The London School of Economics and Political Science

Towards situated anxiety: Taking an experience-sampling approach to exploring the influence of situation and individual tendency on momentary stress

Adam John Barrett Christmas

A thesis submitted to the Department of Methodology of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, February 2022

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

(a) The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. In accordance with the Regulations, I have deposited an electronic copy of it in LSE Theses Online held by the British Library of Political and Economic Science and have granted permission for my thesis to be made available for public reference. Otherwise, this thesis may not be reproduced without my prior written consent.

(b) I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of **52,163** words.

Acknowledgements

I would not have been able to do any of this without the support of family, friends, and colleagues. Special thanks go to my supervisors Jouni and Martin, to my wife Marianne, and to my son Jacob. And SB; thanks for all the fish.

Abstract

Research on methods is a vital part of social science, but many desirable techniques come with high outlays. Adapting mobile technology for research offers new possibilities, including the promise of reduced costs, and access to hard-to-reach groups. However, these claims have been difficult to verify empirically.

This thesis contributes to the literature in two fields. The first is psychology, where I investigate the parts and processes of experience formation in states of stress. Although there have been momentary studies of stress in everyday life, relatively few have looked at the interactions between dynamic and stable influences using the same instrument. I also address the comparability and generalisability of situations across persons, and the function of trait.

My studies also make a methodological contribution to the field of experiential data measurement. The approach is based on experience-sampling, a well-established but traditionally hard-to-implement technique for assessing events as they occur. I describe the process of developing, administering, and evaluating a mobile version through case studies on divers populations.

Throughout the studies, I examine whether technical advances can compensate for the increased burden of intensive self-report. I look at implications for researchers interested in working with experience sampling, and I aim to broaden access to such methods by setting out practical guidelines. I encourage researchers to consider the suitability of such methods for their work by demonstrating usability and flexibility.

To do this, I carry out two exploratory studies, two multiple-subject substantive studies, and one single subject investigation. The evidence presented builds on research on state and temperament, and supports the idea that careful use of mobile technology can improve formerly cumbersome techniques, and apply them to varied populations. It demonstrates that adding momentary and situational information need not add substantially to costs. Finally, it implies cautions and recommendations for future development.

Contents

List of Tables

List of Figures

Introduction and literature review

Introduction

Retrospective data is often influenced by hindsight biases like self-image, self-beliefs, and mis-estimations of intensity (Robinson & Clore, 2002; Schwartz, 2012; Verduyn, Mechelen, et al., 2009). Researchers can allow for these issues in analysis, where this is appropriate; or, they can try to avoid them by taking measurements in the moment (Shiffman et al., 2008; A. Stone & Shiffman, 2002). In this thesis and the concurrent individual studies I adopt the second approach, focussing on experience-sampling as a method of collecting direct observations about stressful experience. I also develop and examine methods of doing such measurement.

Stress has been recognised as a significant influence on health since the first half of the 20th century (Costa, 1996). This has made stress and anxiety a natural area of study for social scientists. Over the years, researchers have discovered that stress and anxiety not only impacts on long-term quality of life, but has a significant effect on moment to moment states of experience as well. Because of this importance, the number and scope of investigations of stress by government and academia are substantial (Szabo et al., 2012).

Statistics from the UK, from before the pandemic, emphasise these concerns. Work-related stress accounted for 40% of work-related illnesses, or 526,000 cases, in 2016/2017 (Health and Safety Executive, 2017). This is "49% of all working days lost due to ill health" (Health and Safety Executive, 2017), and 40,000 more cases than just three years earlier (Health and Safety Executive, 2014). It caused affected workers to miss an average of 23.8 working days on sick leave, 12.5 million in total. What is more, according to the Office of National Statistics (ONS) 'Personal Well-being' dataset from the 2017 Annual Population Survey (APS), one in five (20.09%) of UK adults rated their general anxiety as 'high' or above in 2017 (Office for National Statistics, 2017a); at least six points on a ten-point scale.

Causes of stress can be purely physical, but in the modern world they are more likely to be social-psychological; a confrontation with a co-worker rather than a brush with a hungry carnivore (Sapolsky, 2004, pp. 4–8). Yet regardless of origin, any experience of stress has physical and psychological consequences. These can be seen through changes in behaviour, such as reducing social interaction, and in physical symptoms: difficulty sleeping, increased risk of heart disease and diabetes. Extreme or repeated stress also has long-term detrimental effects on human development.

So what are the processes that influence states of stress? And why are some people more vulnerable to stress? This chapter investigates how these processes unfold, the history of stress, and variations in individual experience in later sections. This has repercussions for methodological design, and measurement choices.

To begin with, however, I will be looking at the construction of stress, and the foundational concepts of momentary experience. I will use these discoveries to lay the foundation for a model of the processes of experience. I examine factors such as situation and temperament, which are fundamental to variation in states of stress and anxiety. I try to trace these influences on individual differences in experience, and I also explore some of the methodological implications for my studies. After that, I discuss the experience-sampling method in detail, and summarise the work to come in the rest of the thesis.

The thesis is structured as follows: two feasibility studies, the first to check the technology and the development method, the second the deployment. The third and fourth studies were designed to be confirmatory, in an exploratory sort of way, testing the model hypotheses in the same way as the two earlier studies tested the methodological ones. The fifth study places the most emphasis on trait as a moderator of anxiety, and fits a number of hierarchical models as suggested in the later literature review.

Theoretical framework

Stress

What is stress? Research shows that stress manifests in three ways: physiologically, in behaviours, and in individual experience. All three form a feedback loop that may ameliorate or exacerbate anxiety in a given situation.

At the physical level, stress and anxiety has a “well-established” (Hunter et al., 2014) effect on brain functions and cell structure: from changes in sleep quality and duration (Nota & Coles, 2014; Zunhammer et al., 2014), to incidence and intensity of psychotic symptoms (Grammenos & Barker, 2014). Childhood exposure to sustained or repeated stress affects growth and behaviour, which in turn affects regulation of situational reactions across an individual’s entire lifespan (Hunter & McEwen, 2013).

Humans are not entirely unique in suffering from psychological stress, but animals tend not to suffer from uniquely human stressors like ‘repetitive negative thinking’ (Nota & Coles, 2014). For example, Bolger and Amarel (2007) report that just the thought of having to give a speech and be assessed by an audience of peers (an “anticipated speech paradigm") can be used for reliably raising stress levels in experiments.

It has taken considerable effort to reach a general agreement on a standard for conceptualising stress; in fact, to agree on the utility of modelling it at all. The early breakthrough that led to modelling stress as we know it today was made by Dr. Hans Selye (Genter, 2014). His research at McGill University in the 1930s was some of the very first to categorise stress as a measurable variable, formed by responses to stimuli such as environment and activity (Selye, 1976).

Selye concluded that stress should be recognised in its own right, defined by non-specific responses of the body, as a result of his research into endocrine systems (Selye, 1936). He realised that stress was an independent syndrome, a reliable response that was nevertheless produced in reaction to many different stimuli and hence, non-specific to any particular agent (Selye, 1976, pp. 15–56). At the time, stress was considered a "purely hypothetical thing" (Selye, 1976, p. 49) by most scientists. His opponents insisted it was abstract and unobservable; even today, some call it a product of "cultural expectations" (Jones & Bright, 2001, pp. 10–12). But Selye showed that stress responses could be grouped, categorised, and therefore modelled empirically. General stress could be measured. Scientists of the time were focussed on these concepts of action and reaction, stimulus and response (P. Wilhelm et al., 2012), and so by situating stress within this empirical tradition, but widening its implications, Selye laid the groundwork for the concept of stress that we still recognise today.

Despite this advance, early research still persisted in using the 'black box' (Bechtel, 2009) model of experience espoused by behaviourism. Internal, mental processes were dismissed as insignificant in experiences of stress. But increasingly, this analytical focus made it difficult for researchers to understand variations in actual experience: why some situations were found stressful only by certain individuals. In particular, the effects of World War II highlighted shortcomings in the behaviourist approach to stress and anxiety (Genter, 2014). Soldiers failed to respond uniformly, as behaviourism predicted, to the various stimuli of battles and prolonged campaigning. This left many looking for better explanations, and after the war, governmental institutions began to join forces with academics to explore new theories.

By the 1950s and 1960s, therefore, there was a significant decline in the popularity of the behaviourist approach to stress (P. Wilhelm et al., 2012). Researchers came to recognise that interactions between the mind and body - emotions, personality, and the environment - are key in determining individual vulnerability to stressors (Sapolsky, 2004, p. 4). This would see research take a turn to cognition; a return to focus on the "unique contribution of thoughts and feelings to emotional reactions" (Spielberger &

Reheiser, 2009). This ultimately brought about an increased appreciation of the factors involved in explaining individual differences, and led us to the concept of stress and experience most agree upon today.

Defining stress and anxiety

Definitions of stress and anxiety can vary a good deal (Fink, 2016). Selye was convinced that stress reactions were always a "nonspecific response" (McCarty, 2016) to any demand. Cannon countered that specific responses were necessary for adaptation. In a way, both were right.

Current state of stress (the external pressures) and state of anxiety (internal reaction) are balanced; as one changes, so does the other (McCarty, 2016). High stress on the organism results in high anxiety as a reaction, and in turn the subject seeks to reduce the pressure and again achieve equilibrium. However, external influences are not as changeable and so differences in reactivity alleviate or exacerbate feelings of anxiety.

Influences on experiences of stress

In studies of stress in daily life, and more broadly in social science generally, researchers are often interested in states, which are processes within a person, and traits, which reflect stable differences between persons (Hamaker, 2012). Historically, some scientists defined trait-level structural factors and momentary states of experience in opposition (Krohne & Hock, 2011), believing that processes and structure did not interact. In this thesis, it is argued that a framework for explaining individual differences in momentary experiences of stress can only be satisfactorily explained by accounting for factors on both levels. This analysis will first discuss the person-level concept of trait, and explore how we can define its relationship to immediate states of stress and anxiety. Broadly, trait temperament can be said to be a "dispositional signature" (McAdams, 1995), which captures "characteristics of the individual's responsivity to changes in stimulation" (Rothbart & Posner, 2006). For example, "emotional reactivity" and speed of reaction to stimuli are "strongly related" to trait anxiety (Strelau & Zawadzki, 2011).

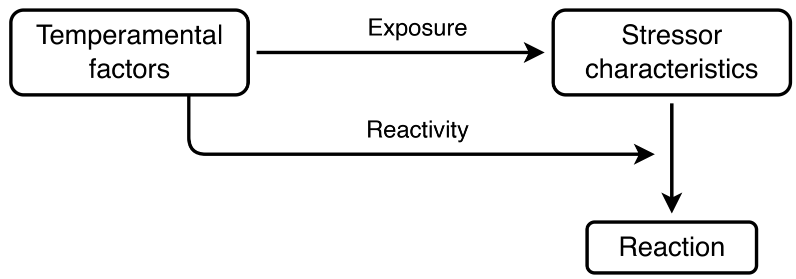

Figure 1: Stress processes. Adapted from Almeida (2005)

Figure 1 shows what a moderating relationship between trait anxiety and state, or momentary, anxiety might look like. The starting point of this framework is based on the widely-accepted view of temperament as a moderator of reactivity (Kagan, 2010; Kagan & Snidman, 2009; Zuckerman, 2012). However, some work will be needed to extend this concept to represent the particular factors of stress and anxiety, with insights from the regulatory theory of temperament (RTT) by Strelau (2010) and others. å

Temperament as a moderator

A moderator, in this model, refers to a factor outside of the event itself that tempers or buffers (Jose, 2013, p. 6) the effect of the event on momentary anxiety. Expanding on this idea of reactivity as a moderator of stress, and therefore an influence on momentary anxiety, this thesis uses a line of research that complements this model of experience formation. The Regulatory Theory of Temperament (RTT) is set out by Strelau (2010), who specifically investigates temperament as a regulator in experiences of stress, defining psychological stress as a state of strong negative emotions such as fear and anxiety (Strelau, 2010, p. 121). His research, therefore, is partly concerned with traits, which he categorises as broad tendencies toward specific behaviours. He also agrees that these tendencies are relatively stable, and change only slightly over time. He argues that we cannot measure traits by reducing the model to behaviour alone, but that we can nevertheless draw inferences from their formal expression in observed behaviours.

Strelau's idea was to "develop a theory of this component of personality in which human behaviour plays the leading role" (Strelau, 2010, p. 48). He outlines more clearly the relationship between temperament and stressful phenomena; these are framed as demands on capacity that may be objective, i.e. physical environment or actions, or subjective, i.e. psychological self-appraisal. His research is based on the observed stability in differences in reaction and behavioural outcome when repeatedly measured across individuals, which agrees with the methodological approach already touched on. While his model places a primary emphasis on reactivity as a moderator in differences in intensity of response, it also includes the concept of external independent stressors arising from the environment. This is a particularly useful addition, in that it allows me to better account for interactions between the individual and the environment.

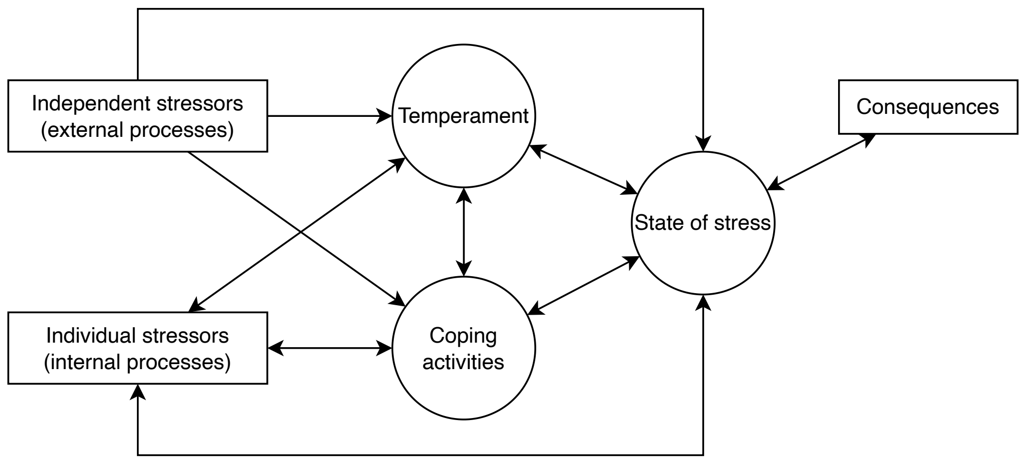

Figure 2: The description

In figure 2, these interactions are more clearly outlined. Emphasis is placed on the reciprocal interchanges between external and internal stressors (i.e. contextual and cognitive events), and temperament and behaviours. It specifically shows how we can frame these stressors, and also their interactions with temperament and activity, to determine the current state of stress and subsequent actions.

This framework suggests that temperament is manifested in both experienced states and behaviours, and that its function remains a regulatory one. In accordance with the earlier discussion, it is seen as moderating the intensity of reactions to situations and experiences.

As we saw, this temperamental mechanism actually serves to increase the effect of stimulation in high-reactive individuals, and consequently raises sensitivity to stressors. Strelau postulates that this may also govern how individuals react in terms of their subsequent behaviour; so when given the choice high reactives may choose to avoid highly stimulating environments like social activities or large gatherings. Van der Doef and Maes (1999) suggest that "misfit" between situation and temperament can intensify the negative effects of high demands, but that some characteristics of personality can also demonstrate a "resistance factor" in such situations.The implication for my studies is that there may be some difference in activities or preferred social contexts.

Separating temperament in the model also implies that there may be some ‘essential’ qualities to experience across the same situations, between individuals. We should be able to see this by comparing repeated measurements at different levels.

Situating experience in its environment

Researchers have long known that people have difficulty with reconstructing experiences (e.g. Gaskell et al., 2000), or predicting their impact (e.g. T. Wilson & Gilbert, 2005). This may be because in recall, we can select and construct our experience the way our self-knowledge tells us it ought to have been (Katz, 2004). Reconstruction is often influenced by factors like desired self-image, or concern for self-presentation (Schwartz, 2012): these are commonly known as hindsight biases (Fahrenberg et al., 2007; Hogarth et al., 2007).

Reconstructed measures may be a misleading guide to the moment to moment experiences of daily life (Kahneman et al., 1993). Therefore, it becomes increasingly difficult to justify using these measurements as a proxy for momentary experience (Shiffman et al., 2008; A. Stone & Shiffman, 2002). Taking measurements in real time can help address some of the concerns with retrospective or ‘domain’ data obtained with interviews or surveys (Gunthert & Wenze, 2012).

As proposed earlier, taking measurements under natural conditions may address this issue by situating data in the ecology of interest in a more representative way (P. Wilhelm et al., 2012). This can enhance the validity of measurements in relation to actual behaviour. Situated cognition is a broadly useful set of approaches, which frame cognitive activities and experience by embedding them within, or extending them into, their physical and behavioural context (R. A. Wilson & Clark, 2009). Situating cognition recognises that processes of experience routinely take advantage of structures in the environment, social or physical, to incorporate contextual factors into experience formation, and direct subsequent behaviours (Robbins & Aydede, 2009). Where exactly the boundary of cognition should be set remains the subject of debate. However, it is generally recognised that there are three main perspectives grouped under the heading of situated cognition: embodiment, embedding, and extension (Robbins & Aydede, 2009). They form a “loose-knit family” (R. A. Wilson & Clark, 2009) of approaches.

Embodiment requires us to recognise that cognition depends on the body as much as the brain. There are parallels with the concept of individual temperament in this thesis, which is partly genetic and embodied. Embedding goes further than this to place an equal emphasis on the environment itself, bringing situation from the background to the foreground of researches. Lastly, the idea of extended cognition, or ‘transcranialism’

(Adams & Aizawa, 2009) is that extended cognitive processes actually take place in the environment themselves; i.e. that "things… undergo cognitive processes" (Robbins & Aydede, 2009).

Taken a step back from transcranialism, this discussion suggests that we need to account for both person and environment to more fully understand experiences of stress and anxiety. And sing the embedded view of experience suggests that a model should be capable of representing the whole interactional 'triad' (Funder, 2006) of person, situation, and behaviour. Figure 3, then, represents the core components I have identified.

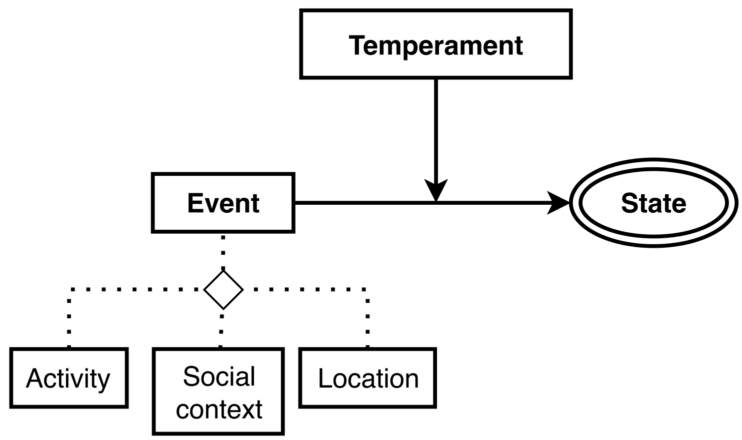

Figure 3: Minimal components for measuring experience

As suggested in other studies (see the later section on ESM in stress and affect), we are specific about placing situation on the same level of importance as temperament. Situations, and states of experience themselves, fluctuate from moment to moment. However, it is proposed that reactions to situation will have an overall stable tendency, i.e. to fluctuate around a mean. Temperament is also considered a stable component, with some essential value set at the level of the individual.

Although this is a stronger view of the importance of environment than that presented in the previous section, this discussion suggests that this change is what is needed to to capture possible interactions between an individual's experiences and their traits or individual preferences (Bolger et al., 2003; Funder, 2006). We need to measure intra-individual variation across situations, locations, and times. This also avoids the "methodological solipsism" (Bechtel, 2009) trap of a purely cognitive approach, accepting that by including context we agree with the implication that it is not just the mind which has a role in processes of experience. This has further implications for the model. In the next section I will talk about how these ideas can be operationalised.

Measurement and method

This section begins with a brief outline of the variables. All have been chosen for consistency and reliability (Anusic et al., 2017; Marteau & Bekker, 1992; Rudaizky & Macleod, 2013).

Independent variables

1. Trait tendency to anxiety, as measured by the Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) (Spielberger et al., 1983), in the form of ten "anxiety-present" and ten "anxiety-absent" items (Marteau & Bekker, 1992), to give a total score of between 20 and 80.
2. Current situation, i.e. activity (15 choices), company (7 choices), and location (3 choices), modelled on proven studies by Kahneman and others (Anusic et al., 2017; Kahneman et al., 2004; Kahneman & Krueger, 2006; Stone et al., 2006). These were chosen for generalisability: in studies using the same measures of situation the results are "generally similar" (Anusic et al., 2017) some 15 years afterward. All possible choices can be seen in the results section.
3. Time, a system-generated timestamp complete with date and, adjusted for time zone.

Response variable

1. Current experience of stress, measured by the Six-item State-Trait Anxiety Inventory (STAI-6) (Marteau & Bekker, 1992) using three "anxiety-present" and three "anxiety-absent" items giving a total score of 6-24. The six-item version (Marteau & Bekker, 1992) of the Spielberger State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983) has similar consistency and reliability to the full 20-item version of the scale (Nilsson et al., 2012; Tluczek et al., 2009).

Review of ESM/diary studies which explored stress and affect

Researchers have been using different types of mobile device to carry out studies of behaviour for more than three decades. For the purposes of this review, it is useful to briefly adopt the typology outlined by Miller(2012), and assign these studies into four general categories.

The first prioritises patterns and networks of users, generally by analysing anonymous, aggregated data from service providers such as call logs or tracking device proximity (Eagle et al., 2009). Whilst these can produce large, sophisticated network analyses, they do not produce the kinds of fine-grained intensive data required to test the theoretical model of temperament and experience developed in this thesis.

The second type does allow precisely this form of data collection. These studies involve specialised devices such as personal digital assistants (PDAs) and electronically-activated recorders (EARs). The FASEM-C (Perrez et al., 2000) is one example of such a study, collecting daily data on stress in family life using custom software running on a pocket computer platform, the HP360LX. It is a flexible concept that collects rich data in a variety of fields, such as diary studies (Bolger et al., 2003). However, this specialised equipment can be a drawback. Participants often forget or lose devices, and can find them too intrusive (Scollon et al., 2003). These systems have largely fallen out of favour with researchers as smartphones have become more capable (Ellis, 2020, pp. 10–14).

The third and fourth type both run directly on smartphones, and are distinguished by Miller by distribution method: the third group hand out a fixed, preprogrammed mobile device. Like the second type of study, this device runs custom software, and the operating system may be bespoke or specialised in some way. This software may be commercial or open-source; a detailed evaluation of the state of available programs follows this section.

The fourth type of study relies on software alone, and as smartphones have become more universally capable Miller's distinction has become less meaningful. In the studies contained in this thesis I started by distributing devices, but as time passes the population of users expands so much that distribution shifts to the software alone, running essentially the same code. In this discussion it is useful to chiefly differentiate between custom devices or hardware (the second type) and custom smartphone software (the third and fourth types).

The following table contains a mix of research recommendations and study findings.

Author	Title	Administration	n	Days	Study characteristics	Software	Aims/hypotheses
Bolger et al. (2003)	Diary methods: Capturing life as it is lived				Contains research recommendations		
Bolger & Zuckerman (1995)	A framework for studying personality in the stress process	Paper diary	94	14	Situation: Interpersonal conflicts Controls: None Differences: Levels of neuroticism	-	The effect of reactivity and exposure to conflicts on daily distress is moderated by neuroticism
Conner & Barrett (2012)	Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine				Contains research recommendations		
Courvoisier et al. (2010)	Psychometric properties of a computerized mobile phone method for assessing mood in daily life	Automated calls	307	7	Situation: Cell phone users Controls: None Differences: Compliance	SmartQ	Test the psychometric properties of the tool
Epstein et al. (2009)	Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use	PDA	114	130	Situation: Drug craving and use Controls: None Differences: Mood and drug-use triggers		Drug-abusing individuals can provide behavioural data using EMA
Fahrenberg et al. (2007)	Ambulatory assessment - monitoring behavior in daily life settings: A behavioral-scientific challenge for psychology				Contains research recommendations		
Fisher & To (2012)	Using experience sampling methodology in organizational behavior				Contains research recommendations		
Hufford et al. (2002)	Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers	PDAs	33	14	Situation: Undergraduate problem drinkers Controls: None Differences: Behavioural reactivity, motivational reactivity	Palm Pilot	Is motivation to change behaviour affected by monitoring Does drinking behaviour change as a function of real-time monitoring
Lucas et al. (2021)	A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM)	Mobile phone	401	1	Situation: Students Controls: None Differences: Affect and activity ratings	Qualtrics Web-based	Exploratory only
		Mobile phone	419	1	Situation: Single app for all components Controls: None Differences: Affect and activity ratings	Open Science Framework	Expectations moderate affect
Mehl et al. (2001)	The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations	EAR	52	2	Methodological study		
Myrtek et al. (2005)	Emotions in everyday life: An ambulatory monitoring study with female students	Custom palmtop	50	1	Situation: Normal activities Controls: Randomized with dummy signals	Custom	Heart rate as an indicator of emotional arousal Emotion perception is not linked to physiological changes

Figure 4: Main studies identified in the literature part 1

					Differences: Heart rate levels and emotion perception		
Perrez et al. (2000)	How to assess social regulation of stress and emotions in daily family life? A computer-assisted family self-monitoring system (FASEM-C)	Palmtop	311	7	Situation: Two-parent families with at least one adolescent Controls: None Differences: Coping, anxiety, life satisfaction	FASEM-C	Functional coping is positively associated with emotional and social well-being
Raugh et al. (2019)	A comprehensive review of psychophysiological applications for ecological momentary assessment in psychiatric populations		128		Systematic review		
Reininghaus et al. (2016)	Stress sensitivity, aberrant salience, and threat anticipation in early psychosis: An experience sampling study	Device	150	6	Situation: Daily life Controls: Three groups; psychosis episode, at-risk, and no history of psychosis Differences: Stress sensitivity, aberrant salience, threat anticipation	PsyMate	Stress sensitivity, aberrant salience, and threat anticipation are positively associated with intensity of psychotic experiences These associations are weakest in controls
Sherman et al. (2015)	The independent effects of personality and situations on real-time expressions of behavior and emotion	Smartphone	209	7	Situation: A 'typical week' Controls: None Differences: Personality and situation, state	Web survey	How much variability in state expressions is between versus within persons How much variability in situation experiences is between versus within persons How are personality traits related to situation experiences How are person and situation characteristics related to state expressions
Shiffman (2009)	Ecological Momentary Assessment (EMA) in studies of substance use				Contains research recommendations		
Verduyn et al. (2009)	Predicting the duration of emotional experience: Two experience sampling studies	Palmtop	60	14	Situation: Daily life Controls: None Differences: Personality traits, situation		Duration of negative affect to correlate negatively with extraversion and positively with neuroticism Duration is positively associated with reappraisal and negatively with suppression Importance of situation is positively associated with duration Onset intensity is positively associated with episode duration Increased physical appearance with the stimulus is positively associated with duration
		Web survey	43	14	Situation: Daily life Controls: None		Physical and mental reappearance of the stimulus extends duration Interpersonal emotional episodes will differ in duration

Figure 5: Main studies identified in the literature part 2

					Differences: Personality traits, situation, data collection method		Results would remain robust despite the change in data collection method
F. H. Wilhelm & Grossman (2010)	Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment				Contains research recommendations		
P. Wilhelm & Schoebi (2007)	Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood	Palm Tungsten PDA	147	7	Situation: Daily life Controls: None Differences: Mood	IzyBuilder	Evaluate the psychometric properties of a short mood measure
Zelenski & Larsen (2000)	The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data	Paper form	82	30	Situation: Daily life Controls: None Differences: Moods		Examine the distribution of primary emotions in daily life Examine the extent and nature of emotional blending Examine the covariation between emotions within and between subjects for evidence of a discrete or dimensional emotions perspective

Figure 6: Main studies identified in the literature part 3

Several important guidelines emerge from an analysis of these studies. A number of investigations (for example, Sherman et al. (2015); Verduyn, Delvaux, et al. (2009); P. Wilhelm & Schoebi (2007); Zelenski & Larsen (2000)) place emphasis on examining variation, between and within person and occasion. Perhaps because custom and customised software is so common, exploratory studies are frequent (Lucas et al., 2021). Measuring situation is frequent. And finally, many of the analyses recommend and themselves use hierarchical modelling. All of these things will be examined in the studies.

Experience sampling applications

Furr (2009) notes that although direct observation may be the strongest way of collecting behavioural measurements, its high costs in terms of time often limit the amount and breadth of data collected. Using acquaintance reports partially compensates for this by concatenating many observations in one sitting, but there is a corresponding reduction in information about individual occasions. Moreover, researchers have to be wary of introducing additional biases by way of the reporting subject's interpretation of observed behaviours. A middle way, he concludes, is experience sampling.

Three of the major approaches are the experience sampling method (ESM), ecological momentary assessment (EMA), and ambulatory assessment (AA). ESM is generally credited to Hektner, Schmidt, and Csikszentmihalyi (Hektner et al., 2007) and at least initially, was inspired by the idea of 'sampling' daily life, with the resulting ability to make statistical inferences. EMA, codified by Shiffman, Stone, and Hufford (Shiffman et al., 2008) places more of an emphasis on ecological validity and recall biases. AA, pioneered by Fahrenberg (Fahrenberg & Myrtek, 1996; F. H. Wilhelm et al., 2012), has its origins in clinical and physiological research, and the desire to bring monitoring methodologies such as biotelemetry to behavioural assessment.

The idea of sampling experience has been around for a long time, in many forms (Conner et al., 2009). Early researchers, noting the rich insights into experience from diary methods, naturally had the desire to extend this into daily life. A new focus on sampling, as a way of estimating frequency and distribution of events, suggested that the two techniques could complement each other (P. Wilhelm et al., 2012). For example, in the 1930s English statistician Leonard Henry Caleb Tippett experimented with random observation of workers in a textile factory to estimate how often their textile machines were idle. The experience sampling method was conceived as a way of leveraging technology to produce randomly sampled self-reports (Hektner et al., 2007).

In the 1970s researchers used digital watches to signal sampling times, and pagers in the 1980s. Participants initially carried stacks of paper to record their responses (Barrett & Barrett, 2001; Scollon et al., 2003). Eventually integrated computer-assisted systems were developed that could run on personal digital assistants (PDAs), and most recently smart phones have increased access once again.

Bolger and Laurenceau (2013, pp. 14–21), in their comprehensive handbook, characterise four ways of approaching experience sampling, seen historically (Wheeler & Reis, 1991). Signal-contingent is perhaps the best known. Participants are prompted to take measurements at random, or semi-random, intervals, usually by a device like a smartphone or pager. Interval-contingent sampling, on the other hand, captures experiences at preset, regular times. For example, researchers could ask for daily reports of stressors, or record resting heart rate on awakening.

Another method is to use an event-contingent approach, for example asking subjects to record measurements following a particular type of social interaction. This can be an additional burden for participants, however, who have to remember to report relevant phenomena. There is also no way to check compliance. Finally, taking the device-contingent route uses advanced mobile technologies to collect data from any number of different modalities simultaneously. Embedded measurement devices are not just limited to self-report; researchers can capture sounds, visuals, even locations and routines (Eagle & Pentland, 2009; Hasler et al., 2008). They can easily include automatic data from built-in sensors, such as time, heart rate, or GPS positioning data. As mobile phones are so widespread, this has the potential to give researchers access to a very large subject pool.

