5 - 8 km/h)	4	No
Treadmill cooling down (3.7 – 4 km/h)	3	Yes
Rest	3	Yes

**The timeline only presents task relevant for the present paper in their presentation order. The full timeline of all experimental conditions can be found in* (van der Mee, Gevonden, Westerink, & de Geus, 2021).

Demographics

A structured interview was conducted before the start of the experiment to ensure participants were eligible to partake in the study. In addition, the interview included questions regarding their age, gender identity, physical activity behavior (for details see section 2.6 Physical Activity), subjective mental health and physical health on a scale of 1 (very poor) to 5 (very good), and experienced work and home stress on a scale of 1 (never) to 5 (very often).

Physiological measures

The physiological measures IBI, RSA and PEP and ns.SCR on the palm of the hand were obtained with a VU-AMS device (version 5-wire 5fs). The VU-AMS is a lightweight

portable device that has been used to measure ANS activity in over 300 scientific studies (see: http://www.vu-ams.nl/research/publications/ for an overview). It records electrocardiogram (ECG) and impedance cardiogram (ICG) from five adhesive 55 mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Eindhoven, Netherlands) placed on the chest and back of the participants with a recording frequency of 1000 Hz (de Geus et al., 1995; Willemsen et al., 1996). VU-AMS data was analyzed with the Vrije Universiteit Data Acquisition and Management Software (VUDAMS version 4.6, available at: http://www.vu-ams.nl/support/downloads/software/). For each experimental condition average values for each ANS measure were calculated. The IBI is calculated based on the time difference between two successive R peaks. RSA is calculated by means of a peak valley method, which combines the R-peaks time series with the impedance derived respiration cycle. In this method the shortest IBI during each inspiration and the longest IBI during each expiration are detected. Then the former is subtracted from the latter. When the calculation of the RSA results in zero or negative values they are coded as zero (de Geus et al., 1995; Goedhart et al., 2007). RSA is well validated measure of PNS activity (Katona & Jih, 1975; Berntson et al., 1993; Migliaro, 2020) and has been frequently studied in relationship to stress (e.a. Beauchaine, 2015; Beauchaine et al., 2019; Campbell et al., 2019; Lane et al., 1992; Tonhajzerova et al., 2016). The PEP is obtained by calculating the time difference between the start of ventricular depolarization (Q onset) in the ECG and the time the aortic valve opens (B point) in the ICG (Nederend et al., 2018; Willemsen et al., 1996). For each time segment of interest, a single averaged ICG complex was derived by means of ensemble averaging of the ICG signal over all R-peaks in the conditions, as explained by Riese and colleagues (2003). Extensive construct and criterion validity has been demonstrated for this method (Nederend et al., 2018, Willemsen et al., 1996). Various studies have shown that the PEP is response to stress, in which a shorter PEP (due to increases SNS activity) is indicative of more stress (Brindle et al., 2014; van der Mee et al., 2020; van der Mee et al., 2021; Rahman et al., 2018). The VU-DAMS software automatically detects and scores the various attributes necessary to calculate the IBI (R-peaks), RSA (R-peaks and respiration) and PEP (Q-onset and B-point). All data scoring of the VU-DAMS was manually checked and if necessary corrected.

For the laboratory section of the study ns.SCRs were obtained from electrodermal activity (EDA) as measured with the VU-AMS on the palm of the hand and on the wrist by a wristwatch. During the daily life section EDA was only obtained with a wristwatch. The wristwatch was a CE approved wearable skin conductance wrist sensor, type DTI5 (Discreet Tension Indicator version 5, Philips), and was used to measure ns.SCR frequency. This wristwatch has been shown to sufficiently capture SNS activity (van der Mee et al.,

2021). The ns.SCRs frequency is defined as the number of peaks per minute. For palm EDA during the laboratory recording, the ns.SCR frequency was obtained using the EDA master toolkit (Joffily, EDA Master Toolbox, 2012) in MATLAB. For wrist EDA during laboratory and daily life recording, the ns.SCR frequency was obtained by an internal method of peak detection that makes use of a curve fit method. For more details on EDA scoring see van der Mee and colleagues (2021). Ns.SCR frequency has been shown to relate to negative emotions (Nikula, 1991; van der Mee et al., 2021), arousal (Nikula, 1991), and stress (Miller & Shmavonian, 1965; Kelsey, 1991) We recently showed this measure to perform even better than the widely used skin conductance level to index changes in SNS activity across a wide variety of stressors (van der Mee et al., 2021).

Ambient temperature and humidity were continuously measured with a thermosensor (Hygrochron iButton, UK) worn on the outer clothing. In addition, skin temperature was continuously measured from a thermosensor (Thermochron iButton, UK) placed directly onto the skin under the left clavicle bone using double adhesive rings (20 * 5 mm) for cup electrodes. In addition, continuous passive sensing through a triaxial accelerometer, embedded in the VU-AMS, was used to detect activity levels. Average activity level was computed by the root of the mean of the squared the X-, Y-, and Z-axis accelerations.

During the structured interview, SBP and DBP were measured twice with the Omron M4-I, HEM 752A. Resting SBP and DBP were calculated by taking the mean of the two measurements.

Affect

An iPad containing an in-house built electronic diary application was provided to participants to report their affect at set times in the laboratory. For the daily life section participants received an iPod containing the same electronic diary application. In both settings, affect was rated with a shortened version of the Maastricht Questionnaire (Myin-Germeys et al., 2001). Positive affect scores were obtained by asking the participants to rate on a scale of 1 (not at all) to 7 (very) whether they felt relaxed, cheerful, enthusiastic, and content and averaging the score over the 4 items. Negative affect was obtained by averaging the scores for 5 items: insecure, lonely, anxious, irritated, and down. In daily life, participants also rated the degree of liking the activity they were engaged in at that moment in time (work/study, leisure, household chores, transportation, relaxing, sleeping). Participants indicated whether or not they would rather be doing something else (on a scale of 1 (strongly like) – 7 (strongly dislike). The like-dislike item was recoded into a binary variable based on the grand median score, in which a score <= median indicated

they liked the activity and a score > median indicated they did not like the activity they were doing.

VO_2max

The maximal volume of oxygen uptake (VO_2max) is derived from a submaximal test. Participants engaged in a treadmill exercise at 3 incremental stages of speed (males: 5, 6.5, and 8 km/h; females: 4.5, 6, and 7.5 km/h), each lasting 4 minutes. After a 3-minute cooling-down on the treadmill (males: 4 km/h, females: 3.7 km/h) participants sat down for a 3-minute recovery stage.

Volume of oxygen (O2) uptake and carbon dioxide (CO2) production were recorded breath-by-breath with a telemetric gas exchange system (Cosmed K5 , Rome, Italy). During the course of the experiment, the main sample unit and the battery pack were attached to the back of the subject. Before each test, the O_2/CO_2 analysis system was calibrated with ambient air and a gas mixture that had an O_2 concentration of 16% and a CO_2 concentration of 5%. The calibration of the turbine flowmeter was performed by via a 3-liter syringe (Crouter et al., 2019).

The last minute of each incremental treadmill stage was included in a linear regression between O_2 uptake and HR (derived from the VU-AMS) for each participant separately to derive their individual regression equation. The last minute was chosen to ensure participants had reached a steady state. Maximum oxygen uptake was then calculated by entering the maximal heart rate, defined as 220 minus the participant age, into their individual regression equation. The resulting value was divided by the participant's weight resulting in VO_2max as measured in milliliter per kilogram per minute. The validity of a graded submaximal test to predict an individual's VO_2max has been shown to correlate strongly with the actual VO_2max (Ekblom-Bak et al., 2014; Grant et al., 1995; Schutte et al., 2016).

Physical Activity

During a structured interview detailed information regarding the participants physical activity was collected. The interview included the following questions on exercise behavior: Do you exercise regularly? What type of exercise do you partake in? For how many years? How many months a year? How many times a week? How many minutes per time? Only exercise activities performed at least six months a year and at least once a week were included (thereby excluding ski holidays, sailing camps, swimming only during the summer, and similar). When the reported number of occasions, or session lengths were variable, an average number of occasions or session length was calculated. There

was no limit on the number of different exercise activities participants could report, and all were included in the study. For each exercise activity the total minutes spent on exercising per week was calculated (number of occasions x session length) and multiplied by their metabolic equivalent score (METscore) value derived from the 2011 Compendium of Physical Activities (Ainsworth et al., 2011) to obtain exercise activity in metabolic equivalent hours (MET-hours).

With regard to other types of common moderate-to-vigorous physical activity (MVPA), the following questions were included: How many minutes in total do you spend walking during the workweek? How many minutes in total do you spend walking during the weekend? How many minutes in total do you spend cycling during the workweek? How many minutes total do you spend cycling during the weekend? This excluded walking or cycling mentioned under the exercise activities but included walking/cycling as a means of transportation, walking a dog or walking/cycling for relaxation. The total minutes per week spent walking and cycling (sum of weekdays and weekend days) was multiplied with their METscore and added to the exercise METhours to obtain total energy spent on MVPA.

The questions included in the interview are obtained from the questionnaires used by the Netherlands Twin Register to quantify, amongst others, leisure time exercise behavior and MVPA (van der Mee et al.2018, Willemsen, et al., 2013; van der Zee et al., 2019). Quantification of exercise and MVPA in terms of their METs does come with a limitation (Byrne et al., 2005; Franklin et al., 2018), but is currently the only metric available to take into account exercise intensity in addition to exercise time when information on heart rate and oxygen consumption during physical exertion in daily life are not available. The exercise METhour construct has been related to amongst others well-being (Stubbe et al., 2007) and mental health disorders (de Moor et al., 2006; de Moor et al., 2008). Furthermore, this construct has been shown to have high (> .82) test-retest reliability and high temporal stability, even across periods of 20 years (de Geus et al., 2014; Stubbe et al., 2006; van der Zee et al., 2020).

Stress reactivity and recovery

Laboratory

The mental stressors used in this study are the Tone avoidance (TA) task and the Paced Auditory Serial Subtraction (PASAT) task. The TA task aims to induce effortful active coping in which participants have to avoid a loud tone by pressing a button on the opposite site of an "X" presented on one of the four corners a computer screen (de Geus et al., 1990; van der Mee et al., 2020). The PASAT is a calculus task with a staircase algorithm to measure capacity and rate of information processing and sustained and divided attention.

Single digits are presented every 3 seconds and the respondent must add each new digit to the one immediately prior to it before the next digit is presented (Tombaugh, 2006; Sampson 1958; Sampson & MacNeilage, 1960).

The social evaluative stressors used in this study are the short Sing-a-Song-Stress-Test (SSSTshort) and the Raven Progressive Matrices IQ (RPM) test. In the SSSTshort participants unexpectedly have to sing a song out loud in front of a camera and the experimenter (van der Mee et al., 2020). The RPM test (Raven, 2003) was timed, participants had to solve as much matrices as possible in a 4-minute time window. They were informed that the more correct answers they gave the higher their IQ score would be.

Laboratory ANS stress reactivity values were calculated by subtracting the mean value during the baseline condition from the mean value during the respective stress tests for each ANS measure (IBI, RSA, PEP, and ns.SCRs 2x) and respiration rate. Recovery values were calculated by subtracting the mean value during the 2-minute rest period following each stress task from the mean value during that respective stress task. Reactivity in positive and negative affect was obtained by subtracting the scores of the affect questionnaire filled in after the baseline sitting condition from the scores immediately after the stress test. Since no affect was measured after resting periods no affect recovery could be calculated.

To obtain a single mental stress reactivity and recovery score per ANS measure, the calculated reactivity and recovery scores from the TA task and the PASAT were averaged. Similar, to obtain a single social-evaluative stress reactivity score, the calculated reactivities from the SSSTshort and RPM were averaged. Since the RPM did not have a recovery period following the task, the social-evaluative recovery score is equal to the SSSTshort recovery score.

The mean ANS and affective values during the baseline condition were also considered variables of interest.

Daily life

For each diary entry the average ANS values during the 5 minutes preceding the entry were calculated. Regarding sleep, hourly averages were created from the reported moments of going to sleep to the reported moment of getting up (both verified by the accelerometer signal). For each participant, the average ANS values were calculated across all valid 1-hr sleep epochs ('sleep') and across all valid 5-minute periods segments that met the category criteria: work/study, leisure time, liked activity (< median), not liked activity (>= median), awake (irrespective of activity performed). Due to the large influence of major body movements on ANS activity (Fu & Levine, 2013) during the awake period, only

the 5-minute periods with accelerometer values < 50 milli G acceleration, consisting of minor body movements, were included. Only if a participant had at least 3 observations for a given activity category, his/her data for that category were included in the analyses.

From the daily life data two physiological stress reactivity measures were derived and one recovery measure. Stress reactivity was defined as 1) the difference between work and leisure activities (calculated as work – leisure) and 2) the difference between liked and disliked activities (calculated as dislike – like). Daily life stress recovery was defined as the difference between awake and sleep (sleep – awake).

A similar approach was applied to the positive and negative affect scores to obtain subjective stress reactivity measures. A mean value was calculated over all diary entries for which a valid ANS value was available, and these were also averaged across work/study, leisure time, liked activity, disliked activity, and total time awake. Subjective stress reactivity scores were computed for positive and negative affect separately, one by contrasting affect during work vs. leisure and one by contrasting affect during disliked vs. liked activities.

Covariates

Several variables were of interest as possible covariates. First, there is evidence that males and females respond differently to stressors with regard to their ANS response, with males being "vascular" reactors and females' "cardiac" reactors (Huang et al., 2013). In addition, males have, on average, a higher VO_2max than females (Wang et al., 2010). A second covariate is age, since with age VO_2max decreases (Wang et al., 2010), ANS activity changes (Peters et al., 2020), and physical activity decreases (van der Zee et al., 2019; Sallis, 2000).

In addition to age and biological sex, a few other covariates were considered that may impact ANS reactivity/recovery. First, the electrodermal activity measure ns.SCR frequency could be influenced by the ambient temperature and/or humidity and body temperature due to involvement of sweating in thermodynamics (Boucsein, 2012). Though our previous work in the laboratory has shown that the ns.SCR frequency measure is less sensitive to thermodynamic effects as compared to skin conductance levels (van der Mee et al., 2021), in a daily life setting these factors are much more dynamic. Second, changes in respiration rate may drive changes in RSA independent of changes in PNS activity (Grossman & Taylor, 2007) and we therefore recorded changes in respiration rate using the thorax impedance signal as outlined previously (Houtveen et al., 2006).

Analytical strategy

Before analyses all variables were checked for outliers. A value was considered an outlier if it deviated more than 4.5*SD from the grand mean. Next, we performed a manipulation check to assess whether the tasks and recovery periods induced significant changes in ANS activity. Paired-samples t-tests were performed for all stress reactivity (task vs. baseline) and recovery (task vs. recovery) contrasts.

Testing of the main hypothesis of cross-stressor adaptation revolved around establishing an association of the aerobic fitness and regular physical activity traits with the stress reactivity and recovery scores, across multiple tasks and settings. Separate linear regression models were run with either VO_2max, MVPA, and exercise as predictors and either ANS (IBI, RSA, PEP, ns.SCR) and affect (NA, PA) baseline, reactivity, and recovery values as outcomes. A total of 93 (31 for each physical activity measure) linear regression model were performed for the laboratory data and a total of 78 (26 for each physical activity measure) regression models were performed for the daily life data.

For the laboratory, the ANS and affect outcomes were baseline stress levels, mental stress reactivity, mental stress recovery, social stress reactivity and social stress recovery. For daily life, the outcome variables were average levels during sleep, average levels during general wakefulness, work-leisure reactivity, dislike-like reactivity, and awake-sleep recovery. Because of sex and age differences observed in VO_2max and ANS reactivity, we checked whether the inclusion of sex or age in the regression analyses changed the results, which they did not. However, they were still included in all analyses. Finally, the analyses were rerun for RSA and EDA with variable-specific covariates added to the respective models (i.e., temperature/humidity for EDA, and RR for RSA).

The relative explained variance in physiological stress reactivity by the exposure measures on the outcomes within each model was based on the partial R^2, calculated with the rsq.partial() function of the "rsq" package in R. The partial correlation coefficient reflects the strength of the relationship between two variables after the correlation of both the outcome and the predictor variable with the covariates is taken into account. To ease comparison with meta-analytic results we additionally report the Cohen's d based on the partial r (obtained by taking the square root of the reported R^2) with the formula:

d = 2*r / √(1-r2)."

To account for multiple testing while taking into account that the number of effective tests is lower than the total tests, we used the correction for non-independent tests implemented in the R package *meff* (Nyholt, 2005; Salyakina et al., 2005). Separately for the

laboratory and daily life data, the zero-order correlation matrix among all stress reactivity and recovery and the three physical activity variables was used to compute the number of effective tests (35 for the laboratory and 30.6 for the daily life data). Significance levels for the laboratory and daily life tests were adjusted from nominal 0.05 to p = .05/35 = .0014 and p =.05/36 = .0016. Power analyses (performed with R package "pwr") showed that with a df(3,116) and a significance level of .0014 we had a power of .02 to detect a small effect (f2 = 0.02/(1 – 0.02)), a power of .78 to detect a medium effect (f2 = 0.15/(1-0.15)) and a power of .99 to detect a large effect (f2 = 0.35/(1 - 0.35)). The minimal effect size (quantified as cohen's d) that could be identified in the current sample with a power of 70%, df(3,116), and a p-value set at a nominal p = .05 was medium (d = 0.57). This indicates that the power of the current study to detect the small effects for HR reactivity (d 95% CI = 0.05 – 0.11) and recovery (d 95% CI = -0.35 – -0.19) reported by the meta-analyses of Jackson and colleagues (2006) was likely low. However, as mentioned in the introduction, these analyses were plagued by heterogeneity issues which might have reduced the meta-analytical estimates for the effect sizes. Indeed, the daily life studies by von Haaren and colleagues (2016) and Chovanec and Gröpel (2020), show far larger effect sizes ranging from d = 0.34 – 0.66. Using more strict definitions of the fitness/physical activity predictors and a variety of homogenous stressors, including negatively valued daily life activities, we expected to find at least medium effect sizes

RESULTS

Study population

Usable data were obtained in 116 participants out of 121 participants originally recruited in the study. Two participants were excluded because they were outliers in terms of age (they were > 45, while all other participants were ≤ 30). One participant was excluded because their data was an outlier on all ANS measures, one participant had insufficient data quality, and one participant withdrew from the study and requested their data to be removed. The final sample had a mean age of 22.48 (SD = 3.56) and 57.76% were female. The majority of the participants were students (81.0%), had good self-rated mental health (21.55% very good, 62.93% good, 12.07% intermediate, 3.44% poor, 0% very poor), and good self-rated physical health (13.79% very good, 69.82% good, 13.79% intermediate, 2.58% poor, 0% very poor). Experiences stress at work (3.44% never, 52.58% sometimes, 29.31% frequently, 12.93% often, 1.74% always) or at home (17.24% never, 56.03% sometimes, 18.96% frequently, 6.89% often, 0.86% always) was low, with most participants reporting

less than frequent stress. For 23 participants VO_2max could not be calculated due to device malfunction (10), missing data (7), too few valid data points for analysis (3), or outlying VO_2 value (3). The mean VO_2max was 43.95 ml/kg/min (*SD* = 9.29), mean MVPA was 73.92 MET-hours (*SD* = 48.96), and mean exercise was 37.19 MET-hours (*SD* = 40.11). The mean SBP was 116.08 mmHg (*SD* = 10.88) and mean DBP was 71.20 mmHg (*SD* = 8.31). Consistent with previous findings males had a higher VO_2max compared to females ($\Delta M = 8.55$, $t(63.66) = 4.61$, $p < .001$) (Wang et al., 2010), but there were no differences with regard to MVPA ($\Delta M = 10.66$, $t(94.01) = 1.13$, $p = .26$) and exercise MET-hours ($\Delta M = 14.22$, $t(83.83) = 1.82$, $p = .072$).

Table 2. Overview of number of participants per daily life category.

Activity	Total	IBI	RSA	PEP	Wrist ns.SCR	Positive affect	Negative affect
Awake	113	113	112	111	111	105	105
Like	107	106	105	104	103	101	101
Dislike	105	105	104	103	103	97	97
Dislike reactivity	105	98	97	96	94	93	93
Leisure	88	88	87	87	84	82	82
Work/study	57	57	57	56	56	52	52
Work reactivity	37	37	37	37	36	34	34
Sleep	108	107	108	104	106	-	-
Sleep reactivity	108	107	107	104	106	-	-

Compliance for e-diary entries during the daily life part of the experiment was good. On average participants had 13.5 entries out of the 15 prompts (range: 4 - 19, *SD* = 2.98). Two participants did not have at least three observations for any activity category and were excluded from the analyses. One participant was removed due to poor overall data quality. For the wrist ns.SCR analyses one outlying value was removed in liked activities, one in leisure and two for dislike reactivity. The other variables contained no outliers.

Some participants only had three or more valid observations for a subset of conditions, leading to different numbers or participants included in each condition. Due to the exclusion of data segments that contained a high amount of movement, we lost data for half of the participants with regard to the work category. An overview of the number of participants per category per measurement is given in Table 2.

An overview of the mean values and SDs for our outcome measures of interest for the laboratory part is presented in Table 3 and for the daily life part in Table 4.

Table 3. Descriptive of stress measures in the laboratory

	Baseline		Stress		Rest		Reactivity		Recovery	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mental Stress										
IBI (msec)	843.42	132.75	779.37	125.59	839.66	127.65	-64.04	93.52	59.70	77.44
RSA msec	84.72	44.32	61.87	27.86	79.50	36.40	-22.25	37.73	17.44	30.16
PEP msec	112.75	15.45	105.53	15.88	111.98	14.77	-7.21	10.19	6.48	7.73
ns.SCR palm (p/m)	5.18	3.50	14.58	3.94	7.40	3.64	9.29	4.57	-7.21	3.82
ns.SCR wrist (p/m)	0.41	0.68	1.58	2.25	1.00	1.46	1.00	2.12	-0.54	2.07
Positive affect	4.41	0.93	3.44	0.96	-	-	-0.96	0.90	-	-
Negative affect	1.54	0.79	2.17	0.97	-	-	0.62	0.62	-	-
Social Stress										
IBI (msec)			773.66	115.71	840.27	120.44	-69.28	81.12	117.50	101.26
RSA msec			68.47	29.45	82.00	37.93	-14.54	37.85	18.64	35.50
PEP msec			106.34	16.14	112.13	15.75	-5.76	10.37	8.52	13.57
ns.SCR palm (p/m)			11.69	3.35	7.59	4.26	6.30	3.51	-5.64	5.28
ns.SCR wrist (p/m)			1.39	2.53	1.12	1.90	0.80	2.15	-0.46	4.03
Positive affect			3.88	0.89	-	-	-0.54	0.81	-	-
Negative affect			1.90	0.89	-	-	0.35	0.58	-	-

Table 4. Descriptive of stress measures in daily life

	Sleep		Awake		Sleep recovery	
	Mean	SD	Mean	SD	Mean	SD
IBI (msec)	1005.39	125.92	786.14	92.11	222.08	90.49
RSA (msec)	86.85	38.89	66.79	24.43	21.10	27.55
PEP (msec)	112.72	17.05	104.33	16.49	8.98	19.49
ns.SCR wrist (p/m)	2.56	2.04	2.48	1.95	-0.12	2.58
Positive affect	-	-	4.47	0.70	-	-
Negative affect	-	-	1.80	0.63	-	-

	Leisure		Work		Work-reactivity	
	Mean	SD	Mean	SD	Mean	SD
IBI (msec)	823.91	103.37	804.72	99.40	-27.87	81.39
RSA (msec)	73.16	29.30	72.78	26.29	-4.05	20.18
PEP (msec)	103.73	15.74	104.95	17.24	1.28	11.39
ns.SCR wrist (p/m)	2.19	1.94	2.12	1.90	-0.05	2.44
Positive affect	4.45	0.81	4.45	0.85	-0.08	0.81
Negative affect	1.68	0.61	1.91	0.74	0.17	0.49

	Like		Dislike		Dislike-reactivity	
	Mean	SD	Mean	SD	Mean	SD
IBI (msec)	791.99	96.14	776.57	99.11	-7.45	75.54
RSA (msec)	68.94	30.23	65.49	25.25	-3.29	22.48
PEP (msec)	102.89	16.60	105.59	17.50	1.80	9.72
ns.SCR wrist (p/m)	2.45	1.94	2.40	2.36	-0.15	1.72
Positive affect	4.72	0.73	4.20	0.77	-0.52	0.61
Negative affect	1.64	0.59	1.97	0.69	0.26	0.44

Stress reactivity

The manipulation check showed that both mental and social stress elicit a significant ANS stress response, lowered positive affect, and increased negative affect (Table 5). In the recovery period after the mental stressors the SNS activity (IBI, PEP, ns.SCR) significantly decreased and PNS activity (IBI, RSA) significantly increased again. A similar pattern is observed for social stress recovery, except for the palm ns.SCRs which remained elevated during the recovery period. Supplementary table 1 contains the full correlation matrix of variables measured during the laboratory session.

