Biological Psychology 84 (2010) 552–569

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment

Frank H. Wilhelm [a,*], Paul Grossman [b]

[a] University of Basel, Institute for Psychology, Department of Clinical Psychology and Psychotherapy, Missionsstrasse 60/62, CH-4055 Basel, Switzerland
[b] University of Basel Hospital, Division of Internal Medicine, Department of Psychosomatic Medicine, CH-4055 Basel, Switzerland

ARTICLE INFO

Article history:
Received 25 April 2009
Accepted 26 January 2010
Available online 2 February 2010

Keywords:
Emotion
Affective sciences
Ambulatory assessment
Ecological momentary assessment
Autonomic nervous system
Social and affective neuroscience
Anxiety
Stress
Methodology

ABSTRACT

Questionnaire and interview assessment can provide reliable data on attitudes and self-perceptions on emotion, and experimental laboratory assessment can examine functional relations between stimuli and reactions under controlled conditions. On the other hand, ambulatory assessment is less constrained and provides naturalistic data on emotion in daily life, with the potential to (1) assure external validity of laboratory findings, (2) provide normative data on prevalence, quality and intensity of real-life emotion and associated processes, (3) characterize previously unidentified emotional phenomena, and (4) model real-life stimuli for representative laboratory research design. Technological innovations now allow for detailed ambulatory study of emotion across domains of subjective experience, overt behavior and physiology. However, methodological challenges abound that may compromise attempts to characterize biobehavioral aspects of emotion in the real world. For example, emotional effects can be masked by social engagement, mental and physical workloads, as well as by food intake and circadian and quasi-random variation in metabolic activity. The complexity of data streams and multitude of factors that influence them require a high degree of context specification for meaningful data interpretation. We consider possible solutions to typical and often overlooked issues related to ambulatory emotion research, including aspects of study design decisions, recording devices and channels, electronic diary implementation, and data analysis.

1. Introduction

Theories of emotion and accompanying research methods to evaluate different aspects of emotional experience have increasingly penetrated various fields of psychology and neighboring sciences like biology, economy, and neurosciences. Given the rapid expansion, it is important to examine essential features of this research enterprise. In this article, we first consider basic assessment paradigms that lie at the heart of almost all emotion research and theory, namely, the very common laboratory experimentation that still constitutes the basis of this research and the self-report data, in form of questionnaires and interviews, which are typically acquired. We then introduce alternative paradigms of ambulatory assessment and field experimentation, which, to date, remain relatively rarely used in psychology and emotion research. The differential benefits of these approaches, as well as their unique difficulties and pitfalls, will be considered. Consequently we also attempt to explain how specific methods may help to overcome important and overlooked assessment obstacles and advance scientific knowledge about emotion in important ways. We argue, from the perspective of theories of emotion, that laboratory and field approaches are fundamentally complementary and not opposing strategies: they offer data and insight from different angles. We propose that the affective sciences urgently need to develop clear, well thought-out programs of research that combine laboratory and naturalistic strategies.

2. The prevailing emotion research paradigms

Almost all current scientific knowledge about emotions is based on laboratory research or inferred from retrospective reports, although understanding real-life emotions is clearly the principal aim. Laboratory and self-report approaches were the most feasible strategies in the past, if not almost the only ones, and still possess distinct advantages: questionnaires and interviews provide easy access to emotion experience and self-observed behavior, and can be used quickly and cost-effectively to evaluate stable characteristics of personality, attitude and cognitions relevant for understanding emotional phenomena. The obtained data are preprocessed and integrated over time by the individual, and thus provide summary information on complex and varying

* Corresponding author. Tel.: +41 61 267 05 93; fax: +41 61 267 0648.
E-mail address: frank.wilhelm@unibas.ch (F.H. Wilhelm).

0301-0511/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.biopsycho.2010.01.017

emotional experiences. Test theory provides sophisticated psychometric tools for developing assessment instruments and analyzing these kinds of data. Additionally, data collecting is usually simple, subject burden is low, and normative comparisons are often available for many widely-used questionnaires and structured interviews–especially in the case of clinical phenomena like anxiety disorders. Good questionnaires and structured interviews are reliable and demonstrate at least some concordant and discordant validity with other instruments.

The laboratory experimental approach complements questionnaire and interview data well. It can provide systematic and reliable information on functional relations between particular stimuli and elicited reactions in the domains of experience, physiology and behavior under well-controlled conditions. This approach is more demanding than retrospective self-report assessment since emotion-inducing stimuli need to be created and implemented in a laboratory context. A large number of laboratory assessment paradigms have been utilized to elicit emotions and probe varying aspects of emotion on different time scales. One of the most frequently used stimulus set is the International Affective Picture System (IAPS, Lang et al., 1995), which contains a large number of pictures with normative ratings for their level of experienced valence and arousal. Subsets eliciting specific emotions can be selected for specific research questions. Similarly, sounds of several seconds duration have been assembled for the same purpose (International Affective Digital Sounds, IADS, Bradley and Lang, 2000). Such stimuli have been widely used to probe basic affective response tendencies (Lang et al., 1990), and much work has been devoted to understanding the short-term reactions to these stimuli in terms of psychological and physiological processes (Lang and Bradley, 2010). Another frequently-used method employs viewing brief video film clips taken from commercial movies (Gross and Levenson, 1995), and has proven highly effective in eliciting various emotions (e.g., anxiety, anger, sadness and joy; Kreibig et al., 2007; Rottenberg et al., 2002), in relation to longer-term dynamic emotion processes. Similarly, scripted imagery has also been used for this purpose (Lang et al., 1983).

Arguably closer to real life are behavioral emotion-induction procedures such as mental stress tasks. For example, the Trier Social Stress Test (Kirschbaum et al., 1993) reliably induces not only stress but also social anxiety and embarrassment. Shorter public speaking or math tasks are also quite effective (Grossman et al., 2001). Stimuli like animals (e.g., spiders, snakes) or needles can be used to induce anxiety among phobic patients (Coldwell et al., 2007). Frustration and anger can be induced by standardized obnoxious behavior of the experimenter (Mauss et al., 2006). A range of emotions can be induced in an almost real-life social context by social interaction on controversial topics (Butler et al., 2003).

These laboratory emotion elicitations work reliably and undoubtedly induce important aspects of genuine emotions, evidenced, for example, by the self-report of participants and objective indicators like crying during sadness or physiological changes during phobic reactions. Furthermore, the constrained environment allows for tight experimental control and reliable elicitation and measurement of emotion response. Even demanding assessment methods like fMRI are feasible. Compliance of participants with experimental instructions is assured through continuous supervision; and measurement in the physiological domain contains relatively few artifacts from participant movement or technical problems. Most importantly, laboratory research allows for systematic variation of independent variables to identify causal factors. For example, by instructing participants to either suppress their emotion reaction or reappraise the situation, the differential causal effects of these emotion-regulation strategies on experience, physiology and behavior can be studied (Gross and John, 2003). Because experiments can be repeated in exactly the same fashion several times, reproducibility of results can be evaluated. The statistical analysis of results follows well-established models like factorial ANOVA. In summary, laboratory emotion research has high internal validity and is particularly good at addressing questions on causes and mechanisms within the constrained setting of the laboratory. However, by exclusively relying on laboratory and retrospective self-report data, researchers are likely to miss important aspects of emotional functioning that may be central to life outside the lab.

3. The ambulatory emotion research paradigm

Many issues, undoubtedly, can only currently be investigated in the laboratory in a tractable way. Nevertheless, there are clearly other phenomena, such as emotion or stress at the workplace or in the family context, that require examination under naturalistic conditions. Unsurprisingly, an increasing number of studies consequently demonstrate added value of ambulatory approaches. Ambulatory assessment has progressed most rapidly in medical applications where the practical benefit is evident, including monitoring of ECG, blood pressure, and activity patterns (Bussmann and Stam, 1998; Haynes and Yoshioka, 2007; Verdecchia et al., 2004). More recently, electronic diaries and other forms of ambulatory "patient-reported outcome" data have become increasingly used in clinical trials, with the intention to overcome retrospective bias and measurement error in the evaluation of treatments (Garcia et al., 2007).

3.1. Is ambulatory assessment important?

Data from our recent study on laboratory mental stress can serve as an illustration of the limitations of a laboratory emotion research approach. Heart rate (HR) is a good index of intensity of mental stress and anxiety. We measured it with a LifeShirt system (Wilhelm et al., 2003a,b) during a laboratory protocol comprised of five mild-to-moderate mental stressors, including, reading, simulated public speaking, an auditory attention task, a memory comparison reaction time task, and a computer math task. Following this morning laboratory session, participants went off to follow their normal daily activities. We controlled HR data for physical activity levels (this method is later described in detail). Figure 1 depicts the course of HR in the lab and in real life, adjusted for physical activity, for a female participant, aged 27 years. The

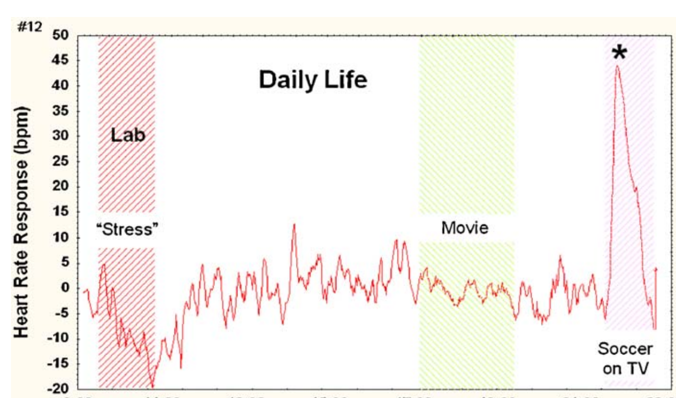

Fig. 1. Heart rate responses (1-min averages) of a study participant (subject 12) monitored with the LifeShirt during and after a laboratory stress protocol consisting of five resting baselines and mild-to-moderate mental stressors ("Lab"). Laboratory stress responses and responses to the movie were small compared to responses to a soccer game the participant watched at home. Note: heart rate was adjusted for ongoing physical activity.

brief task-related peaks seen during the "stress" conditions were typically small, on the order of 5–10 bpm, and were clearly offset in significance to the effects of sitting still from 9–11 a.m., which reduced HR by about 20 bpm. However, when an emotional stimulus specifically relevant to the participant, namely a game of her National soccer team on TV, was broadcast at home, her HR was elevated remarkably (up to 50 bpm) during a period of over an hour (in a semi-supine sitting position!). This example demonstrates that real-life stress can be much more intense than observed in the laboratory and casts doubt on the assumption that laboratory stress results can be generalized to real-life stress.

Another study of ours examined anxiety in daily life. A patient with a diagnosis of social phobia showed highly elevated HR and respiratory irregularities during a business meeting. The patient rated his anxiety on the electronic diary as small to intermediate, while biological reactions were clearly highly elevated (Figure 2). Also note that HR (and ventilatory) changes in anticipation and during the presentation were far out of proportion to the minimal ongoing physical activity (i.e. accelerometry in lowest tracing) and thus clearly emotion-related.

Several other recent studies can also serve as an illustration of how the ambulatory approach provides added and perhaps crucial value for understanding real-life emotion. A study by Ebner-Priemer et al. (2007) assessed emotional experience in borderline personality disorder (BPD). Affective instability is an essential core criterion for this diagnosis. Using electronic diaries, affective states of patients with BPD and healthy controls were repeatedly assessed during a 24-hour period of everyday life. Heightened affect instability was clearly demonstrated in BPD, and self-injuries and skills were identified as emotion-regulation strategies. These results indicate the utility of electronic diaries for emotion-related questions in clinical research.

A study by Gunthert et al. (2007) additionally shows the added value of ambulatory assessment at the interface of psychology and genetics. Participants (n = 345) were genotyped for variation in the serotonin transporter gene. They also reported their feelings daily for 30 days. Participants with the risky genetic variant reported greater anxiety, particularly on high stressor days, compared to low genetic risk individuals. No genetic differences were found using traditional trait questionnaires of anxiety and mood. Findings indicate that ambulatory assessments can be more sensitive than traditional approaches for detecting phenotypic expression of genetic vulnerabilities.

A study by Grossman et al. (2008) assessed patterns of objective physical functioning and perception of mood and fatigue in posttreatment breast cancer patients and healthy controls in their daily lives. Time since treatment was related directly to both momentary mood and perceived energy level assessed by electronic diaries. These measures were also related to objective parameters of activity and cardiorespiratory functioning assessed by the LifeShirt system: importantly, ambulatory, momentary levels of mood and fatigue were related to concurrent physiological activity, providing cross-validating evidence for the relevance of these findings. In contrast, standard questionnaire data of impaired mood and symptoms persisted in patients, independently of time since treatment, indicating a retrospective bias different from moment-to-moment subjective assessments of daily functioning. These three studies are only a few examples of a growing number of reports demonstrating supplementary value for the ambulatory assessment approach.