While traditional longitudinal designs are "often limited by a few repeated measurements taken over long time intervals" (Bolger & Laurenceau, 2013, p. 5), experience sampling methods usually focus on intensive sequential measurements over shorter periods selected to be frequent enough to characterise change processes. This allows assessment of theories which expect more densely packed fluctuations, within days or hours, such as intimacy in marriage (Bolger et al., 2003). More frequent measurement also allows us to better characterise temporal processes, and also has the potential to capture events or situations that are unethical, unsavoury, or simply impractical to reproduce in a laboratory (Barrett & Barrett, 2001).

There is a rich history of development dating back several decades (for a more detailed history see Hektner et al., 2007; and Bolger & Laurenceau, 2013). Experience sampling-type methods address many of the limitations of breadth and validity by collecting data recorded by the participants themselves, embedded in their natural environment (Smyth & Stone, 2003). These are usually also recorded in real time to limit reconstruction biases introduced by distance from the phenomena (Bolger et al., 2003). However, whilst less demanding than direct observational methodologies, experience sampling techniques have nonetheless also been unpopular with subjects, as they are "taxing" (Christensen et al., 2003), and time-intensive (Barrett & Barrett, 2001; Kahneman & Krueger, 2006; Scollon et al., 2003; Shiffman et al., 2008) to take part in. Experience sampling approaches traditionally suffer from high attrition (drop-out rates) and noncompliance (failure to respond) (Lucas et al., 2021). In a set of studies undertaken by Feldman Barrett (2004), for example, attrition was between 15 and 30 percent.

In recent years researchers have been able to offset some of these problems by taking advantage of developments in mobile technology. Experience sampling has a long history of drawing on technological innovations, from early studies using digital watch alarms (P. Wilhelm et al., 2012) to the latest mobile phones (Eagle & Pentland, 2009; Hektner et al., 2007; Mehl & Conner, 2012).

These studies using mobile technologies have demonstrated that they not only reduce participation costs (Raento et al., 2009), but that the lower barrier to participation in itself has opened up new areas of research by widening access. Scientists interested in momentary behaviour have already begun to see the benefits in terms of large, continuous data sets about people's daily lives (Eagle & Pentland, 2009), and this new approach has generated some surprising insights. For example, Eagle and Pentland (2009) claim to be able to predict the following half a day of behaviour with 79% accuracy, based on their analyses of data for the first half of the day. Use of these technologies has also increased access to hard-to-reach populations; Verdoux et al. (2003) were able to take unique measurements of the experiences of habitual cannabis users, and link the progress of psychotic symptoms to their vulnerability.

The apps

Below are a number of apps, taken from reviews of the literature and updated recently. In this review non-native applications were excluded, as these are really web pages and rely on an internet connection. And if a web survey is required there are many platforms to choose from there are many to choose from such as SurveyMonkey and Survey AnyPlace.

Some of these apps run on platforms that are no longer available or updated. For example, openXdata runs in Java which is not supported by modern devices (Android apps are built in Java but compiled for Android, so cannot run any Java app). Similarly Palm OS was discontinued in 2007, but as seen above, many have legacy devices.

The analysis also excluded apps that do not allow administering studies to other participants; 'Momento' is useful for self-tracking, for example, but is private to the user. Similarly I discarded frameworks like Apple's ResearchKit, which is not an app but a set of modules that can be utilised by an app. (ResearchStack is one Android equivalent.)

In the table itself, 'Active' means the application is still receiving updates; inactive applications may still work but cannot be relied on. They can be both customisable (inputting your own data), and programmable (setting your own contingencies for alerts or prompts). In addition, I have tried to rate the amount of technical skill needed to customise the output: low, moderate, or high. Most people have at least the ‘low’ level of skill, and I would imagine social scientists are often moderate or higher.

For space reasons I do not include most of the generic survey services, which can be adapted with some effort but are not really suited for continuous or intensive measurement more typical of ESM. Most of these applications have not been designed with social researchers in mind (Geyer et al., 2019) but many of them are actually quite suitable for the type of research typically done in social science (see the review of studies), with survey-style questions. As seen in the previous section, PDAs have remained in use, perhaps due to sunk costs.

Table 1: An updated overview of apps for experience sampling

Name	Platform	Active	Programmable	Customisable	License	Programming skill
DroidSurvey/iSurvey	Android/iOS	Yes	Yes	Yes	Commercial	Very low
ESM Capture	iOS	Yes	Yes	Yes	Commercial	Low
LifeData	Android	Yes	Yes	Yes	Commercial	Low
mEMA	Android & iOS	Yes	Yes	Yes	Commercial	Moderate
MetricWire	Android & iOS	Yes	Yes	Yes	Commercial	Unknown
mQuest	Android & iOS	Yes	Yes	Yes	Commercial	Low
Psymate	Android & iOS	Yes	Yes	Yes	Commercial	Moderate
Qualtrics	Android & iOS	Yes	Yes	Yes	Commercial	Low
Aware	Android & iOS	Yes	Yes	Yes	Free	High
Beiwe Research Platform	Android & iOS	Yes	Yes	Yes	Free	High
EpiCollect	Android & iOS	Yes	Yes	Yes	Free	Low
LAMP/mindLAMP	Android & iOS	Yes	Yes	Yes	Free	Moderate
OhMage	Android & iOS	Yes	Yes	Yes	Free	Moderate
OpenDataKit	Android	Yes	Yes	Yes	Free	Mixed
Paco	Android & iOS	Yes	Yes	Yes	Free	Moderate
Piel Survery	Android & iOS	Yes	Yes	Yes	Free	High
ExpiWell	Android & iOS	Yes	Mixed	Mixed	Mixed	Moderate
MovisensXS	Android	Yes	Yes	Yes	Mixed	Mixed
Open Science Framework	Mixed	Yes	Yes	Yes	Tiered	High
DataField	Android & iOS	No	No	Yes	Discontinued	Unknown
EmotionSense	Android	No	Yes	Yes	Discontinued	Unknown
Funf	Android	No	Yes	Yes	Discontinued	Moderate
iDialogPad	iOS	No	Yes	Yes	Discontinued	High
iHabit	iOS	No	Yes	Yes	Discontinued	Low
IzyBuilder	Palm and Pocket PC	No	Yes	Yes	Discontinued	Moderate
ContextPhone	Symbian	No	Yes	Yes	Free	High
Experience Sampling Program (ESP)	Palm OS	No	Yes	Yes	Free	Moderate
MyExperience	Windows Phone	No	Yes	Yes	Free	High
Psychlog	Windows Phone	No	Yes	Yes	Free	High
RealLife exp	Android & iOS	No	Yes	Yes	Mixed	Low

However, there is a clear gap in 'one piece' solutions that measure trait as well as state.

Summary: Tool gaps, desirable characteristics, hypotheses

From earlier in the thesis we can see that the requirements are an app that measures variation, between and within person, and allows for situational factors to be recorded. Exploratory studies to test the tool appear to be recommended: the first two studies feature some iterative design and feedback from the early adopters of my application.

Mobile approaches to ESM

Experience sampling methods have been evolving to use new tools since the early twentieth-century (P. Wilhelm et al., 2012). For example, early researchers saw the potential in pagers, PDAs, and then mobile phones (Hofmann & Patel, 2015). In the past decade we have seen the proliferation of mobile computing devices (Hofmann & Patel, 2015; Kubiak & Krog, 2012; Miller, 2012; Vigo et al., 2017; P. Wilhelm et al., 2012) until they are near-ubiquitous. These devices have powerful processors, multi-tasking operating systems (OS), and large high definition touchscreens (Miller, 2012). These features make them very useful for psychological studies, giving researchers the capacity to run embedded measurement platforms without specialised hardware, and the opportunity to record sounds, visuals, and other types of data such as location, speed, and altitude (see for example Ebner-Priemer & Trull, 2009). Even the camera flash could be repurposed (Sorrel, 2010).

In mid-2009, when I started planning an experience-sampling study for the first time, there were very few studies using self-contained applications; that is, downloadable smartphone software which conducts the study autonomously (Miller, 2012), bringing together the entire process of data gathering, upload, and debriefing.

The idea came about as almost a demonstration that this kind of thing could be done entirely on a mobile phone: something that you would naturally have with you, rather than a static application requiring a computer or a dedicated device. I was not the first to think of this (e.g. Eagle & Pentland (2009) - Nokia; MacKerron - Mappiness (2013)), but owing to my previous background as a systems analyst, with a bachelor's degree in computer science and two master's degrees in systems design and analysis, I was able to be amongst the early adopters.

The tool was developed early on, but as the thesis progressed it underwent significant evolution, and the substantive psychological research I was using it for broadened in scope and narrowed in focus. Although there were few alternatives when I began development in 2009, today there are various options. Over the next few pages I will look at some of them, consider the advantages and disadvantages, and try to answer the question of what a researcher should do today if they want to undertake a similar intensive longitudinal study.

Smartphone adoption

We now take them for granted, but in 2007 when Apple introduced the first iPhone (Morrissey & Campbell, 2011, p. 3), consumer adoption of advanced mobile internet services, such as the Multimedia Messaging Service (MMS), video streaming, and high-speed internet browsing (Carlsson et al., 2006) were low. Users appeared to be more comfortable with the old services like the Short Message Service (SMS, or 'texting'), even on advanced devices. As mobile service providers such as Vodafone had been persuaded to pay large sums of money [£22.47 billion in total; Cable et al. (2002)] for third generation (3G) network licenses based on the prospect of up-selling consumers additional, pricier 3G services, this was a major problem for the market (Mackley, 2008).

The iPhone was a major catalyst of change in public perception of mobile phones (Kenney & Pon, 2011; Miller, 2012). With a large touch-screen, and no bulky hardware keyboard, it was the size of a feature phone but ran a sophisticated mobile operating system (OS) with many of the capabilities of a desktop OS. It provided a full web browser, unlike the restricted "silos" (Kenney & Pon, 2011) most preceding phones permitted, where content was constrained by the network provider. Adding support for downloadable applications literally untethered devices from the desktop. These advantages helped promote widespread adoption of smartphones, in part by redefining them as powerful wireless computers (Miller, 2012), not just portable telephones.

More pertinently, from a researcher's perspective, producing software tools for iOS (originally 'iPhone OS') using Apple's integrated development environment (IDE), Xcode, makes it easier to develop custom applications using the full range of the device's capabilities (Hofmann & Patel, 2015; Miller, 2012), instead of being restricted to the often limited functionality of web page or SMS.

Discussion of my technical approach

As a consequence of the issues identified above, I decided to implement my data collection method as a mobile application. Specifically, I developed versions for the Apple iOS platform, concentrating on the iPhone. This is one of the most popular smartphones in the world: by 2018 (when they stopped publicly disclosing exact sales figures), Apple had sold more than two billion iPhones (Apple Inc, 2018a), enough for a quarter of Earth's population (Population Reference Bureau, 2020). In January 2021, Apple

CEO Tim Cook claimed that around 1.65 billion Apple devices are in active use, and more than 1 billion of them are iPhones (Kastrenakes, 2021).

Because Apple control both hardware and software, technical development is easier and individual device compatibility is guaranteed. Developing for iOS also allows the app to be installed remotely, which is much more convenient for the user and the researcher: the ‘ad hoc’ distribution method securely certifies apps through Apple’s own servers, and makes installation a very straightforward process even for naive users. The touch-operated interface offers a number of advantages such as visual affordances, and increased speed of interaction.

The system that I have developed is not over-complicated. It measures all the variables identified; both temperamental propensities, and current state of anxiety (see the three tables below). It also asks participants to categorise, very broadly, their activity, their location, and their social company for the situational aspect of the studies.

Each measurement is time stamped automatically, and the app automatically collects demographic information at the start of participation, and feedback at the end. Data is stored locally and can be downloaded by the participant, as well as accessed by researchers. The app also allows data to be uploaded remotely to a secure database server, as the “most fragile link” (Raento et al., 2009) in computerised data collection is usually data loss. A complete copy of the data is kept locally (storage demands are minimal, bytes rather than megabytes) and this copy is synced remotely on an ‘as needed’ basis when a network connection is available (see figure 12 below). So if a participant is underground, in the remote countryside, or even without network access, a brief sync is all that is needed.

Much of the burden of participation is handled by the application, which directs and supports subjects through the whole process. In addition, a graphical interface allows participants to set repeating reminders, choosing their own dates and times as appropriate. Considerable development time has been spent on the code, visual interface, and in requirements capture, as will be detailed in the individual studies. Over the course of these studies, the software has had to be maintained and updated for compatibility, but most changes have been behind the scenes and reflect operating system requirements.

iOS was designed to support modularity and mobile from the ground-up. It is relatively easy to specify and adapt for a range of iOS devices.

It is secure - apps are checked and certified by the App Store, which makes the process of installation a bit more straightforward for the user. For the developer, it can be complex to go through the distribution process, although Apple have streamlined this a fair amount over the years. However, once these processes have been undergone, developers know the software will run consistently, on a limited set of devices, and

Android is a very different platform for development, based on different programming languages and approach. Because of this, it requires much more investment by developers to support all the types of device. One advantage is that apps can be 'side-loaded', that is installed without having to be verified and secured by the Android equivalent of the App Store ("Google Play"). But this is very much balanced out by not knowing who is going to be running an application, and which of several dozens of devices, each with differing capabilities.

On top of this, they are very likely to be running differing versions of the operating system. Manufacturers usually make tweaks, both minor and major, both to differentiate their products and to add value to their own software ecosystem. It is rare for a device to run with a 'vanilla' (unmodified) operating system. Even Google, who guaranteed universal support on their own devices, have begun phasing these out.

Evolution of the design

As phones have evolved, so has the operating system and with it the programming requirements (figure 7).

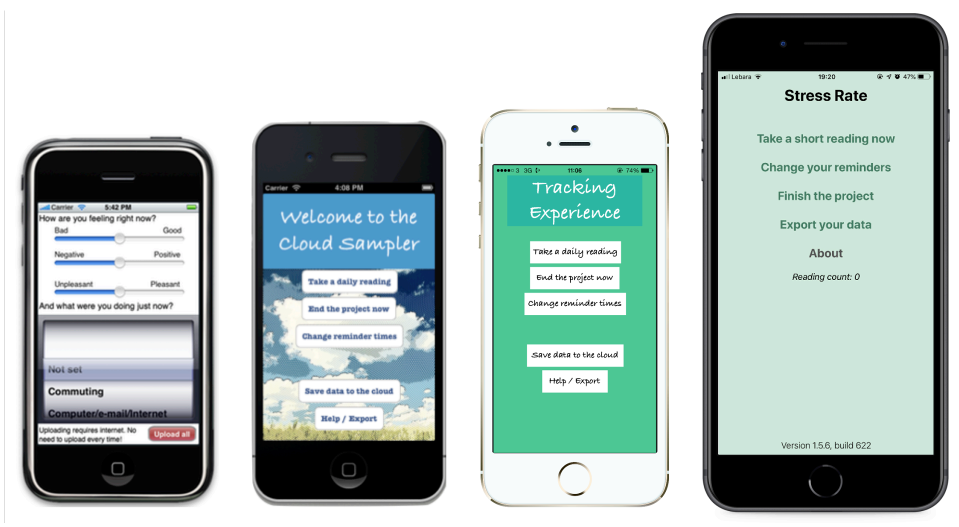

Figure 7: App designs from old to new

Here this section gives a walkthrough of using the app.

Figure 8: Starting the project

First the app collects the Worker ID to match results with submitted tasks. Then simply tap through to give the answers to the STAI-T; each screen should take a second or two and automatically saves and continues when you select your answer.

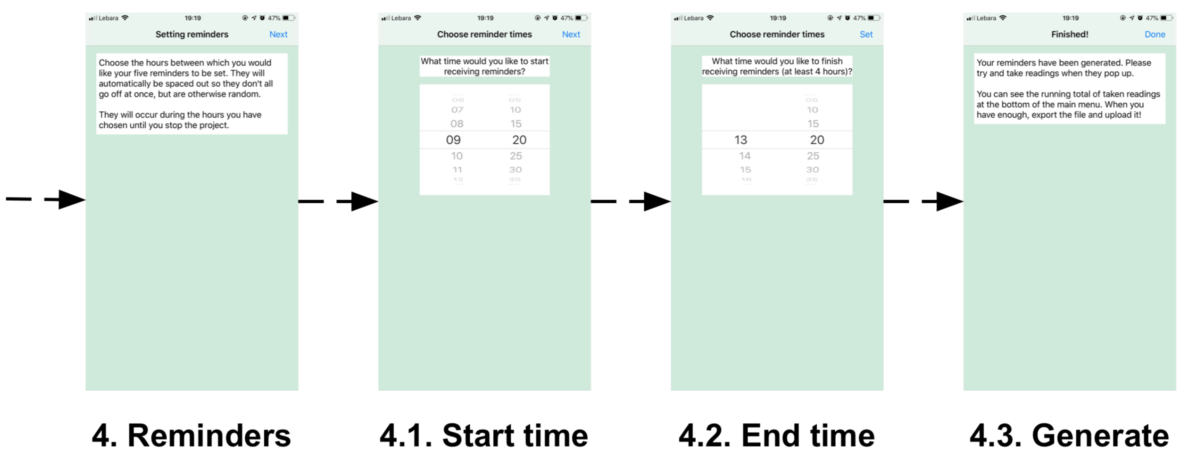

Figure 9: Setting reminders

Then reminders can be set; the latest (studies four and five) version of the app enforces five readings per day, spread over at least four hours to get a selection of activities (so no doing five one after another); new reminders every day at different random times within the block of time chosen. The older versions allowed the number of reminders, and total number of days taking part to be set. I found users often wanted to continue for longer however, so the current version supports basically infinite participation. This was also useful for the year I spent on the self-tracking study.

Figure 10: Taking a short reading

Receive a notification and tap through to launch the app in ‘take a reading’ mode. If it is ignored, the user will be prompted next time they launch the app (‘missed reading’). Possible to track reaction time and responses to notifications, but not necessary for studies carried out. Then six STAI-S screens (they ask about ‘now’ instead of ‘in general’); activity, social context, and location. The total time is just a few seconds, especially when you’ve used it a bit. Readings are saved and time-stamped as they are taken; again, running the app locally reduces chance for data loss e.g. due to internet connection going down or passing out of mobile data range (for example on a train).

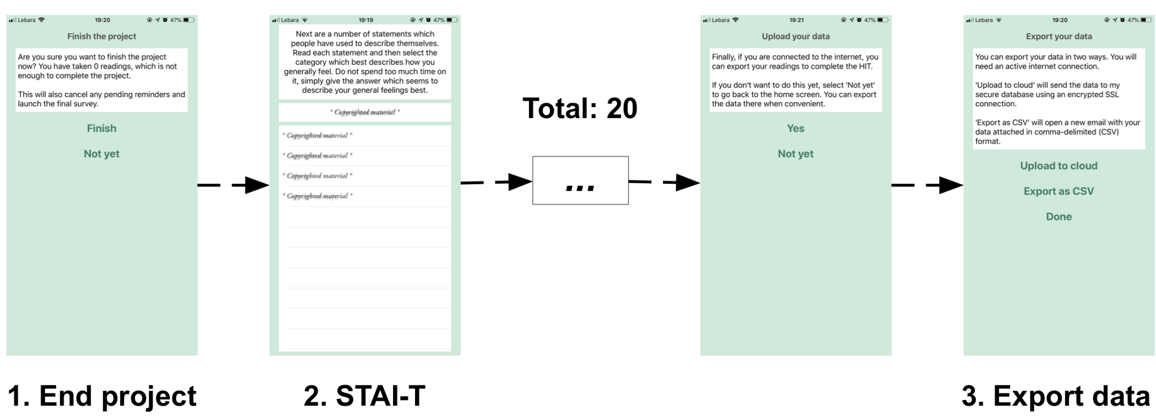

Figure 11: Finishing up

The ‘Are you sure?’ prompt is to make it more difficult to stop by accident. Then the user is prompted to do a final trait reading, and to upload their data. They can do this later if there’s no internet or if it’s inconvenient; they can also send an email of the data to their own email, or to a researcher.

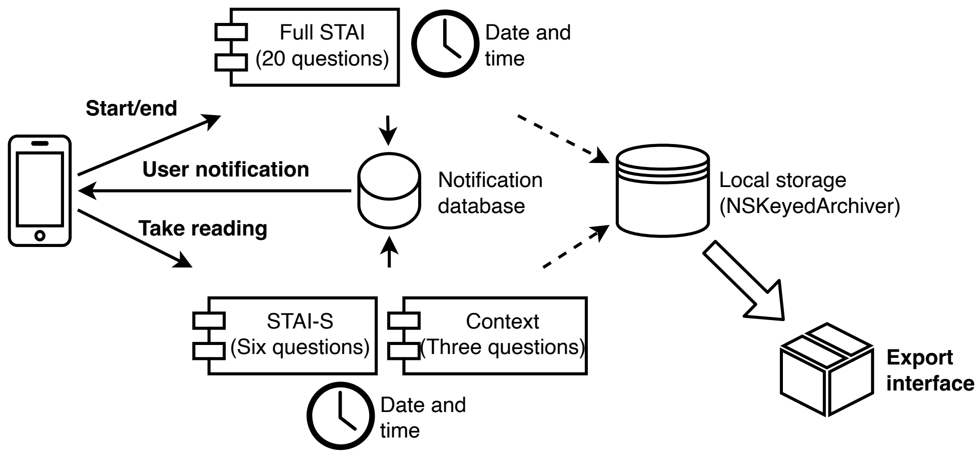

Figure 12: Summary of operation

Programming over the years required a good knowledge of SQLite, Objective-C, Swift, and the https protocol. The code underwent two major re-writes to account for updates to the operating system, including refactoring into Swift to support new language features and security. Despite these backend changes, the core functionality of the application has remained identical, if faster and more reliable.

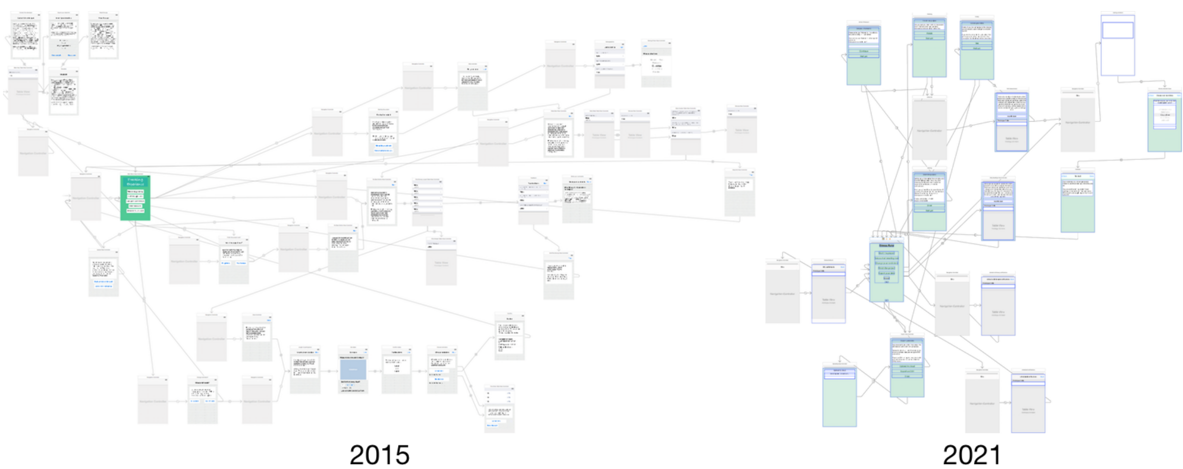

Figure 13: The Xcode SDK's graphical interface builder and storyboard

To aid with this, the Xcode SDK's graphical interface builder and storyboard lets you design a user interface without having to write initial code (Apple Inc, 2018b). Then you define the actions to be taken by the relevant parts of the interface, for example when the user selects the 'upload' button, you want the application to run the uploading routine. Then you feed back the results to the interface to tell it what to do next (in this case, to go to the next screen). The storyboard allows you to set out interactions and relations between elements of your application - for example, going from the main menu to the set-up screen - graphically, greatly reducing the amount of code you need to write.

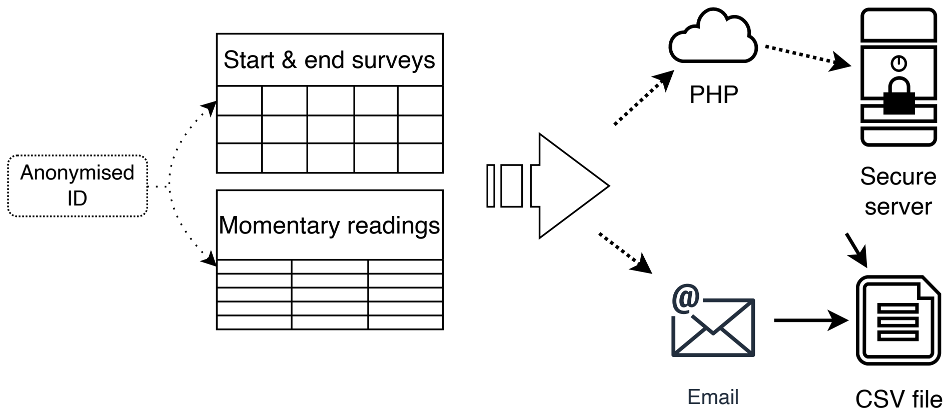

Figure 14: Exporting data, and storage

Overall summary

To sum up, this thesis contributes to the literature in two substantive areas. The first is psychology, where I investigate the components and processes of experience formation in states of stress. Although there have been some momentary studies of stress in every day life, none have looked at the interactions between dynamic and stable factors in this way. I address the comparability and generalisability of situations across person, and the function of trait as a regulator of individual reactions.

My studies also make a methodological contribution to the field of behavioural data measurement. I examine whether technological advances can compensate for the increased burden of intensive self-report, and look at implications for researchers looking to work with experience sampling. I aim to broaden access to experience sampling style methods by setting out practical guidelines in my own work, and to encourage researchers to consider the suitability of such methods for their work by demonstrating usability and flexibility.

Summary of research questions

Having examined new models, and measurement methods, this thesis will address the question of whether it is possible to change the way stress is commonly measured. For example, as we will see in paper one, medical and mental health organisations still often rely on single occasion, summarised recall measures to try to judge levels of stress and choose appropriate interventions. As I have discussed, this leaves a range of factors unexamined, including context and processes of change. I see developing and implementing a new method, based on a finer-grained understanding of the components and ecology of stress, as the main contribution of this thesis. It is both methodological and substantive.

Table 2, details the settings for all four studies that comprise this thesis.

Table 2: Summary of the studies

Study	Population	Subjects (n)	Observations
One	Health professionals	2	153
Two	Students	9	709
Three	Researcher	1	1265
Four	Mechanical Turk workers	24	683
Five	Mechanical Turk workers	30	854

Paper one takes a pragmatic approach to making this contribution. The research question which begins my thesis, and from which the others flow, asks whether this new method and model, based on momentary methodology, can be implemented in practice. If the answer is no, this will be a considerable roadblock; however, I draw on similar development work to address these concerns. I look at how the model and the method meet, and how to make this work in the field, with a series of exploratory interviews, usability metrics, and an iterative piloted development.

This study outlines the development process, with a focus on customising the instrument for individual applications. I work directly with mental health professionals who specialise in enhancing coping methods for unemployed people with mental health issues. I explore how this method can increase participation in hard-to-reach populations, and produce data of use to health professionals as well as academics.

The second paper follows on from this, looking at further generalising the method, and examining the fit of the technology and the method more closely. The research question for this paper complements paper one by testing if this model, method, and technology work as claimed outside of a controlled environment. It is concerned with a number of success metrics, including measurement reliability, participant retention, and effects analysis. I also test the atomicity, and reliability, of the application when deployed remotely with a population with no prior knowledge of the technique through feedback and technical statistics.

Further discussion examines the characteristics of the data gathered, and the implications for my model. I look at remote design and deployment, and explore the kinds of data that can be gathered with an absolute minimum burden for participants. These two studies are by nature exploratory.

The third paper looks outward again to ask how the results of my model and method fit into the wider world. I do this by examining how my hypotheses fit with real- time physiological measures. This situates my work at the junction of medical and psychological research, attempting to combine the advantages of both through integrating real-time and long-term influences. It focusses on a single participant to enable both broad and deep analyses.

The fourth research question, which will be addressed by the fourth paper, explores the idea of the theoretical model in more detail, to test my psychological theories. This study applies the methodology to a larger population of Mechanical Turk workers, allowing me to more rigorously test theoretical propositions and carry out a more involved analysis. Following the methodological tests, the overall research question for this paper asks how the data backs up my model in a larger sample, specifically through hierarchical models which will support a more rigorous analytical approach.

Finally the fifth study addresses the gap between trait measures and state measures, by collecting a verified number of participants and fitting suitable models to test for a moderating effect of temperament discussed above.

Taken together, this means that my research is partly qualitative, and partly quantitative. Earlier studies draw more on sources such as interview data, and lay the groundwork for my later studies which concentrate on empirical investigation and more challenging statistical analyses. While individual papers contain more specialised appraisals of the relevant literature in the area of study, each should be considered to take place in the context of the general framework I have set out in this review.

Limitations and threats to validity

In any study, it is important to consider potential limitations. Some overall limitations are addressed in the conclusion to the thesis. But this section deals with those particular to individual studies, each of which may be impeded by specific threats to validity. The discussion will follow the model of Campbell and Cook (Shadish et al., 2002). They suggest that four, related, components are of importance when exploring cause and effect (Brewer & Crano, 2014). These are statistical conclusion validity, internal validity, external validity, and construct validity. This section examines plausible threats (Campbell, 1988, p. 317) and discusses how, and to what extent, these are addressed through study design and statistical methods. The comments are grouped under headings for the four types of validity.

Statistical conclusion validity

Statistical conclusion validity depends upon the appropriateness of statistical methods and sufficient data collection. While papers one and two are exploratory in nature, paper three uses regression analyses, paper four mixed-effects models, and paper five mixed-effects with interactions.

Statistical power is most important in studies three, four, and five. Because paper three uses longitudinal data within a single participant, there are concerns about autocorrelation. To address this during analysis, several lagged indicators were fitted, but each had a large p value suggesting the null hypothesis could not be rejected. Durbin-Watson and Breusch-Godfrey tests of time-series data (StataCorp, 2009b) also found no indicators of autocorrelation, with large p values. Papers four and five account for dependencies within participants (and days within participants) by fitting mixed-effects models with random effects for participant and day (within participant). In addition, a power analysis was carried out for paper five and the appropriate number of subjects recruited.

Internal validity

Internal validity concerns questions about conclusions about causal relationships between variables. Can we say, with a degree of certainty, that conclusions about estimated associations in the study also give evidence of causal effects?

In this thesis, papers 1 to 3 focus largely on method, tool development, and feasibility. Papers 4 and 5, on the other hand, deal with research questions which have at least partly causal motivation. Because the data collected in these studies are not experimental, causal conclusions are nevertheless limited. One strength of these data is that they are longitudinal. It is then possible to observe the same individuals in different contexts and over time, which increases confidence in the association of changes in specific predictor variables on the outcome. However, the primary focus of the studies is to examine associations between the variables (context and temperament vs. stress), and causal conclusions from them are at best tentative.

External validity

External validity asks how generalisable our particular findings are to persons, and populations, outside of the study group. Again, these studies are not making strong claims about such generalisability. Papers one and two, in particular, are focussed on measurement development rather than generalisability and recruit very specific groups, and paper three is most notably limited to a population of one.