Table 5. Laboratory stress reactivity and recovery

		N	ΔM (SE)	*t*	*p*	*d*
Mental stress reactivity	IBI (msec)	116	-64.04 (8.68)	-7.37	< .001	-0.68
	RSA (msec)	109	-22.23 (3.61)	-6.15	<. 001	-0.60
	PEP (msec)	104	-7.22 (1.00)	-7.22	<.001	-0.71
	ns.SCR palm (p/m)	96	9.29 (0.40)	19.92	<.001	2.36
	ns.SCR wrist (p/m)	104	1.19 (0.19)	5.97	<.001	0.60
	Positive affect	115	-0.96 (0.08)	-11.50	< .001	-1.07
	Negative affect	115	0.62 (0.06)	10.70	< .001	1.00
Mental stress recovery	IBI (msec)	115	59.70 (6.77)	8.26	< .001	0.83
	RSA (msec)	104	17.44 (3.15)	5.90	< .001	0.55
	PEP (msec)	102	6.48 (0.76)	8.46	< .001	0.84
	ns.SCR palm (p/m)	97	-7.21 (0.45)	-18.53	< .001	1.61
	ns.SCR wrist (p/m)	103	-0.60 (0.28)	-3.03	.003	0.21
Social stress reactivity	IBI (msec)	115	-66.18 (6.95)	-9.16	< .001	-0.86
	RSA (msec)	105	-14.54 (3.69)	-3.93	< .001	-0.43
	PEP (msec)	103	-6.43 (1.02)	-6.37	< .001	-0.62
	ns.SCR palm (p/m)	87	6.30 (0.37)	16.76	< .001	1.85
	ns.SCR wrist (p/m)	103	0.99 (0.21)	4.51	< .001	0.45
	Positive affect	114	-0.51 (0.07)	-6.18	< .001	-0.64
	Negative affect	114	0.35 (0.05)	6.36	< .001	0.62
Social stress recovery	IBI (msec)	115	66.18 (7.03)	10.12	< .001	0.89
	RSA (msec)	104	12.93 (3.54)	4.65	< .001	0.37
	PEP (msec)	102	5.71 (0.97)	6.63	< .001	0.59
	ns.SCR palm (p/m)	83	-3.80 (0.59)	-8.65	< .001	-0.76
	ns.SCR wrist (p/m)	104	-0.26 (0.33)	-1.29	.19	-0.08

Less consistent results were found when comparing the different daily life activities. Participants had comparable ANS and affect values during leisure time and work, although negative affect was higher and IBI was lower at work (Table 6). Participants also had comparable ANS activity during liked and disliked activities, although with significantly higher positive affect and lower negative affect during activities they enjoyed with a strong effect size. However, stress reactivity values did show large individual differences as shown by the standard deviations in table 6. Results were more in line with expectations for the contrast between sleep and awake. This was highly significant for IBI, RSA and PEP. On average participants had a higher IBI, RSA and PEP during sleep, with a very strong effect for IBI and a medium to strong effect for RSA and PEP. Again, substantial individual

variation was seen in this recovery. Supplementary table 2 contains the full correlation matrix of variables measured during daily life recording.

Table 6. Ambulatory stress reactivity and recovery

		N	ΔM (SE)	*t*	*p*	*d*
Work - leisure reactivity	IBI (msec)	37	-27.87 (13.38)	-2.08	.044	0.23
	RSA (msec)	35	-4.05 (3.41)	-1.22	.23	0.02
	PEP (msec)	37	1.28 (1.87)	0.68	.49	0.10
	ns.SCR wrist p/m)	35	0.21 (0.29)	0.66	.51	-0.03
	Positive affect	34	-0.08 (0.14)	-0.62	.53	0.00
	Negative affect	34	0.17 (0.08)	2.11	.042	0.00
Dislike - like reactivity	IBI (msec)	98	-7.46 (7.73)	-0.96	.33	0.20
	RSA (msec)	97	-3.30 (2.28)	-1.44	.15	0.15
	PEP (msec)	96	1.80 (0.99)	1.82	.072	-0.28
	ns.SCR wrist p/m)	95	0.15 (0.17)	0.67	.50	-0.02
	Positive affect	93	-0.52 (0.06)	-8.26	< .001	0.84
	Negative affect	93	0.27 (0.04)	5.82	< .001	-0.74
Sleep recovery	IBI (msec)	107	222.08 (8.75)	25.38	<. 001	2.42
	RSA (msec)	107	21.10 (2.66)	7.92	< .001	0.72
	PEP (msec)	104	8.98 (1.91)	4.70	< .001	0.43
	ns.SCR wrist (p/m)	106	0.12 (0.25)	0.50	.61	0.03

Predicting reactivity and recovery from fitness or physical activity

For all analyses, addition of the temperature and humidity covariates had little influence on the relationship between fitness/MVPA/exercise and ns.SCR frequency. Therefore, results are reported without addition of these covariates. The linear regression analyses for the laboratory stressors (Table 7) and daily life (Table 8) show a significant positive relationship of aerobic fitness with laboratory baseline IBI and small positive relation with daily life IBI when awake, with moderate to strong effect sizes. Aerobic fitness also showed a small negative relation with daily life negative affect dislike reactivity and an unexpected positive relation with wrist ns.SCR when sleeping, suggesting higher SNS activity in sleep in more fit subjects. These were the only associations to emerge between all ANS baseline or reactivity measures and our measures of either aerobic fitness when corrected for multiple testing. Even when inspecting results at a nominal $p < .05$

only very scant support for the cross-stressor adaptation hypothesis is found. Only a moderate positive relation with laboratory PEP at baseline and a small negative relation with mental stress PEP reactivity could hint at lower basal SNS activity and attenuated SNS reactivity in more fit participants, while a moderate negative relation with RSA work reactivity suggests attenuated PNS withdrawal during work. The observed small negative relationship with wrist ns.SCR sleeps recovery is likely due to the higher ns.SCR when sleeping in fit individuals.

MVPA showed a moderate to strong significantly positive relationship with laboratory baseline IBI and daily life IBI when awake. No reactivity or recovery effects were found when correcting for multiple testing. At nominal $p < .05$, higher baseline RSA was observed both in the laboratory and during awake time in daily life, hinting at a role for the PNS in explaining the lowered heart rate in more physically active participants. Lower laboratory baseline palm ns.SCR in more active participants suggests this to be paired to lower SNS activity at rest, and the slightly increased palm ns.SCR mental and social laboratory stress reactivity may simply reflect the baseline – reactivity correlation. For MVPA, there also was a small positive relation with positive affect when awake, and a small negative relation with negative affective reactivity to disliked activities.

Exercise showed the least evidence for an effect on stress reactivity or recovery. For IBI and RSA it followed the pattern showed by MVPA, i.e. a moderate to strong significantly positive relationship with laboratory baseline IBI and daily life IBI when awake, paired with higher RSA.

Table 7. Linear regression predicting laboratory stress from fitness

		VO_2max							MVPA							Exercise						
		N	b	SE	t	*p*	R^2	*d*	N	B	SE	t	*p*	R^2	*d*	N	b	SE	t	*p*	R^2	*d*
Baseline	IBI (msec)	93	6.08	1.55	3.91	**<.001***	.38	1.56	115	0.89	0.24	3.68	**<.001***	.32	1.37	115	1.35	0.28	4.71	**<.001**	.40	1.63
	RSA (msec)	91	0.77	0.56	1.38	.17	.15	0.84	112	0.19	0.08	2.31	.022#	.21	1.03	112	0.22	0.10	2.15	.033#	.19	0.97
	PEP (msec)	84	0.51	0.21	2.44	.017#	.26	1.18	105	0.04	0.03	1.42	.16	.14	0.80	105	0.06	0.03	1.69	.093	.16	0.87
	ns.SCRp (p/m)	83	-0.06	0.04	-1.35	.18	.15	0.84	102	-0.016	0.007	-2.19	.024#	.21	1.03	102	-0.01	0.008	-1.91	.058	.18	0.94
	ns.SCRw (p/m)	85	0.01	0.01	0.98	.33	.06	0.50	104	-0.001	0.001	-0.87	.38	.13	0.77	104	-0.001	0.001	-0.56	.57	.12	0.74
	Positive affect	92	0.01	0.01	0.85	.40	.02	0.28	114	0.003	0.002	1.62	.11	.08	0.59	114	0.003	0.002	1.31	.19	.09	0.63
	Negative affect	92	-0.004	0.01	-0.37	.71	.02	0.28	114	-0.002	0.001	-1.72	.088	.08	0.59	114	-0.002	0.002	-1.37	.17	.09	0.63
Mental stress reactivity	IBI (msec)	93	-1.42	1.06	-1.33	.18	.13	0.77	115	-0.12	0.18	-0.68	.50	.06	0.50	115	-0.24	0.22	-1.12	.26	.11	0.70
	RSA (msec)	87	-0.61	0.46	-1.34	.18	.02	0.28	108	-0.12	0.07	-1.68	.096	.16	0.87	108	-0.13	0.09	-1.44	.15	.14	0.80
	PEP (msec)	83	-0.29	0.12	-2.38	.020#	.10	0.66	103	-0.003	0.02	-0.17	.86	.01	0.20	103	-0.02	0.02	-0.84	.40	.08	0.59
	ns.SCRp (p/m)	76	0.05	0.06	0.90	.37	.14	0.80	93	0.02	0.008	2.78	.006#	.28	1.24	93	0.02	0.01	2.26	.026#	.23	1.09
	ns.SCRw (p/m)	83	-0.01	0.02	-0.55	.58	.02	0.28	102	0.002	0.003	0.48	.63	.04	0.41	102	0.002	0.005	.49	.62	.05	0.46
	Positive affect	92	0.003	0.01	0.30	.76	.02	0.28	114	0.002	0.001	1.29	.20	.11	0.70	114	0.003	0.002	1.59	.11	.14	0.80
	Negative affect	92	-0.003	0.008	-0.39	.70	.02	0.28	114	-0.001	0.001	-0.79	.43	.08	0.59	114	-0.002	0.001	-1.78	.077	.17	0.90
Mental stress recovery	IBI (msec)	93	0.91	0.75	1.20	.23	.11	0.70	114	0.10	0.14	0.70	.48	.06	0.50	114	0.13	0.17	0.75	.45	.07	0.55
	RSA(msec)	87	0.27	0.40	0.67	.50	.07	0.55	105	0.04	0.06	0.62	.53	.06	0.50	105	0.01	0.07	0.16	.86	.01	0.20
	PEP (msec)	83	0.08	0.08	1.03	.30	.11	0.70	102	0.008	0.01	0.51	.60	.05	0.46	102	0.01	0.02	0.74	.46	.07	0.55
	ns.SCRp (p/m)	79	-0.02	0.06	-0.30	.76	.04	0.41	98	-0.01	0.009	-1.71	.090	.19	0.97	98	-0.001	0.01	-0.06	.95	.02	0.28
	ns.SCRw (p/m)	85	0.008	0.03	0.24	.81	.03	0.35	103	0.001	0.005	0.29	.77	.03	0.35	103	0.005	0.007	0.77	.44	.07	0.28

Table 7. Linear regression predicting laboratory stress from fitness (continued)

		VO_2max							MVPA							Exercise						
		N	b	SE	t	*p*	R^2	*d*	N	B	SE	t	*p*	R^2	*d*	N	b	SE	t	*p*	R^2	*d*
Social stress reactivity	IBI (msec)	92	-0.12	1.00	-0.12	.90	.01	0.20	114	-0.04	0.15	-0.29	.77	.02	0.28	114	-0.16	0.19	-0.83	.41	.08	0.59
	RSA (msec)	85	-0.16	0.45	-0.37	.71	.04	0.41	104	-0.11	0.07	-1.56	.12	.15	0.84	104	-0.08	0.09	-0.84	.40	.08	0.59
	PEP (msec)	83	-0.12	0.13	-0.90	.37	.01	0.20	102	0.02	0.02	1.16	.25	.16	0.87	102	0.01	0.02	0.64	.53	.09	0.63
	ns.SCRp (p/m)	68	0.06	0.05	1.39	.17	.17	0.90	86	0.02	0.007	2.44	.016#	.24	1.12	86	0.01	0.009	1.19	.23	.12	0.74
	ns.SCRw (p/m)	83	-0.05	0.02	-1.87	.064	.20	1.00	102	0.0006	0.004	0.14	.88	.008	0.18	102	-0.002	0.005	-0.45	.65	.05	0.46
	Positive affect	91	-0.004	0.01	-0.42	.67	.05	0.46	113	-0.0007	0.001	-0.44	.66	.008	0.18	113	0.0006	0.002	0.30	.76	.05	0.46
	Negative affect	91	0.000	0.007	0.07	.94	.01	0.20	113	0.0003	0.001	0.25	.80	.007	0.17	113	-0.004	0.001	-0.27	.78	.04	0.41
Social stress recovery	IBI (msec)	93	0.66	0.83	0.80	.42	.07	0.55	115	0.04	0.14	0.27	.78	.01	0.20	115	-0.07	0.17	-0.40	.68	.04	0.41
	RSA (msec)	87	0.56	0.45	1.23	.22	.12	0.74	108	0.04	0.07	0.64	.52	.05	0.46	108	0.03	0.08	0.32	.75	.03	0.35
	PEP (msec)	83	0.01	0.10	0.12	.90	.01	0.20	103	0.007	0.02	0.38	.70	.03	0.35	103	0.01	0.02	0.68	.49	.07	0.55
	ns.SCRp (p/m)	77	0.08	0.07	1.09	.28	.11	0.70	93	0.0004	0.01	0.04	.97	.04	0.41	93	0.02	0.01	1.46	.15	.16	0.87
	ns.SCRw (p/m)	85	0.003	0.04	0.08	.93	.00	0.00	104	0.002	0.006	0.28	.78	.006	0.15	104	0.003	0.008	0.38	.70	.05	0.46

Note: R^2 reflects partial r squared of the fitness measure on the outcome from the model with age and gender included as covariates.
*Significant after multiple testing correction ($p < 0.0016$) in bold; # significant at nominal $p < 0.05$.

Table 8. Linear regression predicting daily life stress from fitness

		VO_2max							MVPA							Exercise						
		N	b	SE	t	*P*	R^2	*d*	N	B	SE	t	*p*	R^2	*d*	N	b	SE	t	*p*	R^2	*d*
Sleep	IBI (msec)	82	3.98	1.59	2.50	.014#	.08	0.59	106	0.65	0.24	2.77	.006#	.07	0.55	106	0.86	0.29	2.91	.004#	.07	0.55
	RSA (msec)	82	0.52	0.52	1.01	.31	.03	0.35	107	0.10	0.08	1.25	.21	.01	0.20	107	0.08	0.10	0.79	.43	.006	0.15
	PEP (msec)	79	-0.23	0.21	-1.07	.28	.10	0.66	103	-0.02	0.03	-0.67	.50	.004	0.13	103	-0.05	0.04	-1.29	.20	.01	0.20
	ns.SCRw (p/m)	91	0.09	0.02	3.44	**<.001***	.13	0.77	105	0.007	0.004	1.84	.069	.03	0.35	105	0.009	0.005	1.82	.071	.03	0.35
Awake	IBI (msec)	87	4.06	1.07	3.77	**<.001***	.14	0.80	112	0.56	0.17	3.28	**.0014**	.09	0.63	112	0.86	0.20	4.16	**<.001**	.14	0.80
	RSA (msec)	86	0.36	0.31	1.16	.25	.04	0.41	111	0.09	0.04	2.08	.039#	.04	0.41	111	0.13	0.06	2.23	.028#	.04	0.41
	PEP (msec)	85	0.18	0.21	0.86	.39	.003	0.11	110	-0.005	0.03	-0.16	.87	.00	0.00	110	-0.03	0.04	-0.70	.48	.004	0.13
	ns.SCRw (p/m)	86	0.02	0.02	0.99	.32	-.001	0.06	110	0.005	0.003	1.35	.18	.01	0.20	110	0.003	0.004	0.76	.45	.005	0.14
	Positive affect	83	0.007	0.01	0.83	.41	.000	0.00	104	0.003	0.001	2.53	.013#	.06	0.50	104	0.003	0.002	1.61	.11	.02	0.28
	Negative affect	71	-0.008	0.01	-0.93	.35	.01	0.20	104	-0.002	0.001	-1.43	.15	.02	0.28	104	-0.003	0.001	-1.63	.10	.02	0.28
Sleep recovery	IBI (msec)	82	-0.17	1.14	-0.15	.88	.02	0.28	106	-0.06	0.17	-0.38	.70	.001	0.06	106	-0.24	0.21	-1.18	.26	.01	0.20
	RSA (msec)	81	0.22	0.38	0.60	.55	.007	0.17	106	-0.008	0.06	-0.15	.88	.00	0.00	106	-0.06	0.07	-0.88	.38	.007	0.17
	PEP (msec)	79	-0.19	0.24	-0.80	.42	.10	0.66	103	-0.007	0.04	-0.19	.84	.00	0.00	103	-0.005	0.05	-0.10	.92	.00	0.00
	ns.SCRw (p/m)	81	0.08	0.03	2.67	.009#	.08	0.59	105	0.003	0.005	0.66	.51	.004	0.13	105	0.006	0.006	0.99	.32	.009	0.19
Work-leisure reactivity	IBI (msec)	27	-2.45	1.87	-1.31	.20	.08	0.59	36	-0.22	0.29	-0.73	.47	.01	0.20	36	-0.59	0.34	-1.73	.092	.08	0.59
	RSA (msec)	27	-1.40	0.43	-3.28	.003#	.33	1.40	36	-0.09	0.07	-1.18	.24	.04	0.41	36	-0.06	0.09	-0.75	.46	.01	0.20
	PEP (msec)	27	0.44	0.28	1.56	.13	.14	0.80	36	0.008	0.04	0.18	.85	.001	0.06	36	0.008	0.05	0.16	.87	.001	0.06
	ns.SCRw (p/m)	27	0.03	0.04	0.73	.47	.04	0.41	35	-0.01	0.008	-1.87	.071	.10	0.66	35	-0.01	0.01	-1.37	.18	.05	0.46
	Positive affect	25	-0.01	0.02	-0.62	.54	-.06	0.50	33	-0.001	0.003	-0.51	.61	.008	0.18	33	-0.002	0.004	-10.46	.65	.007	0.17
	Negative affect	25	-0.01	0.01	-0.91	.37	-.07	0.55	33	-0.002	0.001	-1.29	.20	.05	0.46	33	-0.002	0.002	-0.84	.40	.02	0.28

Table 8. Linear regression predicting daily life stress from fitness (continued)

		VO_2max							MVPA							Exercise						
		N	b	SE	t	*P*	R^2	*d*	N	B	SE	t	*p*	R^2	*d*	N	b	SE	t	*p*	R^2	*d*
Dislike-like reactivity	IBI (msec)	77	-1.04	1.05	-0.99	.32	-.006	0.15	97	0.14	0.16	0.87	.38	.008	0.18	97	-0.03	0.21	-0.13	.89	.00	0.00
	RSA (msec)	76	-0.29	0.30	-0.96	.34	.02	0.28	96	0.03	0.04	0.78	.43	.006	0.15	96	0.0003	0.02	0.00	.99	.00	0.00
	PEP (msec)	76	0.12	0.14	0.85	.39	-.001	0.06	95	0.03	0.02	1.72	.089	.03	0.35	95	0.006	0.03	0.24	.81	.00	0.00
	ns.SCRw (p/m)	75	0.02	0.02	0.85	.39	.01	0.20	93	-0.005	0.003	-1.39	.16	.02	0.28	93	-0.009	0.004	-2.00	.048	.04	0.41
	Positive affect	75	0.01	0.008	1.40	.16	.03	0.35	92	0.002	0.001	1.46	.15	.02	0.28	92	0.003	0.001	1.74	.084	.03	0.35
	Negative affect	75	-0.02	0.006	-3.29	**.0015***	.15	0.84	92	-0.002	0.001	-2.97	.0038#	.09	0.63	92	-0.001	0.001	-1.33	.18	.02	0.28

Note: R^2 reflects partial r squared of the fitness measure on the outcome from the model with age and gender included as covariates
*Significant after multiple testing correction ($p < 0.0016$) in bold; # significant at nominal $p < 0.05$.

DISCUSSION

To date no consensus has been reached on the validity of the cross-stressor adaptation hypothesis, positing that adaptation to a physical stressor in response to repeated exposure (training) also reduces reactivity to psychological types of stressors. Reviews and meta-analyses on this topic arrived at different conclusions but all unanimously point to the large heterogeneity in study design, plaguing the extant literature (Huang et al., 2013; Mücke et al., 2018, Jackson & Dishman, 2006; Forcier et al., 2006). In the current study we aimed to address these issues by 1) defining fitness as both aerobic fitness and physical activity, 2) including ANS branch-specific, i.e. SNS and PNS, reactivity as the outcome measures, 3) including both laboratory and daily life data, 4) using mental and social stressors, and 5) including stress reactivity and recovery measures. The overarching finding is that the cross-stressor adaptation hypothesis was not supported by the data.

Our separate use of aerobic fitness and physical activity finds justification in the low to moderate correlation between these two exposure variables, which repeats previous findings (Minder et al., 2014; Emaus et al., 2010; Morrow & Freedson, 1994; Siconolfi et al., 1985; Aadahl et al., 2007). Little gain was achieved, however, by separating moderate-to-vigorous activity from activities specifically related to voluntary leisure time exercise behavior. These variables showed a high correlation, likely due to the large overlap in reported activities for these variables. However, based on our results a measure containing all moderate to vigorous physical activity engaged in is favored, as opposed to a measure only including sports and exercise activities. This supports the conclusion of the review of Jackson and Dishman (2006), because it yielded more often a significant relation with stress reactivity or recovery. Therefore the discussion will focus on the results pertaining to aerobic fitness and MVPA.