3.2. Ambulatory assessment can overcome limitations of laboratory research

There is growing concern about low external validity of laboratory and questionnaire findings. The current "Decade of Behavior" (American Psychological Association) calls for a more widespread use of ambulatory assessment techniques that allow examination of psychological phenomena in real life, overcoming many validity problems (Fahrenberg et al., 2007). It has been

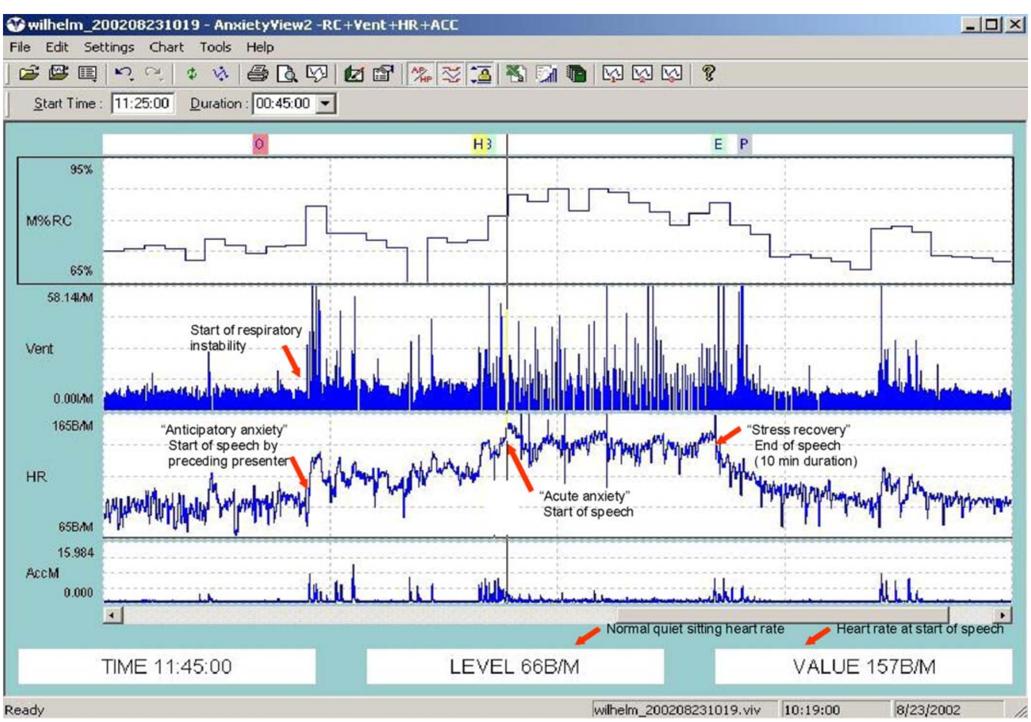

Fig. 2. Annotated ambulatory LifeShirt recording of patient with a diagnosis of social phobia before, during and after a 10-min business meeting presentation. A small subset of recorded and analyzed parameters is diplayed: M%RC: minute-by-minute median of rib cage contribution to tidal volume; Vent: breath-by-breath continuous estimate of minute ventilation; HR: beat-by-beat continuous estimate of heart rate; AccM: motion component of accelerometry.

argued that external validity is an important basis for the public's understanding and support of the academic endeavor, and that lack of external validity in much of contemporary social psychology may lead to reduced funding in the long-term (Rozin, 2009). Serious problems related to validity have also been described in the study of clinical problems that implicate emotion dysfunction, e.g., anxiety and mood disorders (Wilhelm and Roth, 2001). Scientific assessment of these clinical entities commonly occurs in the laboratory or clinic during sedentary baseline periods characterized by extremely low levels of motor activity that are tacitly assumed to reflect quiet rest. Deviations among clinical groups, as compared to the responses of normal individuals, are typically interpreted as evidence of a pervasive trait or diagnostic feature, essentially as an enduring constitutional abnormality. Indeed, many physiological registrations (e.g., blood pressure or blood draws) under immobile laboratory conditions are acquired exactly for such diagnostic purposes.

Laboratory baseline conditions may, however, not closely represent real-world functioning because of specific patterns of emotional reactivity to the setting. This can lead to an overestimation of clinical features. An example of this effect is "white-coat hypertension" in the medical literature, i.e. high blood pressure demonstrated in the clinic but not at home. Indeed, several studies have found rather modest lab-to-life generalizability of blood pressure levels during baseline periods or stress reactions (Gerin et al., 1998; Kamarck et al., 2003).

A similar, perhaps related, laboratory bias exists in anxiety research. It should seem obvious that patients with clinical anxiety disorders may be more prone than others to respond with anxiety to the unfamiliar, potentially threatening situation of being in a clinic or laboratory. Furthermore, the fact that baseline 'resting' measurement is typically carried out immediately before stressful procedures – to which patients know, beforehand, that they may be particularly vulnerable – is further likely to elevate anxiety levels during the preceding "rest" phase. Thus, it is not surprising that often pre-challenge baseline values are systematically elevated compared to post-challenge values (Alpers et al., 2005; Wilhelm and Roth, 1997). Such considerations fundamentally call into question the underlying assumption that the baseline condition reflects a neutral resting state for patients with anxiety or mood disorders. Anticipation of impending tasks, empirically selected to elicit the undesired emotional state, can in no way be considered a baseline resting state. Nor can a brief sedentary measurement period be employed as a neutral baseline – even when not followed by a specific emotional challenge, given likely differences in perception of the laboratory between people with and without psychological disorders.

On the other hand, patients may frequently experience neutral and calm states during everyday life under conditions that are perceived as secure and safe (conversely healthy individuals certainly also show extreme emotional reactions to specific stimuli they find stressful). Hence, any elevated physiological or behavioral indices of anxiety or other emotions may merely be an artifact of the laboratory situation. It has, therefore, been argued that real-life assessment over longer intervals of time is required to establish the validity of laboratory results and to disentangle the effects of trait, temperament or diagnosis from state-trait interaction unique to the laboratory (Blechert et al., 2007; Grossman et al., 2001; Kraemer et al., 1994; Pfaltz et al., 2009; Wilhelm et al., 2006a,b; Wilhelm and Roth, 1996).

An important additional point to consider is that basic emotion research is most commonly concerned with reactivity to specific classes of emotion-eliciting stimuli rather than with individual differences in baseline functioning. Measurements during stimulus presentations are typically controlled for pre-stimulus baselines to reduce the influence of between-individual variation in emotion reactivity scores. Elevated baseline activation can, of course, bias results if the emotional reactivity range is very narrow and produces ceiling effects. Nevertheless, perhaps the greatest risk to the validity of laboratory measurement may be the inherent artificiality of the assessment conditions. A multitude of factors are likely to interfere with the natural unfolding of emotion in the laboratory, including the following: (1) subjects' prior knowledge of experimental stimuli, (2) the unfamiliarity and potential threatening nature of the lab environment in general, (3) the presence of specific and alien technical apparatuses in proximity to the subject, (4) the actual instrumentation of study participants for purposes of physiological or behavioral registration (e.g., being placed in a body plethysmograph or fMRI tube, or tethered to an ECG), (5) the content and presentation of experimental instructions, (6) the awareness of being continuously observed, (7) the repetition of queries about moods and emotions, (8) the unusual requirement of remaining almost completely immobile for at least several minutes (often during situations in which physical activity might be a natural and adaptive response), (9) the caricature-like nature of laboratory stressor conditions that effectively trades off aspects of real life for experimental control, and/or (10) the awareness among many subjects that the experimental situation, in fact, is not real and that an actual threat is proscribed by ethical requirements placed on the researcher.

These factors may have several contrasting and even interactive effects that can result in affecting the overall intensity of elicited emotion reactions, the typical pattern of their manifestation (e.g., reduce facial expressions of emotion), or even the kind or combinations of emotions that are elicited. One could argue that just as in quantum physics, where methods of quantification influence the characteristics of what is being observed (i.e. particles *vs.* waves; Greiner, 2001) – what is commonly known as Heisenberg's uncertainty principle – laboratory research into emotions employs methods that may profoundly alter what researchers are trying to comprehend (i.e., emotions as they commonly occur outside the laboratory). Thus, even the most sophisticated experimental design and measurement may not accurately reflect emotional activity and underlying processes during daily life, although they may precisely characterize what takes place in a rarified realm that sometimes bears little, or only modestly overlapping, resemblance to ordinary situations in the real world. These considerations should make it apparent that without comparison of data from real life, one simply cannot be sure of the relevance of laboratory assessment!

Ambulatory assessment, on the other hand, is less constrained and offers the possibility to provide naturalistic data on emotion in daily life, with the potential to evaluate external validity of more stringently controlled laboratory findings. Ambulatory assessment therefore has the potential to demonstrate the inadequacy of specific laboratory findings and call into question particular theories based on laboratory findings. Consequently, ambulatory approaches may elaborate emotion models in important ways. For example, this has been the case for theories on the origin of panic attacks. Physiological, and particularly respiratory, alterations have reliably been found in laboratory panic induction studies and therefore biological theories have postulated a primary role of respiratory factors (Klein, 1993; Ley, 1985; Wilhelm et al., 2001). However, ambulatory studies have shown that real-life panic attacks are often much less severe compared to what had been observed in the laboratory, both in terms of experienced symptoms and physiological changes (Garssen et al., 1996; Margraf et al., 1987). This fact has subsequently strengthened cognitive accounts of panic (but see Roth et al., 2005). Additionally, very recent ambulatory evidence seriously challenges the idea that respiratory factors play a major role at all in panic disorder (Pfaltz et al., 2009).

One major difference between laboratory and real-world studies is that in the laboratory, conditions are prescribed to participants, while in the real world, individuals exercise natural preferences and choices about situations they seek and avoid. In fact, people differ greatly in terms of the range of situations they encounter in their normal lives, and this is, at least, due in part to personality differences (Ickes et al., 1997). For example, it is increasingly recognized that people regulate their emotions in daily life by situation selection (Gross and Thompson, 2007). This becomes obvious when considering a clinical syndrome like agoraphobia, where certain places and situations are avoided because they induce intense anxiety. The laboratory research paradigm is often unable to tap into this important source of information on emotion-regulation strategies. On the other hand, avoidance of certain negative emotion-eliciting situations in real life may make it difficult to study relevant emotion responses. In the later section on typical problems and solutions for an ambulatory emotion research, we present ambulatory study designs that can circumvent this problem.

Yet another major difference between laboratory and real-world research is the time scale that is able to be captured. Mood states that are temporally fluctuating over the course of a day or week are an integral part of life. Laboratory studies typically do not permit the effective induction of emotional changes over such long durations and are thus not the ideal venue for studying mood alterations.

3.3. Ambulatory assessment of experience vs. retrospective questionnaire data

Beyond the laboratory, the data base on major issues relating to the emotional life of healthy and clinical groups is notably scant: to what degree do people experience, express and regulate emotions and/or moods in their daily lives, what quality and intensity do emotions, moods and cognitions have, and what are the situational contingencies that elicit them? When these issues have been considered, they have often been addressed by means of questionnaires, leaving considerable room for retrospective bias (Pohl, 2004). Notable exceptions are an increasing number of studies that utilize the experience-sampling method (Csikszentmihalyi and Larson, 1987), also called ecological momentary assessment (EMA, Shiffman et al., 2008; Bolger et al., 2003). Some important work on emotions in daily life has recently been done by several groups (Conner and Barrett, 2005; Nezlek et al., 2008). This method has also been applied to assessing mood and emotion in conjunction with clinical phenomena like pain, addiction, anxiety and eating disorders. Recent studies evaluated concordance of EMA data with retrospective questionnaires or interviews and reported notable discrepancies between near-real-time reporting of experiences and later recall of the same events (Shiffman et al., 2008). Importantly, discrepancies were also found in core clinical characteristics such as type and intensity of symptoms (Stone et al., 2004; Solhan et al., 2009). In studies with pain patients, often higher pain ratings were found by retrospective assessment (e.g., Stone et al., 2004).

Both interviews and questionnaires rely on retrospective memory recall and are known to have several shortcomings. The primary problem is forgetting information when the recall period is long. This is accentuated when memory retrieval occurs in a different setting from that of memory acquisition, which is typically the case. Furthermore, the most salient and most recent events at the time of recall will influence overall ratings of emotion of the time period being recalled (Redelmeier and Kahneman, 1996). Most importantly, converging experimental data on recall and hindsight biases, autobiographical memory and daily life functioning indicate that retrospective recall favors mental representations that fit into idealized self-concepts, social expectations and after-the-fact information, rather than those that represent actual experiences or the target behavior examined (Pohl, 2004). When the specific aim is to accurately assess immediate experience, these factors induce error variance at best and systematic bias at worst.

Ambulatory psychological assessment with electronic diaries allows greater proximity, both in terms of time and space, to real-life situations and thereby reduces retrospective recall effects. This method has been successfully applied to self-assessment of behavior and psychological state in a range of disorders and natural settings (Shiffman et al., 2008).

Previous studies on questionnaire-based self-monitoring have shown that systematic, repetitive recordings of behaviors may lead to temporary decrease of problematic or socially undesirable behaviors (Kazdin, 1974), a benefit to dysfunction but also a possible bias for assessment. However, other recent findings, related to studies of alcohol abuse and eating disorders, have shown no such influences of systematic responses to multiple repeated electronic assessment of target behaviors (Hufford et al., 2002; Stein and Corte, 2003). Further research, however, is necessary to determine the extent to which repetitive self-ratings of behavior produces its own class of biases (e.g., response shifts, Schwartz et al., 2006).

An interesting hybrid approach at the intersection of ambulatory experience sampling and retrospective questionnaire is the Day Reconstruction Method (DRM, Kahneman et al., 2004). It consists of a questionnaire that first asks respondents to remember all distinct episodes of the previous day defined by specific activities, persons and contexts, and to estimate the times when they began and ended. Subsequently, respondents are asked to provide affective and other psychological ratings for each episode. The systematic reconstruction of activities and experiences of the preceding day is thought to reduce recall biases. Although correspondence between DRM and experience-sampling data has not been examined, affective circadian patterns based on DRM were similar to results from ambulatory studies (Kahneman et al., 2004). The method represents a compromise to stringent experience sampling, and may have certain advantages: it does not disrupt normal activities (thus reduces the problem of noncompliance and reactivity), may impose less respondent burden, and may provide a more global assessment of contiguous episodes over a full day, rather than a sampling of moments. In addition, the DRM provides detailed information on the timing and duration of activities, which is not practical with experience sampling.

A final point of discussion is warranted concerning the relation of ambulatory momentary assessment of self-reported experience to retrospective approaches. Although retrospective measures may be influenced by a number of factors besides the immediate experience of a situation – e.g., self-concept, social expectations, autobiographical memory and more recent events – such factors apparently play important roles in how individuals characterize their general moods and states of well-being over time. Retrospective measures may thus reflect a basic human tendency to employ memory and social context in order to define and characterize experience, both distal and proximal. Certainly, people who fill out retrospective inventories generally believe their reconstructions of experience to be authentic and true, and their attitudes and future decisions may be more guided by this retrospective "knowledge" than by the sum of momentary experiences (Kahneman, 2000). On the other hand, momentary assessments reflect a more immediate perception – although, of course, also always retrospective to a degree (albeit in terms of minutes, not hours or days) and certainly influenced by prior experience and social context. Within this framework, it may be interesting to view the two types of self-reports of experience as complementary, yielding different types of information, rather than as one superior to the other. Exploring relations between them may improve understanding of subjective experience of emotions as they occur in daily life (Grossman et al., 2008), and early disparagement of either method by opposing camps is not likely to be productive.