Papers four and five do include participants drawn from a general pool of potential respondents, that of Amazon Mechanical Turk users. They are not a representative random sample from any clearly defined population. Instead, any arguments about external validity are limited, as they would be for any study that uses this method of recruitment. To a large extent, the most we may be able to say is that the results are suggestive of similar patterns in other contexts: there are no strong reasons to think that general qualitative conclusions about the effects of situations, or the presence of moderation, would not be broadly applicable to similar populations. Furthermore, some previous research suggests that the population of Mechanical Turk users is at least likely to be more heterogeneous than some other possible subject pools, especially the college population often used in other studies(Buhrmester et al., 2018).

Construct validity

Construct validity invites us to consider how well the variables measured represent the concepts of interest. This depends in part on how well the construct of interest has been defined, as seen earlier in this chapter. But it also depends upon the validity of the measures, in these studies the STAI for state and trait anxiety, and the DRM-derived measures of situation. While these have been well-tested (Anusic et al., 2017; Marteau & Bekker, 1992; Rudaizky & Macleod, 2013), they can show overlap with other mood states such as tension or arousal (P. Wilhelm & Schoebi, 2007), or reactivity and perseveration (Rudaizky & Macleod, 2013).

We also have to consider possible unintended consequences of measurement, in this case if the process of measurement itself might introduce deviations in momentary anxiety. The application is designed to be fast and unobtrusive, with each measurement taking only a second or two. In addition the participants were able to choose their own hours, within certain limitations detailed in the studies, and could ignore a prompt with no consequences as long as they completed the required total number. This may, however, have led to a lower frequency of sampling particular events. Papers two and three investigate reactivity, with paper three discussing it in some detail. Papers two, three, and five also deal with retention and attrition.

However, the experience-sampling approach has advantages in this area, allowing researchers to study processes as they happen, within daily life (Bolger & Laurenceau, 2013, p. 12). Taking measurements online, in their natural environment, gives ecological validity (Perrez et al., 2000), and can produce insights that laboratory or retrospective studies cannot. This also increases the likelihood that the variables are measuring the real reaction to the event, rather than a reconstructed or filtered impression.

Preface: Paper One

This initial study worked to turn the theoretical review into a practical methodological approach. Drawing on experience in the design of information systems, and supporting literature, a consultative, iterative process was used in designing the software tool. I was lucky enough to have collaborators who sat for several interviews and prototyping sessions.

This study sets up the instrument used in the rest of the thesis, which gathers data on momentary experiences of stress, their levels, and the situational context. The emphasis is on the methodological aspects of the study, and as such, much of the paper is concerned with establishing the validity of the method and laying a foundation for use of the tool in later studies. Some of the language is more technical than later papers, and there is more explicit discussion of design, but understanding how the application was coded is not a prerequisite for the rest of the study.

The challenges here were both social and technological. Could I communicate my theoretical ideas to the experts, and could they get their practical aims across to me? I had to turn the results of our discussions into a feasible software design that also satisfied theoretical constraints. The resulting application will have to work well enough to be relied on in later studies.

Mobiles and mobility: Assessing the potential of personal devices in online momentary approaches to experience sampling

Abstract

This paper explores an application of the experience-sampling technique in a health charity. It describes a situational approach to experiences of stress in daily life, and the lessons learned from an iterative development approach with interested participants.

Introduction

Experiences of stress and anxiety are an important influence on quality of life. In the UK, workers affected by stress, depression, or anxiety missed an average of 23 working days during 2013/14 due to illness (Health and Safety Executive, 2014). Researchers from the Health and Safety Executive (HSE), the body responsible for oversight of health and safety in the UK, recorded 244, 000 new cases of stress-related sickness in that period. This brought the total up to 487, 000 cases, or 39% of all work-related illnesses. One in five adults in the UK in the Annual Population study (APS) rate themselves as having 'high' anxiety (Office for National Statistics, 2014), which on the APS means scoring six or more out of ten.

The HSE notes that the influence of stress leads to physical consequences such as "heart disease and the metabolic syndrome" (MacKay et al., 2004) as well as mental health issues such as anxiety and depression. Mackay et al. (2004) also draw attention to the difficulties of demonstrating causal associations between factors and outcomes, citing the "methodological problems" of studying real-life conditions empirically. This is a long-standing problem in psychology (Cialdini, 2009), but fortunately tools for addressing such issues have become much more available.

This study explores the possibilities of momentary approaches in capturing these kinds of measurement, and outlines the process of developing, deploying, and retrieving a momentary data collection tool. It considers whether these methodologies can satisfactorily address the concerns of over-reliance on survey data for empirical inference, reification of laboratory results, and over-generalisation in interpretation.

Following methodological suggestions from a number of sources, this is a pilot study with a mental health charity which looks at these relationships: between trait and state, and situation and state, in momentary experiences of anxiety and stress. This suggests some general development guidelines for researchers, intended for those needing to swiftly tailor their psychological instrument through an embedded design approach: one that conforms as much as possible to a potential subject's environment. The study also considers the place of technology in reducing the burden of longitudinal studies, particularly experience sampling methods (ESM), for researchers and participants alike (Bolger & Laurenceau, 2013, pp. 14–21).

Therefore, the methodological approach is based on research on experience sampling (ESM), ecological momentary assessment (EMA), and intensive longitudinal methods (ILM) (see Bolger & Laurenceau (2013); Hektner et al. (2007); A. Stone & Shiffman (2002) for more information on the respective approaches). This paper begins by discussing why experience-sampling approaches can be effective, and what they can be used for.

Experience sampling methodologies

By collecting measurements as they occur, in the environment they are naturally placed within, ESM and similar methods help address many of the problems of limited scope and validity mentioned above (Smyth & Stone, 2003). However, development can be complex, and participation time consuming. This section covers the development of ESM, and how these burdens have been addressed.

The idea of sampling momentary experiences in daily life has been a concern of researchers for a very long time (P. Wilhelm et al., 2012); statisticians were attempting to estimate frequency and distribution of work activities as early as the 1930s. Historically, researchers using experience sampling techniques are often found in the vanguard of technology uptake, having used pagers, PDAs, and even digital watches (Eagle & Pentland, 2009; Hektner et al., 2007; Mehl & Conner, 2012; P. Wilhelm et al., 2012) in a quest to get measurements as closely as possible to the real-time phenomena of interest (Bolger et al., 2003). Studies using modern mobile technologies have already shown benefits both in terms of larger data sets and increased participation, as well as lowered cost of implementation (Eagle & Pentland, 2009; Raento et al., 2009).

In general, the approaches that have emerged generally fall under one of four categories, distinguished by the trigger that prompts a measurement (Bolger & Laurenceau, 2013, p. 12). Triggers can be randomised, using a pager alert or smartphone message (Signal Contingent), to approximate a true sampling approach to the variables of interest. Equally, such alerts can be set to prompt a measurement after a regular period of time (Interval Contingent).

Researchers can also ask participants to record data after particular events occur (Event Contingent), to ensure they get the experiences they are interested in. However, this places more of a burden on the participants, who are in turn more likely to forget or deliberately miss recording.

The final approach, and the one used in this study, relies on advanced mobile devices (Device Contingent) to collect both automatic and subject-composed readings, from traditional survey style questions to sounds, pictures or video, and data like GPS location (Eagle & Pentland, 2009; Hasler et al., 2008). The Device Contingent approach has become more common as technology has spread, but development must still be approached with caution, particularly with complex subjects. After discussing my model, some ideas for success are briefly explored.

A situated model of stress

As discussed, momentary methodologies help to address the common problem of differences often found between recalled assessment measures, and those taken in actual situations. Recalled data is often criticised for not reflecting immediate experiences (Shiffman et al., 2008; A. Stone & Shiffman, 2002), emphasising reconstruction and introspection (Baumeister et al., 2007) and ignoring the ecological influences of environment and situation (Van Dijk et al., 2015). This does not make recalled measurements wrong; however, researchers need to be aware that they are measuring something different to immediate daily experience (Schwartz, 2012). In particular, "covert" aspects of behaviour (Lahlou et al., 2015) - hard to observe phenomena such as interpretations, intentions, and emotional experiences - are difficult to access in any way other than immediately, and from the subject's perspective.

These issues also make it hard to unpick the influence of processes in experience formation. For example, in many traditional longitudinal studies a series of three, four, or five measurements over the study period is considered sufficient (Bolger & Laurenceau, 2013, p. 5). Yet for theories which postulate dynamic processes of change, which may be temporally contiguous, such as ongoing appraisals of closeness in relationships, these few readings are not sufficient to illustrate effects as they unfold over time (Bolger et al., 2003). This means it can be impossible to assess the hierarchy of cause and effect - what determines what - not to mention the actual order of events in time. Situation, personality, and behaviour are inextricably linked in a "triad" (Funder, 2006) that forms the basis of processes of experience. In this study, the environment of stressful experiences (Robbins & Aydede, 2009) is constructed by taking elements from both temperamental and situated models of event experience and moderation.

An important trait is broad reactivity to stimuli. Kagan (2010, p. 11) suggests that reactivity, or arousal, is one of the most salient aspects of differences in experience.

Whether a person is 'low reactive', i.e. a high threshold for arousal, or 'high reactive', with increased response to stimuli, can be determined as early as four months old, and tends to remain stable over time (Kagan, 2010, p. 15). Studies show that high reactives are much more prone to suffer from anxiety. Kagan's model describes experienced feelings as passing through moderating cognitive biases, both learned and innate, which combine with the context of an event to frame the experienced state. So temperament and learned reactions both play a role in determining experience, but we also need to consider the issue of ecological context.

To do this, the theoretical frame draws on a second model developed by Jan Strelau (2010), which he calls the Regulatory Theory of Temperament (RTT). The RTT also includes temperament as a moderator, but places greater emphasis on independent stressors which arise externally from the environment. Emphasising both sets of components suggests some variables (figure 15), which will be explored in more detail in the methods section to come.

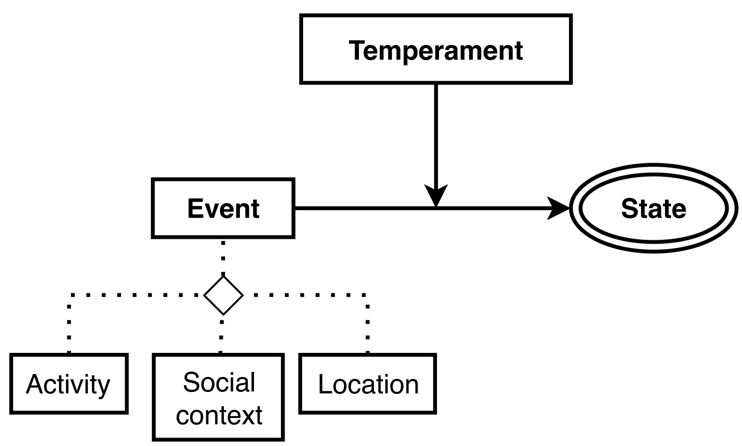

Figure 15: Core components of stressful experiences

Matching the theme of intensive longitudinal measurements (e.g. Bolger & Laurenceau, 2013), Strelau further suggests that research should frame stressors and their interactions with temperament and activity as 'interactive and reciprocal' (Strelau, 2010) processes. This makes them best explored by repeated measurements both within and between individuals.

In fact some researchers (e.g. Hurlburt, 1993) propose that it is only by taking such measurements of experiences, at the time that they occur, "in the subject's own natural environment" (Hurlburt, 1993, p. 23), that we can begin to explore the processes that differentiate inner experiences. Hurlburt proposes that there are certain characteristics that shape the thinking of affected individuals, especially in disturbed affect. By using an experience sampling approach, capturing events from the subject's "own situated perspective, in real situations" (Lahlou, 2011) we can truly see how the influence of particular characteristics and events varies between subjects.

This brings new challenges in designing and implementing these kinds of rich systems (Lahlou, 2010a). The later sections of this paper contain some suggestions for addressing these issues.

Design approach

In this study, two organisers from a mental health charity were recruited. Through an iterative process of interviewing and prototyping, we developed an application to run on Apple's iOS devices. We then ran an experimental field test designed to assess the utility of the approach under their typical usage circumstances.

I worked with them for several hours of interviews: exploratory requirements elicitation, prototyping and design, and finally feedback after the project was completed. They were extremely generous with their time, and this meant that this study could take an step-by-dstep approach, incorporating their ongoing feedback to improve the fit of the final product.

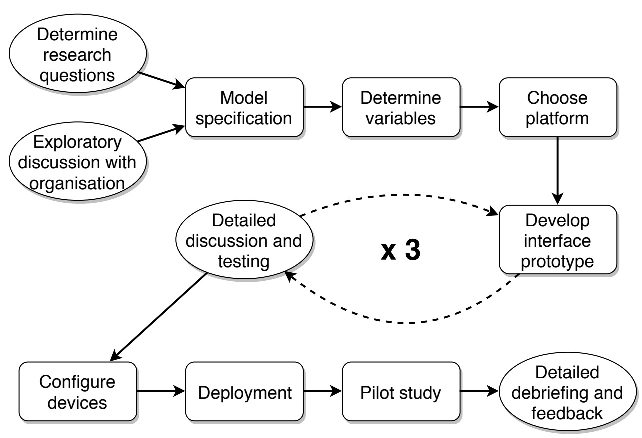

Figure 16: Development process

As figure 16 shows, there were five discussion sessions with the two experts, beginning with an exploratory discussion where mutual aims were discussed, along with the possible goals of a mutual collaboration. These ideas were fed back into the interface design, and then a prototype. The process was repeated three times before the prototype was ready for the pilot study. Finally, there was a debrief and feedback session after the study was completed, to talk about the results and their opinions on their experiences.

A subsidiary aim of this study is to contribute to the development of momentary measurement approaches, by looking at the issues that arise during design and development. As Saadi Lahlou (2010a) says in the opening chapter of 'Designing User Friendly Augmented Work Environments', the difficulty of designing systems is not in writing excellent code, but delivering applications that are usable by the intended users. He observes that technologies designed to be used embedded in the user's environment face a particular set of design issues, peculiar to these kinds of systems. They must be developed differently from technologies that run in isolation, that is, not connected openly to the network and environment. Such systems should fit "graciously" (Lahlou, 2010b) into the subject's environment, without creating new problems for them to solve. This means that designers have to think about the costs of operation from the user's perspective, avoiding overhead wherever possible.

The "zeroes" list (Lahlou, 2010b) is a good way of preserving these aims as development takes place. For example, designing in this mould means aiming for zero training, zero configuration, and zero maintenance for users of their embedded systems. In addition, these systems must keep user's data private, and be highly secure to prevent abuses. In the long run, these minor up-front development commitments are what will keep people actually using the system. These principles were kept in mind during the design process.

Requirements and aims

This section outlines the requirements identified. The design process followed, described in figure 16, meant that the exploratory discussion with the charity intersected with the study's research aims. So an outline of the kind of variables needed was presented to the collaborators, we discussed how this would suit them, and proceeded to prototype and test on that basis.

Many of these requirements overlapped; for example, the charity desired that the software tool should be quick and easy to use for pragmatic accessibility reasons. Speed and simplicity had been identified earlier as a way to maintain proximity to events, and minimise recall effects, as recommended in the experience-sampling literature. Following both principles implies that the interface needs to be easy to understand, and swift to operate. The charity underlined their own need for a tool that is shorter and more accessible than a paper form, which matches the aim of reducing the burden on participants for methodological reasons.

In addition, to support the theoretical aim of testing a situated model, the software must enable capture of information about situation as well as person, time, and experience.

Using an experience-sampling approach necessitates the ability to signal participants, this study is seeking to address a gap in the experience-sampling literature by reducing the burden of participation. This will be clear in the technical design, and through the experience of participation.

Here the success of the approach will be decided primarily through feedback from the participants, but also by exploratory investigation of the suitability and validity of the data gathered.

Methods

This section looks in more detail at the issues involved with the participant organisation, the technical design approach followed, and the specific instruments chosen.

Participant organisation

Engaging with disadvantaged individuals, whose experiences of stress and anxiety may be particularly acute, is often challenging (M. Jensen, 2014). This study involved working with a charity that focuses on helping these kinds of individuals by offering volunteer opportunities on a farm and associated woodlands. The idea is that by making these connections, the participants will experience feelings of social inclusion, improve their health, and perhaps broaden their mental horizons. Individuals involved are from disadvantaged backgrounds, and usually have mental health issues such as depression.

The charity were interested in taking part in the study because they were unhappy with the costs and cross-sectional nature of their current assessment methods. They also felt that the opportunity to see changes in anxiety and stress in specific situations, such as working in the city versus volunteering on the farm, would give them insights into the success of their programmes.

Methodological choices

The software was designed to collect measurements at both person and event level (see the theoretical model in figure 15). This produces ecological data about experienced states, alongside measurements of trait-level constructs, in one application. It is designed to be self-contained and self-reliant, so it collects basic demographic information at the start of participation, and optionally, user feedback at the end. Date and time are associated automatically with every measurement, along with an anonymised user identifier. Context is measured by activity, company, and general location.

Collecting all these data allows this study to address the theoretical questions about the regulatory effects of traits in momentary experience, and to measure contextual variation. The specific scales, below, were chosen for reliability and validity.

Measures

Person-level anxiety: Participants are asked to fill in the Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) (Spielberger et al., 1983) at the start and end of the study (figure **??**). The scale consists of twenty items that measure 'typical' feelings of anxiety, rated on a scale of 1 to 4, to give a total score of between 20 and 80.

Event-level stress: For short repeated readings the Six-item State-Trait Anxiety Inventory (STAI-6) (Marteau & Bekker, 1992) was used. This is a short version of the Spielberger State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983), and measures how anxious the subject is feeling in the moment of observation. It is also scored from 1-4.

Situation: Each observation also asks for contextual information. The three measures of situation were developed these from Kahneman's (2004; 2006) studies of daily life, with additional feedback from the charity during the design process. They measure activity, location, and company, through pre-written lists of choices. Full details are given in the introduction to this thesis in tables **??**, **??**, and **??**.

Time: Exact time and date information is stored at the time of the reading.

Table 3: Variables and categories used

Activity	Social context	Location
Commuting	Alone	At home
Computer/e-mail/Internet	Family	At work
Eating	Friends	Growing Connections Project
Exercising	Strangers	Somewhere else
Napping	Workmates	
On the phone		
Preparing food		
Relaxing		
Relaxing outdoors		
Shopping		
Socializing		
Volunteering		
Watching TV		
Working		

Technical choices

Following the requirements elicitation, a beta version of the app was created for final usability testing. It was built to run on an iPod Touch or iPhone capable of running iOS version six and higher, and does not need to be kept actively running for reminders to pop up as scheduled.

The system runs on a client-server architecture with multiple points of redundancy. Complete data is stored both locally on the device and online, which serves the dual purposes of secure backup for the users, and easy access to the latest data for researchers. It is theoretically possible to undertake the whole process remotely, from installation of the software to feedback and final readings, but this would have removed the collaborative elements that formed part of the design process.

Storing local and online versions of the data allows the software to be used even when there is a lack of internet connection or even infrastructure. In an earlier prototype, a user took readings during a week-long retreat in the remote North Yorkshire Moors and simply uploaded their data when they returned. The app does not require much bandwidth for data upload, which may be expensive for some users.

So the recorded data is accessible by three methods; the online database as described above; export of the data as a set of comma-separated values (CSV) files; and finally, the researcher can physically connect the device to a computer such as a laptop, and copy the database over to their desired storage location. The interface was designed in the Xcode software environment, which allows some visual editing of the flow of logic and the process of using the application (see figure 17; the dark grey screen on the left with white buttons is the starting point). The code for the backend software routines such as score summations, reminder calculation, and so on, are written in Objective-C.

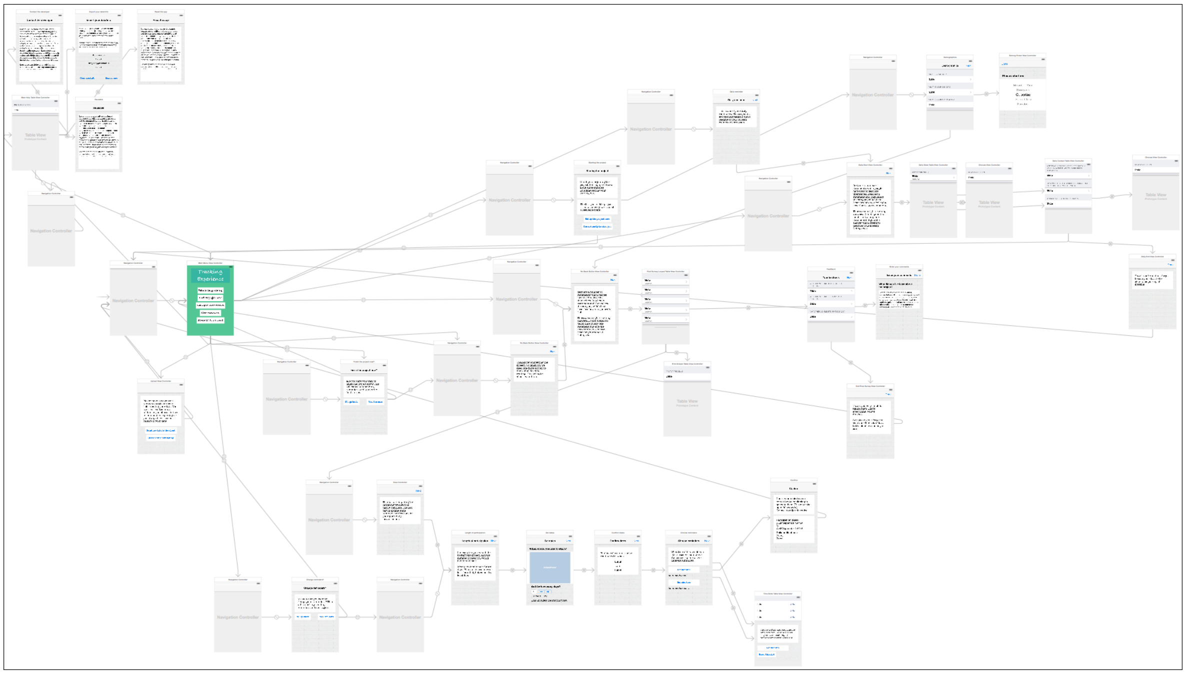

Figure 17: Overview of application logic in Xcode

Reminder alerts are specified by the user, who is able to choose start and end times for up to two sessions during days of participation, e.g. morning and evening, or just one long session for the whole day. The system prompts them to choose between one and ten daily notifications, which are set at semi-random intervals throughout the period specified. For example, with five notifications set for a four hour slot, the time is divided into five 48-minute sub-slots, and then within each sub-slot the exact alert time is generated completely at random. This gives flexibility in choosing participation times, whilst also reducing the burden of taking part by ensuring no reminders pop up at unwanted times.

Recording measurements with the software takes just a few seconds from start to finish. In the final test version, it took about five minutes to completely install and set up the project, and about the same amount for the final measurements at the end of the project.

Procedures

Deployment

The application was distributed on two iPod touch devices pre-configured by the developer. This was because this as an untested beta with just two subjects, it was faster to install on an individual basis. Later projects would use Apple's TestFlight application manager, which is much more integrated and automated.

Study details

The two subjects were administrators for a mental health charity interested in applying momentary techniques to their assessments of the individuals they work with. They each participated for at least twelve days, to get an idea of the system in use (table 4).

Table 4: Participation summary

Participant	Days participated	Total observations	Average daily observations
1	32	114	5.2
2	12	39	4.5

Results

This part of the paper begins with an analysis of the reactions and observations of the participants to the instrument. The results are explored, and implications for the adequacy of the model are considered. Finally there is a discussion and suggestions of possibilities for future research.

Feedback and assessment

This section is an examination of how well the aims of this study were satisfied. Were the usability goals met? Were there efforts to match the client requirements? And were the theoretical and methodological objectives achieved? This section combines the quotes and feedback from the iterative testing sessions, and the final debriefing session.

Breifly recapping, the aims were that taking measurements should be quick and effortless, for avoiding recall bias, for reducing the burden of participation, and for addressing participant concerns about accessibility. The software should capture contextual data as well as experiential information, supporting the charity's custom situation requirements. And it should require no maintenance or configuration by the participant, so that it fitted comfortably into the user's environment.

From a usability standpoint, the feedback was positive. They found it “easy to use”, and thought other participants would find it easy to use too. They thought it would be good for people who don't like “writing and forms”. They enjoyed the self-reflection, and taking this all together, the general conclusion was that the software tool was "comfortable". They were particularly interested in the possibility of receiving "feedback in real time" (Ebner-Priemer & Trull, 2009); that is, guidance or guidelines when averages were exceeded or a particular threshold met.

> [Participant one looking at the timeline]
>
> *Yes, this is just what I'm like!*
>
> [Participant two]
>
> *I think everyone would agree with that!*

Both found taking part "interesting", and they considered that the trait and state level readings matched with subjective assessments of their own characteristics, both by friends and by themselves. On the negative side, they found it hard to remember to charge a separate device and found it kept running out of battery. This was a particular problem for participant two, which we see reflected in lower numbers of measurements (see table 4). They suggested that it would be less of a problem for participants who had their own device, like an iPhone, which they would naturally charge and carry with them. They found they were more worried about taking care of the device (i.e. not breaking it) than they would have been with their own phone or iPod. They also sometimes found categorising experiences difficult, and reflected that perhaps an initial qualitative assessment would have helped with the coding, and made it easier to classify their situations.

> [Participant one]
>
> *Sometimes I ended up choosing a category almost by default, because nothing was quite right.*
>
> [Participant two]
>
> *It was quick, though.*
>
> [Participant one]
>
> *Yes, it was quick.*

Overall, the design and testing phases went smoothly. There was a minor technical hiccup a week into prototyping, which took a day or two to correct. The main drawback with this kind of embedded development was a difficulty in distinguishing 'training and configuration' from design practice, when it is iterative and consultative. There was a lot of discussion which inevitably touched on usage and the technical design, and ideas were passed back and forth.

The next section describes the data gathered, assesses its effectiveness, and sets out some exploratory analyses.

Empirical

The instrument contains questions which ask about both state of stress and relaxation, which are in opposite directions on the same dimension. So a high score on a relaxation question would translate to a low score in momentary stress, and vice versa. Therefore, before analysing measurements, some of the reported scores need to be reversed. Luckily it is straightforward to code the software to do this automatically, so no misinterpretation can result from failing to transform scores for particular questions.

These transformed scores are added up to give a single number from 6 to 24 for momentary stress (six questions), and 20 to 80 for trait (20 questions). Here higher scores indicate more anxiety, or more tendency to anxiety, respectively. In the descriptive statistics which follow, the summed state scores are scaled for readability by dividing by each by six, resulting in a scale of one to four with the same interpretation as the non-scaled scores. The summed trait scores are divided by 20 for the same end. These are the breakdowns over the situational variables.

The size of the markers in these graphs represents the frequency of each type of context; the larger the marker, the more readings had this context. For example, in figure 18, relaxing and working are the most common activities for one participant.

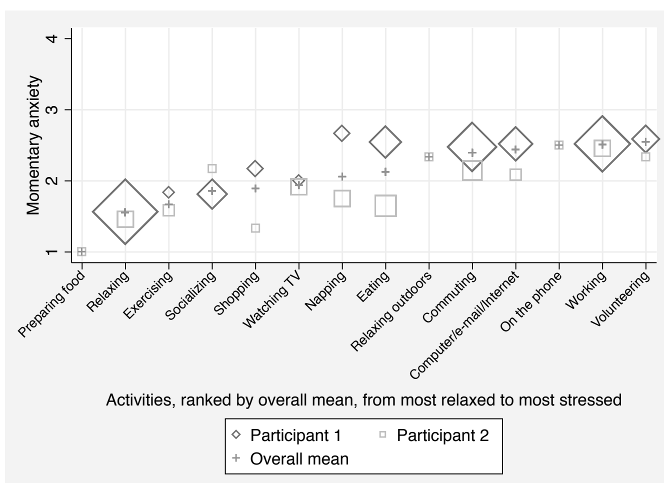

Figure 18: Activity ranked

The activities are ranked left to right by overall mean, indicated by a cross, with each participant's own experiences represented by a diamond or a square.

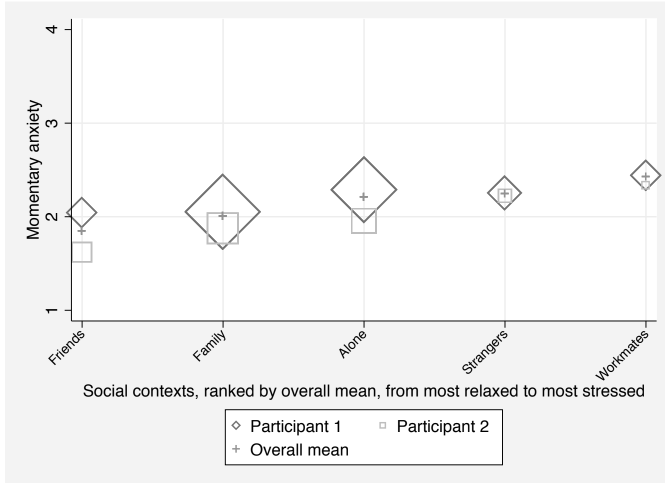

Figure 19: Contexts ranked

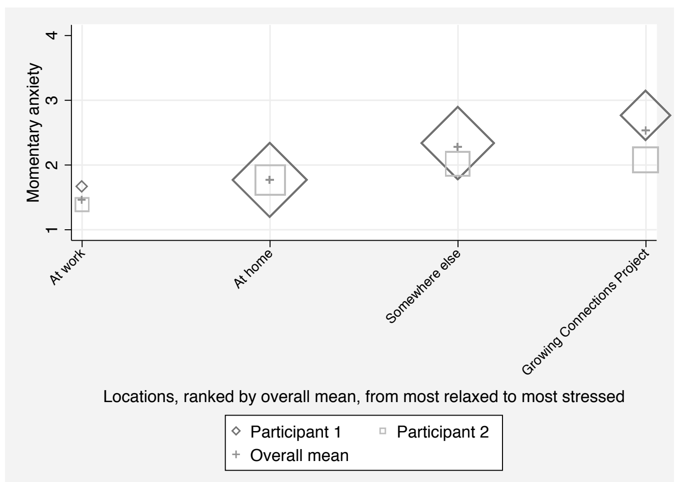

Figure 20: Locations ranked

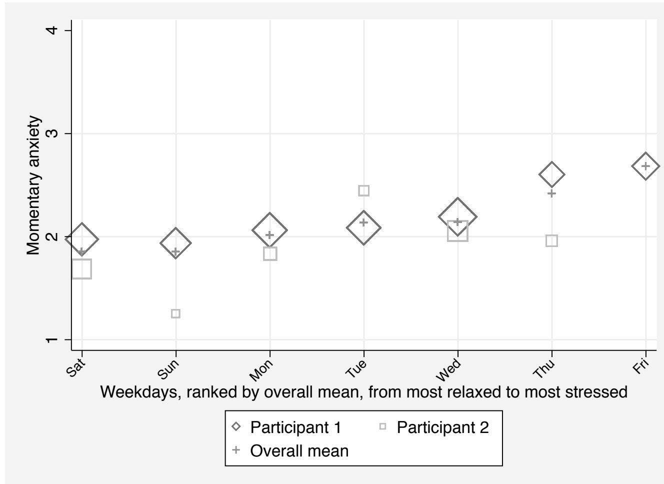

Figure 21: Weekdays ranked

Figures 18 to 21 show notable differences in both frequency and stressfulness of the different situations between subjects. This suggests an influence of temperament, although there are also clearly influences of event and context that are comparable across individuals.