Past results from meta-analyses regarding cardiovascular stress reactivity using aerobic fitness as their fitness measure most consistently reported on faster HR recovery in more fit subjects, paired to a higher HR reactivity (Jackson & Dishman, 2006; Mücke et al., 2018). Studies using regular exercise behavior also showed a faster HR recovery in more fit subjects, but identified a lower HR reactivity (Mücke et al., 2018). When fitness was defined by resting HR, Forcier and colleagues (2006) also found that more fit subjects showed a lower HR reactivity, but no recovery effect was observed. Our experimental stress paradigms evoked ANS reactivity comparable in direction and effect size to previous studies (Brindle et al., 2014), but our laboratory data do not support an effect of aerobic fitness or MVPA on this observed cardiovascular stress reactivity or recovery even after accounting for the main potential moderating factors. In addition, the current

study also did not replicate the findings by Mücke et al. (2018) with regard to a relationship between regular exercise and affective laboratory stress reactivity. More generally, our data conform well to the overarching conclusion from extant meta-analyses, namely that an impact of fitness on stress reactivity and recovery is either absent or too small to survive the plethora of moderators and confounders of stress reactivity.

Whereas the laboratory stressors successfully induced changes in ANS and affect, our daily life analyses showed no significant differences in ANS activity or affect during work compared to leisure time, although there was a trend towards higher HR and higher negative affect at work. This is a limitation that suggest too little stress may have occurred in the daily life part of the study. Previous studies using a daily life design showed lower HRV (Jarczok et al., 2013; Vrijkotte et al., 2000), and higher HR and blood pressure at work compared to leisure time (Vrijkotte et al., 2000). These studies used (white collar) working populations, whereas we mostly used a student population. Though it is estimated that around half of the university student population report moderate levels of stress-related mental health issues (as reviewed by Regehr et al., 2013), and that psychological distress among university student is higher compared to the general population (Adlaf et al., 2001; Stewart-Brown et al., 2000) and their working peers (Cotton et al., 2002; Vaez et al., 2004), our current sample showed little signs of stress as indicated by their good mental health score, low stress experience, and low mood disturbance. Even so, large individual differences were detected that should have allowed a clinically relevant correlation with fitness or PA to surface. Also, they did report to engage in (strongly) disliked activities, but even comparing liked versus disliked activities, no main effect was observed. The only association identified in the current sample is a large effect of aerobic fitness on RSA work reactivity at the $p < .05$ level, in which fitter individuals showed lower RSA work reactivity.

When comparing wakefulness to sleep, we did observe significant increases in IBI, RSA and PEP during restorative sleep, as did others (Burgess et al., 1997; Gonzales et al., 2020; Gregoire et al., 1006; Stein & Pu, 2012; Zoccoli & Amici, 2020). However, using the awake-sleep contrast as our index of ANS recovery in daily life did not show any effects of MVPA or aerobic fitness pointing to cross-stress adaptation. We also generally did not observe an overall relationship between affect and MVPA or aerobic fitness. Sole exception was a negative relationship between aerobic fitness and negative affect dislike reactivity, which was extended by a trend for MVPA. This suggests that fitter or more active individuals had a lower negative appraisal of disliked activities. This finding is in line with a recent study which also identified a positive effect of exercise on the subjective experience of daily life stress (Chovanec & Gröpel, 2020). Surprisingly, our results show a decreased ns.SCR on the wrist in fit individuals during sleep, which should be validated by future research.

The absence of a relationship between aerobic fitness or MVPA and ANS or affective reactivity would suggest little clinical relevance for the improvement of fitness to reduce stress reactivity. However, the absence of a cross-stressor advantage with regard to ANS stress reactivity and recovery does not negate the many clear advantageous effects of fitness. Both cross-sectional and longitudinal studies have shown that a higher resting HR is associated with an increased risk of coronary heart disease and all-cause mortality (Jensen, 2019) and that engagement in any type of exercise reduces resting HR with an average of 4.7% across different studies (Danieli, et al., 2014). In line with previous research, our study does confirm this significant bradycardic effect for both aerobic fitness (Emaus, et al., 2010; Gonzales, et al., 2020; Melanson, 2020) and MVPA (Emaus, et al., 2010; Gonzales, et al., 2020). Fitter and/or more active participants had a lower resting heart rate (HR), with aerobic fitness and MVPA explaining 14.4% and 10.2% of the variance in the controlled laboratory baseline measure, but only 2% and 0.8% of the variance in the average awake HR during the 24 h recording. This latter drops in explained variance is likely due to the inclusion of various activities in the daily life data, while in the laboratory participants were required to sit quietly. Our study also hints at a role for lower SNS activity and higher PNS activity contributing to these HR-lowering effects when adopting a lenient significance threshold. This is consistent with findings of von Haaren and colleagues. (2016), Schilling and colleagues (2020), and part of the studies included in the review of Tonello and colleagues (2014).

The current study was performed in a large enough sample size to find medium to large effects and covered all possible sources of heterogeneity posed by the meta-analyses. As expected, in the laboratory analyses ~30% of effect sizes were large and ~30% where medium. In daily life, however, the majority of the observed effect sizes were small, with only 20% being medium and ~5% large. By adjusting our p-value for multiple testing, our overall positive predictive value (probability that a finding reflects a true relationship; Button et al., 2013) was good. However, the chance to detect small sized relationships between fitness/physical activity and ANS or affective response to stress was low. Despite such small effects being scientifically interesting, one can question their relevance, in terms of allowing us to meaningfully advocate regular physical activity (or exercise sufficiently vigorous to increase fitness) as way to reduce the health impact of repeated cardiovascular stress reactivity. In establishing our sample sizes, we have therefore assumed at least a medium effect size, as this would be more relevant from a public health viewpoint.

The cross-stressor adaptation hypothesis is based on the observation that repeated exposure to exercise allows a person to perform a comparable physical load with a lower

activation of the SNS and a smaller deactivation of the PNS during exercise as well as a more rapid recovery to basal levels of SNS and PNS after exercise (McArdle et al., 2015). These adaptations occur in response to repeated exposure (training) for a large part by increasing the organ responsiveness, such that e.g. larger stroke volume requires less increase in HR to obtain the same cardiac output, and more dense muscle capillarization requires less vasoconstriction in non-muscle tissues and non-active muscles to ensure enough blood is distributed to the active muscles (McArdle et al., 2015). Also changes in exercise-induced feedback from the working muscle and the cardiorespiratory systems (e.g., baroreceptors) may contribute to altered ANS responding to exercise after training. When we experience psychological stress, however, the body only *prepares* itself for the anticipated need for exercise through the so-called fight-or-flight response (Schulkin, 2011; Stefano et al., 2008; Zandara et al., 2018), which is a feed forward mechanism, whereas only mild increase in muscle work is actually seen (Sothmann et al., 1996). Because the extent of the physical activity that the body is going to need to avert the stressor is unknown, the height of this anticipatory response is likely determined by the amount of threat level experienced but may also by a function of the maximal exercise performance capability. In that case, higher rather than lower anticipatory responses could be expected with increased fitness/MVPA. On the other hand, if preparation is always for some fixed amount of fight/flight, then training would reduce the ANS activation needed to attain this cardiovascular readiness state. In that case, lower anticipatory responses could be expected with increased fitness/MVPA. In both cases, by just altering the feed-forward signal and not also using the improved organ responsiveness or changes in exercise induced feedback cues, the cross-stressor adaptation effect could be smaller than detectable by the standard approaches, including the one used here.

Of course, we cannot exclude the alternative explanation that our study had limitations that prevented detection of the cross-stressor adaptation effect. First, specific selection of participants was sub-optimal to detect a cross-stressor adaptation in this feed-forward component. Our population consists of young, and relatively active and fit participants. Over 60% reported to regularly engage in leisure time exercise, with those who did not engage in exercise reporting engagement in at least 105 minutes of non-exercise related MVPA. It might be that the effect of the cross-stressor adaptation hypothesis was obscured by this relatively high physical activity level and can only be observed in a population including true inactive participants. Second, the study is limited by the low levels of experienced stress during the daily life part of the study. It could be that this stress was missed due to the explicit recruitment of healthy participants or due to freedom of participants to choose on which day they took part in the study. It is likely that

participant picked a day in a relatively stress-free week of their lives. A third limitation of the current study is the cross-sectional design, limiting it from shedding light on the effectiveness of exercise intervention studies on stress reactivity, such as those from von Haaren and colleagues (2016) and Chovanec and Gröpel (2020). Last, the current study focused on the validity of the cross-stressor adaption hypothesis with regard to its effect on ANS stress reactivity only. We want to stress that the results of this study can, therefore, not be translated to the effect of physical activity and aerobic fitness on the response of the hypothalamic-pituitary-adrenal (HPA) axis. Future studies should collect data on both ANS and HPA reactivity in more diverse populations and over longer periods of time to increase the variance in experienced affect or select specifics moment of life during which participants know they are going to experience a stressor (i.e., an exam or work deadline).

Taken together our results support the resting HR reducing effect of fitness and exercise engagement both in the laboratory and daily life. It did not provide evidence for the cross-stressor adaptation hypothesis at a multiple testing significance level but gave some indications of a lower basal SNS activity and attenuated SNS reactivity in more fit participants and higher basal PNS and attenuated PNS withdrawal during work. Our study validated the importance to take into account the amount of overall MVPA, rather than only leisure time exercise. More specifically, while aerobic fitness was only associated with reduced SNS activity, MVPA tended to also show associations with increased PNS activity, stressing even further that different measures of fitness should not be used interchangeably.

APPENDIX 1. CORRELATIONS

The results of all correlations are given in supplementary table 1 (laboratory) and 2 (daily life). VO_2max showed a large positive correlation with MVPA and a medium positive correlation with exercise. As expected, MVPA showed a large positive correlation with exercise. VO_2max, MVPA and exercise all showed a medium to large positive correlation with IBI during rest and absolute IBI during stress exposure in the laboratory. Of the fitness measures only, exercise showed a significant positive correlation with subjective physical health of medium size. No relationships between fitness and subjective mental health or experienced stress was identified.

During daily life, all three fitness measures showed large positive correlation with IBI when awake, irrespective of whether activities were liked or disliked. Interestingly only MVPA and Exercise showed large positive correlations with IBI during either leisure or work, while only VO_2max and Exercise showed a medium positive correlation with IBI during sleep. PEP during sleep and ns.SCR on the wrist during liked activities had a medium positive correlation with SBP and PEP sleep recovery a medium negative correlation with SBP. Negative affect reactivity to disliked versus liked activities showed a medium positive correlation with aerobic fitness.

As expected, males had a higher VO_2max compared to females. Furthermore, males had a higher SBP, higher IBI when sleeping, lower IBI sleep recovery and higher ns.SCR on the wrist during liked activities compared to females. No relationships were observed regarding age.

Supplementary Table 1. Laboratory Stressor Correlation matrix

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
Age	-	-0.01	0.03	-0.04	0.10	0.06	-0.01	0.07	0.12	0.01	0.14
Sex		-	0.21*	0.04	-0.17	-0.13	-0.45*•	-0.11	-0.18	-0.42 *•	0.09
Stress work			-	0.36*•	-0.10	-0.06	-0.02	-0.09	-0.14	-0.15	-0.03
Stress home				-	-0.25**	-0.13	-0.09	-0.02	-0.15	-0.17	-0.08
Mental health					-	0.43*•	0.04	0.13	0.23*	0.13	-0.12
Physical health						-	0.24*	0.26***	0.33*•	-0.06	-0.31*•
VO_2max							-	0.48*•	0.39*•	0.02	-0.30***
MVPA								-	0.79*•	0.03	-0.19*
Exercise									-	0.12	-0.18
SBP										-	0.58*•
Baseline Sitting											
IBI	0.12	-0.16	0.07	-0.09	0.22*	0.25**	**0.41*•**	**0.34*•**	**0.43*•**	0.03	**-0.38*•**
RSA0	-0.05	-0.05	0.03	-0.06	0.26***	0.16	0.13	0.22*	0.20*	-0.07	**-0.38*•**
PEP	0.14	0.04	0.13	0.02	0.11	0.15	0.22*	0.14	0.17	-0.24*	-0.24*
SCR palm	-0.11	-0.02	-0.06	-0.13	0.12	-0.17	-0.14	-0.23*	-0.19*	0.10	0.11
SCR wrist	0.14	**-0.30*•**	0.02	0.11	0.14	-0.03	0.21*	-0.05	0.02	0.19	0.12
Positive affect	0.00	-0.05	0.00	-0.25**	0.25**	0.27***	0.09	0.16	0.13	-0.10	-0.23*
Negative affect	0.02	0.09	0.02	0.05	-0.25**	-0.15	-0.05	-0.17	-0.14	0.02	0.26***
Mental Stressor											
IBI	0.00	-0.16	0.02	0.06	0.21*	0.18*	**0.37*•**	**0.31*•**	**0.37*•**	-0.02	**-0.35*•**
RSA0	-0.14	-0.12	0.07	-0.04	0.17	-0.17	0.01	0.04	0.10	0.02	-0.19*

Supplementary Table 1. Laboratory Stressor Correlation matrix (continued)

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
PEP	0.13	0.10	-0.02	-0.01	0.01	0.06	0.06	0.13	0.10	-0.24*	-0.15
SCR palm	-0.18	0.00	0.02	0.03	0.10	0.03	-0.04	0.01	-0.06	0.10	-0.01
SCR wrist	0.13	-0.20*	-0.11	-0.07	-0.04	-0.04	0.01	0.01	0.03	0.10	0.13
Positive affect	0.03	-0.12	-0.01	-0.25**	0.15	0.22*	0.16	0.27***	0.28***	0.00	-0.11
Negative affect	0.03	0.04	0.07	0.11	-0.15	-0.12	-0.06	-0.18	-0.21*	-0.06	0.17
Mental Stress Reactivity											
IBI	-0.16	0.01	-0.07	0.21*	-0.03	-0.10	-0.12	-0.07	-0.12	-0.06	0.07
RSA0	-0.05	-0.07	-0.02	-0.01	-0.14	-0.27***	-0.08	-0.16	-0.13	0.11	0.27***
PEP	-0.03	0.11	-0.24*	-0.09	-0.15	-0.13	-0.28**	-0.04	-0.11	0.01	0.14
SCR palm	0.06	0.02	0.04	0.14	0.02	0.11	0.13	0.29**	0.23*	-0.08	-0.15
SCR wrist	0.08	-0.12	-0.11	-0.07	-0.19	-0.05	-0.04	0.06	0.08	0.03	0.12
Positive affect	0.03	-0.07	-0.02	-0.0	-0.10	-0.05	0.08	0.13	0.16	0.10	0.12
Negative affect	0.02	-0.05	0.09	0.11	0.08	-0.01	-0.02	-0.07	-0.15	-0.12	-0.06
Mental Stress Recovery											
IBI	0.14	0.12	0.11	-0.19*	0.11	0.11	0.06	0.06	0.06	-0.01	-0.01
RSA0	0.04	0.29***	0.08	0.09	0.10	0.19*	-0.06	0.03	-0.04	-0.17	-0.05
PEP	-0.05	-0.03	0.09	-0.05	0.15	0.11	0.11	0.05	0.07	-0.05	-0.16
SCR palm	0.09	-0.03	-0.00	-0.10	0.02	0.08	-0.07	-0.16	0.01	-0.00	0.06
SCR wrist	-0.11	-0.07	0.07	0.16	0.16	0.12	0.02	0.03	0.07	0.03	-0.09
Social Stressor											
IBI	0.09	-0.15	-0.01	-0.02	0.25**	0.19*	**0.45*•**	**0.37*•**	**0.43*•**	-0.03	**-0.35*•**

Supplementary Table 1. Laboratory Stressor Correlation matrix (continued)

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
RSA0	-0.01	-0.12	0.05	-0.17	0.27**	0.00	0.16	0.06	0.09	0.00	-0.22*
PEP	0.16	0.12	0.01	-0.08	0.06	0.07	0.17	0.22*	0.21*	-0.23*	-0.16
SCR palm	-0.14	-0.14	-0.13	0.02	0.05	0.06	0.06	0.02	-0.09	0.01	-0.21*
SCR wrist	0.10	-0.28***	0.07	0.15	0.12	0.06	-0.05	-0.04	-0.06	0.09	0.04
Positive affect	0.10	-0.16	-0.11	-0.29***	0.10	0.15	0.08	0.15	0.20*	0.12	0.02
Negative affect	-0.05	0.07	0.13	0.20*	-0.21*	-0.11	-0.04	-0.13	-0.14	-0.07	0.14
Social Stress Reactivity											
IBI	-0.06	0.06	-0.11	0.13	0.00	-0.13	-0.04	-0.04	-0.09	-0.10	0.12
RSA0	0.01	0.01	0.04	-0.05	-0.05	-0.20*	-0.03	-0.15	-0.08	0.03	0.21*
PEP	0.01	0.06	-0.13	0.01	0.02	-0.11	-0.12	0.11	0.05	0.00	0.08
SCR palm	0.05	-0.09	-0.01	0.15	0.00	0.20	0.20	0.26*	0.15	-0.11	-0.31***
SCR wrist	0.08	-0.19	0.03	0.10	0.07	0.14	-0.14	0.04	0.00	-0.06	-0.05
Positive affect	0.11	-0.10	-0.10	-0.03	-0.13	-0.13	0.01	-0.02	0.06	0.21*	0.25**
Negative affect	-0.11	-0.02	0.16	0.22*	0.04	0.03	0.00	0.02	-0.03	-0.12	-0.15
Social Stress Recovery											
IBI	0.15	0.05	-0.01	-0.25**	0.09	0.06	0.08	0.03	-0.03	-0.03	0.05
RSA0	0.02	0.26**	0.05	-0.07	0.08	0.10	-0.00	0.03	-0.02	-0.14	0.03
PEP	-0.09	-0.09	0.07	-0.07	0.03	0.13	0.01	0.04	0.07	0.08	-0.06
SCR palm	0.17	-0.08	0.06	-0.12	0.04	-0.01	0.09	0.04	0.19	0.04	0.01
SCR wrist	-0.11	-0.10	0.12	0.26**	0.19	0.06	0.04	0.03	0.04	0.04	-0.12

*$p < .05$; **$p < .01$; ***$p < .005$; ***** $p < .0014$ (in bold)

Supplementary Table 2. Daily life stress correlation matrix

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
Age	-	0.02	0.01	-0.03	0.08	0.07	-0.06	0.02	0.08	0.00	0.17
Sex		-	0.19*	0.06	-0.17	-0.11	**-0.40*+**	-0.07	-0.15	**-0.43*+**	0.08
Stress work			-	**0.34*+**	-0.09	-0.03	-0.01	-0.11	-0.11	-0.01	0.05
Stress home				-	-0.25**	-0.09	-0.11	-0.03	-0.13	-0.10	0.02
Mental health					-	**0.44*+**	0.08	0.15	0.24*	0.13	-0.13
Physical health						-	0.22*	0.26**	**0.33*+**	-0.05	**-0.30*+**
VO_2max							-	**0.47*+**	**0.36*+**	0.04	-0.38**
MVPA								-	**0.79*+**	0.08	-0.14
Exercise									-	0.15	-0.15
SBP										-	**0.57*+**
Sleep											
IBI	0.04	**-0.33*+**	0.05	-0.07	0.27***	0.16	**0.36*+**	0.27**	**0.30*+**	0.21*	-0.15
RSA0	-0.11	0.08	-0.01	0.11	0.16	0.09	0.08	0.12	0.06	-0.08	-0.09
PEP	-0.10	-0.21*	-0.05	-0.07	0.11	0.15	0.03	-0.05	-0.10	**0.31*+**	0.15
SCR wrist	-0.11	-0.03	0.06	-0.05	-0.01	-0.07	**0.36*+**	0.18	0.17	0.19	0.00
Awake											
IBI	0.04	0.37	0.04	-0.08	0.24**	0.23*	**0.38*+**	**0.31*+**	**0.38*+**	0.06	-0.29***
RSA0	-0.15	0.04	0.08	0.02	0.20*	0.13	0.09	0.19*	0.19*	-0.02	-0.22*
PEP	0.14	0.08	-0.05	-0.02	0.04	0.12	0.05	-0.02	-0.07	-0.08	-0.08

Supplementary Table 2. Daily life stress correlation matrix (continued)

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
SCR wrist	0.23*	-0.31***	-0.14	0.03	0.19	0.00	0.18	0.14	0.13	0.28***	0.20*
Positive affect	0.06	-0.09	-0.01	-0.18	0.22*	0.15	0.08	0.24*	0.17	0.15	0.06
Negative affect	-0.01	0.03	0.12	0.25*	-0.28***	-0.10	-0.07	-0.14	-0.16	-0.04	0.05
Sleep recovery											
IBI	-0.10	**-0.43*•**	0.04	-0.00	0.15	0.01	0.13	-0.01	-0.05	0.26**	0.09
RSA0	-0.05	0.06	-0.08	0.13	0.06	-0.00	0.05	-0.02	-0.10	-0.05	0.09
PEP	-0.18	-0.31***	-0.00	-0.03	0.07	0.06	0.09	-0.00	0.02	**0.34*•**	0.17
SCR wrist	-0.26**	0.19	0.16	-0.07	-0.13	-0.03	0.23*	0.06	0.06	-0.07	-0.16
Leisure											
IBI	0.14	-0.04	0.01	-0.03	0.13	0.22*	0.29*	0.34**	0.39***	-0.02	-0.32**
RSA0	-0.14	0.20	0.12	0.10	0.05	0.12	-0.02	0.17	0.15	-0.17	-0.26*
PEP	0.10	0.16	-0.07	-0.12	0.01	0.15	0.08	0.07	-0.04	-0.05	-0.09
SCR wrist	0.26*	-0.29**	-0.12	-0.04	0.25*	0.00	0.20	0.18	0.17	0.24*	0.16
Positive affect	-0.02	-0.20	0.06	-0.13	0.21	0.13	0.10	0.19	0.12	0.10	-0.04
Negative affect	0.10	0.07	0.01	0.21	-0.28*	-0.06	-0.08	-0.11	-0.11	0.06	0.18
Work											
IBI	-0.03	0.11	0.15	0.11	0.21*	0.27*	0.33**	0.35**	0.37***	-0.08	-0.34**
RSA0	-0.20	0.08	0.02	0.19	0.29*	0.12	-0.13	0.20	0.26	-0.10	-0.23
PEP	0.25	0.14	0.03	0.04	-0.01	0.13	0.06	-0.03	-0.03	-0.24	-0.08
SCR wrist	0.19	-0.21	-0.17	0.07	0.10	-0.10	0.13	0.05	-0.03	0.28*	0.23
Positive affect	0.17	-0.14	-0.13	-0.41	0.26	0.10	-0.01	0.20	0.16	0.11	0.11

Supplementary Table 2. Daily life stress correlation matrix (continued)