In sum, ambulatory assessment, despite its own possible vulnerabilities, has potential to provide novel normative data on the prevalence, quality and intensity of emotional reactions in real life, not only in the experiential domain, but also in areas of physiological and behavior. Importantly, it may also help to characterize previously unidentified emotional phenomena and provide impetus for implementing novel research ideas into the laboratory.

3.4. Ideographic vs. nomothetic approach

Repeated sampling from individuals across diverse situations and over prolonged periods of time provides key features of ambulatory research. This approach is well suited for idiographic issues largely being neglected by laboratory research, in which data averaged across individuals within a single task or occasion can easily mask underlying differences in processes. In contrast to single-occasion group designs, ambulatory multi-occasion measurement yields a more representative sample of diverse processes of an organism and potential environmental determinants thereof. Many scientific insights have, in fact, been derived from a small number of intensively studied individuals (Garmezy, 1982). Even small-N ambulatory studies may significantly expand current knowledge on emotion (for further discussion and examples of innovative ideographic approaches on cardiovascular reactivity, Grossman et al., 1991; Friedman and Santucci, 2003).

3.5. Collecting ambulatory data for representative laboratory research designs

Egon Brunswik, an early influential cognitive psychologist in the 1940s, first put forward the role of adequate laboratory study design when he proposed an ecological approach to cognition. Brunswik (1955) argued that psychological processes are adapted to environmental properties and that modelling of basic cognitive processes is influenced by the type of stimuli utilized in the laboratory. Thus, laboratory results are fundamentally biased to the extent that the experimental stimuli selected are unrepresentative of cognitive processes operating in real life. We argue here that the same applies to the study of emotion, since cognitive processes of attention are inherently involved, and higher cognitive processes, like appraisal, comprise important ingredients of emotional reactions. For example, if being in the company of the spouse is also strongly associated with the presence of the couple's children, a laboratory study of emotions in couples in the absence of children would violate the principle of representative design and provide biased results. The solution Brunswik offered was the method of representative design. It involves randomly sampling stimuli from the environments of individuals to model the type of cognitive tasks people are faced with in their daily lives, and then creating laboratory stimuli in which environmental properties are preserved. Thus, he advocated that psychology be a science of organism-environment relations.

Although Brunswik's claims have been demonstrated in several experiments, there has been little enthusiasm for adopting what Brunswik called "ecologically valid" laboratory approaches in cognitive psychology or investigations of emotion (Dhami et al., 2004). One may seriously argue that modern technologies (e.g., ambulatory assessment) may be able to overcome practical difficulties associated with representative design that hampered earlier adoption. We believe such ecologically valid applications to be particularly important in the field of emotion research, since emotions are linked to social environments and cognitive processes characterized by complex configurations of stimuli. Within this conception, the first task is to determine and characterize relevant properties of situations in daily life that elicit specific emotional responses. Only afterwards follows the implementation of essential aspects of these elements in the laboratory for the purpose of characterizing underlying mechanisms. As Hammond (1998) wrote:

> "In coming to grips with this problem, it is essential to note that the concept of representative design does *not* call for generalization to "real-world" conditions (a meaningless demand). Rather, representative design calls for, first, a *specification of the conditions toward which the generalization is intended*, and second, a *specification of how those conditions are represented in the experimental conditions*." (Emphasis in the original)

3.6. Summing up the advantages of ambulatory assessment

Technological innovations now permit detailed ambulatory study of emotion across a broader context of domains of experience, physiology and behavior. The ambulatory assessment approach has a number of distinct advantages. Because long-term assessment is feasible and less burdensome than in the laboratory, it allows capturing a wide range of a person's emotional reactions and moods. This means that emotional phenomena may be observed that are so rare they are unlikely to ever show up in a laboratory. Of additional importance, person-situation interactions are captured over time, which allows inferring contextual dependency of emotional reactions, within-individual stability of emotional reactivity in real life, and the degree to which emotionality may be a trait. In long-term recording, the possibility of averaging over multiple data points in varying situational contexts may also provide greater reliability and generalizability of assessment.

Circadian biological processes, e.g., fatigue and mood, are likely to influence emotion in real life. Thus a long-term assessment perspective may enable us to model circadian effects and provide diurnal profiles of emotional activity. In addition, assessment of variability of emotional experience and associated symptoms and their long-term temporal patterns becomes possible. Instead of relying on the accuracy of retrospective memory of emotional states, which may sometimes poorly reflect lived experience, ecological momentary assessment can assess emotional states when and where they occur. Experimental sensitization and habituation effects that often occur in the laboratory pose fewer problems in real life. In summary, ambulatory assessment is a naturalistic approach with high external and face validity for describing human emotion. So far, it has been primarily used to address clinical research questions. Nevertheless, we think the time is ripe to apply ambulatory assessment systematically to basic emotion research and to put positions and evidence derived from laboratory assessment to a "real-life" test. This may be accomplished either with ambulatory studies or by representative design for laboratory studies based on criteria derived from earlier ambulatory assessments.

4. Methods of ambulatory assessment

Progress in psychological and psychophysiological ambulatory assessment has been documented in two volumes (Fahrenberg, 2001; Fahrenberg and Myrtek, 1996) covering a variety of clinical and non-clinical research questions. These methods can be employed to assess a wide range of physiological and psychological parameters. In the following section, we treat specific methodologies in relation to emotion studies. We also discuss difficulties of data collection and interpretation that are unique to each method, as well as possibilities to obviate them.

Our catalogue of issues is large but meant as an aid to developing sound strategies for ambulatory investigations. For any one study, it may be unnecessary or even impossible to consider all the points addressed. Nevertheless, we believe it helpful to know just what the various challenges are in ambulatory assessment, particularly in the psychophysiological domain. This should allow researchers to make the best possible decisions when planning and executing a study. By providing a heuristic of ambulatory assessment, we hope not to discourage but rather to empower investigators in their pursuit of knowledge of emotional processes.

4.1. Experience sampling

Ambulatory experience sampling has traditionally been done using paper-and-pencil diaries. However, since it has been demonstrated that participants tend not to fill in these diaries at the instructed times (Stone et al., 2003), electronic diaries have become standard (Dale and Hagen, 2007). A variety of low-cost device and software solutions exist (http://www.ambulatory-assessment.org) (Ebner-Priemer and Kubiak, 2007; Goodwin et al., 2008). Typical diary software includes a timer that allows time-dependent or random beeping during specified time windows. Among diary software solutions, questions and response items with different types of scales can be implemented relatively easily, including assessment of symptoms, emotions, cognitions, situations and/or events. A balance between detailed query and exact assessment of temporal pattern can be attempted. The electronic format allows implementing questions that are followed by queries for more detailed information only if they are endorsed. Covert time stamping for each entry allows checking "signal compliance" of subjects, i.e. latency to respond to the beep, as well as duration of response to individual items that may indicate aspects of cognitive processing or interference.

Typing in words into a PDA is cumbersome. An interesting addition to electronic diaries is a self-activated sound recorder. Recorded sound snips can be transcribed and analyzed for content relevant to emotion research, e.g., by using text analysis programs such as the Linguistic Inquiry and Word Count software (http://www.liwc.net) (Pennebaker et al., 2003).

Although experience sampling is highly adaptable to specific research questions, one needs to keep in mind that self-reported data on emotions, symptoms, and settings are subject to typical biases associated with this data source. Memory effects may be minimized, but certain other factors can still bias information, including selective attention, low interoceptive acuity, lack of emotional awareness, semantic biases (individual and cultural), anchor and scaling problems, response shift over multiple assessments, expectations, attitudes, and intentional, or unintentional, over- or understatement. This implies that additional data, such as behavior observation and physiological recording, can be used to add relevant information for understanding emotion in daily life (Wilhelm and Roth, 2001).

4.2. Behavior recording

Although psychology is often defined as the science of human experience and behavior, the study of behavior has receded in the last decades, especially in personality and social psychology (Baumeister et al., 2007; Rozin, 2009). Nevertheless, a range of technologies is now available that allows capturing behavior in daily life rather unobtrusively. The least intrusive method, also easiest to operate and analyze and most commonly used, is wrist actigraphy (http://www.ambulatory-monitoring.com; http://www.camntech.com) that allows studying gross motor activity patterns over weeks and months. It permits assessment of daytime activity level, sleep/wake cycle and sleep quality. Numerous studies have utilized actigraphy for evaluating clinical conditions such as chronic pain or sleep disturbance and pharmacological effects on behavior, exercise habits and circadian rhythms (Stanley, 2003). However, currently almost entirely overlooked is the potential of actigraphy for providing objective markers of physical engagement and circadian phase in relation to mood, emotion, or situational context, especially when used together with experience-sampling data. More sophisticated accelerometry systems allow for multichannel analysis of hand or head activity, spatiotemporal gait patterns, and posture using an array of sensors attached to different body parts (Bussmann and Stam, 1998; Fahrenberg et al., 1997). This kind of analysis has been productively applied to emotional disturbances like depression (Todder et al., 2009) and anxiety (Sakamoto et al., 2008), but not to basic emotion questions.

Frequency and distribution of speech behavior across the day is an obviously relevant variable in emotion research due to their relationships with mood, social isolation and depression. However, these simple parameters of speech activity have only recently begun to be systematically collected. A promising new ambulatory assessment method for this purpose is called electronically activated recorder (EAR, Mehl et al., 2001). It uses a portable audio recorder that samples sound segments of a few seconds every few minutes from participants' immediate environments. In tracking moment-to-moment ambient sounds, it yields an acoustic log of a person's day as it naturally unfolds. As a naturalistic observation method, it provides an observer's account of daily life and is optimized for the assessment of audible aspects of participants' social lives. The sound snips can be rated by trained coders for various informative categories, such as situational context (e.g., inside, outside, restaurant, lecture, conversation, etc. are often discernible from short sound recordings), amount of speaking (word count) and source of speech (self vs. other). After transcription, content of speaking can be analyzed, and parameters of emotion or personality inferred (Mehl et al., 2006). Transcripts have been automatically analyzed for emotion word content using the Linguistic Inquiry and Word Count software. This method has been shown to be highly unobtrusive and to produce minimal behavioral reactivity (Mehl and Pennebaker, 2003).

A logical extension of the EAR method is to record photos of the environment automatically every few minutes using a camera with a lens embedded in the frame of eye glasses or clothing (e.g., in a button). The general information content of pictures is higher than that of sound recordings. Investigational devices have been used in a few research studies (Healey and Picard, 1998). Here, as well as with the EAR, ethical and legal issues must be considered. However, these aspects are less problematic within a constrained assessment protocol. The technology industry has discovered the potential for a platform to organize snap-shots of the surroundings periodically taken by an individual and has developed hardware and software for storing and managing the potentially large amount of data. These enable quick retrieval by the user (Sensecam tool used in MyLifeBits project, Gemmell et al., 2006; Hodges et al., 2006). Such innovations may also eventually be effectively modified for ambulatory emotion research purposes.

Another promising technology is video recording, which allows studying dynamic sequences of environmental stimuli or facial expressions, depending on the camera orientation. Video recording is a prerequisite for state-of-the-art assessment of emotional facial expressions using the Facial Action Coding System (FACS, Ekman and Friesen, 1978). A robust face-centered head-mounted camera has been developed for the television show ''Fear Factor'' (NBC Universal, Inc.) for capturing facial expressions of contestants as they perform fear-inducing tasks. However, the amount of data amassed within a day is prohibitive. An alternative approach is to activate the video recording only during periods of interest, e.g., automatically during conversations (using a sound switch), by arousal or startle episodes indicated by skin conductance responses, or manually by the user whenever intense feelings occur. To the best of our knowledge, only investigational devices have, thus far, been developed for such purposes (Healey and Picard, 1998; Teeters, 2007). Of course, an already available ambulatory assessment strategy is the ''Big Brother'' (CBS, Inc.) scheme: a set of strategically placed cameras and microphones are situated in the home of the participant and data are recorded over several hours. The ''Speechome Recorder'' developed by Deb Roy, head of the MIT Media Lab's Cognitive Machines research group, is a self-contained unit that can be deployed for several weeks in the home of study participants. It looks like a floor lamp and contains an overhead fisheye camera, microphone, hard disks, and Ethernet connection for data transmission. This is essentially a hybrid approach putting emotion laboratory equipment into the home of participants, with an aim to provide greater validity to the observation of emotion behavior, especially when observing interactions in couples and families.

Other solutions for identifying speech periods exist besides the EAR, each with their own unique advantages and disadvantages. An external microphone recording can be sampled with a digital recording system (e.g., Vitaport, Varioport or Nexus) at a sample rate below 500 Hz, thus providing little information on speech content and thus bypassing privacy issues. However, this makes it impossible to distinguish participant speech from the speech of others, or from external noise. A throat microphone, e.g., used in polysomnography sleep recordings, is specifically sensitive to the speakers own voice. However, it is rather uncomfortable and may attract undesired attention. An interesting alternative for registering the occurrence of speech involves recording and analysis of a subject's breathing pattern. Since speaking is accompanied by typical changes in respiratory patterns (e.g., a large decrease in the ratio of inspiratory to expiratory time), periods of speaking can be detected automatically from respiratory excursions (Wilhelm et al., 2003a,b). This method only provides information on frequency and distribution of speech behavior across the day, but no information on speech content. However, the lack of recognition of speech content employing this method would circumvent privacy and ethical issues that may be problems when making continuous audio recordings of speaking activity. In addition, respiratory analysis allows automatic and unambiguous assignment of speaking to the participant, and not to someone else, without the need for subsequent ratings.

4.3. Psychophysiological monitoring

Modern portable electronic recording devices are capable of duplicating the typical set of channels employed in an emotion research laboratory. Table 1 provides an overview of channels that can be recorded, variables that can be analyzed based on the recorded biosignals, and available devices that contain these channels (but not necessarily the analysis software to obtain the specific variables). Measures that can be recorded during ambulatory monitoring specifically target the major biological emotion and stress response systems, including the sympathetic-adrenal-medullary system (many cardiovascular measures, e.g., blood pressure, pre-ejection period, pulse wave amplitude; electrodermal activity), the vagal system (a few cardiovascular measures, e.g., RSA), the respiratory system (many measures of respiratory pattern, some measures of gas exchange), and the hypothalamic-pituitary-adrenal axis (cortisol).