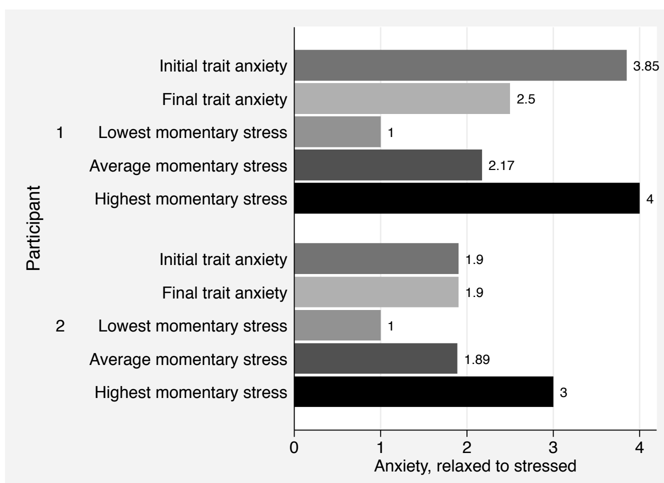

Figure 22: Overall

Supporting this suggestion, it can be seen in figure 22 that each participant had different traits and average experiences, and as we saw earlier, the participants found these readings to be a credible reflection of themselves. These data also appear to be more than adequate for distinguishing situational characteristics.

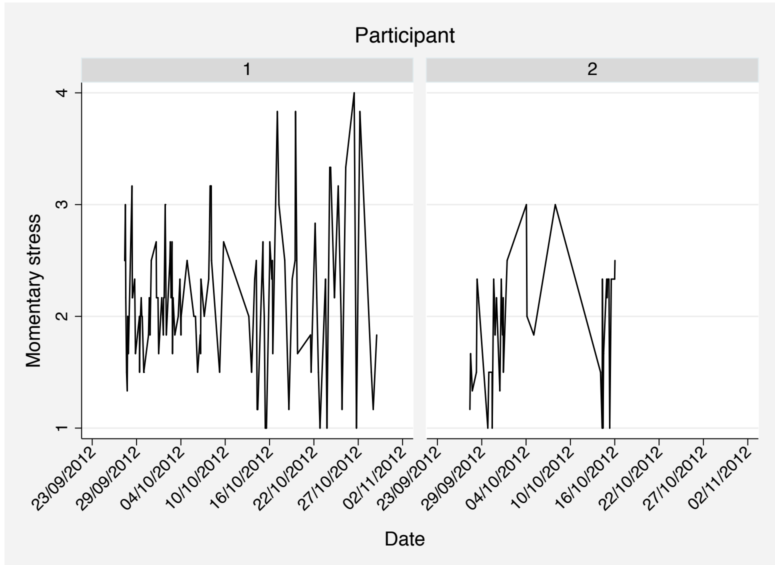

Figure 23: Timeline sample

Finally, we can see considerable fluctuation in this simple timeline of ratings (figure 23). It serves to underline the importance of intensive measurement, and suggests that further investigation of time, perhaps grouping measurements by period, may be useful.

Summary and recommendations

This study investigated two ideas. The first was methodological. Testing a customised, iterative approach met with success in numbers, and good feedback from the participants involved. This may be a promising area for health professionals given the low cost for the organisation. The emphasis placed on simplicity and speed, with a focus on 'zero' (Lahlou, 2010b) training and installation/configuration, worked as designed. Furthermore, it suggests that for researchers with minimal programming skills, for example writing loops or logic (if then else, etc.) in R or Stata, the process of developing an app is not much more complex. It is time consuming, but Apple's Xcode environment makes as much as possible drag and drop. Or for a pre-built solution, there are also many of these available; the introduction to the thesis contains a dedicated section that is not reproduced here.

The second was a psychological question: do situation and trait each play a role in determining feelings of stress in specific experiences? Initial, exploratory analyses seem to back this up, and suggest that this model will support further investigation. However, these data remain exploratory and highly individual, so it is not really appropriate to draw general conclusions.

Within psychological and social science, smartphones and mobile devices have become a fast-moving area of research. This study tries to demonstrate that it is not yet so fast that only the vanguard can benefit. It is tentatively suggested that scope remains for a range of successful studies, from qualitative to quantitative, without hefty investment in technical development.

Preface: Paper Two

Having successfully developed a suitable tool for administering the types of measurement necessary for assessing my theoretical model in the first study, this second paper tests the software and model in a new environment.

In the first study I benefited from one-on-one work with the participants, and was able to tailor my approach according to their feedback. This paper proceeds in a very different way, as I recruited a larger sample from a convenience population of students. Participants had no particular investment in the success of the study, apart from a small financial inducement. Rather than customising the situations, I sought to generalise them, and draw broader conclusions about the utility of the instrument to be used throughout this thesis.

This study also functioned as a trial for remote deployment of the measurement software. Again this was a distinct contrast to the method used in paper one, where pre-configured devices were handed over and then retrieved at the end of the study. In that sense, it is a truer test of the fitness of the design for a general user's environment and an assessment of whether it can be truly embedded, or if mismatches between instrument and environment provide too much friction for success. And trying the technology with a greater load proved a natural test of its robustness.

But some lessons persist from the previous study, such as an emphasis on utilising users' own devices, which are already fitted into their lives (so they would remember to charge the battery, for example). In addition, I kept the data structure and the simple, but effective, interface from the first study. The positive findings in this paper meant that this software would serve virtually unchanged for the third and fourth paper, although the third paper would add some complementary observations.

Embedding measurement with the participant: Using mobile phones to remotely observe daily experiences of stress

Abstract

Mobile technology offers considerable opportunities, especially in the field of momentary measurement, but this may come with a commensurate complexity. This study investigates proposals from experience sampling and other research into ambulatory assessment, and applies them to the problem of measuring immediate states of stress. This enables me to test both the costs and savings of using mobile technologies, and to draw some inferences about the place of ecology in framing stressful experience.

Introduction

Modern technology can be something of a Pandora's box for researchers. With every claim of improved efficiency, researchers often find hidden costs in additional complexity, or expensive devices and software. Some researchers claim that leveraging the spread of the 'smartphone' can offset these technological costs (Eagle & Pentland, 2009), but is this really true?

Smartphones are essentially pocket-sized, touch-enabled computers. Following the introduction of the iPhone in 2007, which did away with legacy technology like keyboards in favour of on-screen controls, competitors swiftly followed Apple into this new market for advanced devices. The power of these devices enables applications previously thought impossible. In addition, their ability to run native applications built by anyone, subject to certain restrictions, has greatly broadened researchers' access to populations that were previously hard to reach, such as those with mental health issues, or drug users (Verdoux et al., 2003).

In this paper I examine the claim of large improvements in implementation, deployment, and participation when using these technologies (Eagle & Pentland, 2009; Raento et al., 2009). The study also makes a contribution to understanding the processes of momentary construction of experiences of stress, by bringing together research on event, context, and person. Using an adaptation of the experience sampling methodology (ESM), I was able to test my psychological hypotheses on a convenience sample of students.

Background

Experience sampling

Experience sampling, as the name suggests, was inspired by the idea of combining statistical sampling techniques with self-reports of experience, taken at the time events occur (Hektner et al., 2007; P. Wilhelm et al., 2012). Hektner and Schmidt (2007, pp. 8–16), for example, were inspired by pagers, which they adapted to signal participants to fill in paper surveys on current experience. Nowadays it is more common to use a mobile phone or personal digital assistant (PDA).

ESM allows researchers to take 'intensive' measurements (Bolger & Laurenceau, 2013, pp. 5–7), capturing their fluctuations and sequences, over a period of days or weeks (Barrett & Barrett, 2001). As the instrument is embedded in the subject's daily life, these measurements are therefore situated in the real-life contexts in which the situations occur. A similar approach is taken in medical science when, for example, a patient wears a portable electrocardiograph (ECG) to closely track changes in heartbeat. These approaches can even be combined (e.g. Fahrenberg et al., 2007; P. Wilhelm et al., 2004) to link physical changes to experiences.

ESM produces high numbers of measurements per participant, usually very rich in event details (Fleeson, 2007). While this is usually seen as a strength of this type of research, researchers new to intensive longitudinal studies have to make sure that they take into account the more complex analyses that are often required for unbalanced, within-person data (Bolger & Laurenceau, 2013, pp. 14–21; Shiffman et al., 2008), as well as the demands they may make of participants. It is with these issues in mind that I look into the claims of modern technology for reducing certain of these research costs.

Claims of improvements

In the past, ESM-type methodologies have been noted for their high costs of participation, for researchers as well as for participants (Gunthert & Wenze, 2012; Shiffman et al., 2008). Expensive and fragile equipment, like PDAs, had to be purchased and carefully husbanded by the researcher; participants could forget or lose them, or even drop them into the toilet (Scollon et al., 2003). Older devices were bulky and intrusive, so few of the benefits of automation and efficient data entry could be realised.

Studies using more modern mobile technologies claim to produce not only richer data and better predictions (Eagle & Pentland, 2009), but improved access to participants and easier deployment (Hasler et al., 2008). They can capture additional modalities like sound, pictures, video, and even Global Positioning System (GPS) location data.

In this study, I test these assertions by examining the quality and quantity of data produced using a mobile phone application. I propose that anything I want to measure will be captured by the application; in addition, setup and participation will require no previous knowledge of the system, or support from the researcher. The application will also collect any feedback subjects may have following participation.

In the next section, I describe the type of data this study needs to collect, and the kind of results I am expecting to see.

Theoretical model

How do events influence moment-to-moment experience? In my analysis, I use a model which allows me to explore these influences at the level of events and persons. Grounding events in their ecological context, or situation, requires an exploration of the place of situation itself (P. Wilhelm et al., 2012). We need to move beyond mental process alone as an arbiter of experience (Bechtel, 2009).

I examine the idea of a stable influence of contextual factors; that is, the idea that events, locations, and social context have specific characteristics, and therefore will show a measurable difference across person.

My model takes the idea of 'embedding' processes of experience formation in the surrounding environment from situated cognitive approaches (Robbins & Aydede, 2009). This elevates ecological factors to the same level as internal processes. Events arising from external stimuli are moderated by the context of the event, as well as an individual's temperamental tendency towards low or high reactivity (Kagan, 2010, p. 11), and their learned biases. For example, an individual with high reactivity is seen as more likely to experience anxiety in a given situation. Anxiety and stress have a substantial impact on health and quality of life (Costa, 1996), and there is a strong body of research applicable to momentary experience (Szabo et al., 2012).

To model activities and contexts, I follow Kahneman's innovative work on experience in daily life (Kahneman et al., 2004; Kahneman & Krueger, 2006). Using a large sample, he extracted broad categories of events that were still specific enough to have a noticeable influence on experiences. As in his research, I therefore expect to see lower ratings of stress and anxiety for social and leisure activities such as relaxing or exercising; conversely, this research suggests that working and commuting will be found more stressful. Locations and social contexts that are work-related should also be rated as more stressful than family and home.

Study aims

My general aims, therefore, are twofold. Firstly, to test the claimed efficiency and self-reliance of mobile data collection. This will become apparent in retention rates, participant feedback, and levels of participation (i.e. number of measurements). Deployment and collection should be as trouble-free as possible.

Secondly, I expect to see support for my model that places an emphasis on situations as well as cognition. Explicitly, there should be differences that are largely applicable across person in experiences of stress in daily life, divided by activity, location, and companion. I expect to see evidence of differences in levels between person as well. The framework for determining success is outlined in the following aims.

Aim one: I expect to see higher numbers of measurements, more than one occasion per day. I suggest three occasions per day, or fifteen measurements in total, as a contrast to reconstructed methods which take one retrospective measurement to stand in for day, week, or month. In a complementary aim, to get these larger numbers of observations will require longer participation than in traditional studies. I suggest at least a week.

Aim two: The study should have low drop-out/non-completion rates, implying less of a burden on participants. This should be lower than typical experience-sampling rates of 50%. Reinforcing this, there should be subjectively positive feedback. In particular, any mentions of burdens or barriers should be noted as evidence that the instrument may not be functioning as promised. There should be a subsidiary matching subjective assessment by the researcher on ease of administering the experiment, and any recurring problems.

Aim three: There should be observable differences in levels (i.e. means across subjects) for measurements of momentary stress by activity, by social context, and by location. Furthermore, I suggest that there will be at least tentative evidence for common-sense rankings of factors; for example, working should be observed to have a higher mean momentary stress than relaxing, being at home should have a comparatively reduced mean stress (Stone et al., 2006), and weekends on average less stressful than weekdays (Harvey et al., 2015; Ryan et al., 2010).

Aim four: There will be differences in mean momentary stress observable between subjects. There will also be differences in trait anxiety between subjects, and some evidence of association between mean momentary stress and trait anxiety (Strelau, 2010).

I explore these ideas in the following sections.

The study

Measures

Trait anxiety: I asked subjects to fill in the Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) (Spielberger et al., 1983). It consists of twenty items about feelings of anxiety in general, each rated on a scale of 1 to 4, to give a total score of between 20 and 80.

Momentary stress: For the short readings I used the Six-item State-Trait Anxiety Inventory (STAI-6) (Marteau & Bekker, 1992), which is an abbreviated version of the Spielberger State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983). It asks the subject to rate how anxious they are feeling at the moment of observation. It is also scored from 1-4, with a total score of 6-24.

Situation: Each observation also collects self-reported contextual information. I developed these from Kahneman's (2004; 2006), which were coded from a large sample. This asks for three answers, and allows you to select your activity, location, and company from a list of choices.

Time: Full date and exact time are stored automatically by the software at the time of the reading.

The application

The application runs on Apple's iPhone, iPod Touch, and iPad devices (iOS). This gives participants a way of verifying the integrity of the study, as iOS applications must be certified by Apple, even in beta, and are distributed through Apple's own TestFlight installation service. Further, using TestFlight makes installation straightforward for end-users: they simply register a device for the beta and are securely added to the approved installation list, then download and install it like any other application.

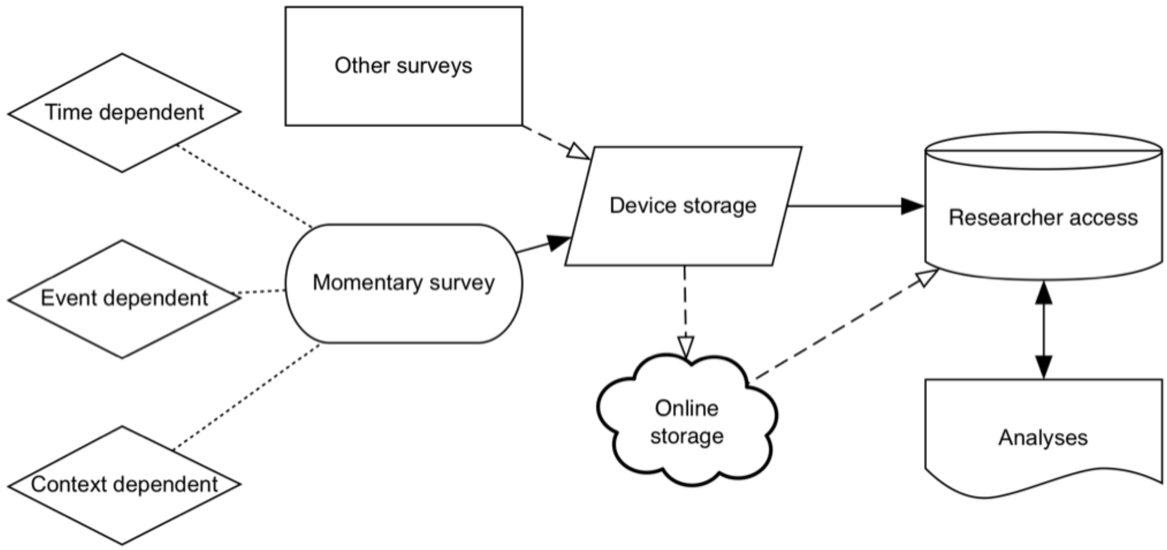

Figure 24: Data collection, storage, and access

Figure 24 shows how this process works. The participant runs the application to take an initial survey and set up preferred times for notification. Reminder notifications are displayed automatically at the appropriate times, whether the application is running or not. Tapping on a reminder loads up the application and prompts to take a short reading. After the required number of readings the final survey is administered.

Recruitment

I invited students to sign up for my study through a dedicated website and mailing list, with the incentive of a five pound Amazon voucher for all participants. Once a participant registers with the list, they are sent instructions on installing the app on their device, as described above. They simply need to upload their data to the cloud at least once after finishing the study, which can be whenever they have internet access.

Participants Of the 110 participants invited at my presentation, 35 signed up for the mailing list. 16 of those joined the beta, and nine fully completed the study. This gave a total of 709 observations, an average of just under 79 per subject. The participants were five females, three males, and one not specified. Ages were between 23 and 56. In the next section, I explore these data, the results, and their implications for my theory of experiences of stress.

Results

Reliability

The scale used, the Six-item State-Trait Anxiety Inventory (STAI-6) developed by Marteau and Bekker (1992), is comprised of three "anxiety-present" and three "anxiety-absent" items chosen for their sensitivity. In the original study, Cronbach's alpha was 0.82; here the alpha (unstandardised as all measures use the same scoring system) was 0.85. We were looking for 0.7 or higher (Nilsson et al., 2012; StataCorp, 2009a), so this demonstrates respectable internal consistency.

An examination of the construct using an exploratory factor analysis gives further support for the scale, with results consistent with a one factor solution. The first factor has an eigenvalue of 2.89, explaining 105% of the variance (allowing for some negative values). The second factor had an eigenvalue of 0.31 and explained 11% of the variance. There is good evidence for the validity of this scale.

Summary statistics

Table 5: Summary of momentary stress over all participants

Average	Std. Dev	Min	Max	Obs
1.94	0.62	1	4	709

As expected, analysis of the ratings of stress by subject reveals different means, lows, and highs. Figure 25 shows how this breaks down for each of the nine subjects.

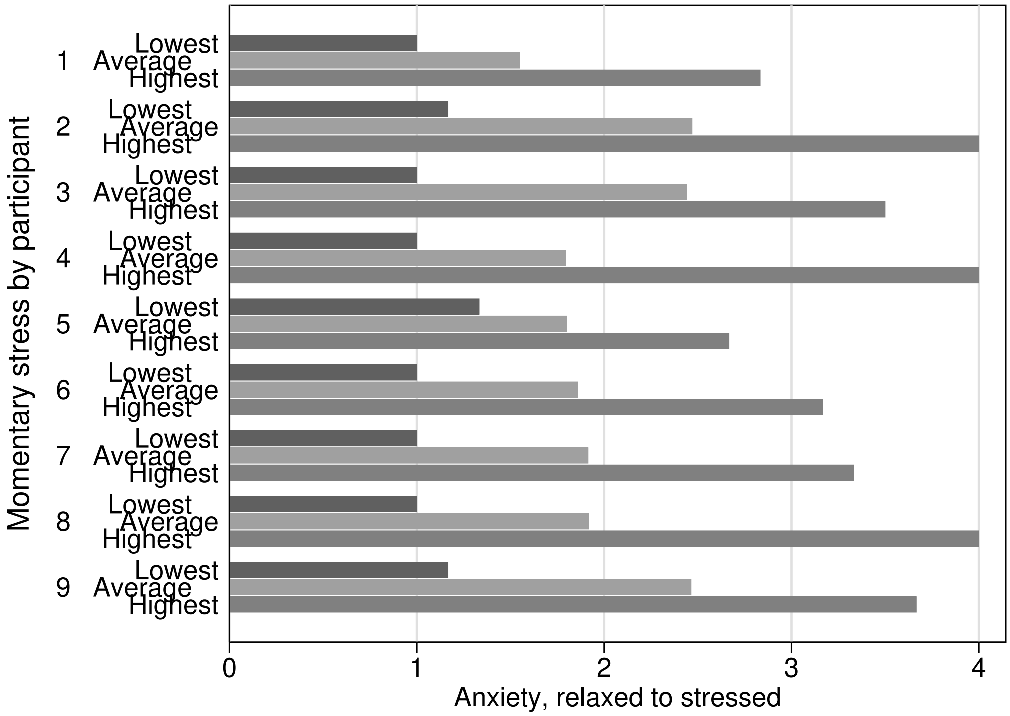

Figure 25: Mean, minimum, and maximum momentary stress by participant

We can see that means and highs vary a good bit between participants. Lows are less variable. Figure 26 shows individual differences in trait anxiety.

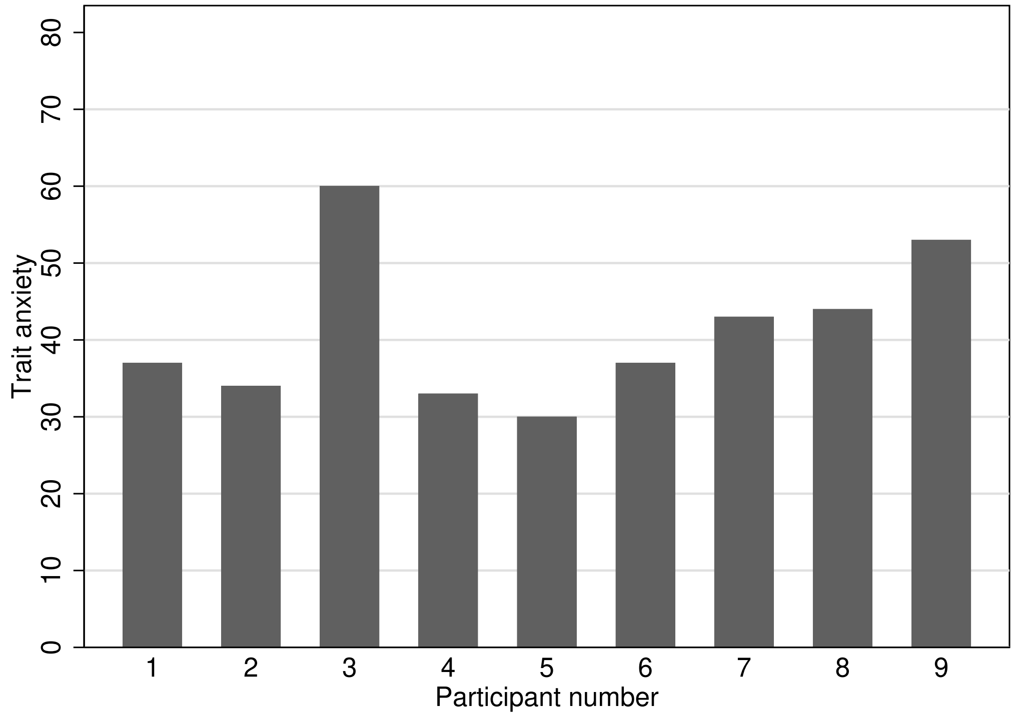

Figure 26: Trait anxiety by participant

There are different levels of general tendency to feel stressed or anxious apparent by subject, which suggests that the instrument is capturing something tangible; something distinguishable. This can be tested against my theoretical model, which implies that trait anxiety should have some observable effect on momentary experiences of stress. Therefore, to explore this more closely, I consider the subject-specific means of momentary stress set against the scores for trait anxiety for each participant (figure 27). For this comparison, the trait scores are on the same scale as the momentary scores - between one and four points - scaled by dividing each score by 20. There is a dashed horizontal line to represent the overall mean stress for all participants at 1.94 points (see table 5 for more summary statistics).

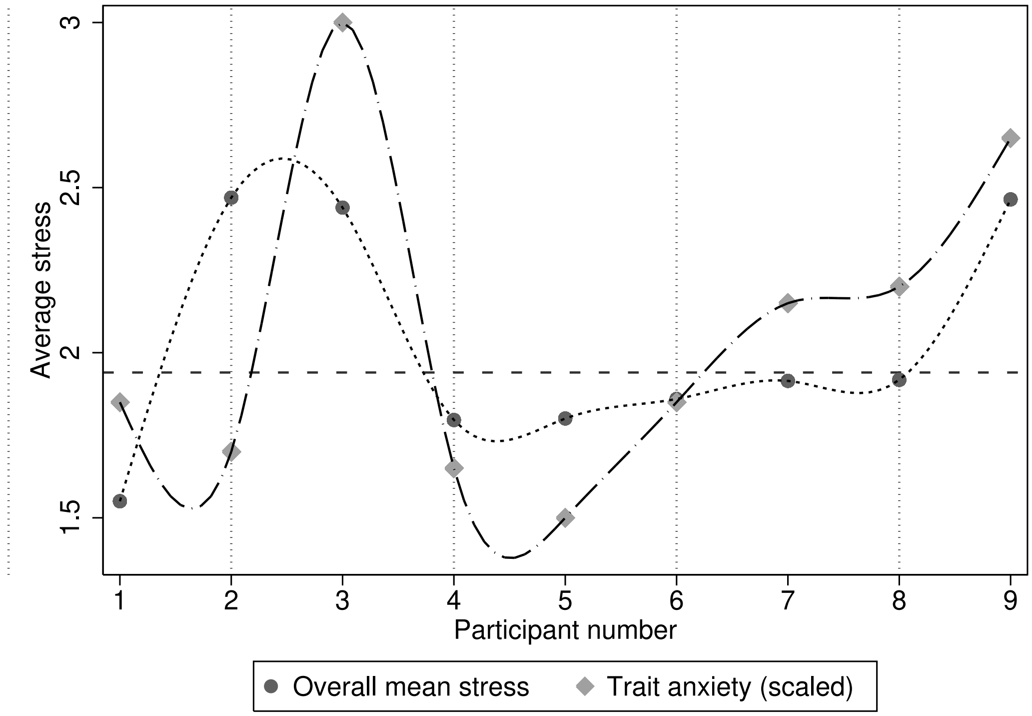

Figure 27: Mean state compared to trait anxiety by participant

With one apparent exception (participant two), trait and subject average stress look like they tend to cluster together. Subjects higher in trait anxiety appear liable to have higher average stress. This is reasonable evidence of an association, and provides support for the validity of my proposals about the association between stress and temperament. With these individual differences in mind, the next section looks at similarities in overall experiences, distributed by self-reported situations.

Analysis of situational factors

In this section the self-reported situational factors are divided into situation-specific mean experiences, and ranked. This enables comparisons with earlier studies (e.g. Kahneman et al., 2004; Kahneman & Krueger, 2006; Stone et al., 2006). In figure 28, for example, we can see that the ranking of activities runs as predicted, from leisure activities as least stressed, to working and commuting as most stressful.

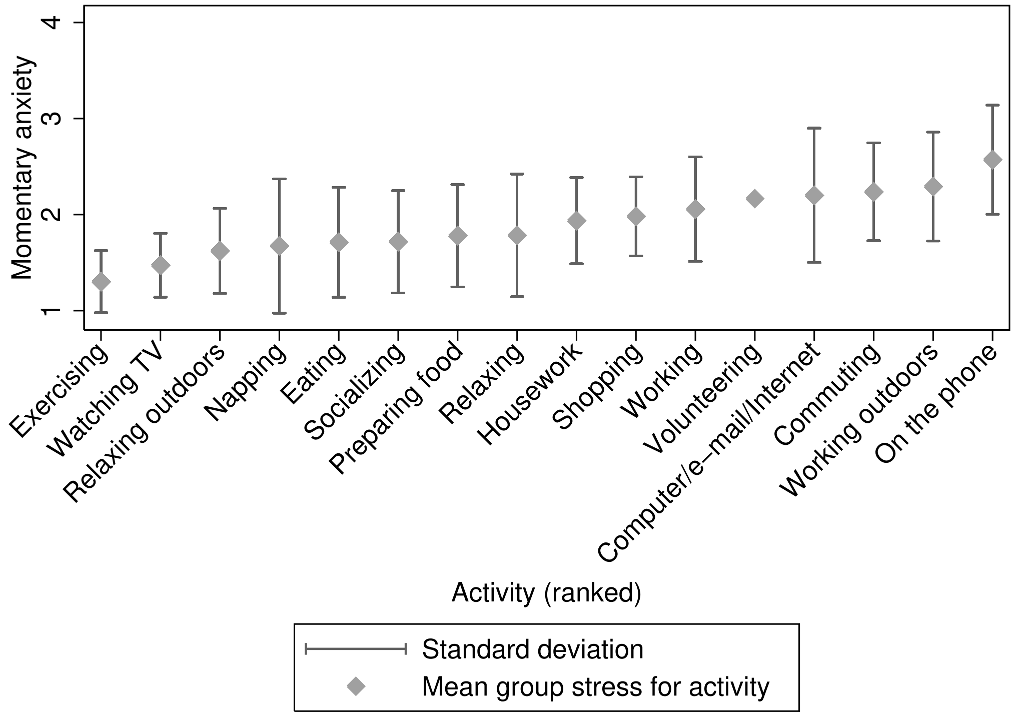

Figure 28: Mean anxiety by activity, with standard deviations

Then in figure 29, we see that 'At work' is the most stressful location, also as expected. Following this, figure 30 shows that being with workmates, alone, and with strangers rank as the most anxiety-provoking company. These results all match up with the predictions made at the start of this study, and support the aims I summarised at the end of the first section.