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
Negative affect	-0.20	-0.08	0.25	0.34*	-0.21	-0.37**	-0.03	-0.26	-0.22	0.17	0.08
Work reactivity											
IBI	-0.04	0.21	0.28	0.28*	0.02	-0.16	-0.28	-0.12	-0.29	-0.22	-0.01
RSA0	0.00	0.02	0.07	0.09	0.07	-0.20	-0.50**	-0.20	-0.13	0.11	0.18
PEP	0.16	0.01	0.41*	0.31	0.00	-0.08	0.27	0.05	0.04	-0.09	-0.05
SCR wrist	0.14	0.02	-0.23	-0.06	0.03	-0.02	0.17	-0.30	-0.23	-0.23	-0.14
Positive affect	0.22	0.18	-0.04	-0.22	-0.28***	0.01	-0.12	-0.06	-0.07	-0.03	0.07
Negative affect	-0.31	-0.24	-0.06	0.28	0.24	-0.19	-0.18	-0.24	-0.15	0.11	-0.07
Like											
IBI	0.02	-0.11	0.05	-0.09	0.26**	0.19	**0.43*•**	**0.32*•**	**0.42*•**	0.04	-0.27**
RSA0	-0.24*	0.09	0.07	-0.02	0.18	0.11	0.12	0.14	0.13	-0.04	-0.20*
PEP	0.09	0.16	-0.04	0.00	0.03	0.07	-0.02	-0.12	-0.17	-0.09	-0.03
SCR wrist	0.20*	**-0.33*•**	-0.13	0.12	0.22*	0.00	0.07	0.17	0.19	**0.31*•**	0.23*
Positive affect	0.16	0.02	0.02	-0.08	0.17	0.18	-0.07	0.19	0.09	0.13	0.15
Negative affect	0.01	-0.03	0.15	0.14	-0.19	-0.10	0.10	0.07	-0.10	0.01	0.00
Dislike											
IBI	0.11	-0.07	0.05	0.01	0.14	0.22*	0.30**	**0.34*•**	**0.36*•**	0.07	-0.29**
RSA0	-0.02	-0.02	0.03	0.07	0.13	0.09	0.04	0.26**	0.23*	0.02	-0.20*
PEP	0.20*	0.08	-0.04	-0.02	0.02	0.02	0.02	0.06	-0.03	-0.09	-0.05
SCR wrist	0.19*	-0.26**	-0.13	-0.11	0.11	0-.07	0.15	0.02	-0.01	0.26**	0.21*
Positive affect	0.07	-0.17	-0.06	-0.18	0.24*	0.18	0.16	0.31***	0.27**	0.09	-0.04

Supplementary Table 2. Daily life stress correlation matrix (continued)

	Age	Sex	Stress work	Stress home	Mental health	Physical health	VO_2max	MVPA	Exercise	SBP	DBP
Negative affect	-0.06	0.04	0.10	0.26**	-0.26**	-0.20	-0.16	-0.24*	-0.22*	-0.04	0.10
Dislike reactivity											
IBI	0.15	-0.04	-0.09	0.11	-0.10	0.06	-0.10	0.10	0.00	0.01	-0.07
RSA0	0.26*	-0.19	-0.06	0.09	-0.10	-0.05	-0.04	0.09	0.04	0.06	0.03
PEP	0.13	-0.12	0.04	-0.05	-0.03	0.04	0.11	0.18	0.05	0.03	-0.05
SCR wrist	-0.04	-0.06	0.06	-0.06	-0.02	-0.04	0.12	-0.14	-0.20	0.07	0.03
Positive affect	-0.23*	-0.23*	-0.12	-0.16	0.08	0.04	0.26*	0.13	0.16	0.00	-0.24*
Negative affect	-0.14	0.16	0.01	0.19	-0.11	-0.10	**-0.38*+**	-0.30***	-0.16	-0.09	0.09

*$p < .05$; **$p < .01$; ***$p < .005$; **+ $p < .0016$ (in bold)

CHAPTER 6

Summary of my findings

A SUMMARY OF MY FINDINGS

In **chapter 2**, I validated a shortened version of the Sing-a-Song Stress (SSST) test, the $SSST_{short}$. The purpose of this test is to induce social-evaluative stress in participants through a simple and brief design that does not require the involvement of multiple confederates. The $SSST_{short}$ involves prompting participants to suddenly sing a song aloud while their video and audio signals are recorded. The study compared the participants' reactivity of the autonomic nervous system (ANS: measures IBI, PEP, RSA and ns.SCR) and affective responses (positive and negative affect) to the SSSTshort with that of a speeded reaction time task that required avoiding aversive loud tones, known as the tone avoidance task (TA). The results indicated that the $SSST_{short}$ was just as effective as the TA task in inducing ANS and affective reactivity. Moreover, the strength of the stress response across different ANS measures was similar for both types of stressors. Participants who showed a stronger response to the TA task also had a stronger response to the $SSST_{short}$. This makes the $SSST_{short}$ a cost-effective alternative to the well-known Trier-Social-Stress task (TSST), which can be easily incorporated into large-scale studies to expand the range of stress types that can be studied in laboratory designs.

In **chapter 3**, I validated a new wrist worn technology for measuring electrodermal activity (EDA). The laboratory protocol involved changes in posture, physical activity, mental load, and social-evaluative stress. As expected, the overall EDA levels measured on the wrist were lower than those measured on the palm, likely due to the lower density of sweat glands on the wrist. The analysis demonstrated that the frequency measure of non-specific skin conductance response (ns.SCR) was superior to the commonly used measure of skin conductance level (SCL) for both the palm and wrist. This was because the ns.SCR measure showed higher responsiveness, correspondence, and predictive validity, and lower sensitivity to thermoregulation. The wrist-based ns.SCR measure was sensitive to the experimental manipulations and showed similar correspondence to the pre-ejection period (PEP) as palm-based ns.SCR. Moreover, wrist-based ns.SCR demonstrated similar predictive validity for affective state as PEP. However, the predictive validity of both wrist-based ns.SCR and PEP was lower compared to palm-based ns.SCR. These findings suggest that wrist-based ns.SCR EDA parameter has a promising future for use in psychophysiological research.

In **Chapter 4** of my thesis, I conducted the first study to directly compare the relationship between affect and ANS activity in a laboratory setting to that in daily life. To elicit stress in the laboratory, four different stress paradigms were employed, while stressful events in daily life were left to chance. In both settings, a valence and arousal

scale was constructed from a nine-item affect questionnaire, and ANS activity was collected using the same devices. Data was collected from a single population, and the affect-ANS dynamics were analyzed using the same methodology for both laboratory and daily life settings. The results showed a remarkable similarity between the laboratory and daily life affect-ANS relationships. In general, the relationship of ANS with arousal was stronger in the laboratory, while the relationship of ANS with valence was stronger in daily life. Interestingly, the laboratory models demonstrated that the direction and strength of the relationship were similar for all individuals. If this finding is confirmed in further studies, it could provide insight into the observed individual differences in daily life. However, this study was limited by the dominance of a high arousal positive affect quadrant. Therefore, further research should be conducted on a sample showing more variance in affect to verify our findings.

In **Chapter 5** of my thesis, I investigated the influence of individual differences in physical activity and aerobic fitness on ANS and affective stress reactivity. Previous research has yielded inconsistent results due to heterogeneity issues in the population studied, stressor type, and the way fitness was measured. My study made a unique contribution to this field by measuring physical activity in three ways: 1) as objective aerobic fitness, 2) leisure time exercise behavior, and 3) total moderate-to-vigorous exercise (including both exercise and all other regular physical activity behaviors). In addition, we measured the physiological and affective stress response in both a laboratory and daily life setting. The total amount of physical activity showed more relationships with stress reactivity compared to exercise behavior alone, suggesting that future research should include a total physical activity variable. However, even in less conservative analyses, we found little evidence for a relationship between aerobic fitness and/or physical activity and stress reactivity. Thus, between-individual differences in ANS and affective responses to stressors were not explained by differences in aerobic fitness or physical activity. Our results did not support the cross-stressor adaptation hypotheses, suggesting that if exercise has a stress-reducing effect, it is unlikely to be mediated by altered ANS regulation due to repeated exposure to physical stress.

CHAPTER 7

Bringing psychophysiology into the 21st century

A discussion of my thesis

INTRODUCTION

The research I performed for my thesis revolved around the question how affect-physiology dynamics can be best measured in daily life. In my thesis I focused on three aspects of this question: 1) Do wearable wristband devices have sufficient validity to capture ANS activity? 2) To what extent is the laboratory design suitable to measure affect-ANS dynamics? 3) Are the affect-ANS dynamics subject to individual differences, both in the laboratory and in daily life? In this final part of my thesis, I discuss how my research has contributed to answering these research questions. I embed my findings in the work of other researchers to provide more complete and conclusive answers to the questions at hand. I end my thesis by providing suggestions on how to improve psychophysiological research and by discussing the impact this research field can have on society.

1) Do wearable wristband devices have sufficient validity to capture ANS activity?

Recent advancements in technology have led to the creation of wearable wristband devices capable of measuring physiological signals during daily life. These devices build upon the previous generation of ambulatory devices, which offered more mobility than traditional laboratory equipment while still using the same techniques to acquire physiological signals. Such devices, like the VU-AMS that was used in my thesis, are referred to as a "lab in a box". However, these devices still depend on the use of electrode stickers, making them unsuitable for long-term measurements of weeks or even months. Participants must frequently replace the electrodes, the electrode adhesive can lead to skin irritation when used for prolonged periods of time, and, specifically for EDA, the location of the electrodes on the hand palm or fingers restricts daily activities. This demand for extended wearability fueled the development of wrist worn devices that apply techniques for ANS measurement that do not rely on adhesive based electrodes. To this end alternatives such as wrist-based photoplethysmography (PPG) to assess PNS activity and dry-electrode EDA to assess SNS activity have been developed. While some devices containing these techniques are developed specifically for research (such as the Empathica), most are aimed at consumers. Researchers can still use these consumer devices, but data collection and cleaning often occur in a proverbial black box, which presents a disadvantage. In addition to the black-box disadvantage, the use of different techniques in the wearable wristband devices compared to the golden standards for measuring ANS raises questions about their accuracy. Since the first wristband devices hit the market, scholars have focused on validating them against the laboratory golden standard devices. Regarding PPG, a review on 18 studies by Georgiou and colleagues (2018) found good to excellent agreement between PPG and ECG during rest, but only moderate agreement during physical activity (8 studies).

In chapter 3 of my thesis, I conducted a thorough validation of a new EDA technology incorporated in a wristband. This study was the first to test the validity of this technology under highly diverse conditions. The overall validity of the wrist EDA was moderate for SCL, consistent with a recent review by Félix and colleagues (2022). However, the validity increased to moderate-large for the EDA ns.SCR measure. Additionally, the wrist ns.SCR measure had a similar predictive validity as the cardiac PEP, favoring this EDA measure over SCL. In chapter 5, I extended these findings by demonstrating that the stress reactivity effect size of wrist ns.SCR in response to mental and social stress induction in the laboratory was comparable to cardiac ANS measures IBI, RSA, and PEP. However, the effect size was smaller in daily life. Similar observations were made in chapter 4. The

confidence intervals around the wrist ns.SCR-affect relationships were larger than those for cardiac ANS measures, possibly due to the wristband technology being dependent on good contact between dry electrodes and skin. In daily life settings, the wristband may move, causing electrodes to lose contact with the skin and increasing measurement error.

The current agreement is that, compared to the laboratory golden standard, the validity of PPG and dry-electrode EDA is good, but decreases during non-stationary conditions. However, comparing validity across different studies is complicated due to the use of different experimental designs and analysis methods. To ensure accurate comparisons, universal validation guidelines, such as the one suggested by Mühlen and colleagues (2021), need to be adopted by the scientific community. Furthermore, a validation pipeline with a standardized set of experimental manipulations needs to be developed and universally applied. Fortunately, such guidelines and a validation pipeline are part of the plans of the recently granted Stress in Action consortium (see https://stress-in-action.nl/).

The answer to my first research question is, therefore, that wearable wristband devices are capable of capturing ANS activity, but do not (yet) reach the quality of their gold standard counterparts under dynamic circumstances. The choice of device should be based on a balance between feasibility and quality. The golden standard lab in a box device remains superior in terms of signal quality making it the preferred option for short laboratory studies. However, wristband technology is the more feasible option for long-term studies lasting from weeks to years. Indeed, during the daily life portion of my thesis study, wristband technology was already shown to be a more viable option than palm- based electrodes for 24-hour EDA measurements. Only half of the participants opted to wear the palm-based electrodes, and ten percent of those who did wear them removed them during the measurement due to interference with their daily activities. By comparison, only one participant removed the wristband due to skin irritation. This sparse number of participants who completed the daily life section with the palm-based electrodes meant that only the wrist-based EDA signal could be used in the daily life analyses of chapters 4 and 5. This highlights the need for high-quality and reliable wristband devices. Continuous improvement in sensor technology will play an important role in achieving better correspondence of such devices with gold-standard counterparts. Furthermore, with advancements in techniques for data cleaning and feature detection (Hossain, 2022; Orphanidou, 2018; Ihsan, 2022; Lan, 2023), data quality will increase. Taking these factors into account, I believe wearable devices will play a significant role in the future study of ANS activity in daily life.

2) To what extent is the laboratory design suitable to measure affect-physiology dynamics?

The development of wearables provided researchers with the opportunity to study the affect-physiology dynamics directly in daily life. This desire was fueled by the concerns on the ecological validity of laboratory induced stress. Various studies comparing stress reactivity to lab stressors with that to daily life stress showed that it is rather difficult to reproduce laboratory findings in a daily life setting, as reviewed by Zandstra and Johnston in 2011. Zandstra and Johnston address two key points that influence this generalizability of laboratory findings to daily life. First, they emphasize that the type of stressor (mental, social-evaluative, etc.) is of major importance, since distinct types of stress involve different psychophysiological processes. In chapter 2, I addressed the need for a larger palette of stress tasks by validating a shortened version of the SSST. This $SSST_{short}$ induces social-evaluative stress in participants through a simple and brief design that does not require the involvement of multiple confederates, which makes it suitable for implementation in large scale studies. Zandstra and Johnston (2011) point out that ecological validity was highest when the laboratory and daily life stress were of the same nature and occurred under similar circumstances, for example when the TSST was compared to a presentation given in daily life. It is therefore of importance to improve laboratory-based stress testing by including a diversity of stressors (mental load, physical, social-evaluative, interpersonal interaction) and creating a situational and contextual setting that is similar to the setting in which these types of stress would occur during daily life. Existing stress paradigms, in short, could be improved by making the applied stressors more lifelike. To do so researchers are exploring the use of new techniques such as virtual reality. Performing experiments in a virtual environment brings the possibility to enhance the affective experience of social interactions with the use of dynamic stimuli while still being in a controlled environment that is similar for all individuals and can be easily replicated (Gorini et al., 2011; Diemer et al., 2015). Such environments provide interesting approaches to test the influence of different 'real life' factors on stress reactivity. For example, stress reactivity to the SSSTshort could be compared across conditions in which individuals are presented with a virtual jury or with (known) peers.

A further issue surrounding the difficulty to replicate past laboratory findings to daily life situations is the difference in the statistical analyses applied. With the development of more sophisticated tools such as multilevel models (also known as linear mixed-effect models or hierarchical linear regression) it has become possible to apply a regression approach to repeated measured designs while taking into account that repeated measures are clustered within individuals. Before this method was available,

ANS and affective responses to stressors would be aggregated across individuals, or the repeated measures within an individual would be aggregated into a single value for each individual. These aggregated ANS and affective responses would then be correlated to study their relationship. However, aggregation has the disadvantage that it leads to a loss of information on time-related changes. Relying on between-subject designs can further lead to ecological fallacy (attributing group characteristic to individuals) or even Simpson's paradox (finding a relationship on group level that cannot be replicated on individual level)(Pollet et al., 2015). Due to these limitations, it is difficult to compare the laboratory results from studies that use a between-subject design to results from daily life studies that apply multilevel modeling with an additional within-subject component. I addressed this issue in chapter 4 of my thesis by performing the first study that directly compared the affect-physiology relationship as measured in the laboratory and in daily life. For this study laboratory and daily life data were collected from a single population and the affect-ANS dynamics were analyzed using the same multilevel methodology across settings. The results of my study look surprisingly promising with regard to the suitability of the laboratory design to predict the affect-physiology relationships occurring in daily life; there was substantial overlap in the 95% CI of the estimated fixed effects of all affect-ANS relationships between the two settings. However, my findings were limited by the low variance in valence and arousal even in the laboratory. It is of major importance for future studies to capture more variance in affective states in both the laboratory and daily life. To increase the chance to capture variance in all four affect quadrants, researchers should focus on improving both measurement settings. The laboratory design needs to be improved by including a broader range of life-like stressors that induce changes across all four affective quadrants. Likewise, in daily life we need to capture more diversity across all four effect quadrants. We can do so by scheduling measurements on days of a known stressor (such as an important work presentation, a tense family dinner, or an exam) or by collecting data over longer periods of time to capture such events as they occur naturally. With such improvements, future psychophysiological research can more robustly quantify the lab–real life generalizability of affect-ANS dynamics.

3) Are the affect-ANS dynamics subject to individual differences, both in the laboratory and daily life?

That individuals differ in the overall strength of their stress response is known from many studies. The findings in my thesis do not contradict this existing body of literature. In all the chapters of my thesis I find substantial variance in ANS and affect levels in both the laboratory and daily life. According to the stress reactivity hypothesis these differences in

stress sensitivity represent a trait-like characteristic: some individuals can be classified as hypo-responders whether others are clear hyperresponders (Turner et al., 2020). In chapter 2, I find support for substantial physiological response stereotypy across distinct types of stressors. Individuals who respond with low physiological arousal to the social-evaluative stressor also respond with low arousal to the reaction time task. It is important to note that this extends to differences between individuals in the pattern of responding, for example within the autonomic space (Berntson, Cacioppo & Quigley, 1993). It is possible for an individual to respond to a stressor by increasing their SNS activity or by decreasing their PNS activity. Our data only shows that when an individual is a predominantly SNS or PNS responder they are so rather consistently across different types of stress.

Various sources and mechanisms for the trait-like individual differences in stress reactivity have been studied under the umbrella of the stress reactivity hypothesis. One such factor is physical fitness, which is believed to affect the autonomic nervous system (ANS) reactivity to psychological stress, as per the cross-stressor hypothesis. In chapter 5 of my thesis, I studied the existence of a cross-stressor effect of physical fitness on affective and ANS stress reactivity to laboratory and daily life stress. The results showed little evidence for a stress-alleviating effect of aerobic fitness or physical activity. This is in contradiction with intervention studies that have shown that physical activity interventions can decrease ANS stress reactivity and mood disorder symptoms (Hearing et al., 2016). If the effect of exercise on stress reactivity and mood is not mediated by altered ANS regulation it is likely mediated by other processes, such as increases in self-esteem. Exercise interventions have been shown to improve self-esteem across all ages (Ekeland et al., 2005; Opdenacker, Delecluse & Boen, 2009; Spence, McGannon & Poon, 2005), which in turn affects our appraisal of stressful events (Galanakis et al., 2016). Further support for this hypothesis comes from interventions studies showing that exercise and mindfulness reduce stress to a similar degree (Morton, Helminen & Felver, 2020; van der Zwan et al., 2015). Like exercise, mindfulness is also known to increase self-esteem (Randal, Pratt & Bucci, 2015).

The results reported in my thesis support the idea that individuals differ widely in their ANS reactivity but the source of these individual differences remains to be determined. They do not, however, necessarily translate to similar individual differences in the coherence between the ANS reactivity and affective response to stress. Let's take the hypothetical situation of the following response pattern of individual A and B to a series of stressors on a completely arbitrary physiological arousal and positive affect scale, both with a theoretical range of 1:10 units. Both individuals start out with the same baseline physiological arousal level and positive affect scores (physiology = 5; affect = 8) and show a

negative response to the stressors on both parameters (Table 1; stress exposure increases their arousal and decreases their positive feelings). Individual A is a hypo-reactor, this individual responds to the stressors with only a slight increase in arousal and a small decrease in positive affect compared to their baseline. Individual B, on the other hand, is a hyper-reactor. This individual responds to the stressors with a strong increase in arousal combined with a strong decrease in positive affect, relative to their baseline. When estimating the linear relationship between physiological arousal and positive affect for both individuals the estimated regression coefficients are highly similar (Figure 1; individual A: $\beta = -0.87$, SE = 0.18, $t = -4.85$, $p = .004$; individual B: $\beta = -0.76$, SE = 0.15, $t = 4.91$, $p = .004$).

The *laboratory* affect-ANS models in chapter 4 of my thesis show a pattern of results that is similar to this hypothetical situation. Allowing the size and direction of the coefficient of the affect-ANS relationship to be random (to be unique for each individual) did not improve the model fit. This indicates that for the majority of the individuals the overall direction and strength of the relationship is similar. As an example based on actual data, Figure 2 shows the relationship between valence and IBI in the laboratory. In this figure each line represents the estimated regression coefficient, with their 95% CI, for the relationship of valence and IBI of each individual participant. Negative relationships are depicted in blue and positive relationships in pink. The gray rectangle indicates the model estimated average slope (solid black line) and 95% CI (dotted gray lines). As can be seen the estimated regression coefficient for most of the individuals falls within the 95% confidence interval of the model's average estimated regression coefficient. Generally, in daily life the model fit <u>does</u> improve when each individual is allowed to have their own regression coefficient. This was also the case for several affect-ANS relationships for the daily life analyses in chapter 4. As an example, Figure 3 shows the same valence-IBI relationship but this time for the daily life setting. It can be clearly seen that now a lot more individuals deviate from the model's estimated average relationship. This is why the fit of this model improved when we allowed the valence-IBI regression coefficient to vary for each individual participant. The difference is most striking when we look at Figures 2 and 3 side to side. If we assume that the relationship between physiology and affect is indeed similar across individuals, like we observe in the laboratory, it raises an interesting question regarding the observed inter-individual differences in affect-ANS dynamics in daily life. What are the driving causes of these individual differences?

Table 1. Hypothetical stress reactivity of two individuals

	Individual A		Individual B	
Condition	**ANS**	**Affect**	**ANS**	**Affect**
Baseline	5	8	5	8
Stressor1	7	7	9	4
Recovery1	4	10	7	6
Stressor2	5	8	9	4
Recovery2	4	9	8	5
Stressor3	6	7	10	5
Recovery3	4	9	6	7

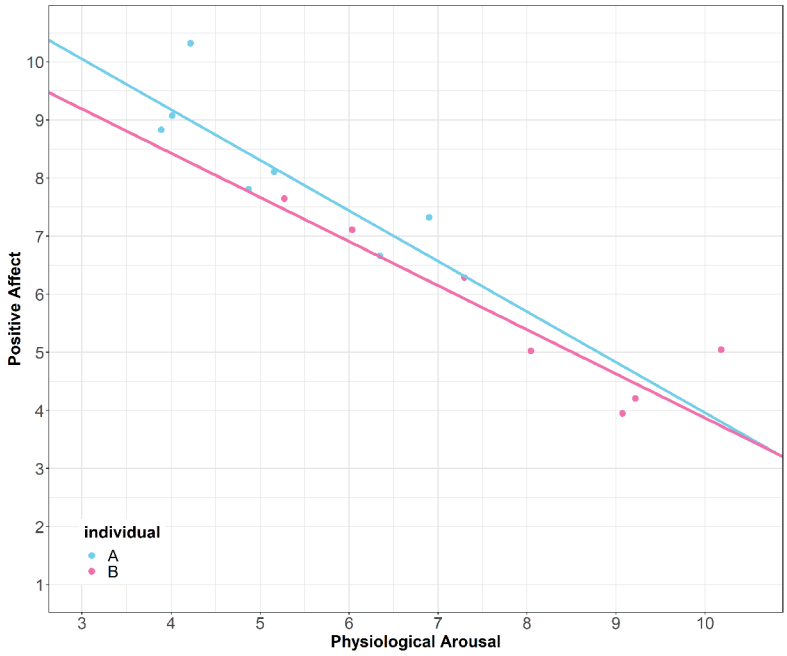


Figure 1. Hypothetical linear relationship between physiological arousal and positive affect for two individuals.