Some of these channels, especially ECG and accelerometry, have been available for ambulatory recording for many years and frequently employed in ambulatory studies. These physiological signals are easy to obtain, relatively free of artifacts and rather straightforward to analyse. Other ambulatory channels are relatively novel (e.g., end-tidal CO_2 capnography), more difficult to obtain (e.g., ICG), and/or quite susceptible to movement or other artifacts (e.g., EDA, pulse plethysmography, ICG, EEG). Thus, they have been recorded less frequently in field studies.

Multichannel devices, capable of registering the above-mentioned signals in different combinations, exist. They are often highly flexible in the number and range of sensors that can be obtained concurrently, and allow a comprehensive assessment of various physiological responses associated with emotional activation during day and night (Wilhelm et al., 2006a,b; Wilhelm et al., 2003a,b). For example, some of the newer systems make possible measurement of the full laboratory polysomnography setup, including EEG, EOG, EMG, ECG, respiration and activity, enabling detailed sleep stage scoring. Modern devices have advanced analysis capabilities for efficient data reduction and built-in sensors that help provide contextual information (e.g., level of physical activity, posture, speaking). With one commercially available system (the LifeShirt), specific behavioral self-report information can be entered into an electronic diary fully integrated with physiological registrations.

A variety of systems are, in fact, on the market. Simple wrist-watch heart rate monitors (e.g., from Polar, Inc. and Timex, Inc.) can be used to track heart rate conveniently outside the laboratory, with beat-by-beat resolution (Halligan et al., 2006). The Actigraph (e.g., from Ambulatory Monitoring, Inc. and Mini Mitter, Inc.), a relatively simple activity monitor, represents a basic tool that can reasonably accurately estimate sleep/wake cycles, sleep efficiency, daytime physical activity level, and circadian rhythm (Teicher, 1995). Multichannel ambulatory monitors exist that offer good accuracy for measurement of impedance cardiography parameters (Goedhart et al., 2006), heart rate variability (Grossman et al., 2004), electrodermal activity (Wilhelm and Roth, 1998a), or calibrated respiratory pattern (Grossman et al., 2006; Grossman et al., in press). Data on physical activity, autonomic functions and respiratory regulation may be supplemented by endocrine measures via special saliva collection swabs, collected over days or weeks and stored in small plastic containers (Salivettes) in the refrigerator. The samples are then sent to the laboratory for analysis of concentration of the stress hormone cortisol and/or other relatively stable biological compounds (Kraemer et al., 2006; Giese-Davis et al., 2006; Alpers et al., 2003).

With some systems (e.g., LifeShirt, Vivometrics, Inc., Ventura, CA), data may be stored internally for later download or alternatively wirelessly transmitted in real time to a computer for display and analysis. Such telemetric capacity may be useful for purposes of evaluating assessment or even targeted treatment. For example, an operant intervention might be designed for children with impulsivity problems, by which reinforcement is determined by thresholds of momentary physical activity. A similar kind of biofeedback application has been built into some wrist actigraphs for operant treatment of attention-deficit hyperactivity disorder (Tryon et al., 2006). For example, the device beeps when activity level exceeds defined levels of intensity and duration to remind patients to calm themselves down.

A considerable number of clinical studies have utilized the ambulatory paradigm and examined physiological parameters associated with emotion outside the laboratory (Conrad et al., 2008; Ebner-Priemer et al., 2008; Wilhelm and Roth, 1997). In addition, studies on work stress and strain have also employed an ambulatory strategy (Rau et al., 2001; Vrijkotte et al., 2000). Physiologically, many emotion and stress responses are often similar in reflecting generalized activation or arousal, thus

Table 1
Biosignals that can be obtained outside the laboratory with modern recording or sampling devices.

Biosignal analyzed variable	Abbrev.	Recording/Measurement	Device
Electrocardiogram	EKG	Electrodes	Vitaport (TEMEC Instruments, Inc., Kerkrade, Netherlands), Varioport (Becker Meditec, Inc., Karlsruhe, Germany), LifeShirt (Vivometrics/Vivonoetics, Inc., Ventura, USA), Nexus (Mind Media, Inc., Roermond-Herten, Netherlands), VU-AMS (Univ. Amsterdam, Netherlands), HeartMan (Heartbalance, Inc., Graz, Austria), SOMNOscreen (SOMNOmedics, Inc., Randersacker, Germany), Bioharness (BIOPAC Inc., Goleta, USA)
Heart rate (or heart period)	HR/HP	RR-intervals	
Heart rate variability	HRV	Standard deviation within HR time-series, better: spectral power within distinct frequency bands (e.g., 0.10 Hz)	
Respiratory sinus arrhythmia	RSA	RSA parameters, e.g., peak-valley-amplitude or HF (high frequency) spectral power (best with procedurally or statistically controlled respiration)	
Type and number of arrhythmias		Specialized ECG quantifications, best by using multichannel ECG	
T-wave amplitude	TWA		
ST-segment	ST		
Pulse plethysmography	PPG	Photoelectric sensor at finger or ear lobe	Vitaport, Varioport, LifeShirt, Nexus, SOMNOscreen
Heart rate	HR	RR-intervals	
Pulse volume amplitude	PVA	Systolic minus diastolic pulse wave	
Pulse wave velocity	PWV	Time from begin of cardiac contraction to maximum of peripheral pulse	
Oxygen saturation		Oxygen content in arterial blood	
Arterial blood pressure	BP	Device with upper arm cuff (oscillometric), finger cuff (Penaz method)	Mobil-O-Graph (I.E.M. Inc., Stolberg, Germany), TensioDay (TensioMed, Ltd., Budapest, Hungary), Portapres (FMS, Inc., Amsterdam, Netherlands)
Systolic blood pressure	SBP	Upper pressure of pulse wave	
Diastolic blood pressure	DBP	Lower pressure of pulse wave	
Impedance cardiography	ICG	Pairwise electrodes at neck and thorax	VU-AMS
Pre-ejection period	PEP	Time from ECG Q-wave to opening of aortic valve	
Left-ventricular ejection time	LVET	Time from opening to closing of aortic valve	
Stroke volume	SV	Calculated by Kubizek formula	
Cardiac output	CO	SV multiplied by HR	
Respiration pattern		Band(s) for inductive plethysmography	LifeShirt, Inductotrace (Ambulatory Monitoring Inc., Ardsley, USA), Vitaport, Varioport, SOMNOscreen
Respiratory rate	RR or f_B	Breath cycles per minute	
Tidal volume	Vt	Depth of breathing	
Minute volume	Vmin	RR multiplied by Vt	
Respiratory variability		Breath-by-breath variability of duration or depth of breathing	
Capnogram	pCO_2	Cannula at the nostril; or diffusion sensor at lower arm or ear lobe	Capnocount (Weinmann, Inc., Hamburg, Germany)
Electrodermal activity	EDA	Electrodes at palm of hand	Vitaport, Varioport, LifeShirt, VU-AMS
Skin conductance level	SCL	Mean value	
Skin conductance response	SCR	Amplitude of stimulus related reaction	
Nonspecific fluctuation	NSF	Number of spontaneous SCRs/minute	
Temperature	TEMP		Vitaport, Varioport, LifeShirt, IButton, CorTemp (HQ)
Peripheral temperature		Surface sensor, e.g., at hand or foot	
Core body temperature		Rectal sensor	
Electromyogram	EMG	Skin elektrodes above muscles:	Vitaport, Varioport, SOMNOscreen
Facial expression muscles		Face (e.g., corrugator supercilii)	
Eye-blink reflex		Underneath eye (orbicularis oculi)	
Motility	MOT	e.g., at upper thigh	
Tremor		e.g., at lower arm	
Accelerometry (actigraphy)	ACC	Calibrated piezoresistive sensor	Actiwatch (CamNtech, Inc., Cambridge, UK), Motionlogger (Ambulatory Monitoring, Inc., Ardsley, USA), Vitaport, Varioport, SOMNOscreen, Bioharness
Motility	MOT	e.g., at upper thigh or trunk	
Posture	POS	At upper thigh (standing) and trunk (upright vs. supine)	
Tremor		e.g., at hand	
Electrooculogram	EOG	Electrodes at head	Vitaport, Varioport, SOMNOscreen
Electroencephalogram	EEG	Head electrodes in standard positions	Vitaport, Varioport, SOMNOscreen
EEG-alpha, etc.		Amplitude of spectral frequencies	
Salivary sampling		Chewing on cotton swab	Salivette (Sarstedt, Inc., Newton, USA)
Cortisol	CORT	Concentration	

allowing an objective quantification of emotion and stress intensity. In clinical studies, a single emotion is often targeted, e.g., anxiety in a phobic population, and generalized arousal can be interpreted relatively straightforwardly as activation of anxiety. For example, in an ambulatory study of flight phobia, different parameters were compared for their effect size (*d*) in distinguishing phobic from control participants during flight (Wilhelm and Roth, 1998a). Heart rate and skin conductance level emerged as the most discriminative measures, with remarkable effect sizes on the order of 1.2. A combination of several measures provided the best emotion discrimination. In general, aggregating across a number of measures associated with an emotion may counteract the error variance inherent in a single measure and thus provide a superior emotion index.

A word of caution is, however, warranted here. Without information on the type of emotion elicited, the interpretation of physiological data can be difficult. Different emotion situations elicit unique patterns of physiological responses ("situational response specificity"). Additionally, situational specificity may be small in comparison to differences between individuals in their preferred physiological response pattern ("individual response stereotypy"; for an explanation of both terms, Myrtek, 1984). Independent of specific emotional effects, individual differences in physiological reactions certainly derive from a number of other factors, including genetics, physical fitness level, general health, medication usage and momentary physical activity (discussed in Wilhelm and Roth, 2001). These influences can profoundly obscure data interpretation related to emotional correlates. Thus, inferring the kind of emotion a person experiences solely from the immediate physiological patterning is typically not possible and requires additional context information, e.g., aspects of the task or situation, facial expression, verbalizations and level of ongoing physical activity. Nevertheless, efforts are currently under way to apply modern pattern recognition procedures to classification of affective state solely on the basis of physiological signals (Christie and Friedman, 2004; Kreibig et al., 2007; Kolodyazhniy et al., 2010; Picard et al., 2001; Stephens et al., 2010). Although this research is currently exclusively done in the laboratory, a distant goal is to develop sensors and algorithms able to detect and categorize emotion episodes in daily life based on physiological changes and certain context variables (Picard, 1997).

Ambulatory physiological monitoring presents unique challenges, as alluded to above. The amount of data collected can be staggering, and without contextual information, the physiological information may be difficult or impossible to interpret. Particularly, the relationship between emotion activation and physiological measures may be distorted or misleading, due to the fact that physiological systems primarily serve homeostatic and metabolic functions, and thus are highly responsive not only to social engagement and emotional processing, but perhaps often more primarily to motor activity, substance intake, energy expenditure and circadian rhythms. In the following section, we propose some approaches to dealing with these issues.

5. Typical problems and solutions for an ambulatory emotion research approach

5.1. Study design decisions

Designing an ambulatory study can be a challenging task, due to the quantity and variety of stimuli experienced in a real-world environment as well as the fact that the literature on ambulatory investigations is relatively sparse and recent. Laboratory studies on emotion, on the other hand, have a long tradition, and there are many examples of study protocols that have proven to work (Friedman, 2010; Kreibig, 2010, for reviews). Consequently, we attempt here to simplify and systematize ambulatory design by providing a conceptual structure of specific design options informed by previous laboratory and field studies. This is followed by a discussion of the unique implications of individual designs.

On a global level, designs can vary along, at least, six relatively independent dimensions: naturalness (ranging from fully naturalistic to fully structured), type and number of channels recorded (ranging from one to over 20), degree of situational context awareness (ranging from no context information to broad context awareness), sampling mode (ranging from continuous monitoring to time- or event-dependent sampling), and assessment duration (ranging from minutes to months). The large range of choices within each dimension and the resulting multitude of combinations of design options make ambulatory studies highly flexible to address different types of research questions. Therefore, design decisions for a particular research question need careful consideration, and the validity, interpretability, and conclusiveness of results largely depends on the choices made.

5.1.1. Naturalness

Naturalness and the associated face validity is, perhaps, the main argument for leaving the laboratory. In fact, many ambulatory studies have been designed to capture life as it is lived. Examples include 24-h psychophysiological monitoring of panic attacks (Margraf et al., 1987) and electronic diary assessment of family interactions (Wilhelm and Perrez, 2004). Such unstructured or "free running" assessment protocols provide data on the time course, frequency and intensity of naturally occurring target phenomena. However, this approach carries the risk that when targeted events relatively rarely occur (e.g., panics attacks or arguments between partners), their infrequency or even absence may jeopardize the value of a study. However, if the processes involved in these events or behaviors are the main aims of the study (rather than the frequency of their occurrence), one possible strategy may be to screen participants for their frequency of occurrence and only to examine those participants who report high frequency of the targeted event. Of course, this also creates a selection bias in which findings only can be generalized to such extreme subgroups.

Another strategy is to prescribe certain activities to participants, or to manipulate situations outside the laboratory, so that relevant events or behaviors are likely to be triggered in a naturalistic context. This type of ambulatory study may be termed semi-structured because it preserves some of the control of laboratory studies. Most of the data are collected under naturalistic conditions, while some activities are prescribed or triggered at certain times. This approach provides a modicum of controlled assessment while only minimally distorting the normal daily routine.

An important variant of this approach is 24-h recording that includes an ambulatory baseline evaluation. As explained above, laboratory baseline findings can be compromised and not representative of real life if subjects react to the artificial setting. The alternative is to conduct assessment in daily life while instructing subjects to occasionally sit alone in a room for a few minutes without speaking or performing any other activity. This approach has been employed recently to study autonomic regulation in major depressive disorder (Conrad et al., 2008). During 24-h monitoring with the LifeShirt system, patients were instructed at six scheduled times to sit quietly for 5 min, followed by 2 min of paced breathing. Instructions and timing signals were provided by Walkman headphones. Electronic diary entries on emotion, mood, and symptoms were elicited after each assessment. This procedure allowed estimating the diurnal pattern of vagal activity in depression in conjunction with salivary cortisol, as well as relationships to positive and negative affect. Importantly, in contrast to previous 24-h ECG monitoring studies on vagal control in major depression (Carney et al., 2001), physical activity and other confounds could be controlled.