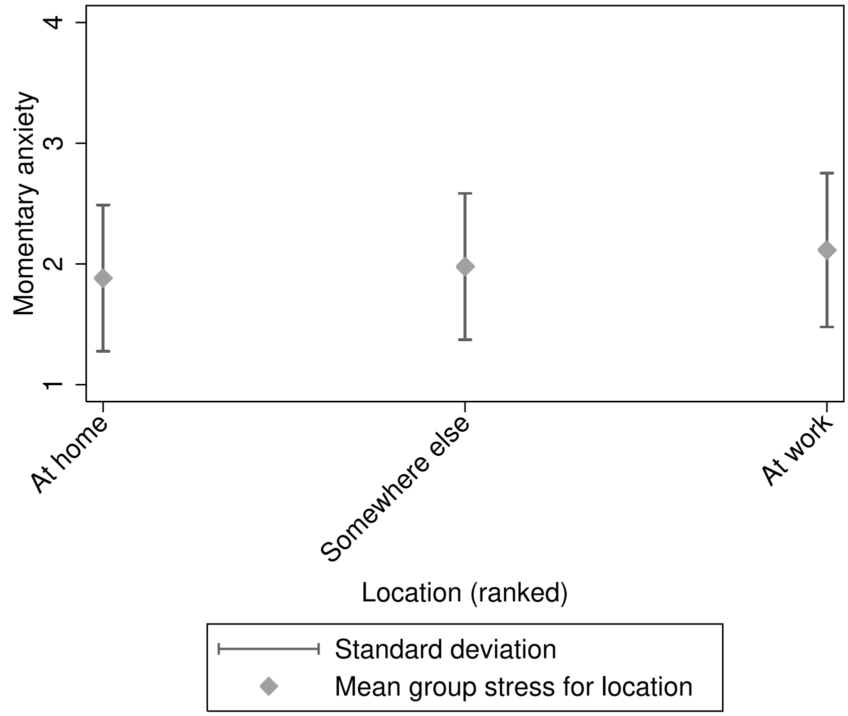

Figure 29: Mean anxiety by location, with standard deviations

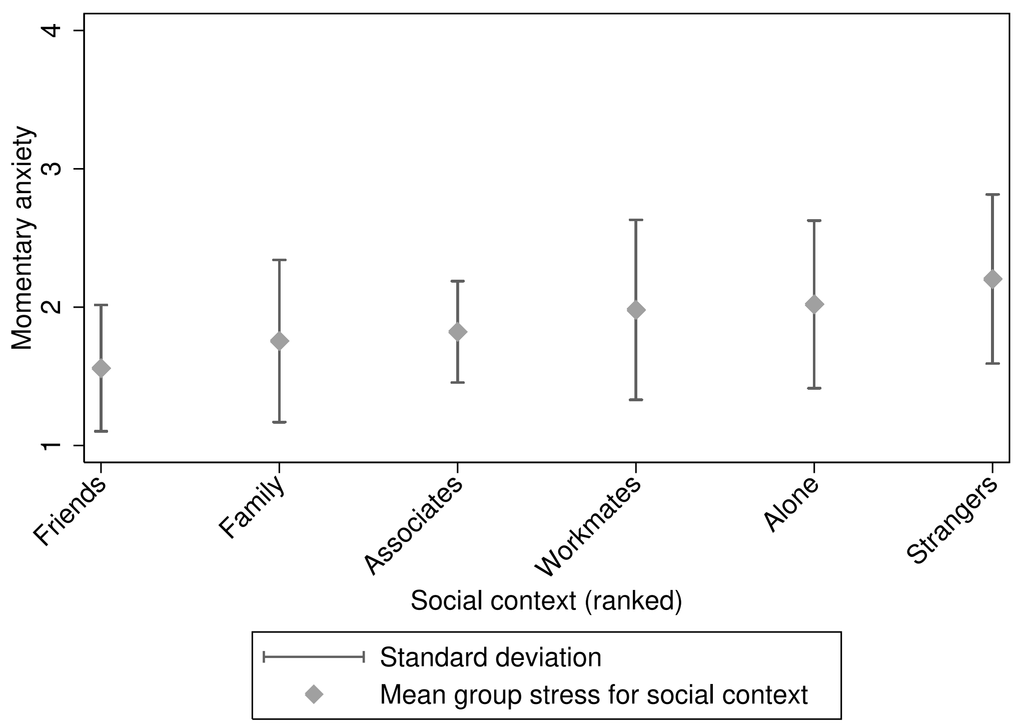

Figure 30: Mean anxiety by social context, with standard deviations

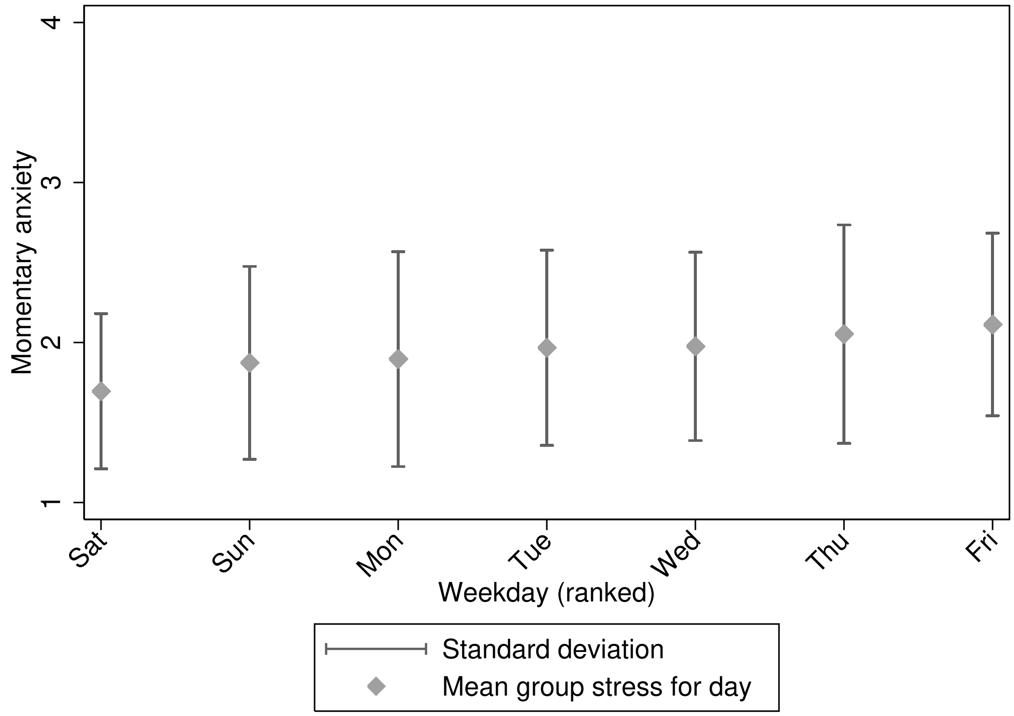

Figure 31: Mean anxiety by weekday, with standard deviations

Finally, there is some suggestion of daily/weekly rhythms in figure 31. Again, these line up as expected, with the weekends rated as least stressful, and stress then rising over the week to peak on Friday (e.g. Stone et al., 2012). Overall, these results show good support for my model and theories.

This means that there is evidence for my assertion that situation deserves to play a role as large as person in processes of momentary experience formation. It seems to suggest that framing experience using this model in a larger study will prove fruitful.

Feedback

After the final readings, the application asks for feedback in the form "liked best, liked least". There is also a section which invites any other comments.

Four participants, or just under half, submitted feedback. Three of their comments reported under "liked best", and were positive, if vague: for example, "the app worked well". Two comments were in "other", and asked about improvements to the range of activities and contexts offered, one saying that they were too general and the other that the categories were too specific and they had trouble finding a match. This was interesting, although not particularly helpful for practical implementation.

Essentially, there is not much data here, so it is difficult to make any inferences from these results. While it is encouraging that the majority of comments were positive, there is no way of exploring the suggestions offered, and they are not detailed enough to pursue independently. The short nature of the survey questions is the main drawback: trying to make the experiment easy to take part in seems to have rebounded unhelpfully in this regard. In particular, since subjects who do not complete the study by definition do not complete the final feedback, there is no way of knowing why they dropped out. I do provide the option to 'finish early' and submit feedback, but this was obviously not satisfactory. However, this indicates that there is scope for investigation into completion rates versus detail of feedback required, I suggest. This would be possible in a more design-focussed study.

Retention

Moving on to the assessment of participation, I look at tables 6 to 8. There seems to be a good spread of observations across the various contexts, although some situations are less common than others. The least common is 'Volunteering', which had only one recorded measurement; this should probably be removed from future studies.

Table 6: Frequency of activities

	Frequency	Percent
Commuting	80	11.28
Computer/e-mail/Internet	114	16.08
Eating	60	8.46
Exercising	16	2.26
Housework	21	2.96
Napping	27	3.81
On the phone	7	0.99
Preparing food	28	3.95
Relaxing	47	6.63
Relaxing outdoors	15	2.12
Shopping	9	1.27
Socializing	47	6.63
Volunteering	1	0.14
Watching TV	49	6.91
Working	184	25.95
Working outdoors	4	0.56

Table 7: Frequency of social contexts

	Frequency	Percent
Alone	349	49.22
Associates	14	1.97
Family	98	13.82
Friends	96	13.54
Strangers	100	14.1
Workmates	52	7.33

Table 8: Frequency of locations

	Frequency	Percent
At home	431	60.79
At work	103	14.53
Somewhere else	175	24.68

More generally, as mentioned earlier, of the 16 individuals who started the study, only nine completed it fully. While this is a retention rate of slightly more than 50% (56.2%), it just clears the bar set in the opening part of this article. It is better than traditional ESM studies, in which retention rates can be as low as ten percent (Barrett & Barrett, 2001; Fleeson, 2007), but not as high as I would have hoped. To give an overview from a different perspective, table 9 breaks down the average days participated for each subject, and table 10 describes the average number of observations per day within participants.

Table 9: Summary of days over all participants

Average	Std. Dev	Min	Max
19	6.13	5	29

Table 10: Summary of measurements taken per day over all participants

Average	Std. Dev	Min	Max
7.5	5.3	1	21

In terms of participated days and measurements taken per day, this study has more than met its targets. Subjects seemed happy to participate for, on average, close to three weeks, and they generally took several measurements per day. This is good evidence for my proposals about usability and utility. However, examining the numbers a bit more closely, in table 11, we can see that there were two subjects with different patterns of participation. This is also worth looking into.

Table 11: Summary of measurements by participant

Participant	Days	Average	Std. Dev	Min	Max	Obs.
1	13	4.9	0.53	4	6	63
2	29	3.95	1.81	1	7	82
3	5	2.64	1.12	1	4	11
4	18	7.64	1.41	1	9	121
5	10	5.74	1.56	1	7	46
6	25	5.03	0.69	1	1	120
7	20	4.33	1.03	1	6	80
8	14	15.77	4.76	1	21	158
9	27	1.07	0.26	1	2	28

There are actually two sequences of lower measurement on display here. Participant three simply didn't take many measurements or take part for very long, whereas participant nine spent nearly a month on the experiment, but only took one measurement on most days. This is another place where it would be interesting to carry out follow-up interviews, so we can see exactly what demotivated these participants. Was it boredom that made participant three reluctant to keep taking measurements? Or a simple misunderstanding of the requirements? This seems quite likely for participant nine. As noted in the feedback section, there was nothing reported about these kinds of problems, i.e. that the requirements were unclear. Although this was not one of the specific aims of this study, I nevertheless count it as a failing.

Reactivity

This section looks into a related area. I examine what evidence there might be for reactivity; that is, whether taking repeated observations with their mobile phones influenced subjects' assessments of events and behaviours (Goldschmidt et al., 2014).

While relevant, a lack of reactivity was not included as a specific aim of the study as the nature of the model I am proposing, and the data I am gathering, makes these sorts of comparisons very difficult. This study proposes that additional, more complex situational factors and personal tendencies are at least partial determinants of momentary states of stress. This makes it difficult to compare like for like.

In this section I see what general patterns can be observed in the data, and if this suggests a change in levels associated with habituation. This could be in a positive direction: repetition making the subject more sensitive to their own feelings (Strelau, 2010); or in a negative direction: acclimatisation or 'hardening' to the events (Bolger & Zuckerman, 1995; T. W. Smith & Anderson, 1986).

To begin with, I picked three participants with high participation rates (see table 11), and compared the patterns in measurements for particular situations over the period of the study, by overlaying scatter plots with a fitted line for each of the three.

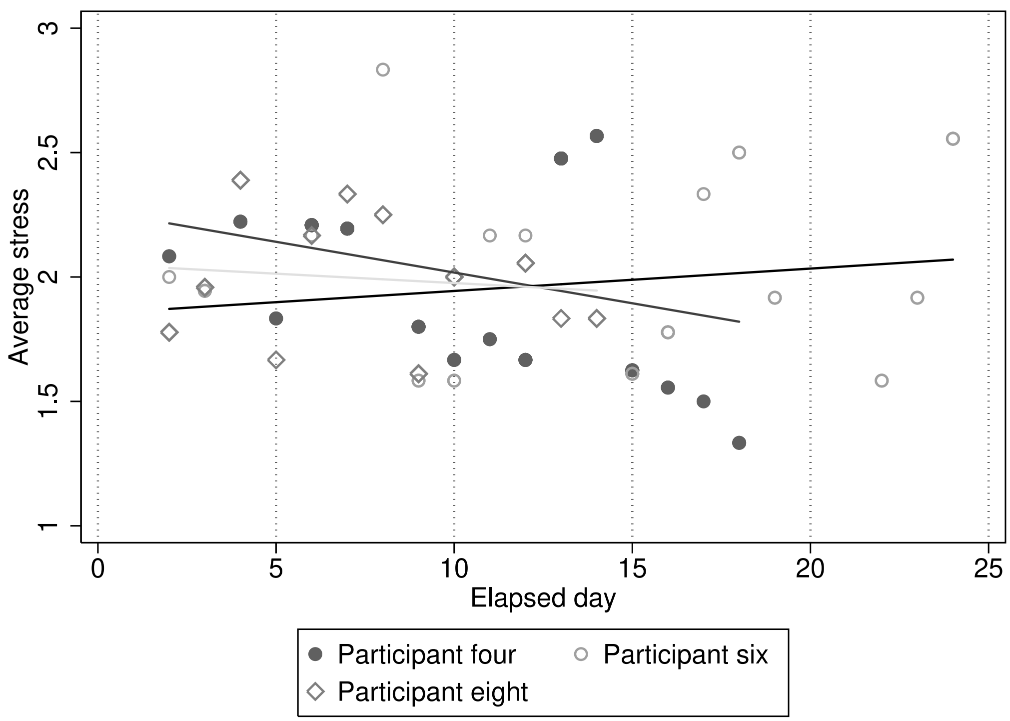

Figure 32: Average stress for 'Working' activity, three participants

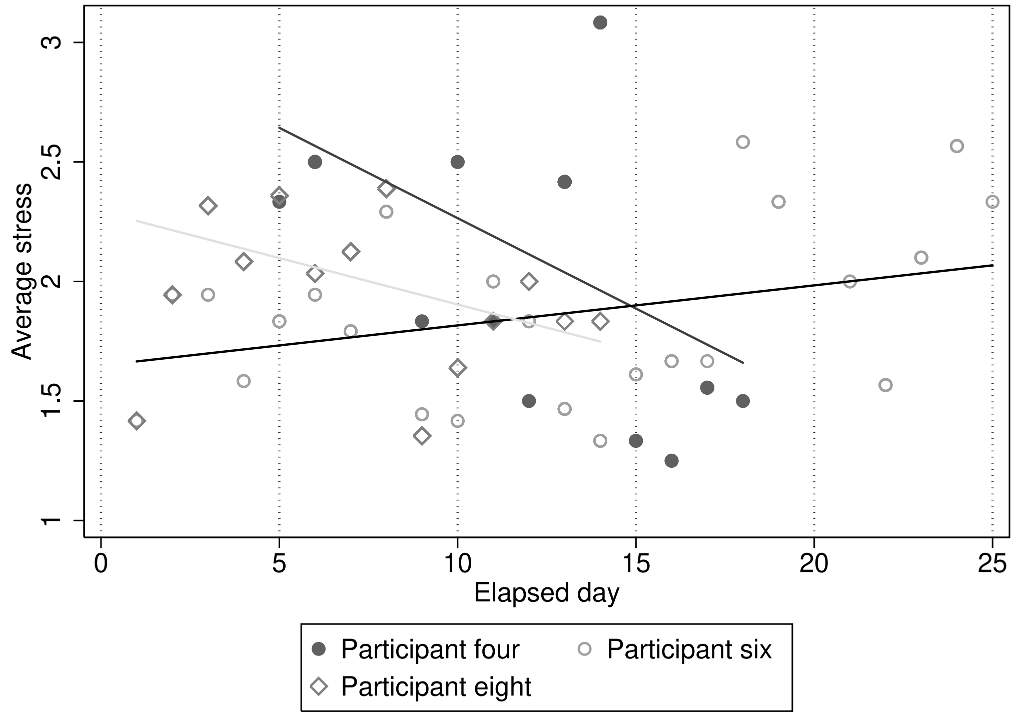

Figure 33: Average stress for ‘Alone’ context, three participants

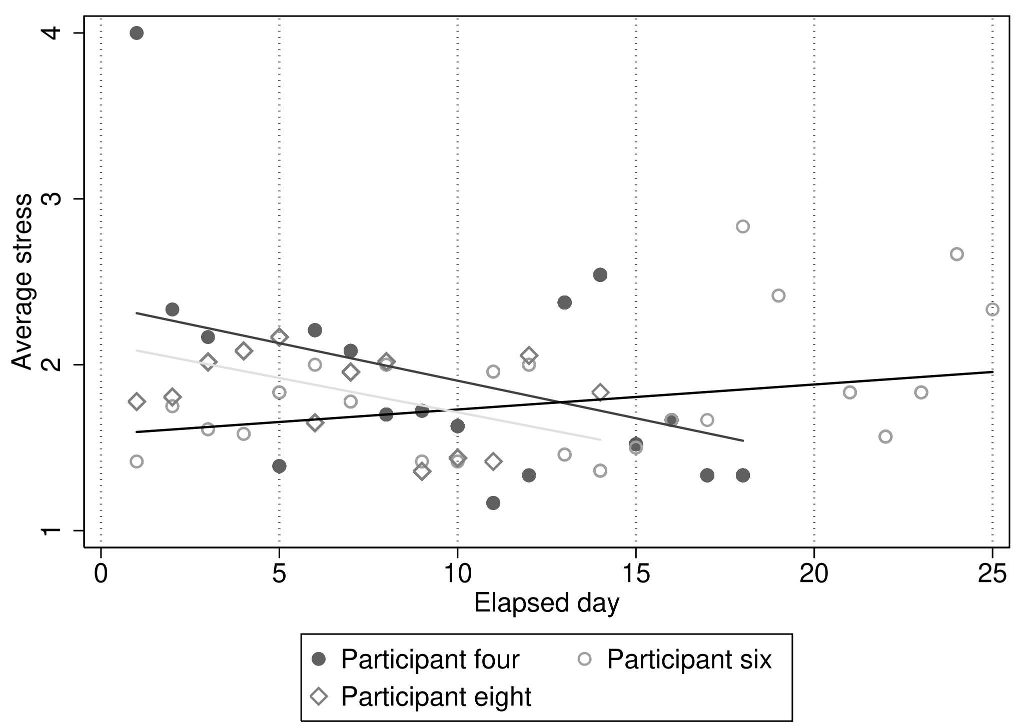

Figure 34: Average stress for ‘At home’ location, three participants

There is no definitive change in one direction, so no obvious pattern of habituation. However, results in other psychological studies generally show no significant differences in observed levels when comparing time periods (e.g. Goldschmidt et al., 2014; Palmier-Claus et al., 2011), so very strong evidence would be needed to suggest reactivity.

Some studies have observed an increase in socially desirable behaviours (and a decrease in undesirable ones) when observations are triggered by events (Stein & Corte, 2003). Although my study asked for observations at random times, this may be be reflected in my results nonetheless. I look for evidence by comparing the frequency counts of two activities for the same three participants. I use working as an example of a socially desirable behaviour, and watching TV as a less desirable one (Stone et al., 2006).

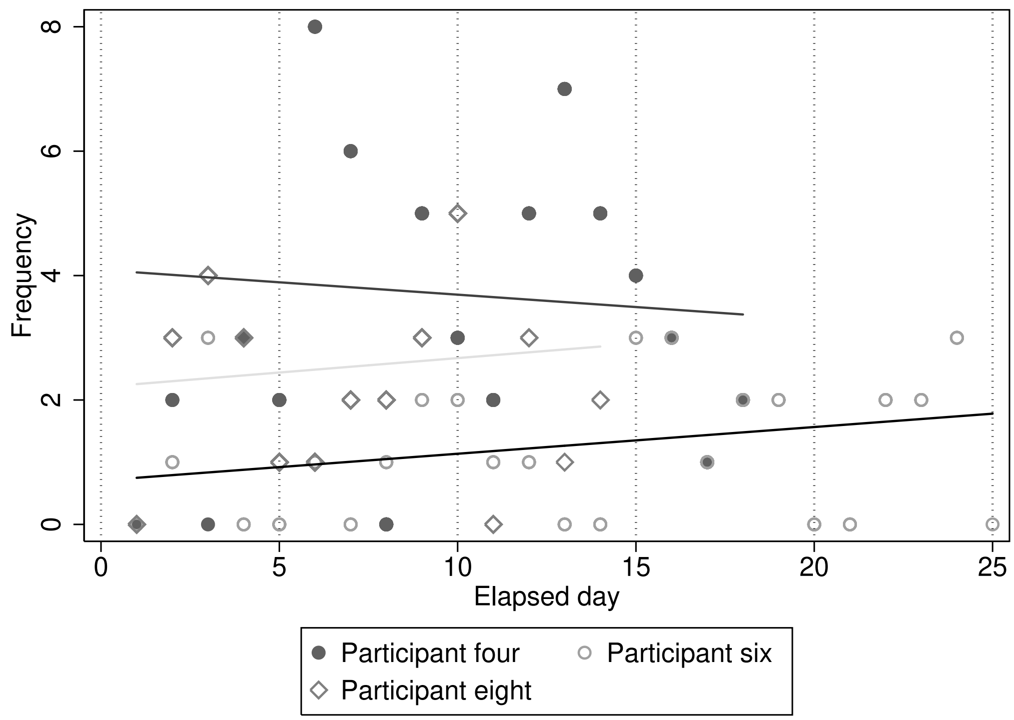

Figure 35: Daily frequency of ‘Working’ activity, three participants

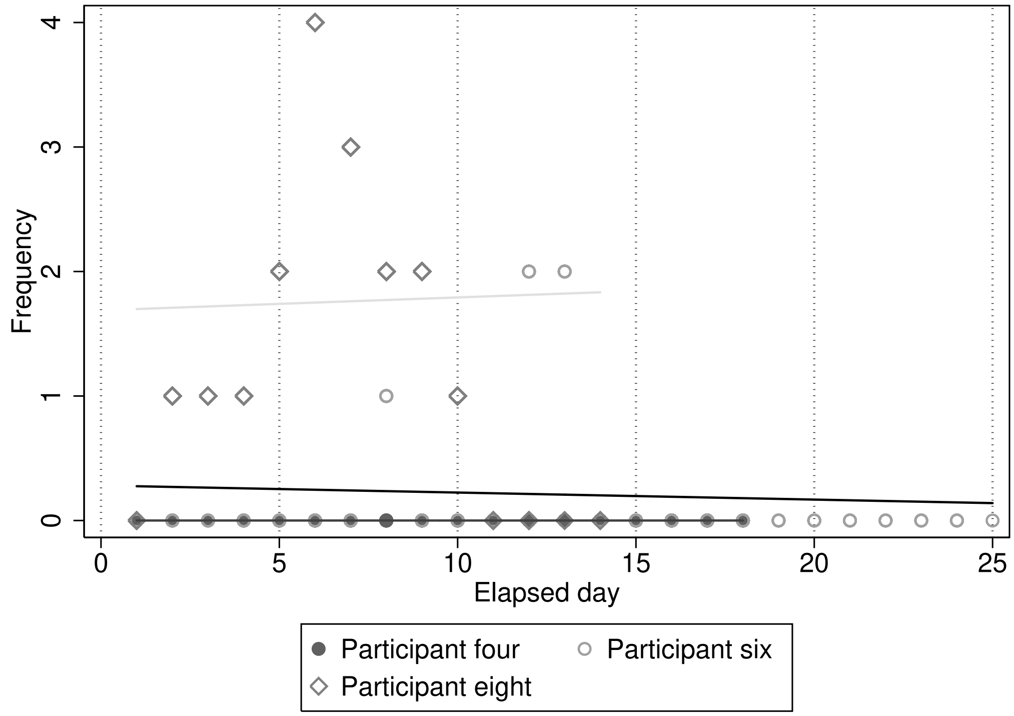

Figure 36: Daily frequency for ‘Watching TV’ activity, three participants

The fitted lines are almost flat, and some of the slopes go in different directions. There is no real appearance of a reactive change here either. A specialised study is really required, most likely one that combines quantitative measures with long-form interviews, and perhaps some method of ensuring repeated situations in the same context. Even so, these are interesting results for an exploratory study. They suggest a range of ideas can be investigated with this methodology.

Researcher assessment

My own evaluation of the procedure, apart from the separate analyses of the data above, concentrates on the cost of operations. In an ideal study, there would be no training and configuration needed, and no maintenance either (Lahlou, 2010b).

In the real world, I provided one-paragraph instructions under ‘Help’ in the menu system. Either these were sufficient, or they were not needed, as no-one reported any trouble using the software, and no-one asked for further instruction or clarifications. The application is designed to be intuitive to operate, so hopefully this is a reflection of that.

As for configuration and maintenance, there were a few minor problems with registering devices correctly for the installation, but these were resolved quickly over email.

There were no issues with security or privacy, and the data were easy to access in an appropriate format.

Summary and conclusions

I feel that this study rates as promising in several ways. Remote deployment and data collection proved highly successful. As far as my conclusions go, the claims of improved efficiency and access held up. Checking these results against the specific aims outlined at the start of this paper, there are several indications of success. Almost all participants completed high numbers of measurements, and the average participation length was 19 days, with a minimum of five days. The non-completion rate was less impressive at 56%, but still within predicted parameters. Feedback was certainly positive but not very useful; there were no complaints about the method or the instrument, but no information on reasons for attrition or low measurement numbers. My own evaluation was likewise positive and in my experience, the administration was surprisingly straightforward given the experimental nature of the software.

There were definite variations in levels between the various situations, and the differences were in line with common-sense and theoretical predictions, such as relaxing being less stressful than commuting, for example. We saw differences in means between subjects for momentary stress, and it was also apparent that subjects' trait anxiety differed. There was further evidence of an association in the trait scores by subject, with clustering observable when compared to the subject's mean momentary stress scores.

However, the process of feedback elicitation was not really satisfactory. A more detailed investigation would be desirable if we want to get better answers about the user experience. This would perhaps entail follow-up interviews with additional incentives; at the least, an opportunity to submit detailed responses.

On balance I feel that it successfully paves the way for larger, or more specialised, studies. The technology appears to be robust, and mature enough to support more participants, who could easily furnish several thousand readings. I would definitely recommend this approach for researchers who are interested in momentary observations, but I would also emphasise that such research needs to make an early choice between specific and general investigations.

Preface: Paper Three

So far, this thesis has presented the development and testing of an experience-sampling technique on mobile devices. In the following study the focus changes to an assessment of model alongside measurements. In order to do so, I spent one year capturing multiple data streams about my own experience. The software developed in the earlier studies was used to record several observations per day of my own stress and situation, augmenting these self-rated readings with a fitness-tracking wristband which measured sleep and heart rate.

The idea of unique insights from intensive single-subject studies was becoming popular, and this suggested a possible contribution to the literature. In addition, there were opportunities offered by affordable wearable actigraphy (movement detection used for sleep and fitness tracking) and portable electrocardiography (heart rate monitoring using miniaturised electro-optical sensors). This opened up the possibility of investigating a more diverse data set, which could serve to validate the psychological data against physical measures.

A subsidiary issue this paper addressed was the lack of conclusive feedback in paper two. Although the study was successful in terms of the aims of the paper, it was apparent that software and method still need a 'stress test', both in terms of seeing how well the software performed over a long period of use, and in gathering and evaluating a great deal of data. As such, it would prove sufficient.

A year of anxiety: An extended individual study of momentary stress using experiential and biometric assessment

Abstract

Experience-sampling techniques can be a compromise between multiple subjects and multiple measurements. In this study, the emphasis on between-subject is reduced by carrying out a year-long self-tracking experiment, which allows a different approach to exploring the kinds of data and inferences this method generates. This also suggests some guidelines for implementation which may aid future researchers.

Introduction

In social science, we are often criticised for relying too much on retrospective reports (Baumeister et al., 2007; Hamaker, 2012), and for failing to take ecological influences into account (Bechtel, 2009; Bolger et al., 2003; P. Wilhelm et al., 2012). Sadly for researchers, the obvious remedy – more informative measurement methods – often turns out to be impractical due to high implementation costs (Furr, 2009). For example, direct observation data can be so time-consuming to analyse that only very small amounts are usable.

This study looks at these issues in the context of anxiety and stress, which has a considerable impact on society (Health and Safety Executive, 2017). The causes and influences of stress have been studied for more than a century (Costa, 1996), and the effects can be clearly seen: in the moment, through reactions to events (Strelau, 2010, p. 121), and over time, through physical and behavioural changes (Sapolsky, 2004, pp. 4–8). Yet much existing research (Baumeister et al., 2007), because of these practical reasons, focusses on one- or two-item retrospective self-reports (Schwartz, 2012).

Here, the research aims to address these problems methodologically by using a longitudinal experience-sampling (ESM) approach. This provides multiple measurements of a variable or variables, grounded in the environment the experience occurs in, preserving vital contextual influences. This paper reports the results of an extended self-tracking experiment using a self-written software tool, running on mobile hardware. By combining ESM with modern portable technology, it is proposed that this approach enables investigation of questions previously limited to observational fieldwork or the laboratory (Cikszentmihalyi, 2012; Hektner et al., 2007, pp. 6–7). It is also suggested that this type of design reduces the cost of these informative measurements, whilst retaining valuable ecological data and allowing examination of "covert" (Lahlou et al., 2015) states of in-the-moment phenomena.

Background

Stress costs money, time, and affects quality of life. In studies by the UK government, for example, stress-related illness accounted for "49% of all working days lost due to ill health" (Health and Safety Executive, 2017) in 2016/2017. 526, 000 cases, with an average of 23.8 days lost for each case; 12.5 million days in total. One in five (20.09%) adults (Office for National Statistics, 2017a) in the most recent UK Annual Population Survey (APS) rated their anxiety as 'high' (defined as six points or more on a ten-point scale (Office for National Statistics, 2017b)).

Studies like these are extremely useful for identifying trends, and for research looking at broad laws, but are less useful for studying individual behaviours. In fact, for a fine-grained understanding of processes in the everyday, we need to look at things as they happen, outside of the mind: in the moment, and in the world (Bechtel, 2009).

Psychologists acknowledge the important influence of environment on immediate experience (Schwartz, 2012; Van Dijk et al., 2015), but concerns remain about the best method of assessment. Tracking behaviours and processes which change over time (Bolger & Laurenceau, 2013, p. 5), embedded in the environment or situation in which they take place (P. Wilhelm et al., 2012; R. A. Wilson & Clark, 2009), can place a heavy burden on the participant and the researcher. Social scientists must choose from a range of techniques, each with their own strengths and weaknesses (Furr, 2009), and supplying different types of behavioural data, from direct observation to retrospective self-report. Often projects must make a compromise between cost and effectiveness.

Approach

Experience-sampling is one method of taking multiple real-time measurements over long periods. It combines the rich data of diary methods with the intensive frequency of momentary sampling (P. Wilhelm et al., 2012), and addresses the complications of retrospective measurements, such as inflation (Broderick et al., 2008) and over-estimation of frequency (Gray et al., 2008). Researchers can send random signals, ask for responses to particular events or times, or rely on device sensors and software (Bolger & Laurenceau, 2013, pp. 14–21). These can capture a wide range of data from sounds to locations (Eagle & Pentland, 2009; Hasler et al., 2008), not to mention automating recording and submitting participant data online. Hektner, Schmidt, and Csikszentmihalyi (2007) give a good overview for beginners, whilst Bolger and Laurenceau (2013) provide a practical handbook.

Although experience-sampling techniques are well-suited for examining daily emotional life more closely (Wichers et al., 2010), and can even incorporate multiple measurement types, they remain under-appreciated in psychology (Conner et al., 2009). This is because historically, they have required very large time commitments at every stage of research: design, participation/administration, and analysis (Christensen et al., 2003; Shiffman et al., 2008). Some of these drawbacks can be ameliorated by specialised equipment, like personal digital assistants (PDAs) (Holt, 2018) or electronically-activated recorders (Mehl et al., 2001), but this can often require a substantial financial commitment.

This study considers the proposal that modern mobile technology can address these issues by reducing up-front investment and participatory requirements, and standing up to a long period of participation. Research with smartphones, now in its second wave (Harari et al., 2016), has seen a lot of progress in the development of tools for augmenting existing methods (Raento et al., 2009). These have increased cost-effectiveness, ecological validity, and participant reach. In earlier studies (see paper one and paper two) a software tool was developed, running on a mobile phone or iPod, that allows multiple grounded measurements of situational variables at random intervals, and saves the data to an online database.

This kind of software has proven particularly useful in self-documentation (Crawford et al., 2015; Lupton, 2014; Swan, 2009), one emerging trend using mobile technology, and some inspiration is taken from these studies. Motivated users take large numbers of measurements, for a wide variety of reasons, including physical measures like cortisol levels (Doane & Adam, 2010) or heart rate (Fahrenberg et al., 2007). Yet again, due to the quantity and detail of observations, researcher-participants often find themselves overwhelmed, with immense data sets to address in which it is difficult to separate signal from noise (Swan, 2013). If they wish to get to grips with this, they must often be prepared to carry out complex analytical procedures (Fleeson, 2007). This is a well-known problem with longitudinal approaches (Bolger & Laurenceau, 2013, pp. 14–21), and unfortunately, experience-sampling data is not the exception. However, in this study, it is hoped that using an abbreviated model of stressful experience (seen in figure 37) will partly remove this issue. This model is based on the 'essential triad' of person, environment, and activity (Funder, 2006), further adapted using Strelau (2010) and Kagan (2010).