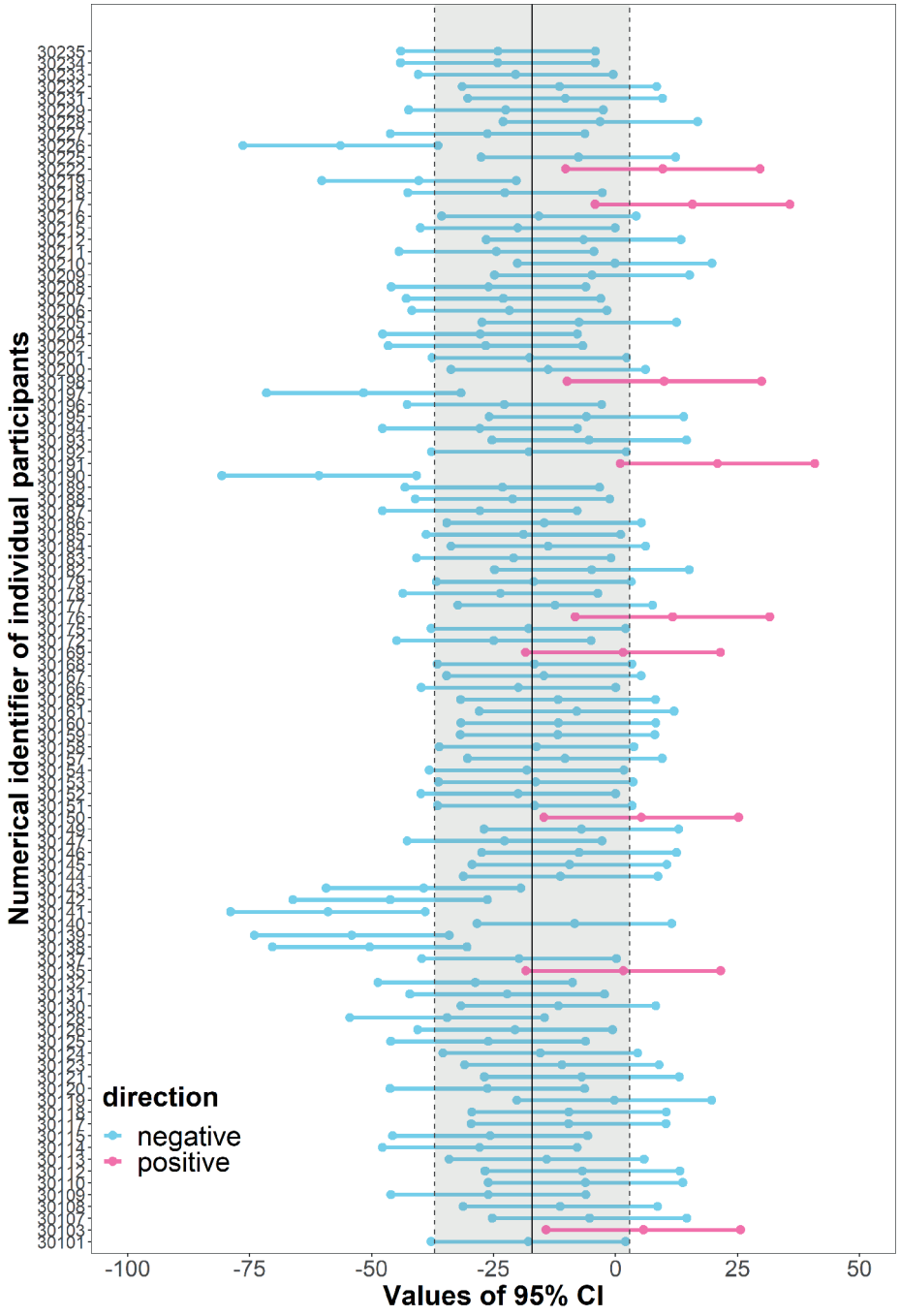


Figure 2. Valence-IBI in the laboratory

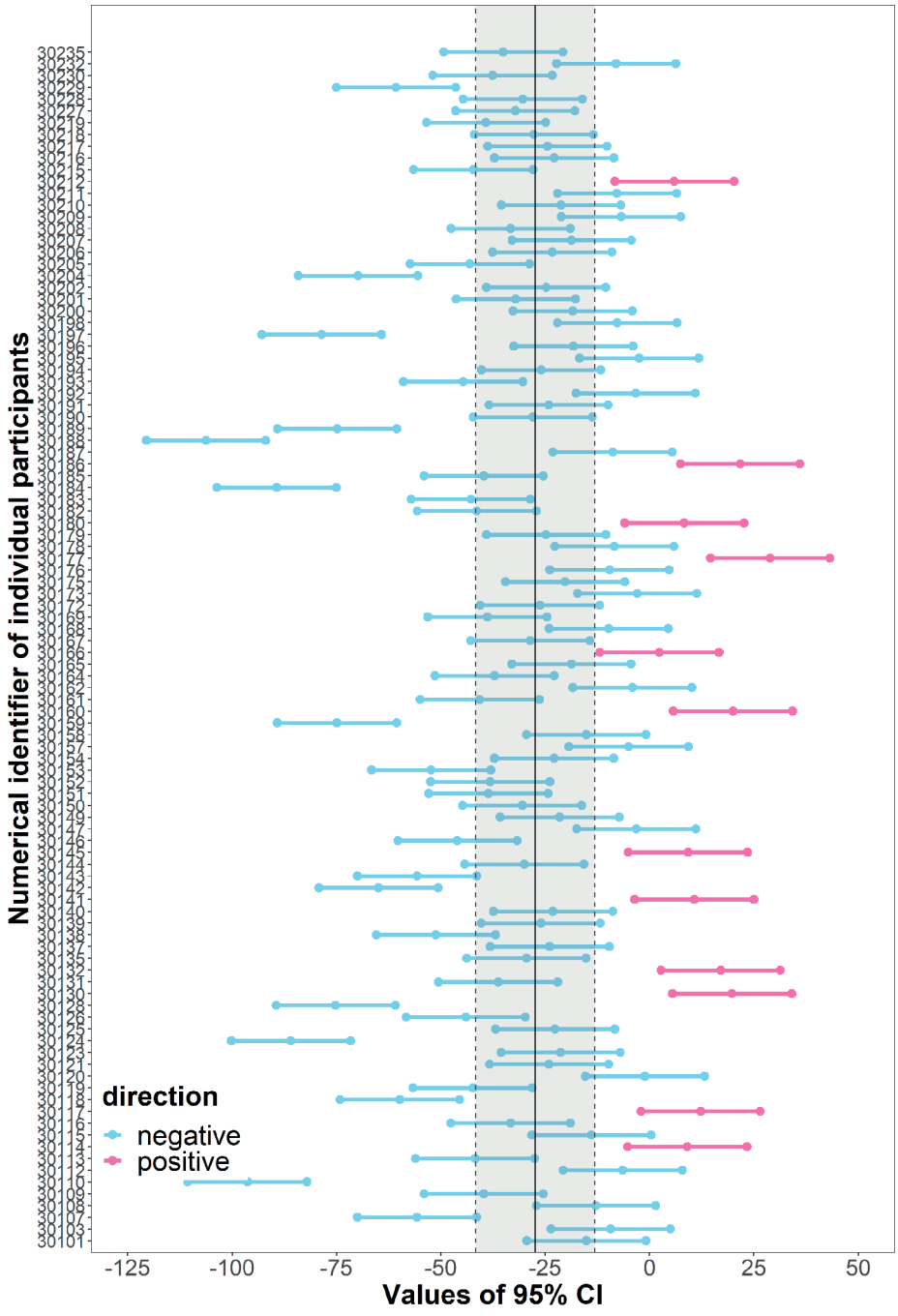


Figure 3. Valence-IBI in daily life

One especially important factor that could explain the individual differences in ANS-affect dynamics in daily life are the individual differences in the frequency and pattern of postural changes and physical activity during ambulatory assessment. Posture change and physical activity are important drivers of ANS activity. They require relatively strong ANS changes to meet the changed gravity effects and the increased metabolic needs. When predicting affect form the ANS these homeostasis-driven changes in ANS can lead to erroneous conclusions. In part this can be avoided by the common practice to only include periods in which the physical activity was low (Broscchot et al., 2007; Pieper et al., 2007; Pieper et al., 2010) or when subjects were in a sitting posture (Vrijkotte et al., 2000). This was also the approach I endorsed in chapter 4 and 5 of my thesis. Despite losing a substantial amount of data, we believe this approach is better than statistically correcting for physical activity and posture as confounders of ANS activity. Addition of physical activity as a confounder, for example, implies a linear relationship with ANS activity across all intensities of physical activity which is often not found (Micheal, Graham & Davies, 2017).

Sadly, the exclude-physical-active-segments approach is not without its flaws. Parts of the data are excluded from the analysis based on the ongoing level of physical activity. Immediately after the physical activity stops, data is allowed back into the analysis. It is well known, however, that it takes time for the ANS to return to resting state levels after physical activity, and that the speed of recovery varies across individuals, amongst others as a function of fitness level (Darr et al., 1988; Imai et al., 1994; Pierpont & Voth, 2004), the intensity of the exercise engaged in (Mann et al., 2014), and type of exercise engaged in (Maeder et al., 2009). Therefore, by simply excluding data that shows current physical activity, it is possible that a recovery process from recent physical activity is still ongoing. By not excluding all changes in ANS activity due to physical activity and ensuing recovery processes, differences in fitness level or the amount and/or type of activity engaged in on the day of the measurement can cause differences in the affect-ANS relationships between individuals. To address this issue, we require a measure that reflects changes in ANS activity above and beyond changes due to metabolic need. This surplus of ANS activity that extends the metabolic need is referred to as non-metabolic or additional ANS activity. Additional ANS activity (addANS) is believed to be a more accurate indicator of changes in cognitive and/or affective states as compared to uncorrected ANS activity.

Currently, there is no consensus in the scientific community on how such addANS measures should be computed. Pioneering work in this field has been limited to the ANS measure HR and HRV. The first line of work applies the addANS framework as a method to trigger EMA prompts. For example, if the HR would increase a predefined cut-off of three beats per minute, without a co-occurring change in physical activity, this would be

indicative of a period of additional HR and trigger an EMA prompt (Myrtek & Brügner, 1996; Myrtek, Aschenbrenner & Brügner, 2005; Ebner-Priemer et al., 2007; Prill & Fahrenberg, 2007, Hoemann et al., 2020). The second line of work aims to develop an algorithm that can remove the part of the HR(V) changes that are driven by changed metabolic demands. Regression models are applied to quantify the relationship of HR(V) with methods that are a proxy of metabolic need such as oxygen consumption (Carroll, Turner & Hellawell, 1986; Carroll, Phillips & Balanis, 2009; Turner, Carroll & Hanson, 1988; Wilhelm & Roth, 1998) or movement (Brown et al., 2020; Johnston, 1996; Linssen et al., 2022; Verkuil et al., 2016). When the HR(V) increase is above what is expected from the regression model it is then considered additional. However, extensive validation of these addANS measures is lacking. Thus far, no study has compared the predictive validity of the different methods to one another or even to non-adjusted ANS activity. It is important for future research to focus on the development and validation of optimal methods for the calculation of additional HR and other addANS parameters. If these additional ANS measures indeed reflect changes in demand due to changes in affective states, they should increase the explained variance in affect by ANS activity. This can aid the specificity and accuracy of emotion prediction algorithms.

To come back to the research question that was central in this section, the research in chapter 4 hints towards comparable affect-physiology dynamics in the lab and in daily life. However, the extent to which this can be replicated across different affect quadrants and/or daily life contexts needs to be verified. The first steps in unraveling the true impact of individual differences on the affect-physiology interplay lie in the improvement and validation of our study designs and methodologies.

The future of psychophysiological research

The research I performed for my thesis has important implications for the chosen methodology and measurements in psychophysiological research. While discussing the three research questions that formed the common thread of my thesis, various important aspects that influenced this decision were addressed. First, the current wrist-worn wearable devices that apply PPG or dry electrodes to measure ANS are sub-optimal in terms of accuracy. However, this disadvantage of PPG or dry electrodes is balanced by an increase in wearability. Second, different research questions require different approaches, and the choice of the right approach includes consideration of the differences between the laboratory and daily life designs.

A specific section of research in which a laboratory design has an important value is the development of wearables and wearable associated emotion prediction algorithms.

Due to the controlled environment the laboratory has to offer, wearable devices can be optimized to work the best they can under these conditions before being tested in the wild. Testing in such a setting has the advantage that the effect of multiple confounders on the accuracy and predictive validity of the devices can be specifically tested. The value of such a setup was shown by a study by Can and colleagues (2020), who aimed to develop a stress prediction algorithm based on the features extracted from the IBI and EDA signals obtained with the Empathica E4 wristband. They investigated the effect of different combinations of features and training environments on the accuracy of their algorithm. Using the laboratory as the training environment to predict stress in daily life provided the best accuracy. When the algorithm was trained on the laboratory data instead of the daily life data itself, accuracy increased with 2.8% to a total of 74.2%. These findings confirm the importance of confounders when predicting the relationship of the ANS with affect. The best results are obtained in a controlled environment, even when the model is to be applied to an uncontrolled daily life setting.

The daily life design, on the other hand, uniquely allows for the study of natural human behavior. For example, this design makes it possible to track the number of times an individual experiences periods of high SNS activity or negative affect states, the dominant physiological or affective profiles, or temporal relations in ANS and affect dynamics. This information can provide valuable insights into between-individual differences that may serve as risk factors, or can be used to monitor disease outcomes. However, the laboratory design will become more hybrid as we develop more wearable tools and experimental tasks that can be performed by individuals from their home environment. It is likely that future psychophysiological research will consist of a structured calibration session, including device attachment and self-application of experimental tasks by the individuals on a tablet or laptop in their home environment, followed by longer term daily life assessment. In such ways, the classical laboratory design in which researchers invite individuals to join their study in a laboratory facility will eventually become far less common. Last, when designing a daily life study, it is important to consider potential individual differences. More research is needed to unravel the key factors of individual differences in the affect-ANS dynamics in daily life. Numerous studies have already shown that in addition to physical activity, factors such as the social environment (Gerteis & Schwerdtfeger, 2016; Schwerdtfeger & Friedrich-Mai, 2009), perceived feelings of safety (Schwerdtfeger, Paul & Rominger, 2022), and differences in emotion regulation (Berna & Nandrino, 2014; Stifter, Dollar & Cipriano, 2011) affect ANS reactivity. The extent to which these factors influence the relationship between ANS and affect in daily life needs to be further explored.

But even if the laboratory design becomes a near perfect representation of daily life, wearables reach a high accuracy, and daily life confounders are identified, we are still not there. There is a final issue that needs to be addressed for the future of psychophysiological research that has not been part of the research I performed: the way we measure affect. The current most frequently used method to collect data on affect is to ask participants to rate a set number of affect items. In the light of the emotional granulation theory, an individual's ability to experience a diversity of emotions (Suvak et al, 2011; Barrett, 2013; Barrett, 2017), we might need to revise this methodology. According to the theory of emotional granulation, individuals with low emotional granulation are unable to differ between different emotional states, for example they cannot distinguish between rage and frustration or irritation and agitation. Individuals with high emotional granulation can distinguish between a vast number of emotions diverging in both relative valence and arousal. The current method of affect assessment likely provides a good fit for the first group of people with low emotional granulation but is ill-fitted for the latter. This is not only important for the study of emotional processes, but also influences the relationship of affective states and physiology. According to the theory of constructed emotion (TCE; Barrett, 2006, 2012, 2013, 2017a, 2017b), individuals with higher granularity will also have more diverse and precise patterns of ANS activity, while individuals with low granularity will only show a more generalized ANS response. The findings of a pioneering study by Hoemann and colleagues (2021) support this hypothesis. They conducted an EMA study in 50 participants lasting 14 days. Each day participants wore a monitor for eight hours that recorded their ANS activity using golden-standard techniques. Participants answered on average around nine diaries a day. During a diary prompt the participants were asked to report their emotions freely. At the end of the day, they received a questionnaire with a summary of their reported emotions and were asked to rate each entry based on 18 preset emotional adjectives on a 7-point Likert scale. Their analyses showed that individuals with higher emotional granularity had more diversity in their pattern of ANS activity. Furthermore, in these individuals the patterns of physiological activity were also distinctive from one another for different emotional adjectives.

If indeed the pattern of physiological activity becomes more distinct with increased diversity in reported emotion, it is important to obtain such information with more detail than the nine affect items included in my thesis. However, there is a trade-off between capturing highly detailed information and the acceptable level of subject burden. If participants feel that they must rate very many items, non-compliance will increase. Future research should explore the best method for collecting this type of data. A good starting point might be to not ask participants to rate individual items, but rather let

them select emotion adjectives from a list of several possible options, for example the 18 included by Hoemann and colleagues (2021). Such a list could be potentially sorted on valence and arousal and each selected adjective would be followed with a question regarding their intensity. Furthermore, to successfully take into account the complexity in the experience of emotions and their relationship with different SNS and PNS parameters, I believe future research should make use of non-linear data analytic methods such as individual network analyses or unsupervised machine learning. Both the affective and ANS parameters are not identities that stand alone, rather they influence themselves and each other over time. It has been shown in EMA studies that emotions experienced during the previous assessment influence the emotions experienced at the current assessment in diverse ways; emotions of the same valence augment one another, and emotion of opposing valence blunt one another (Vansteelandt, Van Mechelen, & Nezlek, 2005; Pe & Kuppens, 2012). Changes in ANS activity also happen gradually, the HR does not go from 80 to 130 and back to 80 in a single second. If such temporal patterns are not considered, they can complicate the connection of distinct affective states to distinct physiological patterns. It is plausible that the physiological activity associated with a current high arousal positive affect states might be different when such a state was preceded by high arousal negative affect than when it was preceded by low arousal negative affect. For example, the bodily response of an individual that is currently cheerful because they just finished a stressful meeting with their boss might be different from the bodily response of an individual that is currently cheerful because their boyfriend gave them a surprise gift.

An example of such a machine learning approach to physiological data is given by Hoemann and colleagues (2020). They processed their daily life ANS data for each individual with gaussian mixed models to identify distinct physiological clusters. This method showed to be very feasible. Across all participants a total of 219 were identified, with a mean of a little under five clusters per person. On average the probability of data points being members of their assigned clusters was high ($M = 0.87$; $SD = 0.06$), indicating that the clusters indeed reflect distinct physiological states. However, applying meaning to these clusters regarding affective states proved to be difficult. Both within and between individuals there was no clear association between self-reported affect and a specific cluster type. However, in this study affect was assessed by self-generated emotions. Individuals could freely describe their emotional state in an open-ended survey item. In this form of assessment, the relative strength of the emotion was not assessed. Furthermore, the separate emotions were not clustered into broader affect constructs or related to one another at the same time point. It could very well be the case that clusters marked by high SNS activity might be associated with high arousal negative affect. Nevertheless, with

their model Hoemann and colleagues provided an exceptionally good framework for future studies to further unravel the affect-ANS dynamics with the use of machine learning.

To take into account the temporal dependency of affective states and ANS activity their relationships could be studied with the use of network analyses (Bar-Kalifa & Sened, 2020; Jordan, Winer & Salem, 2020). From such networks features (density of the network) and centrality measures (strength, closeness, and betweenness) can be derived, which can be studied in relation to various constructs. For example, a study by Bringmann and colleagues (2016) showed that higher trait neuroticism is associated with a higher negative emotion network density and closeness centrality (Bringmann et al., 2016). Another approach is to study the relationship of time-varying constructs with emotions by adding them to the network itself. This approach was taken by Greene and colleagues (2020) to study the relationship between post-traumatic stress disorder symptom clusters and emotions during a period of conflict exposure. They found that the PTSD symptom clusters of arousal and negative alterations in cognition and mood (NACM) showed the strongest connections to negative emotions. The latter approach could be very well suited for continuously measured physiological signals.

It would be interesting to study affect-physiology dynamics by adding the ANS activity measures such as HR, HRV, PEP and EDA to these individualized models and see how they covary with emotions. An interesting expansion of these models is to study at which moment in time and how often the parameters of the affect-ANS network change (Masuda & Holme, 2019; Wilson, Stevens & Woodall, 2019). Not only do such models allow us to study differences in the relationship of ANS activity and affect at moments of different dominant affective or physiological states, but they could also be informative of changes in cognitive, mental, or social processes. For example, an individual can have a different network when experiencing negative emotions in a social setting as compared to being alone. However, this method comes with a severe drawback. The number of repeated measurements per individual that are needed to perform network analyses is estimated upward to 50 (Vrijen et al. 2018). To achieve this number of repeated measures with around 8 EMA's per day translates to a study length of at least 7 days. For studies of this length the current physiological wearables are not a feasible option. For such studies the development of high quality, non-intrusive (wristband) wearables are necessary.

To conclude, there are substantial opportunities to improve psychophysiological research on various domains. Accurate and comfortable wearable devices need to be developed, the way we measure affect needs to be optimized, new methods to take into account confounding effects of physical activity or environmental factors need to be developed, and we need to change the way we analyze our data. Fortunately, I do not

stand alone in these beliefs. I am joined by various scholars from different fields that are already putting these ideas into action. We are living in an exciting time in which innovative technology enables us to expand our knowledge on the connection between the body and the brain. However, knowledge alone is not the goal. We want to apply this knowledge to improve the well-being of society. And with the rapid improvements in the field of psychophysiology I believe we will be increasingly able to do so.