Ambulatory baseline data can also be directly compared to laboratory assessments in order to address the question of lab-to-life generalizability of laboratory findings. Furthermore, multiple baselines during ambulatory recordings allow one to understand the degree to which a certain psychological or physiological feature is a state or a trait (Kraemer et al., 1994): a trait-like feature, e.g., a psychophysiological endophenotype, should be rather constant across multiple standardized baselines, while a state-like feature should vary considerably across time. It is essential for this kind of baseline assessment to be conducted in a standardized way. Any distortion (physical activity, speaking, telephone ringing, person entering the room, ambient noise) may influence results. Thus, distortions should be actively minimized or noted by the participant to allow excluding contaminated baseline periods from the analysis.

A fully structured protocol may also be termed a field experiment because it can include a similar degree of experimental control to that of a laboratory study. It may employ randomized allocation of participants to different conditions (e.g., situations or interventions) or to standardized variations of settings under daily life conditions. In this type of study, the situations or tasks are usually controlled by the experimenter in one of two ways, (1) by instructing participants to engage in certain defined activities, or (2) by arranging specific situations that participants encounter. In emotion research, this may be the instruction for participants to ask someone out for a date or to confront an angry co-worker, situations likely to elicit rather naturalistic emotional responses not previously studied in the laboratory. Nevertheless, knowing that the situation is contrived may still alter participant responses. Consequently, another alternative is actually to arrange specific emotion-eliciting situations that the monitored participant encounters in daily life. For example, provided circumstances remain within clearly determined ethical limits, a confederate might approach a student participant at the university in a predefined manner likely to elicit a certain range of target responses. In the past, this kind of experiment was particularly utilized in social psychology, typically within laboratory or departmental confines, providing important insights on social behavior. However, this ambulatory study approach allows implementing relevant social situations outside the laboratory and studying emotional reactions across several domains of assessment (e.g., physiology and behavior).

One example of a structured protocol is the Stanford flying phobia study that examined phobic patients' reactions to a real flight situation (Wilhelm and Roth, 1996, 1998a). The study protocol involved transportation to the San Francisco Airport and a commercial flight on a small airplane unaccompanied by the experimenter. Subjective emotional reactions to specific segments of the flight (e.g., boarding, taxiing and take off.) were recorded by participants in a diary, and physiological reactions were later analyzed and averaged for these segments. Similarly, in a study with driving phobia participants, the exact course of driving was prescribed and was the same for all participants. Here the experimenter was in the car during the entire study (Alpers et al., 2005). All participants encountered the same situations with the same duration, which makes data especially easy to analyze and unbiased. Such a procedure permits standardized stimulus-based averaging within each individual, and then group statistics across individuals in a fully balanced design that can compare phobic patients with healthy controls. Results indicated that the intensity of anxiety reactions was quite high in phobic participants, both in the subjective and physiological domain. This degree of elicitation of emotional reactions could not have been achieved in a laboratory setting, even with realistic flight or driving simulations (e.g., in virtual reality).

5.1.2. Recorded channels

Type and number of channels recorded are other important dimensions in ambulatory study design. The choices made are largely determined by the dependent and control variables deemed relevant, the availability of miniaturized sensors and recording devices and, most importantly, consideration of subject burden. At the minimum, ambulatory psychological studies normally employ an electronic diary to record setting (see section on context awareness below) and self-reported experiential variables and observations (e.g., positive and negative affect, bodily symptoms, behaviors such as smoking or crying, or partner behaviors). Depending on the research question, implementation of cognitive tests on an electronic diary (e.g., an attention or memory test examining reaction times in response to different items) can also provide important useful information. Thus, the electronic diary can provide several sources of information about various constructs. Still, emotion research can benefit significantly from additional channels that can supplement the subjective nature of diary information and provide an additional perspective on response patterning: a number of control channels and sensors exist that can provide important objective contextual information for improving interpretability of data. These will be discussed in the following section on context awareness.

Some form of cost-benefit analysis should be used to determine the number of channels recorded in an ambulatory monitoring device, the number of items queried in an electronic diary and the number of additional devices utilized (such as the EAR): aside from the cost of the equipment involved, any added channel or item increases subject burden, especially during long-term assessment, and needs to be justified by a significantly increased gain of information. This particularly applies to clinical populations already often severely burdened by their disorder. Adding a low-burden channel (e.g., accelerometer, skin temperature sensor) or single item requires less justification than adding a high-burden channel (e.g., end-tidal pCO_2, blood glucose monitor) or an extra entire questionnaire. Participants experience sensors as especially burdensome if they are visible (often causing embarrassment in social situations), produce discomfort, affect sleep or reduce the range of normal activities like showering or performing sports. These problems are amplified if study duration is long. Any additional subject burden may increase the likelihood of reactivity effects, with the potential of distorting naturalistic responses. Also overburden or discomfort may alter compliance, leading to selective dropout. Such problems may seriously compromise the validity and representativeness of study results to a certain degree.

In emotion research, integration of several relevant channels in one device (e.g., the LifeShirt system that incorporates physiological and control channels, and an electronic diary into a fully integrated device) is the ideal solution since it makes recording easier and reduces subject burden. However, multichannel studies sometimes rely on concurrent recordings from several devices (e.g., the EAR, an electronic diary, and a physiological recorder). In this case, synchronization of data is of utmost importance. This can best be accomplished by synchronizing the internal clocks when the recording is started (e.g., by starting them with the same PC) and making sure that they stay synchronized during the entire recording (e.g., by pressing a marker button or stopping them at the same time, then comparing clock differences between devices). A more sophisticated approach is to provide a synchronization timing signal to all devices. No commercially available ambulatory solution currently exists for this purpose.

5.1.3. Context awareness

One cannot underestimate the importance of information on situational characteristics during ambulatory studies on emotion. In contrast to laboratory studies, the participant is almost always unsupervised, and the sequence and duration of stimuli during registration are not known in advance. Therefore, ambulatory data are especially subject to interpretational ambiguities. Medical ambulatory monitoring studies and older psychological investigations often have neglected this important data source. For certain very limited research or diagnostic questions, recording only one channel may be sufficient. However, the general rule is that a high degree of context awareness is necessary for adequate data interpretation, given the complexity of data streams and the multitude of factors that influence them. Employment of additional channels can partially address this issue. Therefore, during the study design phase, much consideration should be given to (1) potential situational effects (i.e. triggers or moderators of emotion responses), and (2) potential time- or situation-dependent confounders of primary dependent variables (e.g., variations in concurrent physical activity, metabolic activity, ambient temperature and time of day).

At the most basic level, for example, is information about time of day and date. Thus, time stamping of recordings is standard in electronic diaries and physiological recorders. This information allows aggregating data within and across individuals in a systematic way to allow group comparisons and identification of diurnal patterns. The second most basic type of context information is provided by an event marker channel in physiological recorders. This allows registration of the occurrence of predefined events by participants and subsequent meaningful segmenting of recorded data during analysis. For example, panic disorder patients were asked in a 24-h study to indicate onset of panic attacks by pressing a marker button attached to their belt. This allowed for response-locked averaging of heart rate responses (Margraf et al., 1987). In the Stanford flight phobia study, participants pressed a marker button during predefined segments of the flight, then filled out a paper-and-pencil diary on their anxiety and symptoms, and then pressed the marker button twice. In subsequent analyses of the physiological data, the single and dual marker presses could be easily identified and helped in assigning the recorded data to specific flight segments. Additionally, experimenters correctly anticipated that subjects would sometimes forget to press the button. Therefore, context channels were added to the recording that provided important information on correct assignment of data: (1) an air pressure channel, indicating the exact moment of take off, cruising, beginning of descend, and landing, and (2) a physical activity channel, indicating the beginning and end of boarding. By using the redundancy from these additional channels, many subjective marker entries were verified and correct data analysis could be assured.

Event markers, by themselves, are relatively limited when it comes to identifying several specific situations. In contrast, electronic diaries can include elaborate entry menus for type of situation encountered. This is especially relevant in unstructured protocols in which information on the situational contingencies of emotion responses may be important. Categories of setting used in ambulatory studies may, for instance, include home vs. work, physically active vs. inactive, in bed vs. out of bed, and wake vs. sleep. Few studies exist so far that also distinguish important categories, such as indoors vs. outdoors, alone vs. accompanied (and by whom), talking vs. quiet, eating vs. non-eating, and stationary vs. mobile (e.g., walking, biking, driving or being driven).

Assessing all situational categories, whenever they change, would be far too demanding for participants, and lead to unreliable information and lack of compliance. However, there are at least two solutions to this problem: (1) reduction of situational categories, in which only a limited number of relevant situational categories are defined based on *a priori* hypotheses, e.g., negative mood is higher at work than at home, and (2) replacement of self-reported situational categories by continuously recorded context marker channels. Such electronic context markers can be obtained for a large number of situational categories: indoors vs. outdoors (using a light intensity sensor, or better yet a sensor that also records spectral light composition to distinguish indoors vs. outdoors lighting), physically active vs. inactive (using an activity watch or accelerometer sensor), alone vs. accompanied (using the EAR), eating vs. non-eating (using the EAR or special microphones attached to the head), talking vs. quiet (using the EAR, a throat microphone, or respiration pattern sensor, see below), stationary vs. mobile (using the EAR), standing vs. sitting vs. lying (using at least two posture sensors, one on the trunk and one on the thigh), and awake vs. sleep (using an activity watch or accelerometer, or for detailed analysis, polysomnography).

5.1.3.1. The special problem of concurrent metabolic activity variation. A major obstacle to straightforward interpretation of physiological ambulatory data is that time- or situation-dependent confounders can substantially affect primary dependent variables. Particularly, emotion effects in psychophysiological variables, such as heart rate and skin conductance, can easily be masked by circadian and quasi-random variation in physical activity, social interaction and substance intake. A major problem in unconstrained ambulatory assessment is confounding by physical activity (Wilhelm et al., 2006a,b). This is illustrated in Figure 3. The most frequently used autonomic index of emotional arousal is heart rate, shown to be a sensitive index of anxiety under well controlled conditions (Wilhelm and Roth, 1998a). However, it is also functionally coupled with metabolic demand. During daily life recording, most of the variance of heart rate is due to physical activity (Grossman et al., 2004). As a result, heart rate is not a reliable measure of anxiety in ambulatory recordings when level of concurrent physical activity is unknown, because the emotional component may be masked by a substantial metabolic component. The challenge is to unmask nonmetabolical heart rate effects (e.g., emotion). This is feasible using a variety of analytic techniques described below. The prerequisite for these unmasking techniques is that physical activity is recorded reliably and with sufficient time and intensity resolution.

Social interaction is an important situational variable that modulates emotion in daily life. Recording of speaking activity can provide an important context marker for understanding emotion since speaking is frequently associated with social exchanges. However, speaking may also be a confounding factor when assessing physiological emotion channels, since even fully unemotional speaking raises heart rate by 5–10 bpm (Linden, 1987). A relatively simple solution that controls for speech activity

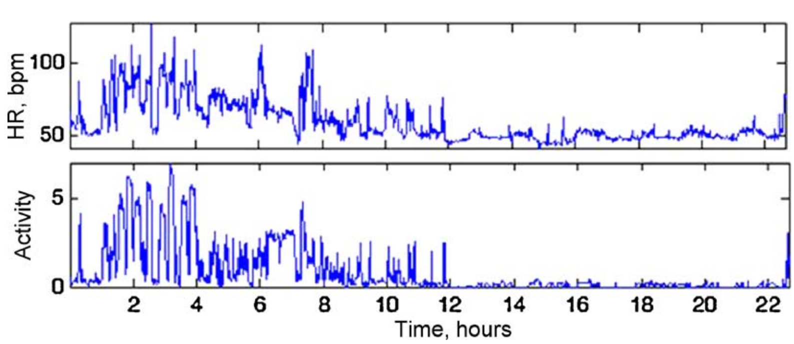

Fig. 3. Data of the relationship of heart rate (HR) to physical activity (Activity) in a single subject during 22 hours of recording (time from start of recording at 10 a.m. is displayed).

is to measure speech activity from the respiratory signal of the person examined (see discussion above).

Ingestion of various substances also will clearly alter the physiological landscape, which may influence physiological measures of emotion. Such intake can include consumption of regular meals, snacks, caffeine, nicotine, alcohol and medication. For example, during and after a high carbohydrate meal, heart rate and blood pressure often increase for several hours (Fagan et al., 1986). Caffeine and nicotine have sympathetic effects on the cardiovascular and electrodermal system. Medications prescribed for allergies, cardiovascular, respiratory or mental disorders often have sympathetic and/or parasympathetic effects. Electronic diary entries of time, type and amount of ingestion are essential to examine effects on emotion or to control for their influence on dependent variables

Diurnal endogenous influences from the circadian pacemaker (the suprachiasmatic nucleus) are especially related to physiological regulation patterns, including the temperature and cardiovascular system (van Eekelen et al., 2004). They may additionally also influence a number of psychological variables such as mood, alertness, and memory (Boivin et al., 1997; Dijk et al., 1992). Since circadian rhythms change slowly over a period of about 24 h, their effects on emotion reactivity can be accounted for, at least to a certain degree, by adjusting for this slowly shifting "baseline" in the analysis. Segmenting by time of day and wake vs. sleep is essential for accounting for circadian influences (Grossman et al., 2008), which can be achieved by appropriate electronic diary entries, accelerometer measurement, or polysomnography.

5.1.4. Sampling mode

The *type of sampling* employed is largely determined by the kind of signal recorded. Physiological and control channels are typically recorded continuously, while self-reports can only be recorded intermittently. The EAR method takes sound snips of 30 sec at evenly spaced intervals of about 12 min to obtain a representative sample of the sounds of daily life. Electronic diary entries can be prompted at random times (random sampling), precluding anticipation effects, or they can be prompted at specific predefined times (time-dependent sampling), reducing subject burden because no unanticipated interruption of activities (such as driving a car or attending a meeting) will occur. Sampling intervals need to be optimized to capture meaningful variation. Spacing between electronic diary prompts is typically in the range of 1 to 24 hours. Three-hour spacing allows obtaining five measurement points per day, which is sufficient to capture gross changes in mood across the waking day (Munsch et al., 2009; Pfaltz et al., 2010). However, in studies concerned with short-term instability of emotion, sampling intervals as brief as 10 min have been used (Ebner-Priemer et al., 2007). Longer sampling intervals allow for a more detailed psychological assessment and higher number of items, while short intervals need to be briefer.