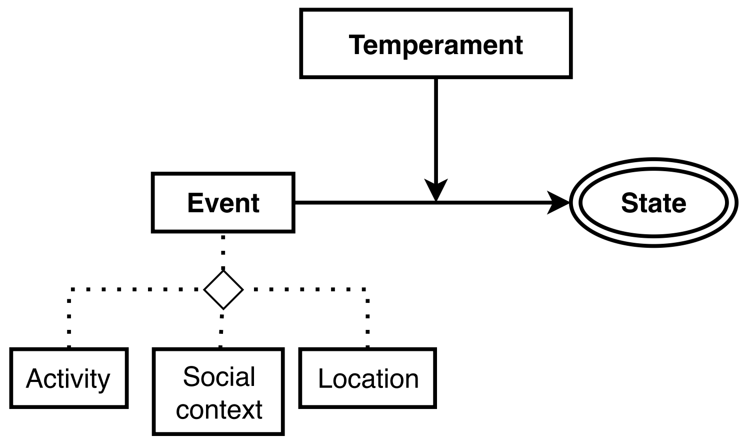

Figure 37: An 'essential' model of stressful experience

Individuals usually vary in their response to stress as a result of traits or temperament (Kagan, 2010; Kagan & Snidman, 2009; Van der Doef & Maes, 1999), but we usually need multiple observations and subjects to make any inferences about effect sizes. Fortunately, using the single-subject approach helps circumvent this issue. Single-subject, or single-case, (Hurtado-Parrado & López-López, 2015) approaches have proven useful for studying many issues in psychology (Reis, 2012), from issues of anxiety to causes of mania. Using this method essentially holds the participant constant between measurements, meaning that they act as a control for their own characteristics (Rabe-Hesketh & Skrondal, 2012, p. 147).

This means that in this study 'person' (figure 37) is essentially being held constant, and therefore situation should be more apparent. This also supports the streamlined data collection, focussing on a few variables, which it is hoped will reduce analytical pressure whilst proving sufficient to test the model of stressful experience. To back this up, the study incorporated physical measurements from a wrist-tracker to complement the self-reported data.

Aims and hypotheses

The purpose of this research is three-fold. Firstly, to establish the validity of the methodological approach through the results, carrying out reliability checks, exploratory analyses, and checking narrative and theoretical concordance. Secondly, to assess the utility of the instrument design. Thirdly, following on from establishing the validity of the measurements, comments can be made on the implications for the veracity of the situational model and its components.

Therefore, the following four questions will be investigated. Model and instrument verification form the larger part of these, whilst participatory burden is a lesser but still significant focus.

Question one: Will the measurements show internal consistency comparable to other studies of stress using the short- and long-form scales? The study is looking for an alpha of between 0.7 and 0.8 (Marteau & Bekker, 1992; Nilsson et al., 2012), which is moderate but not overly ambitious for an exploratory study.

Question two: What is the effectiveness of the system that has been developed? The total participation should be high, reflected in number of measurements overall, number of measurements per day of participation, and low numbers of missed readings compared to traditional experience-sampling approaches. Bolger and Laurenceau (2013, p. 2) suggest five observations per subject, as this allows a minimal within-subject linear model, and will be "descriptively and graphically informative". In this long-form study, this will be extended to five per day. Similarly, the study will be aiming for 'attrition' of less than 15-30% (Barrett, 2004), interpreting this as numbers of missed readings rather than failure to complete the study, to enable a fair comparison for one participant.

Question three: Will exploratory analyses show results consistent with theory, and consistent with the narrative recollection of the year? This means that the broad shape of the results will agree with recollection and records, and that individual components of the theoretical model (figure 37) will stand up to a more granular scrutiny.

Question four: Finally, will the model be verified, with evidence of broad support in the results for the influence of the suggested components? These will be predicted by theory and tested by regression analysis. Analysis will be looking at significance and size of any effects in multiple linear regressions, and comparing this to other results and predictions to form an overall conclusion.

Method

To collect the data required, a year was spent taking multiple readings per day of my own stress, situation, and some physical markers like sleep and heart activity. 1,265 measurements were taken in this time, using the custom application, and a separate record of notable events was kept for comparison.

Measures

The custom software application, running on an iPhone 5, measured the following momentary and ecological variables.

Momentary stress: A six-item version (Marteau & Bekker, 1992) of the Spielberger

State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983), which has similar consistency and reliability to the full 20-item version of the scale (Nilsson et al., 2012; Tluczek et al., 2009), was used for the short daily readings. A participant simply selects one of the four possible answers from a drop-down list, rating their agreement with the statement.

The Six-item State-Trait Anxiety Inventory (STAI-6) contains an even split of six "anxiety-present" and "anxiety-absent" items (Marteau & Bekker, 1992). These are scored from 1-4, with "anxiety-absent" being reversed. (This was handled automatically by the custom measurement software.) This gives a composite score of 6-24, which was rescaled so that 1 stayed the lowest and 4 the highest score possible. The measures proved robust; see the analyses further down for more details.

Context: Each measurement also asks for self-reported situational data. To keep things simple, there are just three questions: what, where, and with whom. These are answered in the same way, from a drop-down list. The set of answers are limited, and intentionally broad, modelled on Kahneman's coded set of situational categories (Kahneman et al., 2004; Kahneman & Krueger, 2006). These were taken from a much larger sample, over multiple studies, and have been verified by many other researchers (e.g. Stone et al., 2006) including this one (see paper two). (The set of possible answers can be seen in full in the results section.)

Time: Computerised automation included the exact time and date of each reading, which were taken at semi-random intervals throughout the time specified by the participant. The daily measurement period (e.g. 10am to 10pm) was divided by the software into 'chunks' by the number of measurements chosen per day, then the exact notification time was generated at a random time within each chunk. To expand on the example, five measurements between 10am and 10pm would result in five chunks of 144 minutes each, with each measurement occurring at a randomised time during that period. This ensures that, for example, all five don't fire one after another, which occurred in testing.

The final two variables were automatically calculated by a fitness tracker wristband. Its operation is described in more detail in the 'Procedure' section.

Sleep: Periods of sleep were registered by analysing movements and heart rate. Originally calculated in seconds, in the analyses these were turned into minutes for better interpretability.

Heart rate: Resting heart rate (RHR) was also calculated and stored by the fitness tracking device. It is measured as the number of heart-beats when a subject is 'still' (M. T. Jensen et al., 2013), usually first thing in the morning or during other periods of inactivity (Fitbit Customer Support, 2018), and stored for each day individually.

Participants

Taking part in the longitudinal study was one middle-aged male, based in the South-East of England. I have a mixed background but have lived chiefly in the United Kingdom.

Procedure

To meet the methodological requirements, an application was built for Apple's iOS operating system from scratch. (iOS looks like an acronym, but does not actually stand for anything.) This was developed using their Xcode integrated development environment (IDE). The underlying code is written in Swift, a C-like language, and the final software utilises a graphical, touch-screen interface for ease and speed of use. Figure 38 shows a portion of the interface with the exact wording blurred, as the STAI is a commercial scale. As mentioned above, all responses are drop-downs, so no typing is needed, and speed and accuracy are maximised.

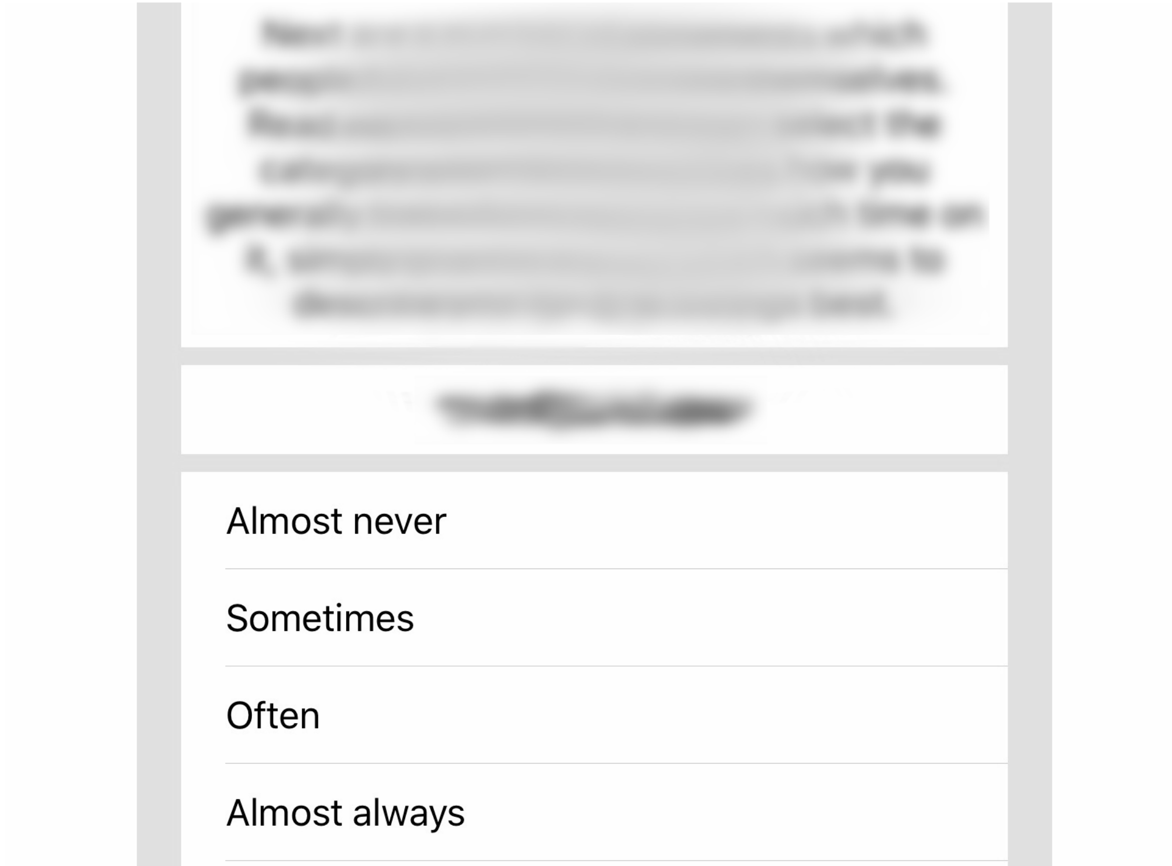

Figure 38: Part of the touchscreen user interface

Measurements are taken locally, with the software running on an iPhone, iPod Touch, or iPad. The results can be uploaded to a secure server over the internet, or read directly from the device by the researcher. One advantage of this approach was a tailored and standardised data format: because the application and its attendant data storage were self-designed, the data require little clean-up after downloading, and come in a standard comma-delimited format.

An automated fitness tracker was also used, consisting of an actigraph to detect sleep, and heart rate sensors. Two models were used over two halves of the experiment. The first, a FitBit Charge HR, had excellent sensors but was rather uncomfortable; the second, a Jawbone UP2, was much more comfortable but, as results will show, proved to have inferior sensors (optical vs. electrical). Data were downloaded from the manufacturers' online storage systems, which are not designed for user convenience, and require a great deal of collation and organisation.

Total time spent taking readings per day was approximately one minute, with each reading taking just a few seconds. Everything else was automatic. The average number of measurements per day was 5.5; much more detail is available below in the results section.

Results

Measurement reliability

The Six-item State-Trait Anxiety Inventory (STAI-6) scale consists of three “anxiety-present” and three “anxiety-absent” items (Marteau & Bekker, 1992), noted by Spielberger (its creator) as being particularly sensitive to stressors. Cronbach’s alpha was 0.744, using unstandardised items as each measure uses the same scale. This result is comparable to other tests of the six-item scale (e.g. 0.82 (Marteau & Bekker, 1992); 0.79 (Tluczek et al., 2009)), and a value above 0.7 shows modest to high internal consistency (Nilsson et al., 2012; StataCorp, 2009a).

An exploratory factor analysis was consistent with a one factor solution, with the first factor having an eigenvalue of 1.93 and explaining 117% of the variance (due to other negative values). The second factor had an eigenvalue of 0.22 and explained 13% of the variance. Some further analysis, particularly validity, is presented later in this section, but this is a positive answer for question one, on instrument reliability.

Tool validation and data summary

The study started on the 25th of December 2015, and the last measurement was taken on the 31st of December 2016, a period of 372 days in total with some missed days. Table 12 gives the total numbers of measurements taken.

Table 12: Summary of measurement numbers

Total observations	Total days	Total weeks	Total months
1265	293	45	12

During this period, an average of 5.5 measurements per day were taken, with the lowest total number being one, and the highest being 12. The standard deviation was 2.3. The total and average measurements show no obvious pattern of favouring one day or another, as broken down by day of the week in figure 39.

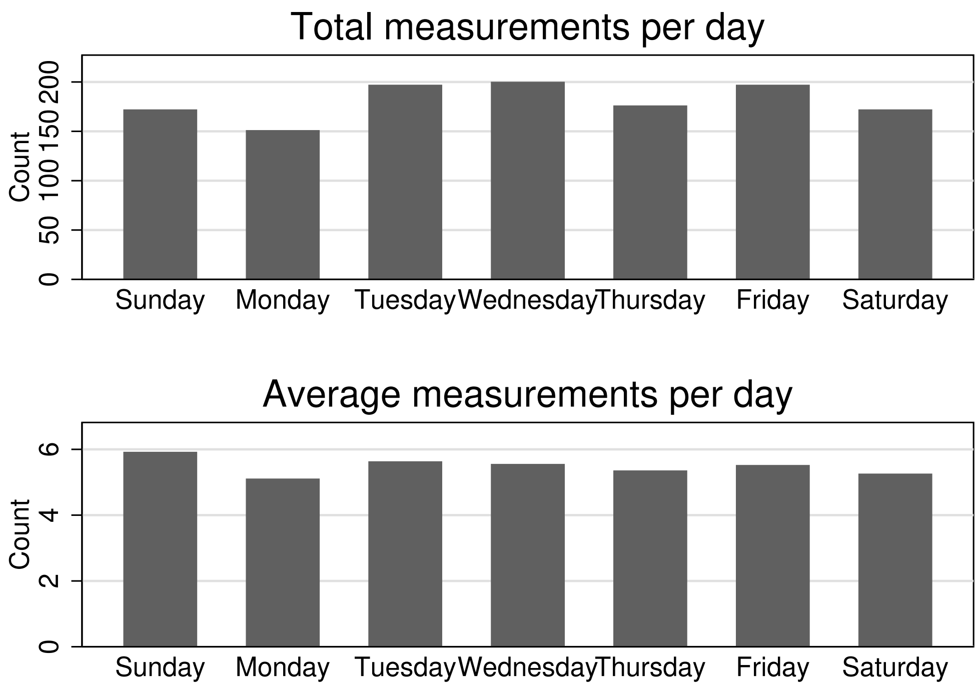

Figure 39: Total and average measurements per day

However, these figures do not accurately reflect contiguous blocks of participation. Six months into the study there was a medical incident which had a significant impact on numbers of readings; this is further discussed in the reflections at the end of the document. Nonetheless, we can use these numbers to estimate attrition. Of the 372 possible days, 79 days - or 21% of possible measurements - were not assessed. This is in fact a reasonable number for an experience-sampling study (Christensen et al., 2003), even with no mitigating factors. On the other hand, 47 of those days (59% of the total) were linked to the medical issue, which was mostly outside of my control. And there were 13 further days (16%) where I could not take readings due to a problem with the software introduced by the device manufacturer. Those 13 days were spent updating the application to work around the issue. If we take these into account, it would suggest that the final figure is more like 19 days, or 5% of possible measurements overall.

Looking back to question two, both 1,265 measurements over 293 days, and 793 measurements over 151 days, represent an average slightly over five measurements per day. This satisfies the first part of the question. Whether this study achieves the second part is more open to debate. The worst interpretation is 20% missed readings, which is lower than the ESM average of 22.5% (Barrett, 2004), and the best is 5%.

Fitness tracking equipment

In some analyses, usable measurement numbers are reduced by a further technical issue, this time with the fitness tracking wristband I used. The first one was a watch-style device which was (as it would turn out) equipped with better sensors, but was bulky and very uncomfortable to wear 24 hours per day. At 151 days I switched to a slimmer, more comfortable wristband-style device which should have had equivalent sensory capacity. Unfortunately, when I began to look more closely at these data, seen in table 13, it became clear that this was not the case. The second set of data are markedly different, with divergence in both the physical measures, resting heart rate (RHR) and sleep duration.

Table 13: Tracking device comparison

Device	RHR mean (bpm)	Slope vs. momentary stress	Sleep mean	Slope vs. momentary stress
One	68.89	0.13	328.72	-0.002
Two	61.89	-0.03	392.1	0.03

Simply using the second device caused my average resting heart rate to drop by seven beats per minute (bpm), and added over an hour to my average sleep duration, which would be an unlikely coincidence. To emphasise this further, the trajectory of both slopes (the correlation with momentary stress) has reversed. The data from the second device suggest that increased stress caused my resting heart rate to go down, and my sleep duration to rise. Not only is this unlikely, it is contrary to theory (e.g. Åkerstedt et al., 2012; Dahlgren et al., 2009) and, coupled with the large discrepancies in levels, I judged that these is a problem with the whole second set of readings. The difference clearly started when I switched devices. As the first set of data produced readings that are in line with my age and fitness level, these remain in consideration, but the second set of sleep and heart rate data are excluded from my analyses.

This leaves 793 data points in these cases, but we still use the full data set where possible, as seen in the summary statistics in table 14.

Table 14: Summary of momentary stress

Mean	SD	Min	Max	Median	Skewness	Kurtosis
2.3	0.44	1.17	3.67	2.17	0.48	2.56

The overall mean was actually slightly below the halfway point in the scale (1-4 points). With the median just below, the data are skewed slightly to the right but remain approximately symmetric (figure 40).

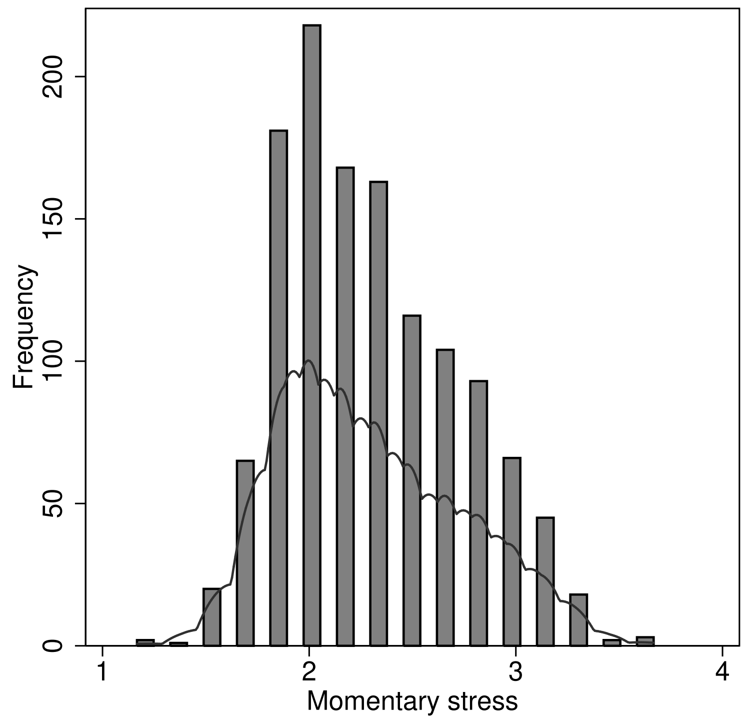

Figure 40: Distribution of momentary stress

In the following sections I look at some of the different ways of framing, and exploring, these data.

Time

In their guide to designing and analysing intensive longitudinal studies, Bolger and Laurenceau (2013) suggest beginning with an analysis of the influence of time on the variable of interest. Not all processes are linear, but patterns of change should correspond with events and inform overall analysis. There are several ways of visualising these patterns. To start with, I present an overview of all measurements over the entire period, with a fitted spline at the median for to represent tendencies of change (figure 41).

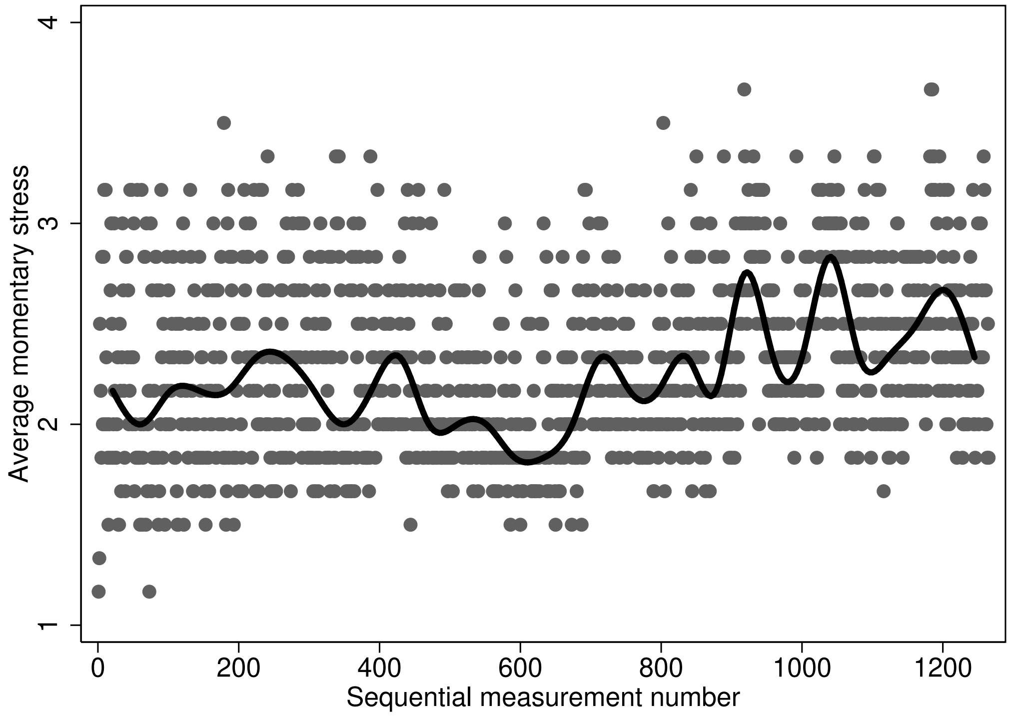

Figure 41: All measurements in order taken

There are clearly ebbs and troughs, but the variability of individual points make it hard to see overall tendencies, if indeed they exist. So I smoothed this view slightly by pooling the data for each day, keeping the fitted median spline, and adding monthly averages for ease of comparison.

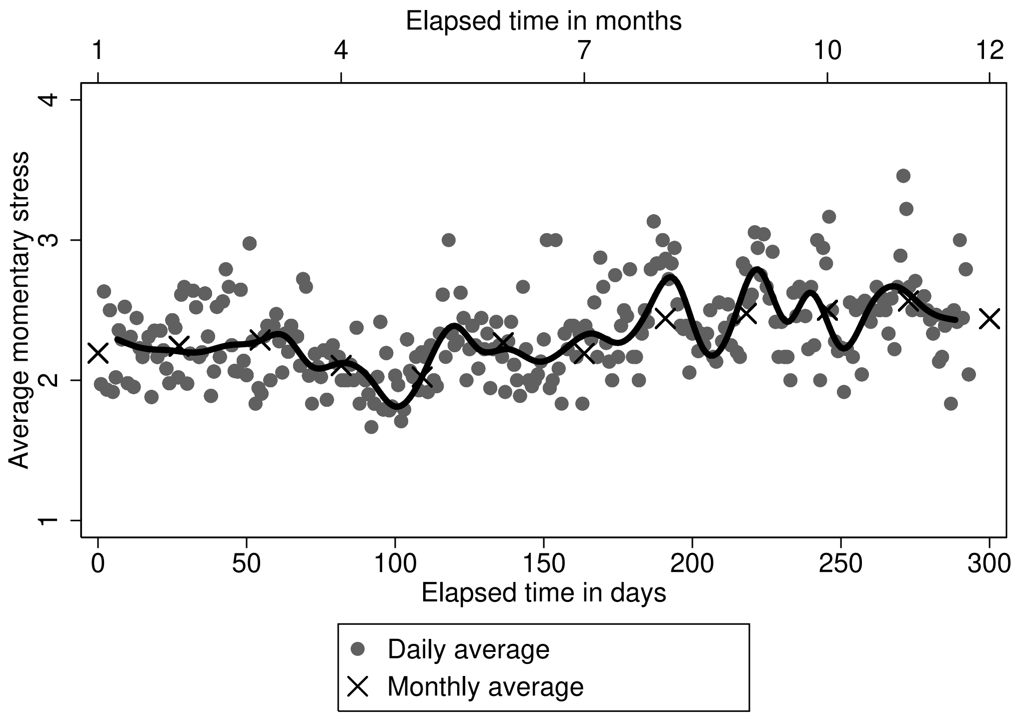

Figure 42: Daily average and trend, with monthly averages overlaid

This high altitude view (figure 42) continues to show the ebb and flow we were expecting, but remains rather unpredictable day to day, despite the averaging.

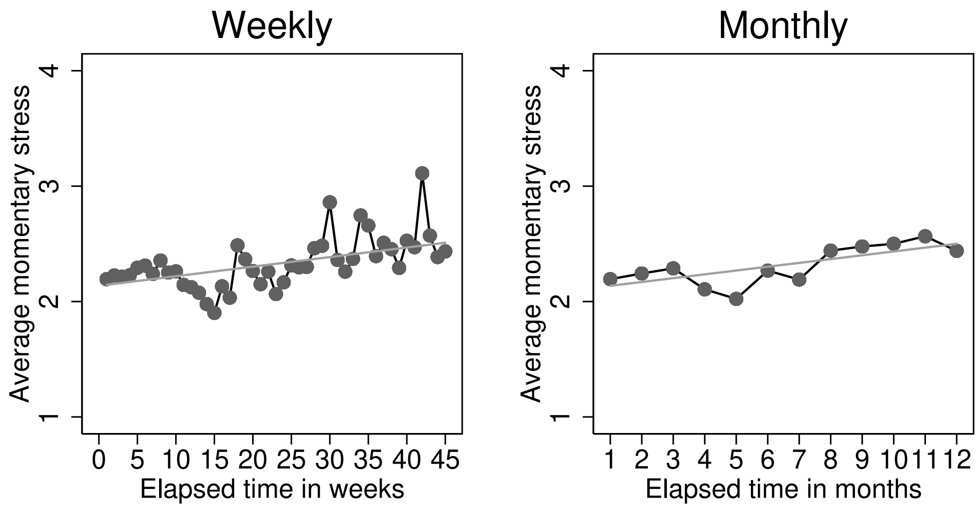

Figure 43: Weekly and monthly averages with fitted line

The next step is to group the data by week and month (fig. 43). This seems to show that there is a broader trend, one revealed by increased temporal averaging, which we can liken to a moving proxy for overall temperament. A year of continuous stress (as discussed narratively in the next section) appears to lead to an increase in this average over time. (With every measurement having exact time data attached, it is relatively simple to convert absolute dates to elapsed time.) I explore this idea more descriptively in the next section.

Narrative validity

One of the benefits of self-tracking experiments is that the participant can always be relied upon for consultation. In addition to recollection, I kept separate records of significant events and consulted them for this analysis. I think this analysis presents a unique glimpse behind the scenes of these types of studies. It is unlikely that a less motivated subject would agree to so much data gathering. Looking at figure 44 reveals correspondence with a number of significant events in my records.

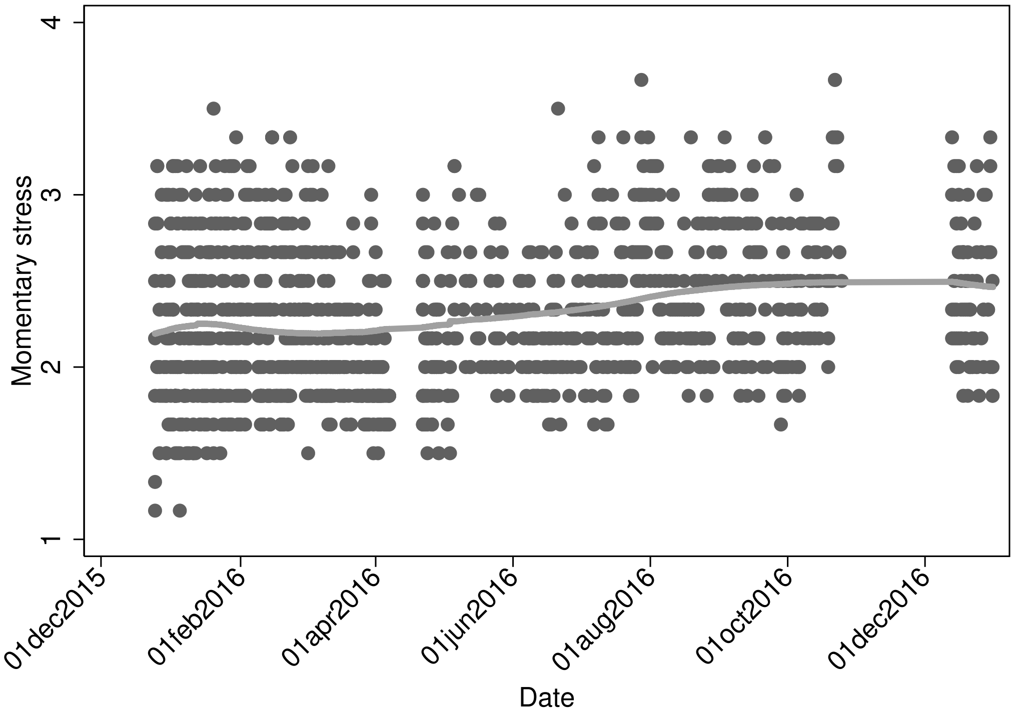

Figure 44: Stress by day of year

There was a 'crunch' period for a work deadline in the first few months of the year, followed by some recuperation as stress goes down again. Technical difficulties with the software, mentioned previously, show up in April as a gap. The second half of the experiment is marked by a period of ill-health, continued pressure of work, and stressful events in society at large. We can see that this leads to rising stress, until I take a hiatus in November, after which things start to settle down a bit. This is epitomised by the average momentary stress on Christmas Day that year, which was 1.83 points on the scale of 1-4, lower than the yearly mean of 2.3. This was even down compared to the Christmas before, when the average was 1.97 points.

The timeline in my recollection would place a rise at the start of the year, a dip, a longer rise, and a final slight dip towards the end of the year. And as we can see, using a locally weighted scatterplot smoothing procedure (LOWESS) to produce a predicted spline (figure 44) shows this flattened 'm'-shaped curve rather clearly, coming compellingly close to my own estimate.

This analysis is, of course, rather subjective. Nevertheless, I view it as good evidence to satisfy question three, which was to verify whether the results showed narrative validity, as well as theoretical. But, to follow this up, I will examine whether specific components are consistent with theory. In the next several sections I carry out individual temporal, physical, and situational analyses.