SOCIETAL IMPLICATIONS

I wholeheartedly believe there is a bright future for the role of hybrid laboratory and daily life psychophysiological research to improve both physical and mental health. I think we can learn a whole lot about human behavior by studying affect-ANS dynamics, specifically in those suffering from affective disorders. A particular field of interest herein is the development of emotion prediction algorithms. The idealistic aim of such algorithms is to predict (changes in) mood without the need of subjective reporting. I believe we are currently lacking fundamental knowledge on the affect-ANS dynamics to do so, but the future looks bright. The newest machine learning models show that even without human knowledge machines can derive features, clusters, and patterns from data. We can use such machine generated output to expand our current knowledge. However, there are also ethical and legal risks attached to the collection of such data, especially if we incorporate environmental and contextual factors obtained by passive sensing. Examples of such passive sensing are the use of GPS data to determine the travel pattern of an individual and the neighborhood quality and urbanicity of the environment travelled. WIFI and Bluetooth detection can be used to indicate how many people, and even which people, are in their social environment. Despite rules and regulation from the General Data Protection Regulation (GDPR) to ensure safety and anonymity of passive-sensing and EMA data, there is a real risk that sensitive personal information collected could be lost or misused, potentially compromising an individual's privacy and confidentiality. Just GPS data alone can give someone who aims to do wrong a lot of information on the lives of individuals such as where they live, work, exercise etc. (Iqbal & Lim, 2010). Furthermore, behavior in traffic inferred from GPS (such as speeding) and information from wearables regarding the ANS activity and physical activity patterns face similar ethical and societal issues surrounding the genetic information and testing in insurance and employment (Godard et al., 2003). Since a data leak can have far reaching and potentially catastrophic consequences for the individual and the society they are a part of, it needs to be determined whether the risks are worth the benefits.

Nevertheless, there is an exciting potential in the use of such data to improve the wellbeing of the human population. If we track the everyday life of an individual across a suitable calibration period with passive sensing by smartphone sensors and wearables, in parallel to self-report data by EMA, and feed this data to a machine learning model, such models could tell us which combination of environmental factors and physiological states is associated with which affective state. When such algorithms are trained well, we would ideally be able to predict (changes in) affective states without the need of self-report. The societal implications of such affect predicting algorithms are huge. Individuals with various conditions could use self-tracking apps to get insight into what triggers unwanted affective states, which could in turn help them and their practitioners treat symptoms better. To illustrate this let's look at the following hypothetical example of John. John is a middle-aged individual with anger management problems. From his data an algorithm could learn that when John's heart rate and number of EDA peaks are rising in the absence of gross body movement, while he is at a specific GPS location with a specific WIFI signal (indicating that he is at the house of his mother-in-law) this is indicative of a risky situation to experience an episode of anger. John's psychologist could use this information to let John's wearable send him a message to engage in certain breathing exercises they practiced together that help John to remain calm. This can tremendously improve the quality of life for John and individuals like him, while asking relatively little effort in return.

Making the tools that the digital age has to offer us work for us instead of against us is what I hope psychophysiological research can achieve in the 21st century.

APPENDICES

References

Publications

Acknowledgement

REFERENCES

Aadahl, M., Kjaer, M., Kristensen, J. H., Mollerup, B., & Jorgensen, T. (2007). Self-reported physical activity compared with maximal oxygen uptake in adults. *European Journal of Cardiovascular Prevention and Rehabilitation, 14*(3), 422-428.

Abdullah SMSA, Ameen SYA, Sadeeq MA, Zeebaree S. 2021. Multimodal emotion recognition using deep learning. *Journal of Applied Science Technology Trends. 2*(2):52–58.

Adlaf, E. M., Gliksman, L., Demers, A., & Newton-Taylor, B. (2001). The prevalence of elevated psychological distress among Canadian undergraduates: Findings from the 1998 Canadian Campus Survey. *Journal of American College Health, 50*(2), 67-72.

Ahola, K., & Hakanen, J. (2014). Burnout and health. In Burnout at work (pp. 10-31). *PsychologyPress.*

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., . . . Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine in Science Sports and Exercise, 43*(8), 1575-1581.

Bach, D. R., Friston, K. J., Dolan, R. J., 2010. Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology, 76*(1), 52-55.

Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PloS one*, 16(3), e0248152.

Bar-Kalifa, E., & Sened, H. (2020). Using network analysis for examining interpersonal emotion dynamics. *Multivariate behavioral research*, *55*(2), 211-230.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46.

Barrett, L. F. (2012). Emotions are real. *Emotion, 12*(3), 413–429.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review, 5*, 379–389.

Barrett, L. F. (2017a). How emotions are made: The secret life of the brain. Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. Social Cognitive and Affective Neuroscience, 12(1), 1–23.

Bassett Jr, D. R. (2000). Validity and reliability issues in objective monitoring of physical activity. *Research quarterly for exercise and sport, 71*(sup2), 30-36.

Beauchaine, T. P. (2015). Respiratory sinus arrhythmia: A transdiagnostic biomarker of emotion dysregulation and psychopathology. *Current opinion in psychology*, 3, 43-47.

Beauchaine, T. P., Bell, Z., Knapton, E., McDonough-Caplan, H., Shader, T., & Zisner, A. (2019). Respiratory sinus arrhythmia reactivity across empirically based structural dimensions of psychopathology: A meta-analysis. *Psychophysiology, 56*(5), e13329.

Benschop, R. J. and Schedlowski, M. (1999)'Chapter 15: Acute Physiological Stress', In M. Schedlowski & U. Tewes (ed.), Psychoneuroimmunology: An Interdisciplinary introduction, pp. 293-306. New York, USA: Springer Science+Business Media, LLC

Berna, G., Ott, L., & Nandrino, J. L. (2014). Effects of emotion regulation difficulties on the tonic and phasic cardiac autonomic response. *PloS one*, *9*(7), e102971.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology, 30*(2), 183-196.

Berntson, G. G., Cacioppo, J. T., Quigley, K. S., 1994a. Autonomic cardiac control. I. Estimation and validation from pharmacological blockades. *Psychophysiology, 31*(6), 572-585.

Berntson, G. G. et al., 1994b. Autonomic cardiac control. III. Psychological stress and cardiac response in autonomic space as revealed by pharmacological blockades. *Psychophysiology, 31*(6), 599-608

Blackwell, E., De Leon, C. F. M., Miller, G. E., 2006. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. *Psychosomatic Medicine, 68*(6), 870-878.

Bosch, J. A., De Geus, E. J., Carroll, D., Goedhart, A. D., Anane, L. A., van Zanten, J. J. V., ... and Edwards, K. M. (2009). A general enhancement of autonomic and cortisol responses during social evaluative threat. *Psychosomatic medicine, 71*(8), 877-885.

Boucsein, W., 2012. Electrodermal activity, 2nd edn., Springer Science & Business Media, New York.

Boucsein, W. et al., 2012. Publication recommendations for electrodermal measurements. Psychophysiology, *49*(8), 1017-1034.

Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*(2), 271–301.

Braithwaite, J. J., Watson, D. G., Jones, R., Rowe, M., 2013. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology, 49*(1), 1017-1034.

Brouwer, A.-M. and Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST), *Frontiers in Neuroscience, 8*:224.

Brindle, R. C., Ginty, A. T., Phillips, A. C., & Carroll, D. (2014). A tale of two mechanisms: A meta-analytic approach toward understanding the autonomic basis of cardiovascular reactivity to acute psychological stress. *Psychophysiology, 51*(10), 964-976.

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., ... & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425-435.

Buijs, R. M. (2013). The autonomic nervous system: a balancing act. *Handbook of clinical neurology*, 117, 1-11.

Button, K. S., Ioannidis, J., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience, 14*(5), 365-376.

Burgess, H. J., Trinder, J., Kim, Y., & Luke, D. (1997). Sleep and circadian influences on cardiac autonomic nervous system activity. *American Journal of Physiology-Heart and Circulatory Physiology, 273*(4), 1761-1768.

Byrne, N. M., Hills, A. P., Hunter, G. R., Weinsier, R. L., & Schutz, Y. (2005). Metabolic equivalent: one size does not fit all. *Journal of Applied physiology, 99*(3), 1112-1119.

Cacioppo, J. T., Klein, D. J., Berntson, G. G., Hatfield, E., 1993. The psychophysiology of emotion. In: M. Lewis, J. M. Haviland-Jones & L. F. Barrett, (Eds.). Handbook of emotions, Guilford Press, New York, pp. 173-191.

Campbell. D. T. (1957). Factors relevant to validity of experiments in social settings. *Psychological Bulletin*, 54, 297-3 12.

Campbell, A. A., Wisco, B. E., Silvia, P. J., & Gay, N. G. (2019). Resting respiratory sinus arrhythmia and posttraumatic stress disorder: A meta-analysis. *Biological Psychology, 144*, 125-135.

Can, Y. S., Gokay, D., Kılıç, D. R., Ekiz, D., Chalabianloo, N., & Ersoy, C. (2020). How laboratory experiments can be exploited for monitoring stress in the wild: a bridge between laboratory and daily life. *Sensors*, *20*(3): 838.

Cannon W. B. (1927). The James–Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, 39, 106–124.

Carbonaro, N. et al., 2013. A mobile biosensor to detect cardiorespiratory activity for stress tracking. *In: 7th International Conference on Pervasive Computing Technologies for Healtcare and Workshops, IEEE*, 440-445.

Carroll D, Turner JR, Hellawell JC (1986). Heart rate and oxygen consumption during active psychological challenge: The effects of level of difficulty. *Psychophysiology, 23*(2), 174-181.

Carroll D, Phillips AC, Balanos GM (2009). Metabolically exaggerated cardiac reactions to acute psychological stress revisited. *Psychophysiology, 46*, 270–275.

Chapleau, M. W., & Sabharwal, R. (2011). Methods of assessing vagus nerve activity and reflexes. *Heart failure reviews*, *16*(2), 109-127.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, *81*(1), 1-8.

Chauntry, A. J., Bishop, N. C., Hamer, M., & Paine, N. J. (2022). Sedentary Behaviour, Physical Activity and Psychobiological Stress Reactivity: A Systematic Review. *Biological Psychology*, 172, 108374.

Chovanec, L., & Gröpel, P. (2020). Effects of 8-week endurance and resistance training programmes on cardiovascular stress responses, life stress and coping. *Journal of sports sciences, 38*(15), 1699-1707.

Conrad, A., Wilhelm, F. H., Roth, W. T., Spiegel, D., & Taylor, C. B. (2008). Circadian affective, cardiopulmonary, and cortisol variability in depressed and nondepressed individuals at risk for cardiovascular disease. *Journal of psychiatric research, 42*(9), 769-777.

Cotton, S. J., Dollard, M. F., & De Jonge, J. (2002). Stress and student job design: Satisfaction, well-being, and performance in university students. *International Journal of Stress Management*, *9*(3), 147-162.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology, 43*(3), 245-265.

Crider, A. et al., 2004. Stability, consistency, and heritability of electrodermal response lability in middle-aged male twins. *Psychophysiology, 41*(4), 501-509.

Crouter, S. E., LaMunion, S. R., Hibbing, P. R., Kaplan, A. S., & Bassett Jr, D. R. (2019). Accuracy of the Cosmed K5 portable calorimeter. *PLoS One, 14*(12), e0226290.

Critchley, H. D., 2002. Electrodermal responses: What happens in the brain. *Neuroscientist*, *8*(2), 132-142

Danieli, A., Lusa, L., Potočnik, N., Meglič, B., Grad, A., & Bajrović, F. F. (2014). Resting heart rate variability and heart rate recovery after submaximal exercise. *Clinical Autonomic Research, 24*(2), 53-61.

Dawson, E., Schell, A. and Filion, D. L. (2000). The electrodermal system, in Dawson, E., Schell, A.., and Filion, D. L. (eds). Handbook of psychophysiology. 2nd edn. Cambridge: Cambridge University Press, pp. 200–223.

Darr, K. C., Bassett, D. R., Morgan, B. J., & Thomas, D. P. (1988). Effects of age and training status on heart rate recovery after peak exercise. *American Journal of Physiology-Heart and Circulatory Physiology, 254*(2), 340-343.

Dennis, P. A., Neal, J. M., Travis, E., Watkins, L. L., Calhoun, P. S., Dennis, M. F., & Beckham, J. C. (2018). Negative affect-related autonomic arousal mediates the association between baroreflex dysfunction and insulin resistance in non-diabetic young adults. *Journal of psychophysiology, 33*(4), 243-253.

Dhabhar, F. S. (2018). The short-term stress response–Mother nature's mechanism for enhancing protection and performance under conditions of threat, challenge, and opportunity. *Frontiers in neuroendocrinology, 49*, 175-192.

Dickerson, S. S., Gruenewald, T. L., and Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology, and health. *Journal of personality, 72*(6), 1191-1216.

Dickerson, S. S., and Kemeny, M. E. (2004). Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychological Bulletin, 130(*3), 355–391.

Diemer, J. E., Alpers, G. W., Peperkorn, H. M., Shiban, Y., and Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*. 6:26.

Doberenz, S. et al., 2011. Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology, 80*(2), 87-95.

van Dooren, M., de Vries, J. J. G. & Janssen, J. H., (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiological Behavior., 106*(2), 298-304.

Ebner-Priemer, U. W., Welch, S. S., Grossman, P., Reisch, T., Linehan, M. M., & Bohus, M. (2007). Psychophysiological ambulatory assessment of affective dysregulation in borderline personality disorder. *Psychiatry research, 150*(3), 265-275.

Edelberg, R., 1967. Electrical properties of the skin. In: Brown C. C. (Eds.). Methods in psychophysiology, Williams & Wilkins, Baltimore, pp. 1-53.

Ekblom-Bak, E., Björkman, F., Hellenius, M. L., & Ekblom, B. (2014). A new submaximal cycle ergometer test for prediction of VO2max. *Scandinavian journal of medicine & science in sports, 24*(2), 319-326.

Ekeland, E., Heian, F., Hagen, K. B., Abbott, J., & Nordheim, L. (2005). Exercise to improve self-esteem in children and young people. Campbell Systematic Reviews, 1(1), 1-52.

Emaus, A., Degerstrøm, J., Wilsgaard, T., Hansen, B. H., Dieli-Conwright, C. M., Furberg, A., . . . Thune, I. (2010). Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromsø study. *Scandinavian journal of public health, 38*(5_suppl), 105-118.

Ensari, I., Schwartz, J. E., Edmondson, D., Duran, A. T., Shimbo, D., & Diaz, K. M. (2020). Testing the cross-stressor hypothesis under real-world conditions: exercise as a moderator of the association between momentary anxiety and cardiovascular responses. *Journal of behavioral medicine, 43*(6), 989-1001.

Feldman, P. J., Cohen, S., Lepore, S. J., Matthews, K. A., Kamarck, T. W., & Marsland, A. L. (1999). Negative emotions and acute physiological responses to stress. *Annals of Behavioral Medicine, 21*(3), 216-222.

Félix, J., Moreira, J., Santos, R., & Sousa, A. S. (2022). Electrodermal Activity Wrist-Based Systems: A Review of Validity, Reliability. *Exploring the Convergence of Computer and Medical Science Through Cloud Healthcare*, 198.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., ... & Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS medicine, 10*(11), e1001547.

Forcier, K., Stroud, L. R., Papandonatos, G. D., Hitsman, B., Reiches, M., Krishnamoorthy, J., & Niaura, R. (2006). Links between physical fitness and cardiovascular reactivity and recovery to psychological stressors: A meta-analysis. *Health Psychology, 25*(6), 723.

Franklin, B. A., Brinks, J., Berra, K., Lavie, C. J., Gordon, N. F., & Sperling, L. S. (2018). Using metabolic equivalents in clinical practice. *The American journal of cardiology, 121*(3), 382-387.

Fu, Q. I., & Levine, B. D. (2013). Exercise and the autonomic nervous system. *Handbook of clinical neurology, 117*, 147-160.

Galanakis, M. J., Palaiologou, A., Patsi, G., Velegraki, I. M., & Darviri, C. (2016). A literature review on the connection between stress and self-esteem. *Psychology, 7*(5), 687-694.

van Genugten, C. R., Schuurmans, J., Lamers, F., Riese, H., Penninx, B. W., Schoevers, R. A., ... & Smit, J. H. (2020). Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of clinical medicine, 9*(2): 322.

Georgiou, K., Larentzakis, A. V., Khamis, N. N., Alsuhaibani, G. I., Alaska, Y. A., & Giallafos, E. J. (2018). Can wearable devices accurately measure heart rate variability? A systematic review. *Folia medica, 60*(1), 7-20.

de Geus, E. J., Bartels, M., Kaprio, J., Lightfoot, J. T., & Thomis, M. (2014). Genetics of regular exercise and sedentary behaviors. *Twin Research and Human Genetics, 17*(4), 262-271.

de Geus, E. J., van Doornen, L. J., de Visser, D. C. & Orlebeke, J. F., (1990). Existing and training induced differences in aerobic fitness: their relationship to physiological response patterns during different types of stress. *Psychophysiology, 27*(4), 457-478.

de Geus, E.J., Willemsen, G.H., Klaver, C.H. and van Doornen, L.J. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. Biological psychology, 3, 205–227.

Gerber, M., & Pühse, U. (2009). Do exercise and fitness protect against stress-induced health complaints? A review of the literature. *Scandinavian journal of public health, 37*(8), 801-819.

Gerteis, A. K. S., & Schwerdtfeger, A. R. (2016). When rumination counts: Perceived social support and heart rate variability in daily life. *Psychophysiology, 53*(7), 1034-1043.

Gnam, J. P., Loeffler, S. N., Haertel, S., Engel, F., Hey, S., Boes, K., . . . Strahler, J. (2019). On the relationship between physical activity, physical fitness, and stress reactivity to a real-life mental stressor. *International Journal of Stress Management, 26*(4), 344.

Godard, B., Raeburn, S., Pembrey, M., Bobrow, M., Farndon, P., & Aymé, S. (2003). Genetic information and testing in insurance and employment: technical, social, and ethical issues. *European Journal of Human Genetics*, *11*(2), 123-142.

Goedhart, A.D., van der Sluijs, S., Houtveen, J.H., Willemsen, G. and de Geus, E.J.C. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recording. *Psychophysiology, 27*(6), 203–215.

Goedhart, A. D., Willemsen, G., de Geus, E. J., In, M. K., (2008). Sympathetic nervous system activity in the heart and the skin: are they comparable? In: Sympathetic Nervous System Research Developments, Nova Science Publishers, New York.

Gonzales, T. I., Jeon, J. Y., Lindsay, T., Westgate, K., Perez-Pozuelo, I., Hollidge, S., . . . Brage, S. (2020). Resting heart rate as a biomarker for tracking change in cardiorespiratory fitness of UK adults: The Fenland Study. *medRxiv* July 2020.

Gordon, A. M., & Mendes, W. B. (2021). A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform. *Proceedings of the National Academy of Sciences, 118*(31), e2105573118.

Gorini, A., Capideville, C. S., DeLeo, G., Mantovani, F., and Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, behavior, and social networking, 14*, 99–105.

Grant, S., Corbett, K., Amjad, A. M., Wilson, J., & Aitchison, T. (1995). A comparison of methods of predicting maximum oxygen uptake. *British journal of sports medicine, 29*(3), 147-152.

Greene, T., Gelkopf, M., Fried, E. I., Robinaugh, D. J., & Lapid Pickman, L. (2020). Dynamic network analysis of negative emotions and DSM-5 posttraumatic stress disorder symptom clusters during conflict. *Journal of traumatic stress*, *33*(1), 72-83.

Gregoire, J., Tuck, S., Hughson, R. L., & Yamamoto, Y. (1996). Heart rate variability at rest and exercise: influence of age, gender, and physical training. *Canadian Journal of Applied Physiology, 21*(6), 455-470.

Gronwall, D. M. A. (1977). Paced auditory serial-addition test: a measure of recovery fromConcussion. *Perceptual and motor skills, 44*(2), 367-373.

Grossman, P., Beek, J.V. and Wientjes, C. (1990). A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. *Psychophysiology, 27*(6), 702–714.

Grossman, P., Karemaker, J. and Wieling, W. (1991). Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: the need for respiratory control. *Psychophysiology 28*(2), 201–216.

Grossman, P., & Taylor, E. W. (2007). Towards understanding respiratory sinus arrythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological Psychology, 74*(2), 263-285.

von Haaren, B., Ottenbacher, J., Muenz, J., Neumann, R., Boes, K., & Ebner-Priemer, U. (2016). Does a 20-week aerobic exercise training programme increase our capabilities to buffer real-life stressors? A randomized, controlled trial using ambulatory assessment. *European Journal of Applied Physiology, 116*(2), 383-294.

Harker, M., (2013). Psychological sweating: a systematic review focused on aetiology and cutaneous response. *Skin Pharmacology and Physiology, 26*(2), 92-100.

Healey, J., Nachman, L., Subramanian, S., Shahabdeen, J., & Morris, M. (2010). Out of the lab and into the fray: Towards modeling emotion in everyday life. In *International Conference on Pervasive Computing* (pp. 156-173). Springer, Berlin, Heidelberg.

Henze, G. I., Zänkert, S., Urschler, D. F., Hiltl, T. J., Kudielka, B. M., Pruessner, J. C., & Wüst, S. (2017). Testing the ecological validity of the Trier Social Stress Test: Association with real-life exam stress. *Psychoneuroendocrinology, 75*, 52-55.

Heron, K. E., Everhart, R. S., McHale, S. M., Smyth, J. M., 2017. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: A systematic review and recommendations. *Journal of Paediatric Psychology, 42*(10), 1087-1107.

Hoemann, K., Khan, Z., Kamona, N., Dy, J., Barrett, L. F., & Quigley, K. S. (2021). Investigating the relationship between emotional granularity and cardiorespiratory physiological activity in daily life. *Psychophysiology*, *58*(6), e13818.

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., ... & Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific reports*, *10*(1), 1-16.

Hossain, M. B., Posada-Quintero, H. F., Kong, Y., McNaboe, R., & Chon, K. H. (2022). Automatic motion artifact detection in electrodermal activity data using machine learning. *Biomedical Signal Processing and Control*, *74*, 103483.

Houtveen, J. H., Groot, P. F. C. and Geus, E. J. C. (2005). Effects of variation in posture and respiration on RSA and pre-ejection period. *Psychophysiology, 42*(6), 713–719.

Houtveen, J. H., Groot, P. F., & de Geus, E. J. (2006). Validation of the thoracic impedance derived respiratory signal using multilevel analysis. *International Journal of Psychophysiology, 59*(2), 97-106.

Huang, C. J., Webb, H. E., Zourdos, M. C., & Acevedo, E. O. (2013). Cardiovascular reactivity, stress, and physical activity. *Frontiers in physiology, 4*: 314.

Ihsan, M. F., Mandala, S., & Pramudyo, M. (2022, July). Study of Feature Extraction Algorithms on Photoplethysmography (PPG) Signals to Detect Coronary Heart Disease. In *2022 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 300-304). IEEE.

Ilies, R., Dimotakis, N., & Watson, D. (2010). Mood, blood pressure, and heart rate at work: an experience-sampling study. *Journal of Occupational Health Psychology, 15*(2): 120.

Imai, K., Sato, H., Hori, M., Kusuoka, H., Ozaki, H., Yokoyama, H., ... & Kamada, T. (1994). Vagally mediated heart rate recovery after exercise is accelerated in athletes but blunted in patients with chronic heart failure. *Journal of the American College of Cardiology*, *24*(6), 1529-1535.

Iqbal, M. U., & Lim, S. (2010). Privacy implications of automated GPS tracking and profiling. *IEEE technology and society magazine*, *29*(2), 39-46.

Izard, C. E., & Tomkins, S. S. (1971). Affect and behavior anxiety as a negative affect. anxiety and behavior Edit. (81-125). CD Spielberger.

Jackson, E. M., & Dishman, R. K. (2006). Cardiorespiratory fitness and laboratory stress: A meta-regression analysis. *Psychophysiology, 43*(1), 57-72.

James, W. (1884). What is an emotion? *Mind, 9*, 185-205.