Event-dependent sampling is contingent on the occurrence of a specified event relevant for the study, e.g., a binge-eating attack, an argument, or a perceived emotion-related bodily symptom. This data entry mode is highly susceptible to poor compliance because subjects, either intentionally or unintentionally, can easily omit entries that cannot later be detected. One solution is automatic event detection, already implemented for online detection of physiological signs of an emotion episode by using a specific detection algorithm based on the model of additional heart rate (Myrtek et al., 2005): the recording device is programmed to detect sudden heart rate increases not explained by concurrent increases in physical activity. Whenever a certain threshold of heart rate increase is exceeded, subjects are beeped to fill out an electronic diary about their current experience. A similar approach would be to monitor skin conductance of subjects. Whenever the signal change (first derivative) crosses a threshold (an index of skin sympathetic responsivity), a camera is activated recording the surroundings. Thus, the contextual information is contingent upon increased sympathetic activation (Healey and Picard, 1998). Other analogous approaches to automatic event detection might also be relevant to emotion research (e.g., facial EMG of the corrugator supercilii and zygomaticus major muscles as indices of the quality and intensity of facial emotion expression, or environmental cues, see the Context-Aware Experience-Sampling Project at MIT, http://web.mit.edu/caesproject).

5.1.5. Assessment interval and overall duration

Total assessment duration in ambulatory studies has varied from minutes (e.g., exposure to a phobic situation using a fully structured protocol; Wilhelm and Roth, 1998a) to months (e.g., the development of circadian rhythm in newborns; So et al., 2007). The appropriate duration needs to be chosen so as to obtain enough data to reliably test a hypothesis while keeping subject burden, study cost and analysis load at a minimum. For example, if the study question relates to circadian variation in mood during seasonal changes, at least one day during different seasons needs to be sampled. To increase reliability of this assessment, sampling of at least three days per season and averaging across these days may make sense, especially for variables known to show considerable between-day variation. Along a similar line, the emotional landscape on weekends is often quite distinct from that of weekdays. To account for this effect, at least one day from the weekend and weekdays need to be examined. In general, it is also useful to consider any special characteristics of the sample investigated. Thus, for student populations often characterized by an unstable life style, several days during the week, as well as the weekend, may need to be sampled in order to collect representative data.

Sufficiently long recording intervals need to be implemented to increase the likelihood of documenting rare phenomena (e.g., panic attacks or fits of rage). In panic studies, one week of registration has been an adequate duration to obtain at least one panic attack from most participants (Margraf et al., 1987).

The overall number of samples per subject determines what kinds of analyses are feasible. Modeling of temporal dynamics in individual response patterns and temporal interrelatedness of different channels is feasible using time-series analysis such as autoregressive modeling. This linear predictive analysis typically requires at least 30–50 sampling points per subject to provide sufficient reliability, with more points increasing reliability. When using mixed-effects models to examine situational dependency of emotion responses in a nested design, the assessment duration should allow for obtaining at least one instance of each situation of interest in most participants.

5.2. Data analytic strategies

5.2.1. Fully structured protocols

Statistical analysis of fully structured ambulatory studies is straightforward since it involves comparisons across prescribed situations or tasks subjects encounter. For this type of study, statistical procedures developed for laboratory research, such as repeated measures analysis of variance, are appropriate to examine effects of tasks or situations on dependent variables, possibly with random assignment of different subject groups to specific conditions in order to evaluate their differential effects in a between-subjects design.

5.2.2. Unstructured or semi-structured protocols

In contrast, typical research questions in unstructured or semi-structured ambulatory studies are often concerned with the relationship between naturally occurring and temporally fluctuating situational characteristics and dependent variables (e.g., the influence of different types of social encounters on experienced anxiety, blood pressure, or speaking activity, as well as the extent to which individual characteristics predict these associations). Because of the unstructured nature of situational characteristics (and the possibility that not all subjects encounter a certain situation), the complexity of multi-modal assessment and the sheer amount of ambulatory data, a detailed analysis plan should be an inherent part of ambulatory study design. Ambulatory data are generally characterized by many repeated measurements of several dependent variables over time and often include predictor or confounding variables varying within-person, producing a nested data structure that needs to be considered in the analysis. In addition, missing data or unequal numbers of observations across individuals are inevitable, because of lack of compliance, different recording durations (e.g., due to differing lengths of the waking day), the fact that not all subjects encounter all situations, artifactual recordings or technical malfunction. Some designs may also include unequally spaced time intervals. Modern statistical methods that can handle these kinds of data are advisable.

Some advanced analytic strategies for ecological momentary assessment data have been described (Schwartz and Stone, 1998). Although they focus on experience sampling, many of their recommendations also apply to the domains of behavior and physiology data, since data structures for these measurements are similar. Using hypothetical study examples and borrowing from experimental approaches, Schwartz and Stone have convincingly demonstrated that common ways of analyzing ambulatory data in unstructured protocols, such as aggregation across repeated measurements, repeated measures analysis of variance, pooled within-person regression, and two-stage estimation procedures for multilevel models, are feasible but usually suboptimal. These approaches neglect the fact that ambulatory data often are unbalanced (in terms of N per cell), show serial within-subject autocorrelation, or violate the assumption of homoscedasticity, possibly leading to incorrect inferences. In addition, they may inappropriately conflate within-person and between-person effects. In their 1998 publication, Schwartz and Stone recommended multilevel models (for an introductory textbook, Singer and Willett, 2003) because they can better address many ambulatory research questions. These types of models have progressively developed to deal more flexibly with the various aspects inherent to data obtained from ambulatory assessment (we recommend Walls and Schafer, 2006) and are often referred to in the literature as (linear) mixed models, hierarchical (linear) models, random coefficient models (or within the structural equation modeling framework, latent growth curve models). By collecting a large number of emotion ratings in daily life, concurrently with relevant situational information, emotion-by-situation analyses are feasible using multilevel models. Interestingly, the results of such global analyses indicate that emotion reactions in daily life seem less subject to situational influences than what participants report in questionnaire studies (Pawlik and Buse, 1996).

5.2.3. Segmenting of physiological and behavioral data

Special consideration needs to be given to physiological and behavioral records. In contrast to experience-sampling data, these data are usually collected in a continuous fashion. A necessary first step is data reduction (e.g., a reduction of data size by a factor of 60,000 is achieved when a one-minute ECG signal sampled at 1000 samples/s [Hz] is reduced to average heart rate value for that segment of measurement). For most channels, this can be done highly automatically using appropriate software. However, physical movement of muscles and tissue near sensors (such as ECG, EDA, and pulse plethysmography) is more of a problem than in the laboratory and can interfere with accurate recordings. Thus, the data reduction software needs to be able to detect and exclude significant signal deviations from physiologically plausible values. Full disclosure analysis is advisable for ambulatory assessment (in which the complete original signal is available offline for inspection), thus allowing comparison of derived values and original raw data, whenever derived parameters are suspect (e.g., unusual values that may or may not be due to true physiologic changes or to movement artifact or sensor malfunction) (Wilhelm and Roth, 1998b). Investment in modern devices that possess full disclosure capabilities is therefore important.

Studies in biological psychology often address interrelationships of psychological, situational, behavioral and/or physiological factors. Thus, the unit of segmentation for these data domains should be similar to allow meaningful cross-domain analysis. For example, when stress ratings are provided each hour for the preceding hour, behavioral and physiological data should also be averaged for each hour. In addition, the experience-sampling interval needs to be chosen appropriately to tap into the process of interest. If meaningful fluctuations of a variable (e.g., positive mood) are expected to occur within the period of a few hours, a rating every few hours is appropriate. If changes are expected to be swift, e.g., during the onset of a panic attack, sampling every few minutes may be appropriate. Physiological and behavioral data need to be aggregated across time accordingly (so called "segmenting" or "binning").

5.2.4. Estimating emotional stability

The serial nature of ambulatory data and the long assessment durations allow estimation of the degree to which an emotional process is stable in daily life. Emotional instability has been examined, for example, in borderline personality disorder and in panic disorder using the van Neuman statistic (Ebner-Priemer et al., 2009; Pfaltz et al., 2010). This statistic effectively indexes the point-by-point fluctuations in the dependent variable and provides a summary statistic for emotional instability.

5.2.5. Detecting points of abrupt change

The serial nature of ambulatory data also allows identification of the specific time points when abrupt and statistically significant level shifts occur that may be associated with psychological or other events. Change point analysis is an approach adapted from statistical control theory that allows the identification of such time points (Basseville and Nikiforov, 1993). By imposing different constraints on the change point parameters, the examination of various kinds of hypotheses on the relation of specific events and preceding, concomitant, or subsequent changes in other variables is possible. A recent study illustrates the application of this analysis for detecting discrete changes in 24-hr ambulatory psychophysiological data leading up to naturally occurring panic attacks (Rosenfield et al., 2010).

5.2.6. Estimating causality

Because of the serial nature of measurements, examination of relationships for temporal concordance (e.g., good mood is related to concurrently being more talkative) may be insufficient when causal or stochastic processes come into play (e.g., increase in bad mood may predict an episode of binge-eating). Special techniques of time-series analysis, e.g., Granger causality analysis (Granger, 1969), provide relations between variables that may have different time courses but nonetheless be predictive (although this relationship is, of course, inherently correlational in nature–a causal relation only provable by conducting experimental laboratory research or field experiments).

5.2.7. Controlling for physical activity variation and other confounding variables

As previously mentioned, a major problem in the interpretation of physiological ambulatory data is that time- or situation-dependent confounders can significantly affect primary dependent variables such as heart rate or skin conductance. Variation in physical activity and posture, social interaction and ingestion across the assessment can mask more subtle emotion effects on dependent variables. Even worse, their influence may be misattributed as emotion effects. Several strategies have been described to adjust for the main confound physical activity. Some older studies relied on self-reported physical activity (every hour or so) and excluded physiological data from the analysis for times when physical activity was reported. With modern recording devices, controlling for physical activity level can more accurately be performed by stratification based upon accelerometry data. Two recent studies, using similar but not identical strategies, stratify physiological variables on the basis of levels of accelerometry activity. Grossman et al. (2008) used bins based on quintiles of activity (five), derived for each individual subject, and time of day (four) during awake hours to make averages of physiological parameters (four times of day by five quintiles of activity is equal to 20 mean values for each physiological variable). Levels of activity were, thereafter, examined to determine whether target groups differed overall or at specific times of day. Provided groups of interest do not vary, any group differences in physiological parameters can be assumed to be independent of activity. Should there be differences, covariance analyses can be applied to control for these to some extent, assuming that relationships are linear and subjects do not differ much in overall variance of activity levels across the day. The other approach also stratifies by time of day and activity (Pfaltz et al., in press). However, it bases its activity strata on absolute threshold levels for six different normative intensities of activity identified using an initial calibration study with the specific device (sedentary, minimal movement, slow/intermediate/fast walking, running). Thus, for example, the lowest level of activity occurs at all those values below the same specific accelerometry threshold for all subjects. The advantage of this more normative approach is that variation in physiological activation can be assigned to bins of predefined and interpretable activity levels. The disadvantage is that some higher levels of activity may be missing for individual subjects or may only rarely occur, requiring more advanced statistics.

For ambulatory heart rate data, the already briefly mentioned quantification of "additional heart rate" has been used to adjust for the confounding effects of activity. This has been based on accelerometry data and an algorithm that identifies heart rate increases disproportional to changes in accelerometry levels (Myrtek et al., 2005). Wilhelm and Roth (1998c) have argued that minute ventilation is a better index of physical workload, since it is a more direct measure of metabolic activity under most circumstances. Figure 4 illustrates the close coupling of heart rate and minute ventilation during daily activities, as well as an emotional phase when heart rate is clearly out of proportion to minute ventilation. Minute ventilation is more difficult to measure than accelerometry but is clearly superior. For example, ascending vs. descending a flight of stairs will produce similar levels in the accelerometry recording, while minute ventilation will accurately distinguish varying metabolic demands (and effects on heart rate). Wilhelm and Roth (1998c) validated this measure as an index of anxiety-related arousal among flying phobia patients as they entered an airplane. The relationship between minute ventilation and heart rate was determined using linear regression analysis during a physical exercise calibration in the laboratory with stepwise workload increases for each individual. The individual regression equation was used to predict heart rate based on minute ventilation during the ambulatory assessment. Results showed that self-reported anxiety mirrored additional heart rate but not conventional heart rate. Thus, this method served to unmask the emotional component of heart rate.

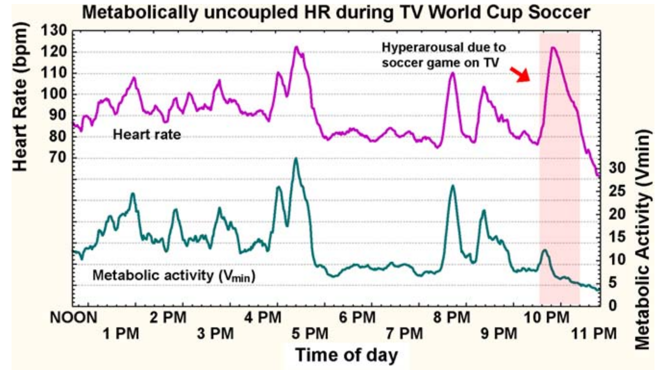

Fig. 4. Metabolically uncoupled heart rate during viewing a World Cup soccer game on TV.

It is unknown to what degree other psychophysiological variables can be adjusted using this method. One prerequisite is that the relationship does not show bottom or floor effects, and the nature of the relationship between measures is close and mathematically characterizable (e.g., a linear or logarithmic relationship). Many cardiovascular variables probably follow this pattern, for example respiratory sinus arrhythmia, an index of cardiac vagal activity (Grossman et al., 2004). However, studies demonstrating the appropriateness of this adjustment method are currently lacking for variables other than heart rate.

A more general strategy for accounting for confounding variables, such as activity, posture, speaking, or meal intake, is to include them into a mixed effects model analysis. For example, when examining the effects of experienced anxiety on heart rate during a week of recording, the confounding variables can be included in the model, thus permitting estimation of their effects.