Day of the week as a temporal frame

I start by looking at the idea of a "weekly mood cycle" (Stone et al., 2012), which is a well-established theory of mood fluctuation. There is good evidence, for example, for the 'weekend effect' of increased positive emotions (Harvey et al., 2015). But, in spite of these similarities, specific observed effects often vary with other factors such as age and working status (Stone et al., 2012). In this study, therefore, I propose day of the week simply as a frame linked to non-specific change in the ebb and flow of mood over time, without making stronger claims about particular weekday effects. This framing device allows for some useful observations about my particular routine. (Table 15 recaps the total measurements I took for each day, demonstrating no particular pattern of favouritism in numbers.)

Table 15: Count of weekdays

Weekday	Frequency
Sunday	172
Monday	151
Tuesday	197
Wednesday	200
Thursday	176
Friday	197
Saturday	172

A scatterplot with standard deviations shows how each weekday differs in figure 45, with a dashed line at the overall mean of momentary stress for the year, which is 2.3.

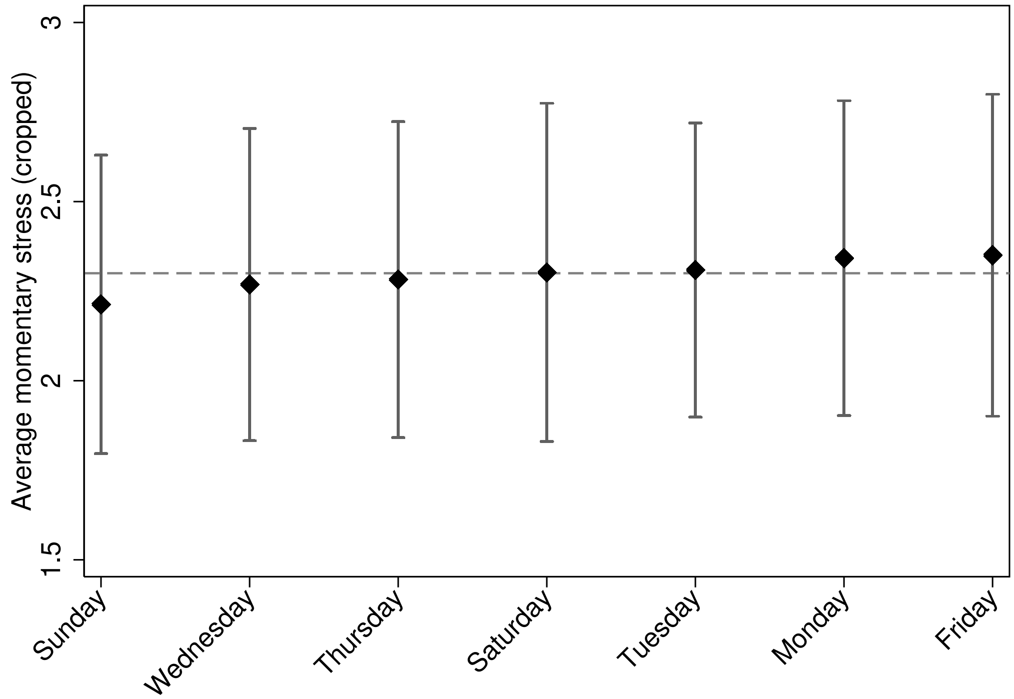

Figure 45: Day of the week ranked from least to most stressful

It is to be expected that Sunday would rate amongst the least stressful days (figure 45), and that Monday and Tuesday would rank amongst the highest. It is unusual, though, for Friday to rank highly in terms of negative affect. Studies have found that positive mood is more associated with Friday and particularly Friday evening, but some evidence indicates a less specific weekend vs. weekday model (Ryan et al., 2010).

In my own routine, Friday is often spent doing household chores in preparation for the weekend, when there will be no break as there is on a school day. This is in addition to my academic work and the usual amount of childcare, which is what I spend most of my time doing, as a full-time carer and part-time student. None of this is particularly relaxing. Saturday is usually fairly busy too, which may account for its presence in the middle of the rankings rather than the low end.

It would be interesting to see if this ranking holds true for other full-time carers, as the schedule is essentially the opposite of an office job. I will talk more about the frequency and effect of activities in later sections. Nevertheless, Sundays do tend to be the quietest day of the week, so I can testify as to the validity of these results. As the influence of time on momentary stress does not seem to be linear, for the purposes of analysis I am representing the effect of day of the week by modelling each day as a non-ordinal category (table 16).

Table 16: Estimates for the influence of weekday on momentary stress

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
0: Sunday†						
1: Monday***	0.13	0.05	0.008		0.03	0.22
2: Tuesday**	0.1	0.05	0.037		0.01	0.19
3: Wednesday	0.06	0.05	0.226		-0.03	0.14
4: Thursday	0.07	0.05	0.142		-0.02	0.16
5: Friday***	0.14	0.05	0.003		0.05	0.23
6: Saturday*	0.09	0.05	0.059		0	0.18
Constant***	2.21	0.03	<0.001		2.15	2.28

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Sunday is the least stressful day (figure 45), so I set this as the reference category. Both Monday and Friday are significant at the 1% level, with similar-sized estimates and confidence intervals. Tuesday is significant at 5%, and finally Saturday is just outside 5% at the 10% level. The interpretation is that, compared to when it is Sunday, readings taken on a Monday are expected to be 0.13 points (on the scale 1-4) higher. Or, to put it another way, average momentary stress on a Monday is predicted to be around 4% higher than on a Sunday.

These results are positive evidence for question three, exploratory evidence of theory, and are preliminary support for question four which deals with confirmatory evidence. An effect of weekday was predicted by the literature, and we saw patterns of ebb and flow that seemed to have some temporal stability. In the next section I examine the evidence for an effect of sleep duration.

Physical correlates: sleep duration and quality

The data in this section, and the following section on resting heart rate, represent only the first 151 days of the study owing to technical difficulties described above. These observations are summarised in table 17.

Table 17: Summary of sleep data

	Mean	SD	Min	Max	No. of observations
Sleep (previous night) in minutes	328.72	70.04	94	504	796

In their overview of the field Hirotsu, Tufik, and Andersen (2015) find general support for a positive correlation between stress and decreased sleep duration (and reduced overall well-being). But as most studies (e.g. Dahlgren et al., 2009) focus on the effect of stress on sleep and not vice versa, it is not clear that sleep variation can be relied on as a predictor of, rather than a response to, stress. I look first at the overall pattern, grouped naturally by day.

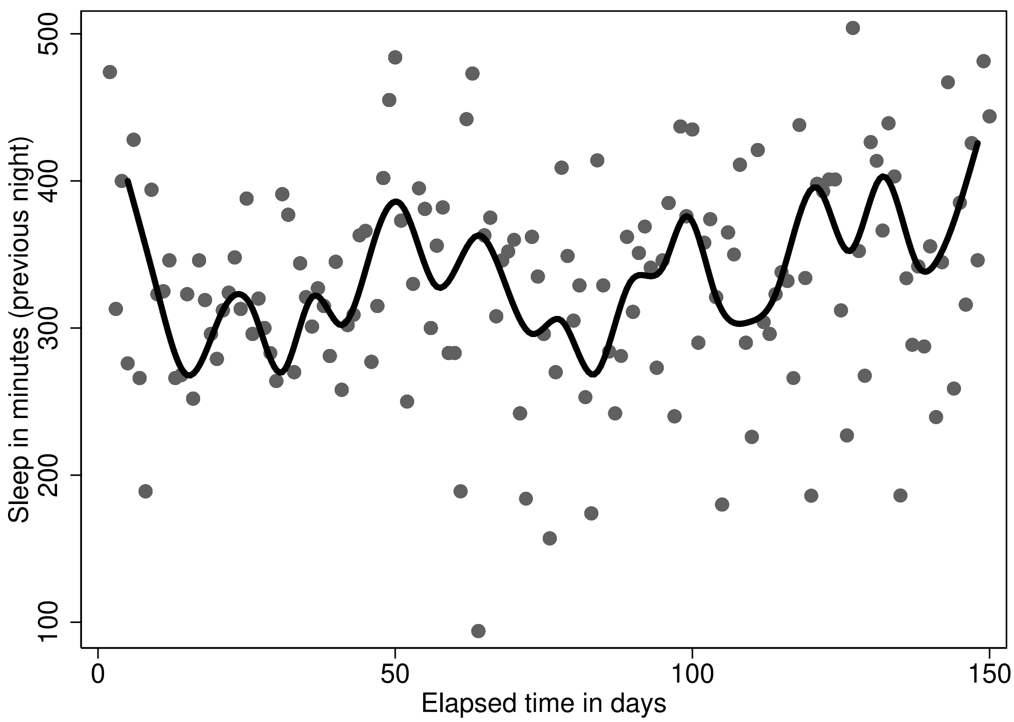

Figure 46: Daily sleep by day of study

It is immediately apparent, in the timeline in figure 46, that sleep is highly variable. The standard deviation, seen in table 17, is one hour and ten minutes against a mean of five and a half hours. Fitting a median spline (the solid line, figure 47) against stress (the dashed line), transformed to fit on the same scale as sleep, does not show any obvious pattern.

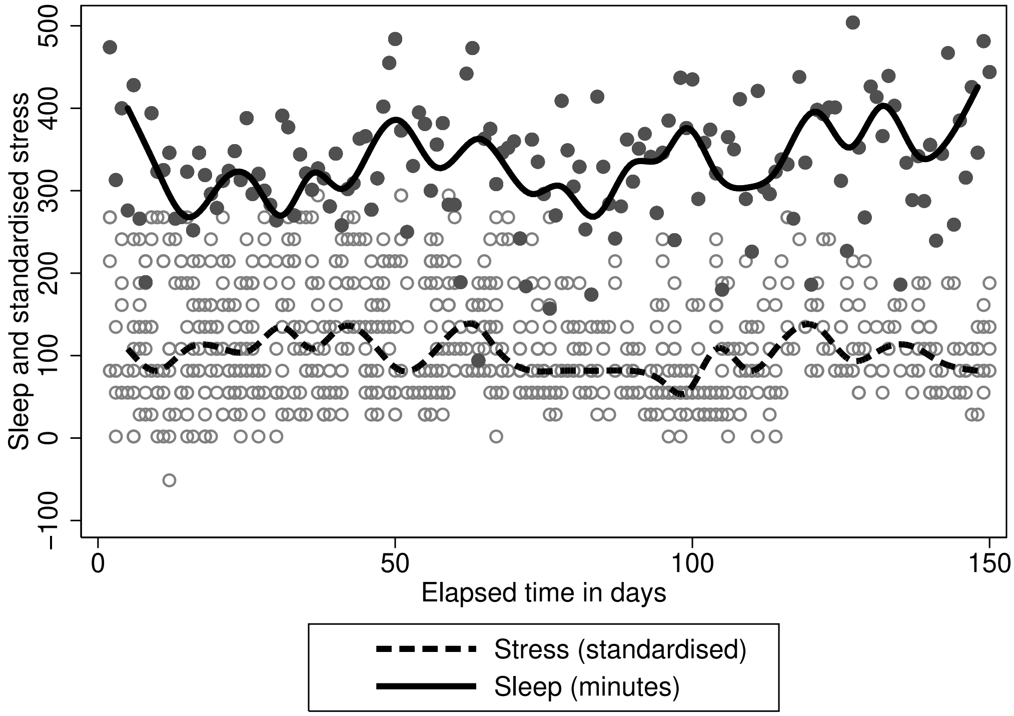

Figure 47: Sleep and momentary stress

Although there are a few correspondences, there are just as often large changes in sleep levels without a visible change in stress levels, and the relationship seems to reverse itself sometimes. In fact describing sleep against both overall and monthly sleep (figure 48) shows no real correlation, and the fitted lines are essentially flat.

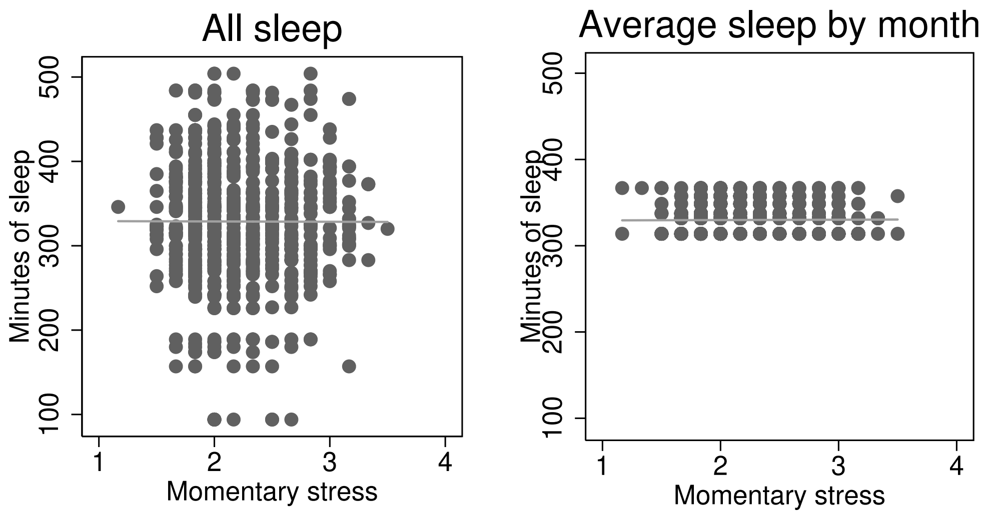

Figure 48: Sleep as a predictor of momentary stress

The monthly average reveals no new pattern, but simply smooths out the variation, which is no more informative. The correlations with momentary stress are therefore expected to be slight, as shown in table 18.

Table 18: Correlation of sleep scores with momentary stress

Measure	Correlation
Sleep (night before)	-0.002
Weekly average sleep (night before)	0.006
Monthly average sleep (night before)	0.007

So far these results are not very revealing. Although there is an association in the right direction between sleep and stress the next day, more sleep being associated with lowered stress, it is not very sizeable; the same can be said of the weekly and monthly averages, with the exception that these suggest an increase in average duration associated with increased stress.

Some research shows that increased stress may result in extra sleepiness in the morning, e.g. reluctance to awake fully, staying longer in bed, and "snoozing" (Dahlgren et al., 2009). Other studies propose that longer sleep duration over time is associated with poorer sleep quality (Åkerstedt et al., 2012). In this study, it is not really possible to say anything with certainty with the results so far.

Nightly sleep is moderately correlated with the weekly average (0.47), and less so with the monthly (0.23), which matches with figure 46 showing that sleep is quite variable day to day. We may see some evidence of reduced sleep quality, suggested in the same literature as an association with increased fragmentation, or more nightly awakenings (Mezick et al., 2009). This further introduces increased variability in sleep duration.

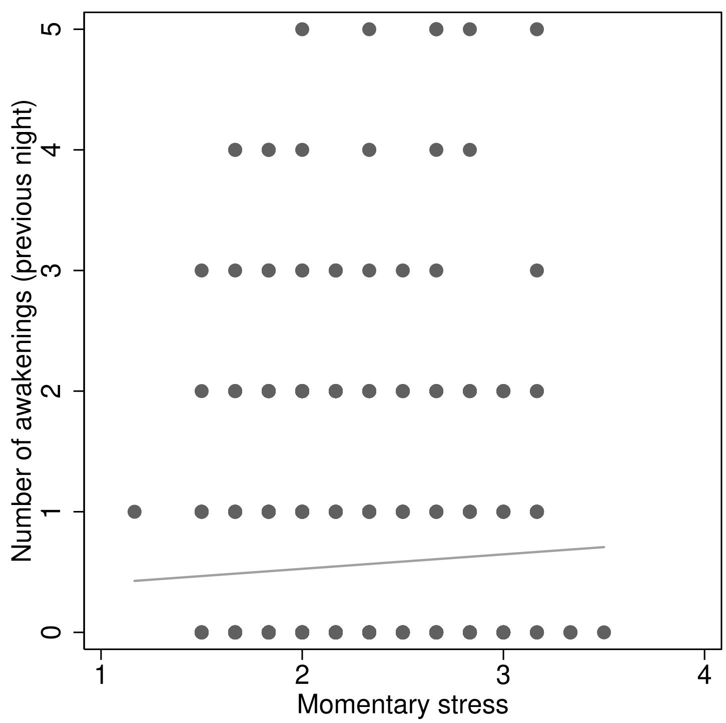

Figure 49: Number of awakenings as a predictor of momentary stress

The appearance of an association of increased momentary stress with number of awakenings is a little more pronounced (figure 49), but the fitted line still has a rather shallow slope. The correlation with stress is 0.05, not quite large enough to be a particularly useful predictor given that nightly awakenings has a mean of 0.55, and a standard deviation of 0.93. The median is actually zero, with a range from zero to five.

A regression on momentary stress using these variables is shown in table 19.

Table 19: Estimates for the influence of sleep variables on momentary stress

	Estimate	(*SE*)	p	95% CI	(*lower*)	(*upper*)
Sleep	-0.00004	0.00024	0.861		-0.00051	0.00043
Weekly average sleep	0.00007	0.00059	0.901		-0.00109	0.00123
Monthly average sleep	0.0005	0.0011	0.651		-0.00167	0.00267
Number of awakenings	0.02662	0.01647	0.106		-0.00571	0.05895
Constant***	2.00684	0.31064	<0.001		1.39706	2.61662

*** $p \leq 0.01$

None of the predictors rate as significant, even at the 10% level. Number of awakenings shows the most promise, but is still under 90%, and all the confidence intervals contain zero. Leaving significance to one side and looking at effect sizes, the estimates are also small. It would take many minutes of change in sleep duration to predict a noticeable effect on momentary stress. For example, four hours (240 minutes) of extra sleep the night before is expected to reduce momentary stress by 0.01 points (on the scale 1-4), or about 0.3%. A large number of nightly awakenings would have a noticeable effect, but the numbers required simply do not show up in the observed data.

These results strongly suggest that sleep duration is not a very good predictor of momentary stress. Sleep quality is more promising, but it is either poorly served by number of awakenings as an indicator or needs better measurement. I suggest that in future studies subjective sleep quality, which has proven useful in predicting stress elsewhere (Åkerstedt et al., 2012), is considered alongside objective measures. But overall sleep, at least as measured by wrist actigraphy, is not useful in this model.

This is not to say that these data are not at all useful. With sleep as the response variable, as is more usual in the literature on sleep and stress (e.g. Hirotsu et al., 2015), there is evidence for the utility of stress as a predictor. This can be seen in the bivariate regression in table 20. When stress is one point higher, sleep is likely to be lower, estimated at about 12 minutes less, at the 5% level. This may indicate a useful avenue for investigations into sleep patterns as a result of stress, but is not useful evidence for sleep as a predictor.

Table 20: Estimates for the influence of momentary stress as a predictor of sleep the following night

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Momentary stress	-11.93	5.91	0.044		-23.53	-0.33
Constant**	352.18	13.18	0		326.31	378.06

*** $p \leq 0.05$*

This is not very strong evidence against question three, which calls for theoretically consistent results. The literature on sleep and stress is mainly concerned with the effect of stress on sleep, and in fact we did see some evidence for that. Overall, though, these results are inconclusive. There is more support for resting heart rate as a predictor, as we will see in the next section.

Physical correlates: resting heart rate

Resting heart rate (RHR) is considered a good proxy for general health by medical professionals (M. T. Jensen et al., 2013). It is seen as an indicator of overall medical well-being, with an elevated RHR associated with a higher general risk of mortality. In this study, I took 151 days' worth of resting heart rate readings, which are associated with 799 individual observations.

Table 21: Summary of resting heart rate data

Summary statistics	Mean	SD	Min	Max	No. of observations
Resting heart rate (beats per minute)	68.89	4.52	52	77	799

In table 21, we see that RHR appears to be a less variable measure than sleep duration, with a standard deviation of only a few beats. The summary statistics are misleading in one respect, though, as they hide a spike and dip seen more clearly in the full timeline (figure 50).

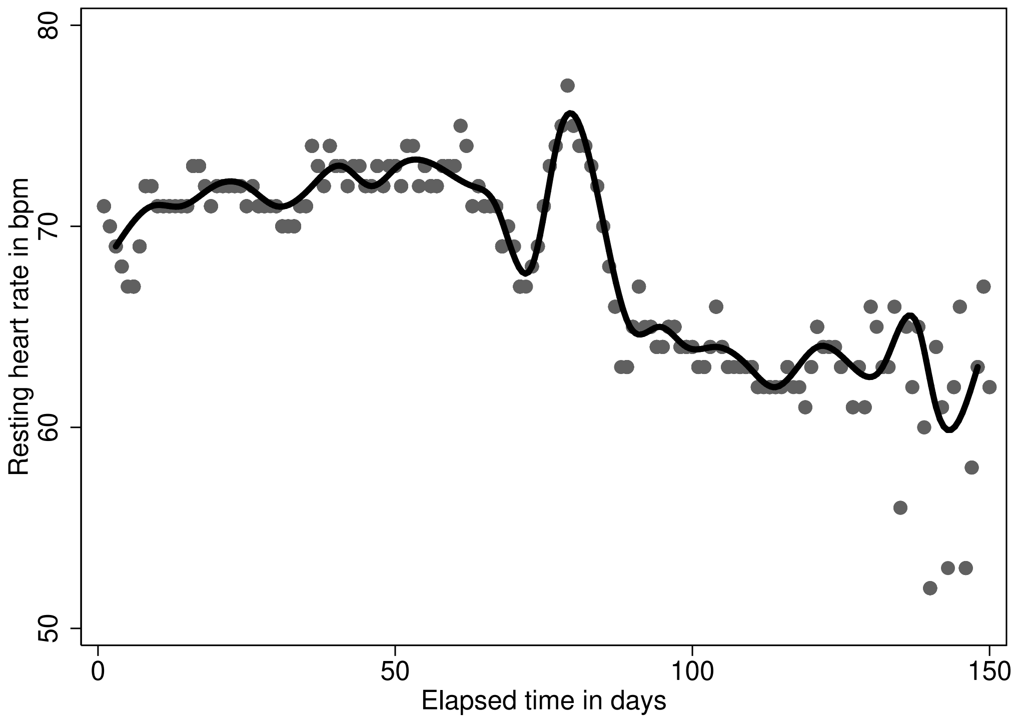

Figure 50: Resting heart rate by day of study

The elevated starting rate and the spike correspond with a 'crunch' period of very hard work in January and February, and a deadline in early March, then a return to a normal routine in April and May; taking the two periods separately would give a lower standard deviation of just one or two beats per minute. This is interesting, and it might be possible to model a factor variable to see if there is a 'switching' effect which modulates reactivity to stress after prolonged exposure (Bolger & Zuckerman, 1995). That is, that physiological effects are suppressed by those more prone to stress with high reactivity (T. W. Smith & Anderson, 1986). That would likely be a whole study in itself, though.

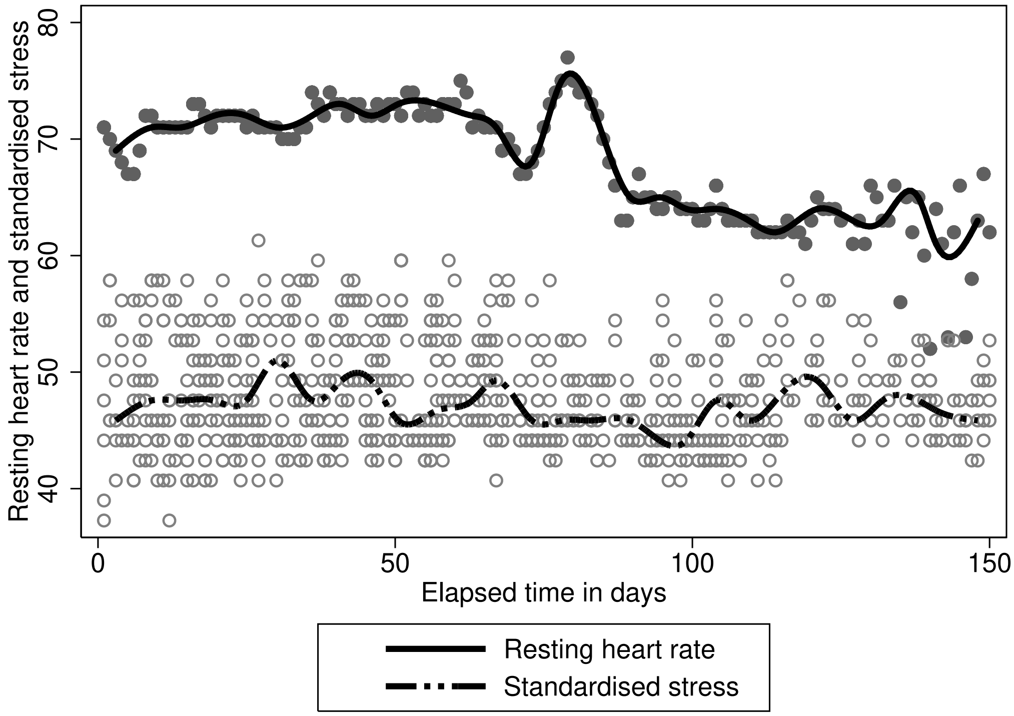

Figure 51: Resting heart rate and momentary stress

In my analyses for this study, figure 51, resting heart rate (the solid line) often appears to track stress (the dotted line) with a slight delay. Observe the period of higher stress on days 30 to 70, with a large, if non-linear, increase in RHR following. But it is equally clear from this graph that stress is not the only correlate of resting heart rate, as it drops substantially in the second half while stress creeps back up.

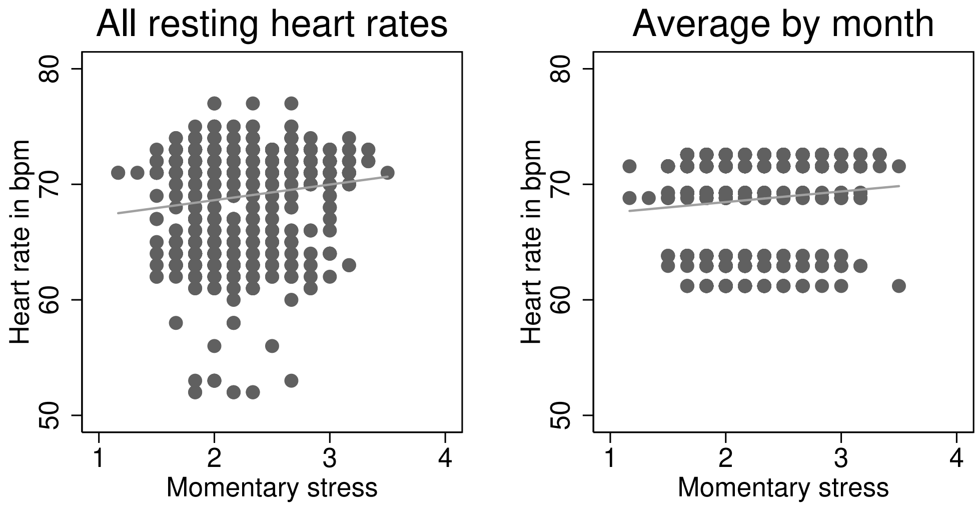

Figure 52: Resting heart rate as a predictor of momentary stress

But overall, there is the appearance of a positive association between resting heart rate and momentary stress. Figure 52 shows the weekly and monthly averages, with a fitted line. This ties in with the conclusions from bivariate analyses.

Table 22: Correlation of resting heart rate measures with momentary stress

	Correlation
Resting heart rate	0.13
Weekly average RHR	0.14
Monthly average RHR	0.12

The correlation in table 22 is not large, as the association is far from perfect, but it is significant when tested in a simple linear regression, seen in table 23.

Table 23: Estimates for the influence of resting heart rate on momentary stress

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Resting heart rate***	0.012	0.003	<0.001		0.005	0.018
Constant***	1.382	0.226	<0.001		0.938	1.826

*** $p \leq 0.01$

Even though it does not have a particularly large predicted effect size at first glance, in reality even an athlete would have an RHR of 40-60 beats per minute (Pasternak, 2018), adding around half a point on the 1-4 point momentary stress scale. And if resting heart rate were to go up by five beats, roughly the standard error, it would predict an increase of 0.06 points in momentary stress, or 2% of the scale. This is quite substantial given that the range is from 52 bpm to 77 bpm over just five months.

I view this as good supporting evidence for question three, testing the theoretical validity of individual components, and another leg for question four, the overall model, to rest on in the analysis later on. In the next section we will take a final look at the individual components of the model alone, which prepares for a multiple regression analysis and a final assessment of question four.

Self-reported situational factors

Like day of the week, these are categorical variables, despite having some advance expectations of the rankings (e.g. Kahneman et al., 2004). In these analyses, the reference categories are simply the most frequently occurring situation (see figure 53), to prevent problems with multi-collinearity. I frame the regression estimates as the influence on reaction of activity and social context, which are an intrinsic part of the environment.

My own experience with the self-reported characterisation of the categories was that it usually felt quite straightforward to decide where my focus was (figure 53). But on occasion, most frequently when engaged in child-care, it was difficult to choose between 'relaxing' vs 'taking care of my children'. In these situations I had to think a little harder, as my focus was often split and I had to rely on my intuition.

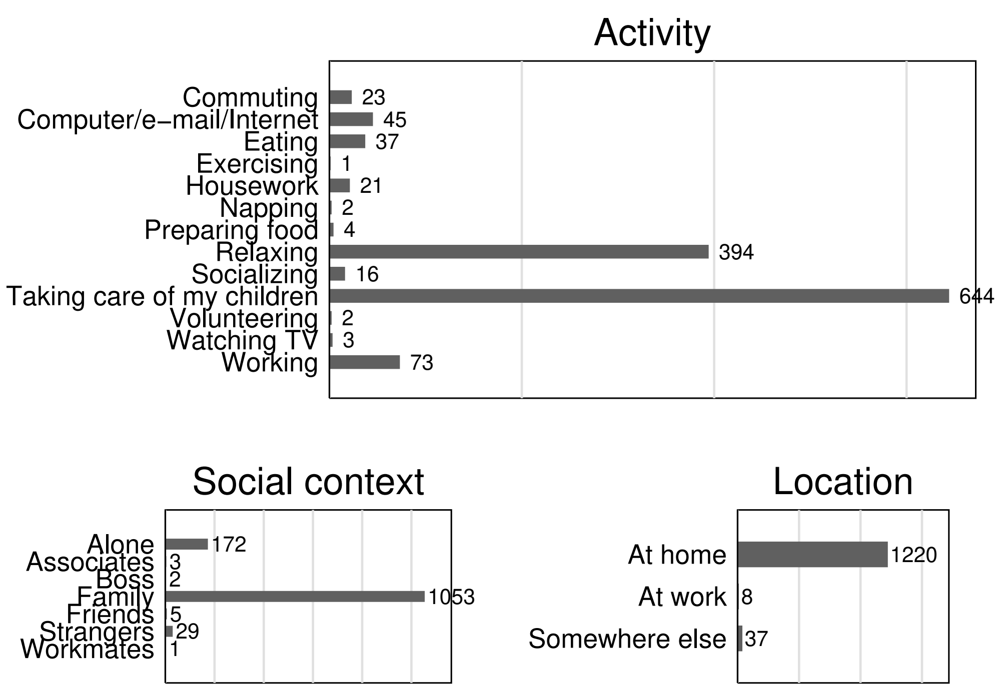

Figure 53: Frequency of situations

These measurements have the added benefit of allowing time-use estimation, using the appropriate modelling assumptions. I will not go into this fully, but, for example, from the frequency counts in figure 53 it is fairly easy to work out that I am a full-time carer for my child, and that I work part-time. Almost all of my work is done at home.

This being so, it seems sensible to exclude the location data from the remaining analyses, as it communicates no real information. For another respondent, of course, these data can be more informative, as seen in other studies such as Eagle and Pentland (2009), and in my own studies with a larger cohort (see paper two and paper four). In the next section, I look at activity and social context by themselves.