Jansen, A. S., Van Nguyen, X., Karpitskiy, V., Mettenleiter, T. C., & Loewy, A. D. (1995). Central command neurons of the sympathetic nervous system: basis of the fight- or-flight response. *Science, 270*(5236), 644-646.

Jarczok, M. N., Jarczok, M., Mauss, D., Koenig, J., Li, J., Herr, R. M., & Thayer, J. F. (2013). Autonomic nervous system activity and workplace stressors—A systematic review. *Neuroscience & Biobehavioral Reviews, 37*(8), 1810-1823.

Jebelli, H., (2019). Wearable Biosensors to Understand Construction Workers' Mental and Physical Stress. In: Doctoral dissertation.

Jensen, M. T. (2019). Resting heart rate and relation to disease and longevity: past, present, and future. *Scandinavian journal of clinical and laboratory investigation, 79*(1-2), 108-116.

Joffily, M. (2012) 'EDA Master Toolbox', Available at: https://github.com/mateusjoffily/EDA.

Johnston, D. (1996). Improving control and psychological event detection during ambulatory cardiovascular recording. *In Ambulatory assessment: computer assisted psychological and psychophysiological methods in monitoring and field studies*. (pp. 129-140). Hogrefe & Huber Publishing.

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of clinical psychology, 76*(9), 1591-1612.

Jung, Y. & Yoon, Y. I., 2017. Multi-level assessment model for wellness service based on human mental stress level. *Multimedia Tools and Applications, 76*(9), 11305-11317.

Kamarck, T. W., Schwartz, J. E., Shiffman, S., Muldoon, M. F., Sutton-Tyrrell, K., & Janicki, D. L. (2005). Psychosocial stress and cardiovascular risk: what is the role of daily experience? *Journal of personality, 73*(6), 1749-1774.

Kamarck, T. W., Shiffman, S. M., Smithline, L., Goodie, J. L., Paty, J. A., Gnys, M., & Jong, J. Y. K. (1998). Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: daily life consequences of recurring stress in a multiethnic adult sample. *Health Psychology, 17*(1), 17-29.

Katona, P. G., & Jih, F. E. L. I. X. (1975). Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control. *Journal of applied physiology, 39*(5), 801-805.

Kashou, A. H., Basit, H., & Chhabra, L. (2017). Physiology, sinoatrial node (SA Node).

Kelsey, R. M. (1991). Electrodermal lability and myocardial reactivity to stress. *Psychophysiology, 28*(6), 619-631.

Kelsey, R. M., 2012. Beta-adrenergic cardiovascular reactivity and adaptation to stress: The cardiac pre-ejection period as an index of effort. In: Wright R.A. & Gendolla G.H. (Eds.). How motivation affects cardiovascular response: Mechanisms and applications, American Psychological Association, Washington, pp. 43-60.

Kennedy, A. P., Epstein, D. H., Jobes, M. L., Agage, D., Tyburski, M., Phillips, K. A., ... & Preston, K. L. (2015). Continuous in-the-field measurement of heart rate: Correlates of drug use, craving, stress, and mood in polydrug users. *Drug and alcohol dependence, 151*, 159-166.

Kenny, G., McGinn, R., 2017. Restoration of thermoregulation after exercise. *Journal of Applied Physiology, 122*(4), 933-944.

Kim, J., Murata, T., Foo, J. C., Hossain, B. M. A., & Togo, F. (2021). A Pilot Study of Temporal Associations Between Psychological Stress and Cardiovascular Response. *In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 7040-7043). IEEE.

Kimhy, D., Delespaul, P., Ahn, H., Cai, S., Shikhman, M., Lieberman, J. A., ... & Sloan, R. P. (2010). Concurrent measurement of "real-world" stress and arousal in individuals with psychosis: assessing the feasibility and validity of a novel methodology. *Schizophrenia bulletin, 36*(6), 1131-1139.

Kirschbaum, C., Pirke, K. M. and Hellhammer, D. H. (1993). 'The "Trier Social Stress Test" - A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81.

Kleckner, I. R., Feldman, M. J., Goodwin, M. S., Quigley, K. S., (2020). Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behavioral Research Methods, 53*, 518-535.

Koch, E. D., Moukhtarian, T. R., Skirrow, C., Bozhilova, N., Asherson, P., & Ebner-Priemer, U. W. (2021). Using e-diaries to investigate ADHD–State-of-the-art and the promising feature of just-in-time-adaptive interventions. *Neuroscience & Biobehavioral Reviews, 127*, 884-898.

Konstantinou, P., Trigeorgi, A., Georgiou, C., Gloster, A. T., Panayiotou, G., & Karekla, M. (2020). Comparing apples and oranges or different types of citrus fruits? Using wearable versus stationary devices to analyze psychophysiological data. *Psychophysiology, 57*(5), e13551.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology, 84*(3), 394-421.

Lan, E. (2023). Performer: A Novel PPG-to-ECG Reconstruction Transformer for a Digital Biomarker of Cardiovascular Disease Detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1991-1999).

Lane, J. D., Adcock, R. A., & Burnett, R. E. (1992). Respiratory sinus arrhythmia and cardiovascular responses to stress. *Psychophysiology, 29*(4), 461-470.

Larkin, K. T., Ciano-Federoff, L. M., & Hammel, D. (1998). Effects of gender of observer and fear of negative evaluation on cardiovascular reactivity to mental stress in college men. *International Journal of Psychophysiology, 29*(3), 311-318.

Larsen, J. T., Hershfield, H., Stastny, B. J., & Hester, N. (2017). On the relationship between positive and negative affect: Their correlation and their co-ccurrence. *Emotion, 17*(2): 323.

Lewis, R. P., Rittogers, S. E., Froester, W. F. & Boudoulas, H., (1977). A critical review of the systolic time intervals. *Circulation, 56*(2), 146-158.

van Lien, R., Schutte, N. M., Meijer, J. H., & de Geus, E. J. (2013). Estimated preejection period (PEP) based on the detection of the R-wave and dZ/dt-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions. *International Journal of Psychophysiology, 87*, 60–69.

van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., de Haan, H. A., Vollenbroek-Hutten, M. 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*, 607-629.

Liu, H., Chen, B., Wang, Y., Zhao, X., & Hu, J. (2022). Social affiliation moderates the link between depressive symptoms and heart rate variability in healthy middle-aged and older individuals: An intensive ecologic momentary assessment study. *Psychophysiology, 59*(2), e13958.

Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., & Lyu, J. (2020). Changes in the global burden ofdepression from 1990 to 2017: Findings from the Global Burden of Disease study. *Journal of psychiatric research, 126*, 134-140.

Lumley, M. A., Shi, W., Wiholm, C., Slatcher, R. B., Sandmark, H., Wang, S., ... & Arnetz, B. B. (2014). The Relationship of Chronic and Momentary Work Stress to Cardiac Reactivity in Female Managers: Feasibility of a Smart Phone–Assisted Assessment System. *Psychosomatic Medicine, 76*(7), 512-518.

Määttänen, I., Henttonen, P., Väliaho, J., Palomäki, J., Thibault, M., Kallio, J., ... & Jokela, M. (2021). Positive affect state is a good predictor of movement and stress: combining data from ESM/EMA, mobile HRV measurements and trait questionnaires. *Heliyon, 7*(2), e06243.

Machado-Moreira, C. A., Taylor, N. A. (2012). Psychological sweating from glabrous and nonglabrous skin surfaces under thermoneutral conditions. *Psychophysiology, 49*(3), 369-374.

Maeder, M. T., Ammann, P., Rickli, H., & Brunner-La Rocca, H. P. (2009). Impact of the exercise mode on heart rate recovery after maximal exercise. *European journal of applied physiology, 105*, 247-255.

Malta, L. S., Giosan, C., Szkodny, L. E., Altemus, M. M., Rizzo, A. A., Silbersweig, D. A., & Difede, J. (2021). Development of a virtual reality laboratory stressor. *Virtual Reality, 25*(2), 293-302.

Mann, T. N., Webster, C., Lamberts, R. P., & Lambert, M. I. (2014). Effect of exercise intensity on post-exercise oxygen consumption and heart rate recovery. *European journal of applied physiology, 114*, 1809-1820.

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology, 33*(10), 1264-1273.

Masuda, N., & Holme, P. (2019). Detecting sequences of system states in temporal networks. *Scientific reports, 9*(1), 1-11.

Matthews, G., Davies, D. R., Stammers, R. B., & Westerman, S. J. (2000). Human performance: Cognition, stress, and individual differences. Psychology Press.

McFetridge, J., & Sherwood, A. (1999). Impedance cardiography for noninvasive measurement of cardiovascular hemodynamics. *Nursing research, 48*(2), 109-113.

McArdle, W. D., Katch, F. I., & Katch, V. L. (2015). Chapter 16: Cardiovascular regulation and integration. In Exercise physiology: nutrition, energy, and human performance. Eight Edition, (pp 325-340). Wolters Kluwer Health.

McCarty, R. (2016). The fight-or-flight response: A cornerstone of stress research. In Stress: Concepts, cognition, emotion, and behavior (pp. 33-37). Academic Press.

McNair, D. M., Lorr, M. & Droppleman, L. F. (1971). Manual for the profile of mood states. San Diego, CA: Educational and Industrial Testing Service.

van der Mee, D. J., Duivestein, Q., Gevonden, M. J., Westerink, H. J., & de Geus, E. J. (2020). The short Sing-a-Song Stress Test: A practical and valid test of autonomic responses induced by social-evaluative stress. *Autonomic Neuroscience, 224*, 102612.

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. (2021). Validity of electrodermal activity-based measures of sympathetic nervous system activity from a wrist-worn device. *International Journal of Psychophysiology, 168*, 52-64.

van der Mee, D.J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. (2022). Cardiorespiratory fitness, regular physical activity, and autonomic nervous system reactivity to laboratory and daily life stress. *Psychophysiology*, e14212.

Melanson, E. L. (2020). Resting heart rate variability in men varying in habitual physical activity. *Medicine and science in sports and exercise, 32*(11), 1894-1901.

Melillo, P., Bracale, M., & Pecchia, L. (2011). Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. *Biomedical engineering online*, *10*(1), 1-13.

Menghini, L. et al., 2019. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. Psychophysiology, 56(11), e13441.

Michael Jr, E. D. (1957). Stress adaptation through exercise. Research Quarterly. *American Association for Health, Physical Education and Recreation, 28*(1), 50-54.

Michael, S., Graham, K. S., & Davis, G. M. (2017). Cardiac autonomic responses during exercise and post-exercise recovery using heart rate variability and systolic time intervals—a review. *Frontiers in physiology, 8*, 301.

Migliaro, E. R. (2020). The mechanical side of respiratory sinus arrhythmia. *Physiological Mini Reviews, 13*(4), 34-44.

Miller, L. H., & Shmavonian, B. M. (1965). Replicability of two GSR indices as a function of stress and cognitive activity. *Journal of personality and social psychology, 2*(5), 753-756.

Milstein, N., & Gordon, I. (2020). Validating measures of electrodermal activity and heart ratevariability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience*, *14*, 148.

Minder, C. M., Shaya, G. E., Michos, E. D., Keenan, T. E., Blumenthal, R. S., Nasir, K., . . . Blaha, M. J. (2014). Relation between self-reported physical activity level, fitness, and cardiometabolic risk. *The American Journal of Cardiology, 113*(4), 637-643.

Milstein, N., Gordon, I., 2020. Validating measures of electrodermal activity and heart rate variability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience, 40*, 148.

de Moor, M. H., Beem, A. L., Stubbe, J. H., Boomsma, D. I., & de Geus, E. J. (2006). Regular exercise, anxiety, depression, and personality: a population-based study. *Preventive medicine, 42*(4), 273-279.

de Moor, M. H., Boomsma, D. I., Stubbe, J. H., Willemsen, G., & de Geus, E. J. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of general psychiatry, 65*(8), 897-905.

Morrow, J. R., & Freedson, P. S. (1994). Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science, 6*(4), 315-329.

Morton, M. L., Helminen, E. C., & Felver, J. C. (2020). A systematic review of mindfulness interventions on psychophysiological responses to acute stress. *Mindfulness*, *11*, 2039-2054.

Mosso, A. (1881). Kreislauf des Blutes im Menschlichen Gehirn [The circulation of blood in the human brain]. Leipzig, Germany: Verlag von Veit.

Mosso, A. (1884). La paura [Fear]. Milano, Italy: Fratelli Treves

Mücke, M., Ludyga, S., Colledge, F., & Gerber, M. (2018). Influence of regular physical activity and fitness on stress reactivity as measured with the trier social stress test protocol: A systematic review. *Sports Medicine, 48*(11), 2607-2622.

Mühlen, J. M., Stang, J., Skovgaard, E. L., Judice, P. B., Molina-Garcia, P., Johnston, W., ... & Schumann, M. (2021). Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. *British journal of sports medicine*, *55*(14), 767-779.

Myin-Germeys, I., van Os, J., Schwartz, J.E., Stone, A.A. and Delespaul, P.A. (2001) 'Emotional reactivity to daily life stress in psychosis,' Archives of general psychiatry, 58(12), 1137-44.

Myrtek, M., & Brügner, G. (1996). Perception of emotions in everyday life: studies with patients and normals. *Biological psychology, 42*(1-2), 147-164.

Myrtek, M., Aschenbrenner, E., & Brügner, G. (2005). Emotions in everyday life: an ambulatory monitoring study with female students. *Biological psychology, 68*(3), 237-255.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine, 52*(6), 446-462.

Nakisa B, Rastgoo MN, Rakotonirainy A, Maire F, Chandran V. 2020. Automatic emotionrecognition using temporal multimodal deep learning. *IEEE Access*. 8: 225463–225474.

Nederend, I., Ten Harkel, A. D., Blom, N. A., Berntson, G. G., & de Geus, E. J (2017). Impedancecardiography in healthy children and children with congenital heart disease: Improving stroke volume assessment. *International Journal of Psychophysiology, 120*, 136–147.

Nederend, I., de Geus, E. J., Kroft, L. J., Westenberg, J. J., Blom, N. A., & ten Harkel, A. D. (2018). Cardiac autonomic nervous system activity and cardiac function in children after coarctation repair. *The Annals of thoracic surgery, 105*(6), 1803-1808.

Newlin, D. B., & Levenson, R. W. (1979). Pre-ejection period: Measuring beta-adrenergicinfluences upon the heart. *Psychophysiology*, *16*(6), 546-552.

Newzoo International B.V. (accessed 2021-02-20). *Top Countries/Markets by Smartphone use*. Newzoo: https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users.

Nikula, R., 1991. Psychological correlates of nonspecific skin conductance responses. *Psychophysiology, 28*(1), pp. 86-90.

Novak, D., Mihelj, M., & Munih, M. (2011). Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica, 29*(3), pp. 367-374.

Nyholt, D. R. (2005). Evaluation of Nyholt's procedure for multiple testing correction-author's reply. *Human heredity, 60*(1), 61.

Ogawa, T., (1975). Thermal influence on palmar sweating and mental influence on generalized sweating in man. *Japanse Journal of Physiology, 25*(4), 525-536.

Oldfield, R. C., 1971. The Assessment and Analysis of Handedness: The Edinburgh Inventory. Neuropsychologia, 9(1), pp. 97-113.

Opdenacker, J., Delecluse, C., & Boen, F. (2009). The longitudinal effects of a lifestyle physical activity intervention and a structured exercise intervention on physical self-perceptions and self-esteem in older adults. *Journal of sport and exercise psychology, 31*(6), 743-760.

Orphanidou, C. (2018). Quality Assessment for the Photoplethysmogram (PPG). In *SignalQuality Assessment in Physiological Monitoring* (pp. 41-63). Springer, Cham.

Orne, M. T., & Holland, C. H. (1968). On the ecological validity of laboratory deceptions. *International Journal of Psychiatry*, *6*(4), 282-293.

Payne, A. F. H., Schell, A. M., Dawson, M. E., 2016. Lapses in skin conductance responding across anatomical sites: Comparison of fingers, feet, forehead, and wrist. *Psychophysiology, 53*, 1084-1092.

Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. Emotion, 12, 1320-1328

Peronnet, F., & Szabo, A. (1993). Sympathetic response to acute psychosocial stressors in humans: Linkage to physical exercise and training. *Exercise psychology: The influence of physical exercise on psychological processes*, 172-217.

Peters, C. H., Sharpe, E. J., & Proenza, C. (2020). Cardiac pacemaker activity and aging. *Annual review of physiology, 82*, 21-43.

Pieper, S., Brosschot, J. F., van der Leeden, R., & Thayer, J. F. (2007). Cardiac effects of momentary assessed worry episodes and stressful events. *Psychosomatic medicine, 69*(9), 901-909.

Pieper, S., Brosschot, J. F., van der Leeden, R., & Thayer, J. F. (2010). Prolonged cardiac effects of momentary assessed stressful events and worry episodes. *Psychosomatic Medicine, 72*(6), 570-577.

Poh, M. Z., Swenson, N. C., Picard, R. W. (2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Bio-Medical Engineering, 57*(5), 1243-1252.

Pollard, T. M., Pearce, K. L., Rousham, E. K., & Schwartz, J. E. (2007). Do blood pressure and heart rate responses to perceived stress vary according to endogenous estrogen level in women? American Journal of Physical Anthropology: *The Official Publication of the American Association of Physical Anthropologists, 132*(1), 151-157.

Pollet, T. V., Stulp, G., Henzi, S. P., & Barrett, L. (2015). Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *American journal of primatology, 77*(7), 727-740.

Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of ourevolutionary heritage. A polyvagal theory. *Psychophysiology, 32*(4), 301-318.

Posada-Quintero, H. F., Chon, K. H., 2020. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors, 20*(2): 479.

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

Prill, T., & Fahrenberg, J. (2007). New methods in ambulatory blood pressure monitoring: Interactive monitoring and detection of posture and movement patterns. Behavior *Research Methods, 39*(3), 390-398.

Prince, S. A., Cardilli, L., Reed, J. L., Saunders, T. J., Kite, C., Douillette, K., . . . Buckley, J. P. (2020). A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity, 17*(1), 1-17.

Rahman, S., Habel, M., & Contrada, R. J. (2018). Poincaré plot indices as measures of sympathetic cardiac regulation: Responses to psychological stress and associations with pre-ejection period. *International Journal of Psychophysiology*, 133, 79-90.

Randal, C., Pratt, D., & Bucci, S. (2015). Mindfulness and self-esteem: systematic review. *Mindfulness, 6*, 1366-1378.

Raven, J., 2003. Raven progressive matrices. In: Handbook of nonverbal assessment, Springer, Boston (MA), pp. 223-237.

Regehr, C., Glancy, D., & Pitts, A. (2013). Interventions to reduce stress in university students: A review and meta-analysis. *Journal of affective disorders, 148*(1), 1-11.

Renaud, P. and Blondin, J. P. (1997). The stress of Stroop performance: physiological and emotional responses to color-word interference, test pacing, and pacing speed. *International Journal of Psychophysiology, 2*, 87–97.

Rickles Jr, W. H., Day, J. L. (1968). Electrodermal activity in non-palmer skin sites. *Psychophysiology, 4*(4), pp. 421-435.

Riese, H., Groot, P. F., van den Berg, M., Kupper, N. H., Magnee, E. H., Rohaan, E. J., ... & de Geus, E. J. (2003). Large-scale ensemble averaging of ambulatory impedance cardiograms. *Behavior Research Methods, Instruments, & Computers*, 35(3), 467-477.

Robinson, B. F., Epstein, S. E., Beiser, G. D., & Braunwald, E. (1966). Control of heart rate by the autonomic nervous system: studies in man on the interrelation between baroreceptor mechanisms and exercise. *Circulation Research, 19*(2), 400-411.

Russel, J. (1980). A circumplex model of emotions. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Sallis, J. F. (2000). Age-related decline in physical activity: a synthesis of human and animal studies. *Medicine and science in sports and exercise, 32*(9), 1598-1600.

Salyakina, D., Seaman, S. R., Browning, B. L., Dudbridge, F., & Müller-Myhsok, B. (2005). Evaluation of Nyholt's procedure for multiple testing correction. *Human heredity, 60*(1), 19-25.

Sampson, H. (1958). Serial addition as a function of stimulus duration and pacing. *Canadian Journal of Psychology, 12*, 179-183.

Sampson, H., & MacNeilage, P. F. (1960). Temporal integration and the serial addition paradigm. *Australian Journal of Psychology, 12*, 70-88.

Sano, A., Picard, R. W., 2011. Toward a taxonomy of autonomic sleep patterns with
electrodermal activity. *In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE*, 777-780.

Sano, A., Picard, R. W., Stickgold, R., 2014. Quantitative analysis of wrist electrodermal activity during sleep. *International Journal of Psychophysiology, 94*(3), 382-389.

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., Picard, R.I., 2018. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *Journal of Medical Internet Research, 20*(6), e9410.

Sapolsky, R. M. (1988). Individual differences and the stress response: studies of a wild primate. *Mechanisms of physical and emotional stress*, 399-411.

Schell, A. M., Dawson, M. E., Filion, D. L. (1988). Psychophysiological Correlates of Electrodermal Lability. *Psychophysiology, 25*(6), 619-632.

Schilling, R., Herrmann, C., Ludyga, S., Colledge, F., Brand, S., Pühse, U., & Gerber, M. (2020). Does cardiorespiratory fitness buffer stress reactivity and stress recovery in police officers? A real-life study. *Frontiers in psychiatry, 11*, 594.

Schipke, J.D., Arnold, G. and Pelzer, L. (1991). Effect of respiration rate on short-term heart rate variability. *Journal of Clinical and Basic Cardiology, 2*(1), 92–95.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419-436.

Schneiders, J., Nicolson, N. A., Berkhof, J., Feron, F. J., Van Os, J., & Devries, M. W. (2006). Mood reactivity to daily negative events in early adolescence: relationship to risk for psychopathology. *Developmental Psychology, 42*(3), 543-554.

Schulkin, J. (2011). Social allostasis: anticipatory regulation of the internal milieu. *Frontiers in evolutionary neuroscience, 2*(111), 1-15.

Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M., & de Geus, E. J. (2016). Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiological Genomics, 48*(3), 210-219.

Schutte, N. M., Nederend, I., Hudziak, J. J., de Geus, E. J., & Bartels, M. (2016). Differences in adolescent physical fitness: a multivariate approach and meta-analysis. *Behavior genetics, 46*(2), 217-227.

Schuurmans, A. A., de Looff, P., Nijhof, K. S., Rosada, C., Scholte, R. H., Popma, A., & Otten, R. (2020). Validity of the Empatica E4 wristband to measure heart rate variability (HRV) parameters: A comparison to electrocardiography (ECG). *Journal of medical systems*, *44*(11), 1-11.

Schwerdtfeger, A., & Friedrich-Mai, P. (2009). Social interaction moderates the relationship between depressive mood and heart rate variability: evidence from an ambulatory monitoring study. *Health Psychology, 28*(4), 501.

Schwerdtfeger, A. R., Paul, L., & Rominger, C. (2022). Momentary feelings of safety are associated with attenuated cardiac activity in daily life: Preliminary evidence from an ecological momentary assessment study. *International Journal of Psychophysiology, 182*, 231-239.

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., and Van Doornen, L. J. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology, 27*(1), 1-23.