6. Summary and outlook into the future

In this article, we have attempted to summarize and describe a wide range of arguments in support of increased use of ambulatory assessment in emotion research, as well as provide a number of design options that may facilitate successful ambulatory research. It is becoming increasingly obvious that naturalistic studies employing ambulatory monitoring techniques can address essential and often ignored issues in emotion research. Ambulatory research strategies have the potential to go beyond the multiple laboratory confounds we describe. Furthermore, only an ambulatory strategy is capable of determining the degree of stability and representativeness of emotional reaction patterns in daily life. Nevertheless, ambulatory assessment until recently has been seldom employed, probably due to common perceptions about pitfalls and challenges of this approach. Particularly, physiological ambulatory assessment does pose special technical difficulties, and sophisticated analysis software is often needed to process the large amount of data. It is also true that participants are typically unsupervised and move freely during monitoring, and this can lead to measurement artifacts and metabolically-related changes that are easily confused with emotion-induced alterations. Nevertheless, we have presented ways to address these obstacles, and we believe that good procedures and technologies are currently available and relatively easy to employ for any serious scientist intent upon investigating emotion.

How important is ambulatory assessment for emotion research? Although the purpose of this article was not to review all ambulatory studies relevant to emotion research, we have provided a number of examples that highlight the unique information gained by ambulatory data. Other ambulatory data clearly underline the value of this approach (Halligan et al., 2006; Heppner et al., 2008; Kuppens et al., 2007; Lopes et al., 2004; Meuret et al., 2001; Meuret et al., 2008). Ambulatory assessment methods provide fresh perspectives on the psychological and biological bases of emotion, appraisal, and behavior among healthy people and those with a variety of mental, physical and psychosomatic disorders. To date, ambulatory psychophysiological research has largely been applied to clinical questions, and has already provided for improved understanding of as number of disorders. We believe that the time is ripe also to apply it to basic research questions on emotion, motivation, attention and appraisal. The increasing availability of inexpensive yet sophisticated measurement systems, as well as novel research ideas and designs, is likely to inform fundamental questions unamenable to laboratory investigation. Such a development would surely contribute to a sounder foundation for understanding human emotion rooted in real life.

Although not all emotion research questions can be addressed by the ambulatory paradigm, it is easy to predict that ambulatory studies will see greatly increased interest in the next decades. For one, technologies for data recording have become remarkably powerful, yet inexpensive and miniaturized, thus providing a lot of "bang for the buck." In the months and years ahead, new miniaturized biological sensors of interest to emotion researchers will become more and more available. Analysis strategies and software for coming to terms with the complex data stream have also evolved in the past years and will continue to do so, as more researchers and engineers put their intelligence to work. Furthermore, the multidisciplinary *Society for Ambulatory Assessment* (http://www.ambulatory-assessment.org/) has just been founded, an organization devoted to developing a platform of exchange for interested researchers, as well as a vehicle for building up a critical mass of knowledge and providing guidelines for standardization.

Medical applications, in the future, will increasingly utilize the potential of ambulatory technologies for diagnosis and patient monitoring. The ability to monitor stress responses and emotional activation may also become of greater interest to the medical community in their own search for understanding, treating and managing disease. Thus, a cross-fertilization of medicine and behavioral science may be expected. It is additionally likely that physiological monitors and their user interfaces will be made so non-intrusive and useful that they become part of normal life, first in certain disease conditions (e.g., in acute diabetes, epilepsy, cardiac diseases, where efficient monitors already exist or are under development), then among early adopters (e.g., athletes, in order to increase their performance), and finally among the general consumer. Development of such applications may further provide new possibilities for emotion research. In sum, we believe there to be a convincing case for an ecologically relevant assessment perspective, with respect to the study of emotion, which approaches biopsychosocial processes as they naturally unfold in the real world.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

Preparation of this manuscript was supported by grant 105311-105850 (FW) from the Swiss National Science Foundation, the EC 6th Framework Project EUCLOCK (No. 018741) (FW), Basel Scientific Society (FW), grant KLS-02038-02-2007 from the Swiss Cancer League (PG) and grant NCT00106275 from the Samueli Foundation (PG). We would like to thank Andrea Meyer for advice on statistical issues.

References

Alpers, G.W., Wilhelm, F.H., Roth, W.T., 2005. Psychophysiological assessment during exposure in driving phobic patients. Journal of Abnormal Psychology 114 (1), 126–139.

Alpers, G.W., Abelson, J.L., Wilhelm, F.H., Roth, W.T., 2003. Salivary cortisol response to exposure treatment in driving phobics. Psychosomatic Medicine 65 (4), 679–687.

Baumeister, R.F., Vohs, K.D., Funder, D.C., 2007. Psychology as the science of self-reports and finger movements. Perspectives on Psychological Science 2 (4), 396–403.

Blechert, J., Michael, T., Grossman, P., Lajtman, M., Wilhelm, F.H., 2007. Autonomic and respiratory characteristics of posttraumatic stress disorder and panic disorder. Psychosomatic Medicine 69 (9), 935–943.

Basseville, M., Nikiforov, I., 1993. Detection of Abrupt Changes: Theory and Application. Prentice-Hall, Englewood Cliffs, NJ.

Boivin, D.B., Czeisler, C.A., Dijk, D.J., Duffy, J.F., Folkard, S., Minors, D.S., et al., 1997. Complex interaction of the sleep/wake cycle and circadian phase modulates mood in healthy subjects. Archives of General Psychiatry 54 (2), 145–152.

Bolger, N., Davis, A., Rafaeli, E., 2003. Diary methods: capturing life as it is lived. Annual Review of Psychology 54, 579–616.

Bradley, M.M., Lang, P.J., 2000. Affective reactions to acoustic stimuli. Psychophysiology 37 (2), 204–215.

Brunswik, E., 1955. Representative design and probabilistic theory in a functional psychology. Psychological Review 62 (3), 193–217.

Bussmann, J.B., Stam, H.J., 1998. Techniques for measurement and assessment of mobility in rehabilitation: a theoretical approach. Clinical Rehabilitation 12 (6), 455–464.

Butler, E.A., Egloff, B., Wilhelm, F.H., Smith, N.C., Erickson, E.A., Gross, J.J., 2003. The social consequences of expressive suppression. Emotion 3 (1), 48–67.

Carney, R.M., Blumenthal, J.A., Stein, P.K., Watkins, L., Catellier, D., Berkman, L.F., et al., 2001. Depression, heart rate variability, and acute myocardial infarction. Circulation 104 (17), 2024–2028.

Christie, I.C., Friedman, B.H., 2004. Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach. International Journal of Psychophysiology 51 (2), 143–153.

Coldwell, S.E., Wilhelm, F.H., Milgrom, P., Prall, C.W., Getz, T., Spadafora, A., et al., 2007. Combining alprazolam with systematic desensitization therapy for dental injection phobia. Journal of Anxiety Disorders 21 (7), 871–887.

Conner, T., Barrett, L.F., 2005. Implicit self-attitudes predict spontaneous affect in daily life. Emotion 5 (4), 476–488.

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.

Csikszentmihalyi, M., Larson, R., 1987. Validity and reliability of the Experience-Sampling Method. Journal of Nervous and Mental Disease 175 (9), 526–536.

Dale, O., Hagen, K.B., 2007. Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. Journal of Clinical Epidemiology 60 (1), 8–17.

Dhami, M.K., Hertwig, R., Hoffrage, U., 2004. The role of representative design in an ecological approach to cognition. Psychological Bulletin 130 (6), 959–988.

Dijk, D.J., Duffy, J.F., Czeisler, C.A., 1992. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. Journal of Sleep Research 1 (2), 112–117.

Ebner-Priemer, U.W., Eid, M., Kleindienst, N., Stabenow, S., Trull, T.J., 2009. Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. Journal of Abnormal Psychology 118 (1), 195–202.

Ebner-Priemer, U.W., Kubiak, T., 2007. Psychological and psychophysiological ambulatory monitoring: a review of hardware and software solutions. European Journal of Psychological Assessment 23 (4), 238–247.

Ebner-Priemer, U.W., Kuo, J., Kleindienst, N., Welch, S.S., Reisch, T., Reinhard, I., et al., 2007. State affective instability in borderline personality disorder assessed by ambulatory monitoring. Psychological Medicine 37 (7), 961–970.

Ebner-Priemer, U.W., Kuo, J., Schlotz, W., Kleindienst, N., Rosenthal, M.Z., Detterer, L., et al., 2008. Distress and affective dysregulation in patients with borderline personality disorder: a psychophysiological ambulatory monitoring study. Journal of Nervous and Mental Disease 196 (4), 314–320.

Ekman, P., Friesen, W., 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto.

Fagan, T.C., Conrad, K.A., Mar, J.H., Nelson, L., 1986. Effects of meals on hemodynamics: implications for antihypertensive drug studies. Clinical Pharmacology & Therapeutics 39 (3), 255–260.

Fahrenberg, J., 2001. Progress in Ambulatory Assessment. Hogrefe & Huber, Gottingen.

Fahrenberg, J., Foerster, F., Smeja, M., Muller, W., 1997. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. Psychophysiology 34 (5), 607–612.

Fahrenberg, J., Myrtek, M., 1996. Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies. Hogrefe & Huber Publishers, Gottingen, Germany.

Fahrenberg, J., Myrtek, M., Pawlik, K., Perrez, M., 2007. Ambulatory assessment-monitoring behavior in daily life settings: a behavioral-scientific challenge for psychology. European Journal of Psychological Assessment 23 (4), 206–213.

Friedman, B.H., 2010. Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. Biological Psychology 84, 383–393.

Friedman, B.H., Santucci, A.K., 2003. Idiodynamic profiles of cardiovascular activity: a P-technique approach. Integrative Physiological and Behavioural Science 38 (4), 295–315.

Garcia, S.F., Cella, D., Clauser, S.B., Flynn, K.E., Lad, T., Lai, J.S., et al., 2007. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. Journal of Clinical Oncology 25 (32), 5106–5112.

Garmezy, N., 1982. The case for the single case in research. In: Fiske, D.W. (Series Ed,), Kazdin, A.E., Tuma, A.H. (Vol. Eds.), New Directions For Methodology Of Social And Behavioral Sciences: Single-Case Research Designs, vol. 13. Jossey-Bass: San Francisco, pp. 5–17.

Garssen, B., Buikhuisen, M., van Dyck, R., 1996. Hyperventilation and panic attacks. American Journal of Psychiatry 153 (4), 513–518.

Gemmell, J., Bell, G., Lueder, R., 2006. MyLifeBits: a personal database for everything. Communications of the ACM 49 (1), 88–95.

Gerin, W., Christenfeld, N., Pieper, C., De, R.D., Su, O., Stroessner, S., et al., 1998. The generalizability of cardiovascular responses across settings. Journal of Psychosomatic Research 44 (2), 209–218.

Giese-Davis, J., Wilhelm, F.H., Conrad, A., Abercrombie, H.C., Sephton, S., Yutsis, M., et al., 2006. Depression and stress reactivity in metastatic breast cancer. Psychosomatic Medicine 68 (5), 675–683.

Goedhart, A.D., Kupper, N., Willemsen, G., Boomsma, D.I., de Geus, E.J., 2006. Temporal stability of ambulatory stroke volume and cardiac output measured by impedance cardiography. Biological Psychology 72 (1), 110–117.

Goodwin, M.S., Velicer, W.F., Intille, S.S., 2008. Telemetric monitoring in the behaviour sciences. Behavior Research Methods 40 (1), 328–341.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.

Greiner, W., 2001. Quantum Mechanics: An Introduction. Springer.

Gross, J.J., John, O.P., 2003. Individual differences in two emotion-regulation processes: implications for affect, relationships, and well-being. Journal of Personality and Social Psychology 85 (2), 348–362.

Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. Cognition & Emotion 9 (1), 87–108.

Gross, J.J., Thompson, R.A., 2007. Emotion regulation: Conceptual foundations. In: Gross, J.J. (Ed.), Handbook of Emotion Regulation. Guilford Press, New York, USA, pp. 3–24.

Grossman, P., Deuring, G., Garland, S.N., Campbell, T.S., Carlson, L.E., 2008. Patterns of objective physical functioning and perception of mood and fatigue in posttreatment breast cancer patients and healthy controls: an ambulatory psychophysiological investigation. Psychosomatic Medicine 70 (7), 819–828.

Grossman, P., Karemaker, J., 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., Spoerle, M., Wilhelm, F.H., 2006. Reliability of respiratory tidal volume estimation by means of ambulatory inductive plethysmography. Biomedical Sciences Instrumentation 42, 193–198.

Grossman, P., Wilhelm, F.H., Brutsche, M. (in press). Accuracy of ventilatory measurement employing ambulatory inductive plethysmography during tasks of everyday life. Biological Psychology.

Grossman, P., Wilhelm, F.H., Kawachi, I., Sparrow, D., 2001. Gender differences in psychophysiological responses to speech stress among older social phobics: congruence and incongruence between self-evaluative and cardiovascular reactions. Psychosomatic Medicine 63 (5), 765–777.

Grossman, P., Wilhelm, F.H., Spoerle, M., 2004. Respiratory sinus arrhythmia, cardiac vagal control and daily activity. American Journal of Physiology 287 (2), H728–734.

Gunthert, K.C., Conner, T.S., Armeli, S., Tennen, H., Covault, J., Kranzler, H.R., 2007. Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: a daily process approach to gene-environment interaction. Psychosomatic Medicine 69 (8), 762–768.

Halligan, S.L., Michael, T., Wilhelm, F.H., Clark, D.M., Ehlers, A., 2006. Reduced heart rate responding to trauma reliving in trauma survivors with PTSD: correlates and consequences. Journal of Traumatic Stress 19 (5), 721–734.

Hammond, K.R., 1998. Ecological validity: then and now. Retrieved March 8, 2009, from http://www.albany.edu/cpr/brunswik/notes/essay2.html.

Haynes, S.N., Yoshioka, D.T., 2007. Clinical assessment applications of ambulatory biosensors. Psychological Assessment 19 (1), 44–57.

Healey, J., Picard, R.W., 1998. StartleCam: A cybernetic wearable camera. Paper presented at the International Symposium on Wearable Computers, Pittsburgh, PA.

Heppner, W.L., Kernis, M.H., Nezlek, J.B., Foster, J., Lakey, C.E., Goldman, B.M., 2008. Within-person relationships among daily self-esteem, need satisfaction, and authenticity. Psychological Science 19 (11), 1140–1145.

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., 2006. SenseCam: a Retrospective Memory Aid. Paper presented at the 8th International Conference on Ubicomp, Orange County, CA.

Hufford, M.R., Shields, A.L., Shiffman, S., Paty, J., Balabanis, M., 2002. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychology of Addictive Behaviors 16 (3), 205–211.

Ickes, W., Snyder, M., Garcia, S., 1997. Personality influences on the choice of situations. In: Hogan, R., Johnson, J., Briggs, S. (Eds.), Handbook of Personality Psychology. Academic Press, London, UK, pp. 165–195.

Kahneman, D., 2000. Evaluation by moments: past and future. In: Kahneman, D., Tversky, A. (Eds.), Choices, Values and Frames. Cambridge University Press and the Russell Sage Foundation, New York, pp. 693–708.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A., 2004. A survey method for characterizing daily life experience: the day reconstruction method. Science 306 (5702), 1776–1780.

Kamarck, T.W., Schwartz, J.E., Janicki, D.L., Shiffman, S., Raynor, D.A., 2003. Correspondence between laboratory and ambulatory measures of cardiovascular reactivity: a multilevel modeling approach. Psychophysiology 40 (5), 675–683.

Kazdin, A.E., 1974. Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. Journal of Consulting and Clinical Psychology 42 (5), 704–716.

Kirschbaum, C., Pirke, K., Hellhammer, D., 1993. The 'Trier Social Stress Test': a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28 (1–2), 76–81.

Klein, D.F., 1993. False suffocation alarms, spontaneous panics, and related conditions. An integrative hypothesis. Archives of General Psychiatry 50, 306–317.

Kolodyazhniy, V., Kreibig, S.D., Gross, J.J., Roth, W.T., Wilhelm, F.H., 2010. Psychophysiological sensing in affective computing: towards subject-independent and stimulus-independent classification of film-induced emotions. (Manuscript submitted for publication).

Kraemer, H., Gullion, C., Rush, A., Frank, E., Kupfer, D., 1994. Can state and trait variables be disentangled? A methodological framework for psychiatric disorders. Psychiatry Research 52 (1), 55–69.

Kraemer, H.C., Giese-Davis, J., Yutsis, M., O'Hara, R., Neri, E., Gallagher-Thompson, D., et al., 2006. Design decisions to optimize reliability of daytime cortisol slopes in an older population. American Journal of Geriatric Psychiatry 14 (4), 325–333.

Kreibig, S.D., 2010. Autonomic nervous system activity in emotion: a review. Biological Psychology 84, 394–421.

Kreibig, S.D., Wilhelm, F.H., Roth, W.T., Gross, J.J., 2007. Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. Psychophysiology 44 (5), 787–806.

Kuppens, P., Van Mechelen, I., Nezlek, J.B., Dossche, D., Timmermans, T., 2007. Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion 7 (2), 262–274.

Lang, P.J., Bradley, M., 2010. Emotion and the motivational brain. Biological Psychology 84, 437–450.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1990. Emotion, attention, and the startle reflex. Psychological Review 97 (3), 377–395.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1995. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. Gainesville, FL: University of Florida, Department of Psychology, Center of Research in Psychophysiology.

Lang, P.J., Levin, D.N., Miller, G.A., Kozak, M.J., 1983. Fear behavior, fear imagery, and the psychophysiology of emotion: the problem of affective response integration. Journal of Abnormal Psychology 92 (3), 276–306.

Ley, R., 1985. Blood, breath, and fears: A hyperventilation theory of panic attacks and agoraphobia. Clinical Psychology Review 5, 271–285.

Linden, W., 1987. A microanalysis of autonomic activity during human speech. Psychosomatic Medicine 49, 562–578.

Lopes, P.N., Brackett, M.A., Nezlek, J.B., Schutz, A., Sellin, I., Salovey, P., 2004. Emotional intelligence and social interaction. Personality and Social Psychology Bulletin 30 (8), 1018–1034.

Margraf, J., Taylor, C.B., Ehlers, A., Roth, W.T., Agras, W.S., 1987. Panic attacks in the natural environment. Journal of Nervous and Mental Disease 175, 558–565.

Mauss, I.B., Evers, C., Wilhelm, F.H., Gross, J.J., 2006. How to bite your tongue without blowing your top: implicit evaluation of emotion regulation predicts affective responding to anger provocation. Personality and Social Psychology Bulletin 32 (5), 589–602.

Mehl, M.R., Gosling, S.D., Pennebaker, J.W., 2006. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. Journal of Personality and Social Psychology 90 (5), 862–877.

Mehl, M.R., Pennebaker, J.W., 2003. The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. Journal of Personality and Social Psychology 84 (4), 857–870.

Mehl, M.R., Pennebaker, J.W., Crow, D.M., Dabbs, J., Price, J.H., 2001. The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. Behavior Research Methods, Instruments & Computers 33 (4), 517–523.

Meuret, A.E., Wilhelm, F.H., Roth, W.T., 2001. Respiratory biofeedback-assisted therapy in panic disorder. Behavior Modification 25 (4), 584–605.

Meuret, A.E., Wilhelm, F.H., Ritz, T., Roth, W.T., 2008. Feedback of end-tidal pCO2 as a therapeutic approach for panic disorder. Journal of Psychiatric Research 42 (7), 560–568.

Munsch, S., Meyer, A.H., Milenkovic, N., Schlup, B., Margraf, J., Wilhelm, F.H., 2009. Ecological momentary assessment to evaluate cognitive-behavioral treatment for binge-eating disorder. International Journal of Eating Disorders 42 (7), 648–657.

Myrtek, M., 1984. Constitutional Psychophysiology. Academic Press, Orlando, FL.

Myrtek, M., Aschenbrenner, E., Brugner, G., 2005. Emotions in everyday life: an ambulatory monitoring study with female students. Biological Psychology 68 (3), 237–255.

Nezlek, J.B., Vansteelandt, K., Van Mechelen, I., Kuppens, P., 2008. Appraisal-emotion relationships in daily life. Emotion 8 (1), 145–150.

Pawlik, K., Buse, L., 1996. Ambulatory behavioral assessment and in-field performance testing. In: Fahrenberg, J., Myrtek, M. (Eds.), Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies. Hogrefe & Huber Publishers, Gottingen, Germany, pp. 29–50.

Pennebaker, J.W., Mehl, M.R., Niederhoffer, K.G., 2003. Psychological aspects of natural language use: our words, our selves. Annual Review of Psychology 54, 547–577.

Pfaltz, M.C., Michael, T., Grossman, P., Blechert, J., Wilhelm, F.H., 2009. Respiratory pathophysiology of panic disorder: an ambulatory monitoring study. Psychosomatic Medicine 71 (8), 869–876.

Pfaltz, M.C., Michael, T., Grossman, P., Margraf, J., Wilhelm, F.H., 2010. Variability of symptoms in daily life of patients with panic disorder and patients with posttraumatic stress disorder. (Manuscript submitted for publication).

Pfaltz, M.C., Grossman, P., Michael, T., Wilhelm, F.H. (in press). Physical activity and respiratory behavior in daily life of patients with panic disorder and healthy controls. International Journal of Psychophysiology.

Picard, R.W., 1997. Affective Computing. MIT Press, Cambridge, MA.

Picard, R.W., Vyzas, E., Healey, J., 2001. Toward machine emotional intelligence: analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (10), 1175–1191.

Pohl, R.F., 2004. Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement, and Memory. Psychology Press, New York.

Rau, R., Georgiades, A., Fredrikson, M., Lemne, C., de Faire, U., 2001. Psychosocial work characteristics and perceived control in relation to cardiovascular rewind at night. Journal of Occupational Health Psychology 6 (3), 171–181.

Redelmeier, D.A., Kahneman, D., 1996. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 66 (1), 3–8.

Rosenfield, D., Zhou, E., Wilhelm, F.H., Conrad, A., Roth, W.T., Meuret, A.E., 2010. Change point analysis for longitudinal physiological data: detection of cardiorespiratory changes preceding panic attacks. (Manuscript submitted for publication).

Roth, W.T., Wilhelm, F.H., Pettit, D., 2005. Are current theories of panic falsifiable? Psychological Bulletin 131 (2), 171–192.

Rottenberg, J., Gross, J.J., Wilhelm, F.H., Najmi, S., Gotlib, I.H., 2002. Crying threshold and intensity in major depressive disorder. Journal of Abnormal Psychology 111 (2), 302–312.

Rozin, P., 2009. What kind of empirical research should we publish, fund, and reward? A different perspective. Perspectives on Psychological Science 4 (4), 435–439.

Sakamoto, N., Yoshiuchi, K., Kikuchi, H., Takimoto, Y., Kaiya, H., Kumano, H., et al., 2008. Panic disorder and locomotor activity. Biopsychosocial Medicine 2, 23.

Schwartz, C.E., Bode, R., Repucci, N., Becker, J., Sprangers, M.A., Fayers, P.M., 2006. The clinical significance of adaptation to changing health: a meta-analysis of response shift. Quality of Life Research 15 (9), 1533–1550.

Schwartz, J.E., Stone, A.A., 1998. Strategies for analyzing ecological momentary assessment data. Health Psychology 17 (1), 6–16.

Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annual Review of Clinical Psychology 4, 1–32.

Singer, J.D., Willett, J.B., 2003. Applied Longitudinal Data Analysis. Oxford University Press, Oxford, UK.

So, K., Adamson, T.M., Horne, R.S., 2007. The use of actigraphy for assessment of the development of sleep/wake patterns in infants during the first 12 months of life. Journal of Sleep Research 16 (2), 181–187.

Solhan, M.B., Trull, T.J., Jahng, S., Wood, P.K., 2009. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychological Assessment 21 (3), 425–436.

Stanley, N., 2003. Actigraphy in human psychopharmacology: a review. Human Psychopharmacology 18 (1), 39–49.

Stein, K.F., Corte, C.M., 2003. Ecologic momentary assessment of eating-disordered behaviors. International Journal of Eating Disorders 34 (3), 349–360.

Stephens, C.L., Christie, I.C., Friedman, B.H., 2010. Autonomic specificity of basic emotions: evidence from pattern classification and cluster analysis. Biological Psychology 84, 463–473.

Stone, A.A., Broderick, J.E., Shiffman, S.S., Schwartz, J.E., 2004. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. Pain 107 (1–2), 61–69.

Stone, A.A., Shiffman, S., Schwartz, J.E., Broderick, J.E., Hufford, M.R., 2003. Patient compliance with paper and electronic diaries. Controlled Clinical Trials 24 (2), 182–199.

Teeters, A.C., 2007. Use of a Wearable Camera System in Conversation: Toward a Companion Tool for Social-Emotional Learning in Autism. Massachusetts Institute of Technology, Cambridge, MA.

Teicher, M., 1995. Actigraphy and motion analysis: new tools for psychiatry. Harvard Review of Psychiatry 3 (1), p18–35.

Todder, D., Caliskan, S., Baune, B.T., 2009. Longitudinal changes of day-time and night-time gross motor activity in clinical responders and non-responders of major depression. World Journal of Biological Psychiatry 10 (4), 276–284.

Tryon, W.W., Tryon, G.S., Kazlausky, T., Gruen, W., Swanson, J.M., 2006. Reducing hyperactivity with a feedback actigraph: initial findings. Clinical Child Psychology and Psychiatry 11 (*4*), 607–617.

van Eekelen, A.P., Houtveen, J.H., Kerkhof, G.A., 2004. Circadian variation in base rate measures of cardiac autonomic activity. European Journal of Applied Physiology 93 (1–2), 39–46.

Verdecchia, P., Angeli, F., Gattobigio, R., 2004. Clinical usefulness of ambulatory blood pressure monitoring. Journal of the American Society of Nephrology 15 (Suppl. 1), S30–33.

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.

Walls, T.A., Schafer, J.L. (Eds.), 2006. Intensive Longitudinal Data. Oxford University Press, Oxford, UK.

Wilhelm, F.H., Handke, E.M., Roth, W.T., 2003a. Detection of speaking with a new respiratory inductive plethysmography system. Biomedical Sciences Instrumentation 39, 136–141.

Wilhelm, F.H., Pfaltz, M.C., Grossman, P., 2006a. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interacting with Computers 18 (2), 171–186.

Wilhelm, F.H., Pfaltz, M.C., Grossman, P., Roth, W.T., 2006b. Distinguishing emotional from physical activation in ambulatory psychophysiological monitoring. Biomedical Sciences Instrumentation 42, 458–463.

Wilhelm, F.H., Roth, W.T., 1996. Ambulatory assessment of clinical anxiety. In: Fahrenberg, J., Myrtek, M. (Eds.), Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies. Hogrefe & Huber Publishers, Gottingen, Germany, pp. 317–345.

Wilhelm, F.H., Roth, W.T., 1997. Acute and delayed effects of alprazolam on flight phobics during exposure. Behaviour Research and Therapy 35, 831–841.

Wilhelm, F.H., Roth, W.T., 1998a. Taking the laboratory to the skies: ambulatory assessment of self- report, autonomic, and respiratory responses in flying phobia. Psychophysiology 35 (5), 596–606.

Wilhelm, F.H., Roth, W.T., 1998b. Trusting computerized data reduction too much: a critique of Anderson's ambulatory respiratory monitor. Biological Psychology 49 (1–2), 215–219 (discussion 221–212).

Wilhelm, F.H., Roth, W.T., 1998c. Using minute ventilation for ambulatory estimation of additional heart rate. Biological Psychology 49 (1–2), 137–150.

Wilhelm, F.H., Gevirtz, R., Roth, W.T., 2001. Respiratory dysregulation in anxiety, functional cardiac, and pain disorders: assessment, phenomenology, and treatment. Behavior Modification 25 (4), 513–545.

Wilhelm, F.H., Roth, W.T., 2001. The somatic symptom paradox in DSM-IV anxiety disorders: suggestions for a clinical focus in psychophysiology. Biological Psychology 57 (1–3), 105–140.

Wilhelm, F.H., Roth, W.T., Sackner, M.A., 2003b. The LifeShirt. An advanced system for ambulatory measurement of respiratory and cardiac function. Behavior Modification 27 (5), 671–691.

Wilhelm, P., Perrez, M., 2004. How is my partner feeling in different daily-life settings? Accuracy of spouses' judgements about their partner's feelings at work and at home. Social Indicators Research 67 (1–2), 183–246.