Siconolfi, S. F., Lasater, T. M., Snow, R. C., & Carleton, R. A. (1985). Self-reported physical activity compared with maximal oxygen uptake. *American Journal of Epidemiology, 122*(1), 101-105.

Simon, S. G., Sloan, R. P., Thayer, J. F., & Jamner, L. D. (2021). Taking context to heart: Momentary emotions, menstrual cycle phase, and cardiac autonomic regulation. *Psychophysiology, 58*(4), e13765.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., ... and Nater, U. M. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology, 51*, 227-236.

Sloan, R. P., Shapiro, P. A., Bagiella, E., Boni, S. M., Paik, M., Bigger Jr, J. T., ... & Gorman, J. M. (1994). Effect of mental stress throughout the day on cardiac autonomic control. *Biological psychology*, *37*(2), 89-99.

Slootmaker, S. M., Schuit, A. J., Chinapaw, M. J., Seidell, J. C., & Van Mechelen, W. (2009). Disagreement in physical activity assessed by accelerometer and self-report in subgroups of age, gender, education, and weight status. *International Journal of Behavioral Nutrition and Physical Activity, 6*(1), 1-10. doi: 10.1186/1479-5868-6-17

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & biobehavioral reviews*, *75*, 274-296.

Sothmann, M. S., Buckworth, J., Claytor, R. P., Cox, R. H., White-Welkley, J. E., & Dishman, R. K. (1996). Exercise training and the cross-stressor adaptation hypothesis. *Exercise and sport sciences reviews, 24*, 267-287.

Sothmann, M. S., Hart, B. A., & Horn, T. S. (1991). Plasma catecholamine response to acute psychological stress in humans: relation to aerobic fitness and exercise training. *Medicine & Science in Sports & Exercise, 23*(7), 860-867.

Spence, J. C., McGannon, K. R., & Poon, P. (2005). The effect of exercise on global self-esteem: A quantitative review. Journal of sport and exercise psychology, 27(3), 311-334. Statista (accessed 2023-02-20). *Smartwatch Devices Unit Sales in the United States from 2016 to 2020*. Technology & Telecommunications; Consumer Electronics: https://www.statista.com/statistics/381696/wearables-unit-sales-forecast-united-states-by-category/.

Stefano, G. B., Stefano, J. M., & Esch, T. (2008). Anticipatory stress response: a significant commonality in stress, relaxation, pleasure, and love responses. *Medical Science Monitor, 14*(2), 17-21.

Stein, P. K., & Pu, Y. (2012). Heart rate variability, sleep, and sleep disorders. *Sleep Medicine Reviews, 1*, 47-66.

Steptoe, A., Leigh, E. S., & Kumari, M. (2011). Positive affect and distressed affect over the day in older people. *Psychology and Aging, 26*(4), 956-965.

Stewart-Brown, S., Evans, J., Patterson, J., Petersen, S., Doll, H., Balding, J., & Regis, D. (2000). The health of students in institutes of higher education: an important and neglected public health problem? *Journal of Public Health, 22*(4), 492-499.

Stifter, C. A., Dollar, J. M., & Cipriano, E. A. (2011). Temperament and emotion regulation: The role of autonomic nervous system reactivity. *Developmental psychobiology*, *53*(3), 266-279.

Stone, A. A., Shiffman, S., Atienza, A. A., Nebeling, L., Stone, A., Shiffman, S., ... & Nebeling, L. (2007). Historical roots and rationale of ecological momentary assessment (EMA). *The science of real-time data capture: Self-reports in health research*, 3-10.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Stubbe, J. H., Boomsma, D. I., Vink, J. M., Cornes, B. K., Martin, N. G., Skytthe, A., ... & de Geus, E. J. (2006). Genetic influences on exercise participation in 37.051 twin pairs from seven countries. *PloS one, 1*(1), e22.

Stubbe, J. H., de Moor, M. H., Boomsma, D. I., & de Geus, E. J. (2007). The association between exercise participation and well-being: a co-twin study. *Preventive medicine, 44*(2), 148-152.

Suvak, M. K., Litz, B. T., Sloan, D. M., Zanarini, M. C., Barrett, L. F., & Hofmann, S. G. (2011). Emotional granularity and borderline personality disorder. *Journal of abnormal psychology*, *120*(2), 414.

Tanosoto, T., Arima, T., Tomonaga, A., Ohata, N., & Svensson, P. (2012). A Paced Auditory Serial Addition Task evokes stress and differential effects on masseter-muscle activity and haemodynamics. *European journal of oral sciences, 120*(4), 363-367.

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015, August). Automatic identification of artifacts in electrodermal activity data. In 2015 37th Annual *International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC) (pp. 1934-1937). IEEE

Thayer, J. F., Hansen, A. L., & Johnsen, B. H. (2010). The non-invasive assessment of autonomic influences on the heart using impedance cardiography and heart rate variability. In *Handbook of behavioral medicine* (pp. 723-740). Springer, New York, NY.

Tombaugh, T.N. (2006). A comprehensive review of the paced auditory serial addition test (PASAT). *Archives of clinical neuropsychology, 21*(1), 53-76.

Tonhajzerova, I., Mestanik, M., Mestanikova, A., & Jurko, A. (2016). Respiratory sinus arrhythmia as a non-invasive index of 'brain-heart' interaction in stress. *The Indian journal of medical research, 144*(6), 815.

Tonello, L., Rodrigues, F. B., Souza, J. W., Campbell, C. S., Leicht, A., & Boullosa, D. A. (2014). The role of physical activity and heart rate variability for the control of work-related stress. *Frontiers in physiology, 5*: 67.

Turner, J. R., Carroll, D., Hanson, J., & Sims, J. (1988). A comparison of additional heart rates during active psychological challenge calculated from upper body and lower body dynamic exercise. *Psychophysiology, 25*(2), 209-216.

Turner, J. R., Sherwood, A., & Light, K. C. (1991). Generalization of cardiovascular response: supportive evidence for the reactivity hypothesis. *International Journal of Psychophysiology, 11*(2), 207-212.

Turner, A. I., Smyth, N., Hall, S. J., Torres, S. J., Hussein, M., Jayasinghe, S. U., ... & Clow, A. J. (2020). Psychological stress reactivity and future health and disease outcomes: A systematic review of prospective evidence. Psychoneuroendocrinology, 114, 104599.

Tuvblad, C., Isen, J., Baker, L. A., Raine, A., Lozano, D. I., & Jacobson, K. C. (2010). The genetic and environmental etiology of sympathetic and parasympathetic activity in children. *Behavior Genetics, 40*(4), 452-466.

Vaessen, T., Rintala, A., Otsabryk, N., Viechtbauer, W., Wampers, M., Claes, S., & Myin- Germeys, I. (2021). The association between self-reported stress and cardiovascular measures in daily life: A systematic review. PloS one, 16(11), e0259557.

Vaez, M., Kristenson, M., & Laflamme, L. (2004). Perceived quality of life and self-rated health among first-year university students. Social indicators research, 68(2), 221-234.

Vansteelandt, K., Van Mechelen, I., & Nezlek, J. B. (2005). The co-occurrence of emotions in daily life: A multilevel approach. *Journal of Research in Personality, 39*(3), 325-335.

Venables, P. H., Christie, M. J., 1980. Electrodermal activity. In: Techniques in psychophysiology, Wiley, New York, pp. 3-67.

Verkuil, B., Brosschot, J. F., Tollenaar, M. S., Lane, R. D., & Thayer, J. F. (2016). Prolonged non-metabolic heart rate variability reduction as a physiological marker of psychological stress in daily life. *Annals of Behavioral Medicine*, *50*(5), 704-714.

Versluis, A., Verkuil, B., Spinhoven, P., van der Ploeg, M. M., & Brosschot, J. F. (2016). Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. *Journal of medical Internet research, 18*(6), e5642.

de Vries, L. P., Baselmans, B. M., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies*, 22(5), 2361-2408.

Vrijen, C., Hartman, C. A., Van Roekel, E., De Jonge, P., & Oldehinkel, A. J. (2018). Spread the joy: How high and low bias for happy facial emotions translate into different daily life affect dynamics. *Complexity, 2018*, 1-15.

Vrijkotte, T. G., van Doornen, L. J., & de Geus, E. J. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension, 35*(4), 880-886.

Wallin, B. (2004). Microneurographic assessment of sympathetic nerve traffic. *Supplements to Clinical Neurophysiology., 57*, 345-351.

Wang, P., Gao, Y., Isen, J., Tuvblad, C., Raine, A., & Baker, L. A. (2015). Genetic covariance between psychopathic traits and anticipatory skin conductance responses to threat: Evidence for a potential endophenotype. *Development and psychopathology, 27*(4pt1), 1313-1322.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.

Weekes, N., Lewis, R., Patel, F., Garrison-Jakel, J., Berger, D. E., & Lupien, S. J. (2006). Examination stress as an ecological inducer of cortisol and psychological responses to stress in undergraduate students. *Stress*, *9*(4), 199-206.

Weise, S., Ong, J., Tesler, N. A., Kim, S., Roth, W. T., 2013. Worried sleep: 24-h monitoring in high and low worriers. *Biological Psychology, 94*(1), 61-70.

Westerink, J. et al., 2009. Emotion measurement platform for daily life situations. *In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, IEEE*, pp. 1-6.

Weiner, H. (1992) 'Perturbing the organism: The biology of stressful experience', University of Chicago press.

Wilhelm, F. H., & Roth, W. T. (1998). Using minute ventilation for ambulatory estimation of additional heart rate. *Biological psychology, 49*(1-2), 137-150.

Willemsen, G. H., de Geus, E. J., Klaver, C. H., van Doornen, L. J., & Carrofl, D. (1996). Ambulatory monitoring of the impedance cardiogram. *Psychophysiology, 33*(2), 184-193.

Willemsen, G., Vink, J. M., Abdellaoui, A., Den Braber, A., van Beek, J. H., Draisma, H. H., ... & Boomsma, D. I. (2013). The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Research and Human Genetics, 16*(1), 271-281.

Wilson, J. D., Stevens, N. T., & Woodall, W. H. (2019). Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Quality and Reliability Engineering International*, *35*(5), 1363-1378.

World Health Organization (2013). Mental health: a state of well-being. [Online] Available at: http://www.who.int/features/factfi les/mental_health/en/.

Wu, W., Gil, Y. & Lee, J. (2012). Combination of wearable multi-biosensor platform and resonance frequency training for stress management of the unemployed population. Sensors, 12(10), 13225-13248.

Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., & Lei, J. (2018). Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. *JMIR mHealth and uHealth*, *6*(4), e9754.

Yang, T., & Zubcevic, J. (2017). Gut–brain axis in regulation of blood pressure. *Frontiers in physiology, 8:* 845.

Yasuma, F., & Hayano, J. I. (2004). Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, *125*(2), 683-690.

Zandara, M., Garcia-Lluch, M., Villada, C., Hidalgo, V., & Salvador, A. (2018). Searching for a job: Cardiac responses to acute stress and the mediating role of threat appraisal in young people. *Stress and Health*, *34*(1), 15-23.

Zanstra, Y. J., & Johnston, D. W. (2011). Cardiovascular reactivity in real life settings: Measurement, mechanisms and meaning. *Biological psychology, 86*(2), 98-105.

van der Zee, M. D., Helmer, Q., Boomsma, D. I., Dolan, C. V., & de Geus, E. J. (2020). An extended twin-pedigree study of different classes of voluntary exercise behavior. *Behavior genetics, 50*(2), 94-104.

van der Zee, M. D., Schutte, N., & de Geus, E. J. (2019). Is physical activity regulated by genetics? Evidence from studies in humans. In J. T. Lightfoot, M. J. Hubal, & S. M. Roth, *Routledge handbook of sport and exercise systems genetics* (pp. 67-79). London & New York: Routledge Taylor & Francis.

van der Zee, M. D., van der Mee, D. J., Bartels, M., & de Geus, E. J. (2019). Tracking of voluntary exercise behaviour over the lifespan. *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), 1-11.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality, 34*(2), 178-197.

Zenker, M., Venz, J., Koenig, J., Voss, C., Beesdo-Baum, K., & Pieper, L. (2021). Evidence for the association between physiological and emotional states in adolescents and young adults without psychopathology under ecologically valid conditions. Psychophysiology, 58(10), e13902.

Zhang, Y., Qin, F., Liu, B., Qi, X., Zhao, Y., Zhang, D. (2018). Wearable neurophysiological recordings in middle-school classroom correlate with students' academic performance. *Frontiers in Human Neuroscience, 12*: 457.

Zoccoli, G., & Amici, R. (2020). Sleep and autonomic nervous system. *Current Opinion in Physiolog, 15*, 128-133.

van der Zwan, J. E., de Vente, W., Huizink, A. C., Bögels, S. M., & de Bruin, E. I. (2015). Physical activity, mindfulness meditation, or heart rate variability biofeedback for stress reduction: a randomized controlled trial. *Applied psychophysiology and biofeedback*, 40, 257-268.

PUBLICATIONS

Overview of publications & author contributions

During my PhD I have had the opportunity to continue my work on earlier projects I was involved in. During my first master internship of the research master Neurosciences at the Vrije Universiteit Amsterdam I studied the influence of Theory of Intelligence (TOI) on the response to errors during a math task under the supervision of Prof. dr. Nienke van Atteveldt. I continued working on this topic as a co-author on two papers of Smiddy Nieuwenhuis. I contributed to both papers of Nieuwenhuis et al. by assisting in the data cleaning of the physiological signals, interpretation of the results, writing of the physiological section of the introduction and discussion, and actively participated in editing and reviewing the first and subsequent drafts.

Nieuwenhuis, S., Janssen, T., van der Mee, D. J., Rahman, F. A., Meeter, M., & van Atteveldt, N. (2023). A Novel Approach to Investigate the Impact of Mindset and Physiology on Effort During an Arithmetic Task. Mind, Brain, and Education, 17(2), 123-131.

Nieuwenhuis, S., van der Mee, D. J., Janssen, T., M., Verstraete, L.L.L., Meeter, M. & van Atteveldt, N. (2023). Mindset and School Burnout Symptoms in Young Adolescents: the Role of Vagal Activity as Potential Mediator. Frontiers in Psychology, section Educational Psychology, accepted.

During my second internship of my master, I got the opportunity to work with data from the Netherlands Study of Depression and Anxiety (NESDA) under the supervision of dr. Mandy X. Hu en Prof dr. Eco J.C. de Geus. I studied the relationship between cortical thickness and ANS activity in depressed and non-depressed individual. The data of this project was included in a pooled mega-analysis led by dr. Julian Koenig. In addition to providing the data for the analyses I actively participated in editing and reviewing the first and subsequent drafts.

Koenig, J., Abler, B., Agartz, I., Åkerstedt, T., Andreassen, O. A., Anthony, M., ... & Quintana, D. S. (2021). Cortical thickness and resting-state cardiac function across the lifespan: A cross-sectional pooled mega-analysis. Psychophysiology, 58(7), e13688.

Before I started my PhD, I got to work on a research project that studied the relationship between dopamine candidate genes and the type of exercise engaged under the

supervision of Prof. dr. Eco J. C. de Geus. In this project I classified all exercise questionnaires of the Netherlands Twin Registry (NTR) on multiple components, namely: metabolic equivalent (MET) score, pacedness level, individual vs. group exercise, and competitive vs. non-competitive exercise. This project resulted in my first publication. This paper was conceptualized by me and Eco de Geus. Co-authors Erik Ehli and Gareth Davies were involved in the collection of the DNA samples and genotyping, Iryna Fedko and Jouke-Jan Hottenga processed the genotype data, Toos van Beijsterveldt and Lannie Lighthart collected and processed the NTR data. I performed all analyses, created the tables and figures, and wrote the first draft. Co-authors Eco de Geus and Matthijs van der Zee were actively involved in editing and reviewing of the first and subsequent drafts. My classification also became the basis of the work by Matthijs D. van der Zee, who tracked the adherence to these different types of exercise over the lifespan. For this paper I partook in editing and reviewing of the first drafts and subsequent drafts.

van der Mee, D. J., Fedko, I. O., Hottenga, J. J., Ehli, E. A., van der Zee, M. D., Ligthart, L., ... & de Geus, E. J. (2018). Dopaminergic genetic variants and voluntary externally paced exercise behavior. Medicine and science in sports and exercise, 50(4), 700.

van der Zee, M. D., van der Mee, D., Bartels, M., & de Geus, E. J. (2019). Tracking of voluntary exercise behaviour over the lifespan. International Journal of Behavioral Nutrition and Physical Activity, 16(1), 1-11.

The work included in my thesis is based on a single study I performed under the supervision of dr. Martin J. Gevonden, Prof. dr. Eco J.C. de Geus and Prof. dr. Joyce H.D.M Westerink. With this teams we conceptualized and designed the experimental procedure that is the basis of my thesis. Together with Eco de Geus I wrote the document for medical ethical approval. Co-author Martin Gevonden was highly involved in the data collection, management, and preprocessing. Each paper was conceptualized by me and Eco de Geus. For all papers I performed the analyses, created the figures and tables, and wrote the first draft of the paper. Co-authors Martin Gevonden, Joyce Westerink and Eco de Geus all actively participated in editing and reviewing of the manuscript. For chapter 1 co-author Quincy Duivestein participated in the collection, processing and analyses of the data and writing of the first draft of the paper.

Chapter 1

van der Mee, D. J., Duivestein, Q., Gevonden, M. J., Westerink, J. H. D. M., & de Geus, E. J. C. (2020). The short Sing-a-Song Stress Test: A practical and valid test of autonomic responses induced by social-evaluative stress. Autonomic Neuroscience, 224, 102612.

Chapter 2

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. C. (2021). Validity of electrodermal activity-based measures of sympathetic nervous system activity from a wrist-worn device. International Journal of Psychophysiology, 168, 52-64.

Chapter 4

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. C. (2023). Comparing the relationship of physiology with affect across laboratory and real-life settings. Psychosomatic medicine submitted Februari 2023.

Chapter 5

van der Mee, D. J., Gevonden, M. J., Westerink, J. H., & de Geus, E. J. (2022). Cardiorespiratory fitness, regular physical activity, and autonomic nervous system reactivity to laboratory and daily life stress. Psychophysiology, e14212.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to the people in my environment who have played a crucial role in enabling me to embark on my PhD research journey and complete this dissertation. Without their help and support, this significant milestone would not have been possible. Therefore, I extend my sincere thanks to everyone who has provided me with advice, much-needed coffee breaks, or distractions from my worries in any way or form throughout my PhD trajectory. While acknowledging the collective support, I would like to mention a few individuals who have made a profound impact on my journey.

First and foremost, I want to express my deepest appreciation to my promoter Eco de Geus and my co-promoters Joyce Westerink and Martin Gevonden. Dear Eco, your multitude of valuable advice, always helpful comments on my articles, and approachable demeanor have contributed immensely to my growth. I have learned so much from you, and I am profoundly grateful for your continuous availability and the freedom you granted me to develop in the direction I desired. Your guidance has shaped me into the researcher I aspire to be. Dear Martin, throughout my PhD research, I experienced a true sense of teamwork between us. It was through our collaboration that all the research in my dissertation took shape. Beyond your academic support in setting up the study and its subsequent stages, your role as an emotional anchor has been invaluable to me. Thanks to our close cooperation and your open nature, I felt comfortable expressing my emotions and seeking solace during challenging times. And indeed, there were difficult moments, particularly as I juggled the responsibilities of growing and giving birth to two children alongside my PhD research. Dear Joyce, although your role may have been slightly smaller than Eco and Martin's, I would also like to express my gratitude for your invaluable input in crafting our articles. I have thoroughly enjoyed your enthusiasm and cheerfulness over the years, and our conversations have always been enjoyable and inspiring. Thank you immensely for that! Furthermore, I always felt like I had an additional co-supervisor in Cor Stoof, despite not being officially part of my promotion team. Whenever I needed a breather or a sounding board for my thoughts, I knew I could count on having coffee with Cor. Cor, your incredible ability to provide perspective and your inexhaustible enthusiasm and cheerfulness have been truly remarkable. I always left from our conversations with a smile.

I want to acknowledge the invaluable contribution of the Philips team, without whom my PhD research would not have been possible. Navin Natoewal and Martin Ouwerkerk, our collaboration has been immensely valuable, and thanks to your presence, I gained profound insights into the differences between academia and working for a company like Philips. I am sincerely grateful for the opportunity to work on a project of this nature. Another

aspect that significantly enhanced my joy throughout my PhD journey is the Department of Biological Psychology and the VU-AMS team. Department like biological psychology, where the sense of community is akin to that of a family are very rare. Despite initially being the sole occupant of my research island, I never felt isolated during lunch, drinks, department outings, or the Christmas party. I am profoundly grateful to everyone I have had the privilege to work with during my 6.5 years in this department. But in particular, I want to express my gratitude to Natascha Stroo, as I believe you are the driving force behind the sense of togetherness and warmth in the department. Additionally, I extend my thanks to the VU-AMS team, one of the smaller groups within the department. From the individuals who have been there from the very beginning, such as Cor, Martin, and Eco, to those who joined gradually like Eric, Quinta, Nicole, and Sjors, as well as those who became part of the team after my PhD research, including Myrte, Melisa, Aniket, and Artemis. I genuinely enjoy our collaboration and eagerly anticipate our future endeavors. Furthermore, being a member of the VU-AMS team has had the added benefit of fostering collaborations with individuals outside the department. This has not only expanded my academic network but has also resulted in a valuable friendship. Smiddy Nieuwenhuis, while you prefer not to be in the spotlight, I want to express my appreciation for the privilege of getting to know you. I am glad our paths crossed!

As I conclude my words of gratitude, I want to dedicate this part to my *rots in de branding* Patrick Hoek, my *leergierig aapje*, Ryan Hoek, and my *eigenwijze framboos*, Zoey Hoek. Patrick, I cannot imagine where I would be in life without you. Your level-headedness and optimism always bring me back to reality. From you, I have learned the importance of reducing stress, especially about matters beyond our control. You have provided unwavering stability in my life for the past 10 years, and that is truly invaluable. Ryan and Zoey, you are beautiful children, radiating with love. You complete me and remind me to maintain a balanced perspective on what truly matters in life. Through you, I have truly understood the art of letting go and embracing the present moment.

Once again, I express my deepest gratitude to everyone who has contributed to my journey, whether directly or indirectly. Your support, both big and small, has been fundamental to my achievements.

The work in my thesis was funded by the VU-AMS talent fund and Philips ltd. Eindhoven.

ADHD? Can that be?

I know that I am different, but that really isn't me.
That's only for boys, insistent on noise.
I am struggling in a very different way,
you see.

Sometimes my head gets left behind,
And my impulsion take over instead.
I'm always forgetting, apologizing, neglecting,
Because there is just to much going on in my head.

I feel emotions intensely, criticism deeply.
Do you know that can be hard?
My ideas are relentless and I dream non-stop.
Maybe I need more inner strength to get started
and not to just discard.

I'm often afraid and I feel like I have strayed,
From the path that others are on.
But I like my brain even though it really does drain,
I don't want those things gone.

Wait, that's ADHD? Well then it must be me.
For so long I have been lost and confused.
Maybe now I can start to understand,
And take part in life enthused.

Accept myself as I am,
Bring this new found power to life.
As we are all unique,
And for happiness and acceptance we strife.