Activity and company The first thing I found when I looked at the data for self-reported activities and social contexts, was that the results in this section (see figure 54) broadly replicate those found in previous studies of daily situations (Kahneman et al., 2004; Kahneman & Krueger, 2006; Stone et al., 2006). For example, family and friends rank as least stressful company, while 'boss' is more stressful than being alone (Kahneman et al., 2004). This bodes well for the rest of the analysis.

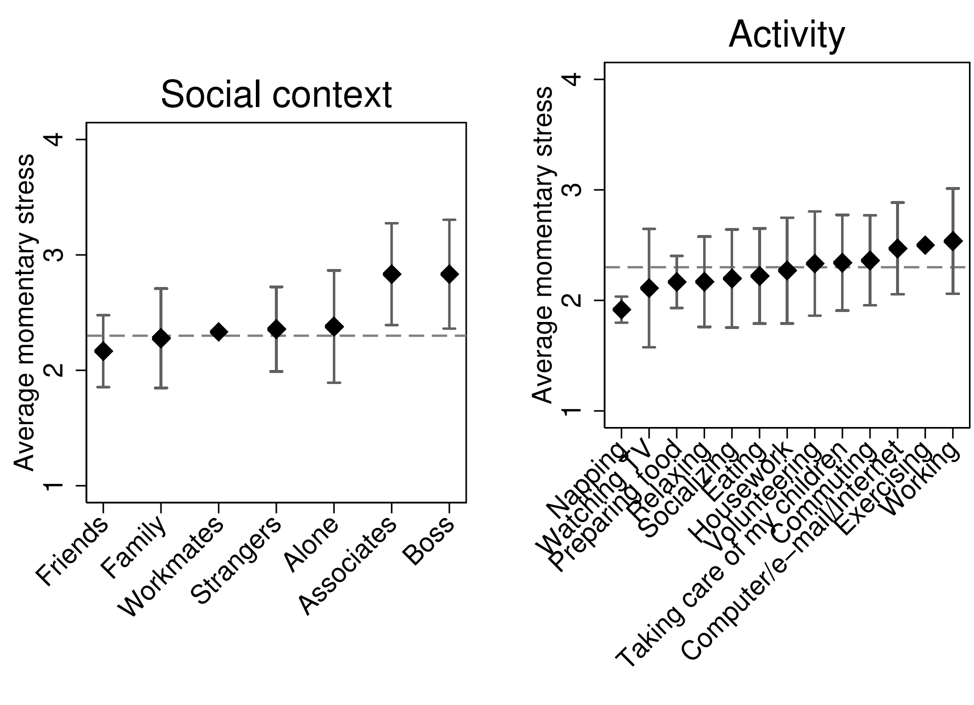

Figure 54: Situations ranked from least to most stressful, with mean and standard deviation

Working, commuting and taking care of children all rank at the high end, as expected

(Kahneman et al., 2004; Stone et al., 2006), whilst napping and watching TV are among the most relaxed. There has been some controversy before (e.g. Shiffman et al., 2008) about the validity of these rankings in reconstructed ratings, but this does not manifest here. For convenience I have added a dashed line representing the overall mean of momentary stress at 2.3 in both these graphs.

These divisions seem fairly natural, with nothing unexpected. I used two bivariate regressions to examine this more closely, tables 24 and 25.

Table 24: Estimates for the influence of social context on momentary stress

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
1: Alone***	0.1	0.04	0.005		0.03	0.17
2: Associates**	0.56	0.25	0.028		0.06	1.05
3: Boss*	0.56	0.31	0.073		-0.05	1.16
4: Family†						
5: Friends	-0.11	0.2	0.57		-0.5	0.27
6: Strangers	0.08	0.08	0.341		-0.08	0.24
7: Workmates	0.06	0.44	0.899		-0.8	0.91
Constant***	2.28	0.01	<0.001		2.25	2.3

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Table 25: Estimates for the influence of activity on momentary stress

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
1: Commuting	0.02	0.09	0.811		-0.16	0.2
2: Computer/email/internet**	0.13	0.07	0.049		0	0.26
3: Eating*	-0.12	0.07	0.098		-0.26	0.02
4: Exercising	0.16	0.43	0.71		-0.68	1
5: Housework	-0.07	0.09	0.456		-0.26	0.12
6: Napping	-0.42	0.3	0.162		-1.02	0.17
7: Preparing food	-0.17	0.21	0.418		-0.59	0.25
8: Relaxing***	-0.17	0.03	<0.001		-0.23	-0.12
9: Socializing	-0.14	0.11	0.188		-0.36	0.07
10: Taking care of my children†						
11: Volunteering	-0.01	0.3	0.981		-0.6	0.59

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
12: Watching TV	-0.23	0.25	0.354		-0.72	0.26
13: Working***	0.2	0.05	<0.001		0.09	0.3
Constant***	2.34	0.02	<0.001		2.31	2.37

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Although the company of the boss (table 24) has one of the highest estimates for increasing stress, adding 0.56 points on a scale of 1-4 compared to being with family, its 95% confidence interval contains zero (-0.05; 1.16) and it is only significant at the 10% level. Without reading too much into this, I suggest that this is still fairly noticeable given the low frequency of these encounters (see figure 53). Of the other categories, alone has the highest significance with $p \leq 0.01$, and an estimate of 0.1 (95% confidence interval 0.03 to 0.17). This is a reasonably large effect size given the stress rating range of between one and four points. If we were to combine working with the company of the boss, for example, that would predict an increase of 0.76 points, about 25 percent of the total scale.

Along with working, relaxing has the highest significance in this category with both at $p \leq 0.001$. Their predicted effect sizes (0.2 and 0.17 respectively) are similar too, and of reasonable size on the stress scale, 1-4. In fact, only commuting, housework, and volunteering have effect sizes smaller than 0.1 in either direction, so despite variations in significance and some in effect size, there appears to be good evidence for an effect of activity, as well as social context.

This is continued support for question three, as we continue to examine individual model components (i.e. figure 37). Yet these are individual exploratory analyses, not yet joined up, and not lending themselves to an overall assessment. So in the next section I carry out a full linear regression on the overall model to properly investigate question four.

Multiple regression

So far, we have seen support for some person-level effects, in the form of resting heart rate, and for those of situation, in day of the week, activity, and social context. Sleep proved to be a less useful predictor, though of some value as a response variable.

Following these analyses, I carried out a multiple linear regression using these components, with momentary stress remaining as the dependent variable. The results are in table 27.

We were expecting to see lower sleep duration and quality associated with increased stress, but the evidence for that was scanty. For extra verification, I fitted an initial model including the basic sleep variables (sleep the night before, number of awakenings the night before), and as predicted they were not significant. The estimate for sleep was 0.0001 (p=0.56; 95% confidence interval -0.0003 to 0.0005), and number of awakenings was 0.02 (p=0.2; 95% confidence interval -0.01 to 0.05). And similar to the model using sleep components alone, this prediction is a very small effect size on a scale of 1-4. In addition, a likelihood ratio test compared to a model with the situational variables and resting heart rate (RHR) was 2.08 at two degrees of freedom, a p of 0.35.

A review of the literature further suggested that higher resting heart rate would be associated with increased stress. Exploratory analysis supports this, with a moderate and significant positive association for RHR that day. But, given the lower variability of RHR over time, it correlates quite highly with itself in compound measurements (table 26). For this reason I only use daily RHR as a predictor, because as can be seen in table 22 earlier, the correlation with momentary stress is very similar for each measure.

Table 26: Correlation of resting heart rate measures

	Correlation with daily RHR
Weekly average RHR	0.94
Monthly average RHR	0.86

Looking at the other suggested components, the broadest support was found in theoretical review for the influence of context and environment, and the analysis above bears this out. Location was discarded as not informative, as I was at home 96% of the time (1,220 out of 1,265), 95.5% in the reduced dataset (757 out of 793).

This leaves social context, activity, day of the week, and resting heart rate as our final components. Including resting heart rate restricted the final model to 793 observations, versus 1,265 for the full dataset, but this did not prove a barrier to analysis.

Table 27: Estimates for the final model

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Social context						
1: Alone	0.08	0.05	0.109		-0.02	0.17
2: Associates**	0.53	0.24	0.027		0.06	0.99
3: Boss***	0.75	0.29	0.01		0.18	1.32
4: Family†						
5: Friends	0.02	0.21	0.935		-0.39	0.42
6: Strangers*	0.29	0.16	0.066		-0.02	0.6
7: Workmates	0.28	0.42	0.51		-0.55	1.1
Activity						
1: Commuting	-0.17	0.18	0.328		-0.52	0.17
2: Computer/email/internet***	0.25	0.07	<0.001		0.12	0.38
3: Eating*	-0.08	0.09	0.339		-0.26	0.09
4: Exercising	0.31	0.4	0.446		-0.49	1.1
5: Housework	0.01	0.1	0.897		-0.18	0.21
6: Napping	-0.39	0.4	0.329		-1.19	0.4
7: Preparing food	-0.17	0.29	0.552		-0.73	0.39
8: Relaxing***	-0.09	0.03	0.006		-0.16	-0.03
9: Socializing	-0.11	0.12	0.34		-0.35	0.12
10: Taking care of my children†						
11: Volunteering§						
12: Watching TV	-0.31	0.29	0.28		-0.88	0.25
13: Working**	0.17	0.07	0.016		0.03	0.3
Day of the week						
0: Sunday†						
1: Monday***	0.16	0.06	0.004		0.05	0.27
2: Tuesday***	0.14	0.05	0.009		0.03	0.24
3: Wednesday	0.02	0.05	0.712		-0.09	0.12
4: Thursday	0.07	0.05	0.223		-0.04	0.17
5: Friday**	0.11	0.05	0.034		0.01	0.22
6: Saturday	0.06	0.06	0.299		-0.05	0.17
Other						
Resting heart rate***	0.01	0	<0.001		0.01	0.02
Constant***	1.29	0.23	<0.001		0.85	1.74

† *indicates reference category*
§ *no observations in this dataset*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

In this model, activity, social context, and day of the week are dummy variables with a number of categories: thirteen, seven, and seven respectively. There is no *a priori* expectation of a particular ranking or interval between the levels. Resting heart rate is a continuous variable. Each category, and resting heart rate, show good evidence of statistical significance. For example, social context has three out of seven possible values significant at 10% or better.

Within this analysis the constant, or intercept, is not particularly meaningful as there will always be a value for situation and, hopefully, heart rate. Nonetheless, it is useful for interpretation. We can calculate predicted values for situations by inserting particular values for the variables. My model implies:

Intercept + Activity + Context + Resting heart rate + Weekday = Momentary experience

$$\text{momentary} = \textit{intercept} + \text{activity} + \text{context} + \text{resting heart rate} + \text{weekday} \qquad (1)$$

If a subject were relaxing, compared to taking care of their children, there is an estimated 0.09 point reduction in momentary stress on the scale 1-4 (95% confidence interval (CI) from -0.16 to -0.03). Being in the company of their boss is estimated to raise their stress level by 0.75 points (95% CI from 0.18 to 1.32) compared to being with family. Assuming a subject is on the low end of the adult average of 60 to 100 beats per minute (Solan, 2018), this is estimated to raise momentary anxiety level by 0.6 points (95% CI from 0.01 to 0.02). And if this were on a Monday, compared to Sunday, momentary stress would be an estimated 0.16 points higher.

1.29 + 0.09 + 0.75 + 0.6 + 0.16 = 2.89

2.89 points on the momentary stress scale is higher than the average of 2.3, and the standard error is 0.44 (see table 14 for full summary statistics). Fitting some predicted values for activity and comparing the means to the actual values (figure 55) shows that the predictions are fairly close to what was actually observed.

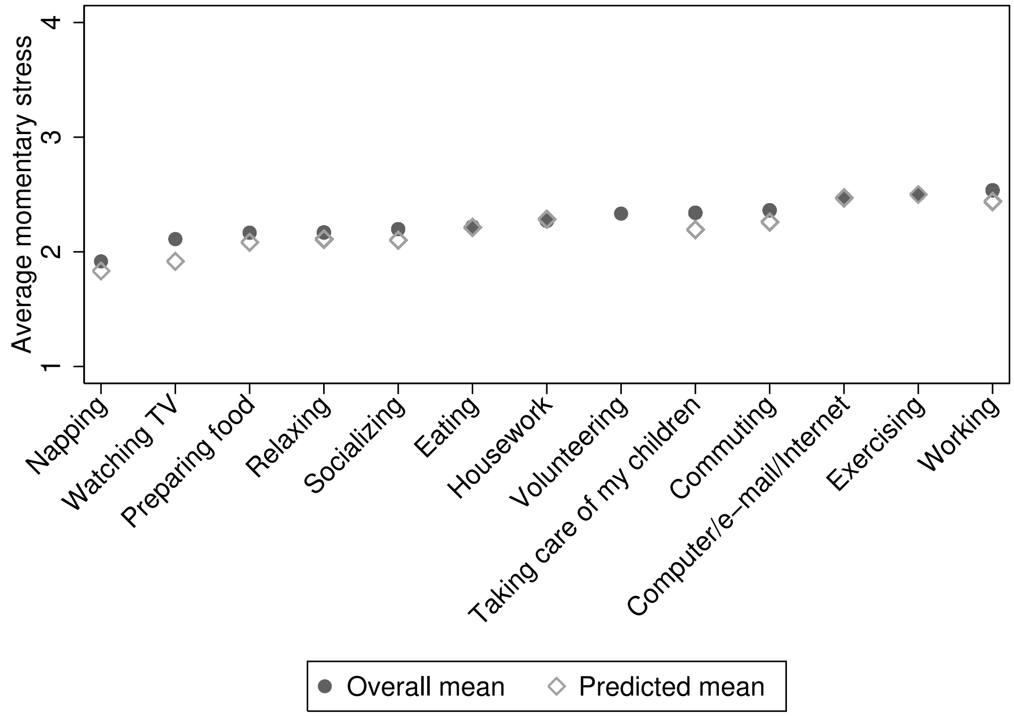

Figure 55: Predicted and actual mean stress by activity

For this full model the R^2 was 0.12, suggesting there is still unexplained variation. This is not surprising, given the simplified model and the complexity of individual behaviours. Rather, this study sought to find and examine relationships and associations between the different variables, without generalising too much, or trying to explain every facet of experience. A more complex model might include moderators and interactions, for example, but this study seeks to demonstrate the practicality and versatility of less involved analyses.

For corroboration, when comparing the model with situational predictors to the model with RHR alone, the likelihood ratio test was 68.13 at 17 degrees of freedom, giving a $p \leq 0.001$ of finding that value or greater. As we see in table 27, all of the categorical effects have some significance, and most predicted effect sizes are of at least moderate size. Resting heart rate is likewise significant, at less than the 1% level, but its predicted effect size is smaller than the average situational prediction.

Question four asked for broad support for the influence of the final chosen components. This analysis has demonstrated that, in significance and effect sizes. In addition, concordance with individual predictions was good: effect sizes are very similar, for example. This suggests a fair amount of confidence in the model is appropriate, and satisfies question four.

Discussion and conclusions

Methodological questions rightfully remain at the forefront of social science research. Experience sampling has been around in different forms for more than 80 years (P. Wilhelm et al., 2012), and has seen considerable improvement in application over this time. Disadvantages still remain, however, and this study has looked at new opportunities for enhancements opened up by recent waves of mobile instrument research (Harari et al., 2016; Raento et al., 2009).

In this paper I have tried to thoroughly explore the possibilities offered by this kind of method. I developed a mobile application for taking momentary measurements of stress, using a short form of Spielberger's State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983). With little extra burden for participants it is also able to broadly measure categories of situation including activity and company. Accompanied by a wrist actigraph device that automatically tracked sleep and heart rate, I took 1,265 readings over one year.

Initial evidence for the practicality of this approach seems good. The measurements were reliable (question one), and numerous (question two). The system seems to be effective, and evidence for both theoretical and narrative validity is good. On the negative side, a switch in activity tracking device showed that technology is no silver bullet, as the second device proved unreliable and measurements had to be discarded. Taking these results together gives rise to a number of conclusions.

Methodological

The first lesson learned is that it is impossible to design a method that works in all circumstances. Although I am a fairly experienced software developer, and carefully planned my interface design, I was not prepared for the impact of extreme circumstances on ability to participate. Even a few seconds per observation became too much. It is hard to avoid the conclusion that the length of the study had an impact, as I was a maximally willing participant: I therefore suggest that researchers be cautious about trying this length of experiment. Perhaps building in a break would have helped, or reducing frequency in response to circumstances. Had I limited the experiment to six months my conclusions about the burden of participation would have been very different.

On the other hand, it is hard to see the experiment itself as anything but a success, given the quantity and richness of the data gathered. The data collection strategy I used, even though streamlined as much as possible, resulted in several thousand data points (800-1,200 observations multiplied by the number of variables, not counting constructed variables.) Although I thought myself prepared, even I was surprised by the number of possibilities presented for analysis. I could have expanded just one or two sections to a complete article on various topics. It is clear that this single paper only scratched the surface of what can be done with mobile technology, both with experience sampling and in self-assessment.

Theoretical

This article has, by necessity, presented an abbreviated overview of the data, and a mostly exploratory analysis. Nonetheless we saw good evidence for the sufficiency of the model of stressful experience proposed in earlier sections. There was a significant effect of situation, as outlined in the descriptive analyses, and seen more clearly in the regression analyses. Resting heart rate, too, showed a moderate predictive power, and this also helped validate the more subjective conclusions. On the other hand, there was little evidence for the utility of sleep duration as a predictor of momentary stress, which was disappointing if not unexpected. Most research focusses on the trajectory of sleep influenced by stress (Hirotsu et al., 2015), so there was not a great deal of evidence going into the study. My suggestion that an investigation of sleep be accompanied by other measures still stands.

More broadly, further investigation can certainly be done with analyses using more specialised techniques for longitudinal and nested data. In particular, using multilevel analysis would allow better modelling of the effects of time, such as a random-effects model with nesting by time which better models distribution of variance, or a lagged-response model, which treats previous responses as a covariate so that we can predict their effect on future responses (Rabe-Hesketh & Skrondal, 2012, p. 228 & p.269). It would also be interesting to look at the evidence for polynomial effects, interactions, latent variable modelling, a mediation model, and many more. Although it is frustrating to leave so many possibilities unexplored, it is satisfying to confirm that my method produces these kinds of results. Future researchers, I hope, will have more time and skill.

Reflections

A major gap in my understanding of the process turned up partyway through the study. I had expected it would be easy to take these measurements. But when my stress got too high, it was like the polarity was reversed.

This is plainly visible in figure 56, which presents the number of measurements per day; there is a gap of 47 days running from November to mid-December. Just before this time my stress was rising (see the dashed line), and it began to feel like the act of assessing my stress, while we were in and out of hospital, was actually increasing it. So I took a break of six weeks. The impact can be further observed by examining the solid line, which is estimated by locally weighted scatterplot smoothing (also known as LOWESS), and shows the decrease in average measurements per day. Meanwhile the dashed line for stress rises, rated on the right-hand scale of momentary stress.

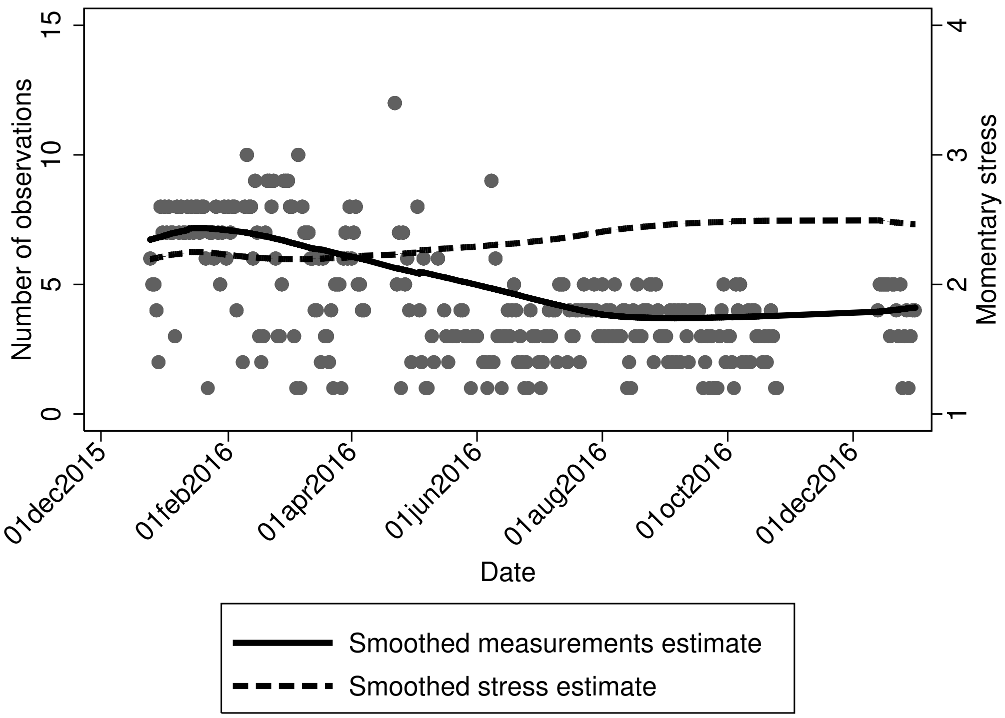

Figure 56: Number of observations with smoothed estimates for observation count and momentary stress

Visual comparison of the estimated lines for observation count and momentary stress suggests some evidence of a measurable relationship between the two, especially in the second half of the year. In fact, a bivariate analysis estimates the effect as -0.90 measurements for every point of stress ($p \leq 0.001$; 95% confidence interval -1.18 to -0.62). There might be more to the lowered count than measurement fatigue: participants in earlier studies (e.g. paper two) often said that they enjoyed the introspection of assessing their feelings. This suggests that under prolonged or extreme stress, the direction of this relationship might flip.

This is further supported by entries from my diary, which record my thoughts and hobbies, such as playing the horn; this gradually disappears until the period when even taking measurements stopped. It was harder than I expected to start them again.

Future directions

There are a number of directions future research could go. We discovered some limitations of single-subject data: in this particular study, a lack of variation in some of the contextual factors, particularly self-reported location, made the analyses less informative. For further exploration of situational effects, it will probably be necessary to recruit a number of participants, although with less data for each. A broader, rather than deeper, study of daily stress and stressors.

Nevertheless, I suggest that it is also possible to use this kind of exploratory investigation as a starting point, or a grounding terminal, for a deeper study of a particular area. For example, rather than broadening the sample range I could expand the category lists through qualitative investigation and iteration, adding person-specific situations based on things like free-form self-report, interview, or even GPS classifications. I feel, however, that this kind of work needs very specific research questions, or the study will get bogged down in the iteration phase trying to satisfy every possibility (Norton, 2001).

As noted in previous studies, investigation of a specific research question or research population would almost certainly benefit from more specialised categories for situational variables. Or researchers could choose to go with a more qualitative approach, asking for written, descriptive answers or even short recordings. Adapting the software to do this would be fairly trivial, and there are commercial options available too.

One possible question, of course, is the previously-discussed effect of stress on sleep, touched on briefly in earlier sections. There is already a very rich field of research, and it would be fairly straightforward to include, say, a reminder and reading for salivary cortisol (e.g. Dahlgren et al., 2009) at specified or random intervals with this kind of measurement application. A dedicated heart-rate monitor could provide momentary readings, and would not be much more cumbersome than a wrist-mounted fitness tracker.

And being able to estimate frequency or duration of situations suggests research going back to the peak-end rule, by Kahneman and others, which states that retrospective evaluations of emotional episodes tend to focus on the most intense part (the peak), and the end of the episode (Fredrickson, 2000; Kahneman et al., 1993). Are repeated situations in proximity more or less effecting, for example? Is there an adaptation effect? This study suggests that the effects of stressful situation do not, in fact, decline over time, but an analysis at the day or situation level could examine this further. Diurnal patterns can prove illuminating (Stone et al., 2006): figure 57 shows time, activity, and stress level for a single day.

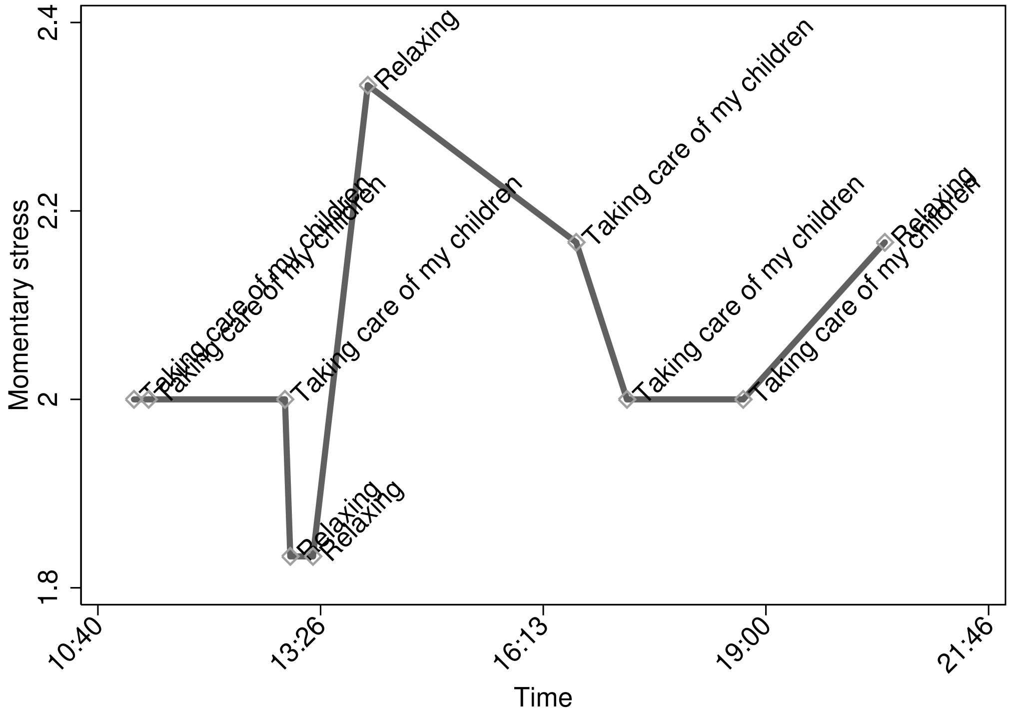

Figure 57: Stress timeline: 3rd March 2016

Finally are methodological considerations, ranging from issues of instrument administration that can be addressed by usage analysis, to the possible correlation between self-reflection, situation, and experienced intensity. For example, custom software could randomise user interface elements, or question order, or even administer a shorter or longer form of the instrument to evaluate the impact of different inputs on the experience and measurements. This study found some evidence of a link between a rise in stress and a reduction in measurement numbers, for instance; it would be very useful to see if there is some way of addressing this methodologically. For example, as an illustration of the adaptability of a software instrument, we could easily vary the number of questions inversely to the level of cumulative momentary stress, going from the 20-item to the 12-item, the six-, or even the four-item version (Tluczek et al., 2009). I believe that the possibilities in this area have only just begun to be explored.

Conclusion

This study proposes both a method and a model for taking fine-grained momentary measurements. I attempted to address design, measurement, and analysis, with some success. I think that this study has demonstrated the feasibility of this approach, and tested the validity of even an abbreviated model with multiple measurements over time.

Although some technical knowledge was needed, experience suggests that sufficient programming skill for this kind of work can be found by any researcher with experience of command line interfaces or scripted commands in statistical software. And there are plenty of pre-written tools available on many platforms, if time is an issue: a recent list is available from https://www.otago.ac.nz/psychology/otago047475.pdf (Conner, 2015).

My investigation was focussed on the components of momentary stress, but the flexibility of experience-sampling allows addressing of most theoretical interests (Hamaker, 2012; Shiffman et al., 2008; A. Stone & Shiffman, 2002), and is not limited to quantitative data. Any kind of information, from text to sound to video, can be collected in response to whatever kind of signal and interval the researcher decides on. Here I used a software approach, but notebooks, tape recorders, and landline phones have all been used, with pagers, watches, automated call reminders... the list goes on (e.g. Cikszentmihalyi, 2012; Hektner et al., 2007; P. Wilhelm et al., 2012). There is much that can be done using these approaches, and I hope my work goes some way towards encouraging further developments in this rich and useful field.

Preface: Paper Four

While paper three tested the effectiveness of highly detailed single-subject data, this paper conducts a more thorough analysis of many multiple-subject observations over a shorter time period. It also showcases a new approach to recruiting subjects for intensive longitudinal study, using Amazon's Mechanical Turk.

There are many studies on crowdsourcing data, and a number on mobile applications for momentary and longitudinal research. Yet there is little to nothing on crowdsourcing for mobile, momentary longitudinal studies. This felt like a gap that I could make use of, with a small amount of additional development.

The study confirmed the viability of this approach, and analysis of the results demonstrated a good level of validity. This suggests possibilities for further research, supplying data that is of sufficient richness for hierarchical/multilevel analysis.

Crowdsourcing momentary experiences of stress: A longitudinal study using Mechanical Turk and a bespoke software application

Abstract

This paper takes, as its substantive research question, the way in which everyday occurrence of stress is grounded in situations, persons, and times. I examine this by gathering momentary data on experiences of stress, and analyse them by using hierarchical models. Three essential elements of situation are proposed: what you are doing (activity), who you are with (social context), and where you are (location). As my approach to collecting these data is novel, a secondary aim is to assess the practicality and feasibility of the method.

Introduction

Research shows that momentary states are best explained by situating experience in its environment (R. A. Wilson & Clark, 2009), but also by differences between and within person and time (Bolger & Laurenceau, 2013). In this paper, I use a mixed effects model to try and outline some of these sources of variation, without forgetting situational factors that may be present at different levels. I used crowdsourcing to recruit 24 participants, with a median participation length of 6 days.

This study looks at experience through the lens of stress and anxiety, which, like all human behaviours, requires forethought and careful design to observe *in situ*. Chiefly cognitive stressors, such as anticipation (Bolger & Amarel, 2007) or negative thoughts (Nota & Coles, 2014), are hard to measure. Not only that, but they also vary in their individual impact, due to tendencies at the subject level such as learned reactions (see Strelau, 2010; and Kagan & Snidman, 2009, for much more detailed analyses). Environmental factors can push some individuals into "maladaptation" (Cicchetti, 2010), whilst others display resiliency in the face of similar difficulties. To study these processes, we need to define a general model that is sufficient to capture what is happening.

Stress in the moment

Stress has been extensively studied, but is difficult to describe with precision (Beehr, 2014). It can be measured with physical markers like cortisol levels (Adam et al., 2006), or with instruments like the State-Trait Anxiety Inventory (Spielberger et al., 1983), which has substantial confirmation for the validity of its construct.

Implicit to these approaches is the idea of stressors, factors including events and environment which precipitate changes in stress levels (Goldschmidt et al., 2014). In general, it is usually accepted that data about a person and an event's context are needed to understand the formation of momentary experiences (Fleeson, 2004). At its simplest level, a context or situation requires some element of activity, and these activities unfold over time (Augustine & Larsen, 2012). Reactions are often cued by the environment (Busemeyer et al., 2009), and so location becomes a variable of interest. Finally, there is a social context to behavioural processes (E. R. Smith & Conrey, 2009); the experience of being alone is appreciably different to being with family or work colleagues (Kahneman et al., 2004).

In this model I rely on Strelau (2010) and Kagan (2010; 2009), greatly simplified, to describe the core components of experience, seen in figure 58. A central idea in this model is the idea that experience is grounded in the environment it occurs in, at the time it occurs (Reis, 2012).

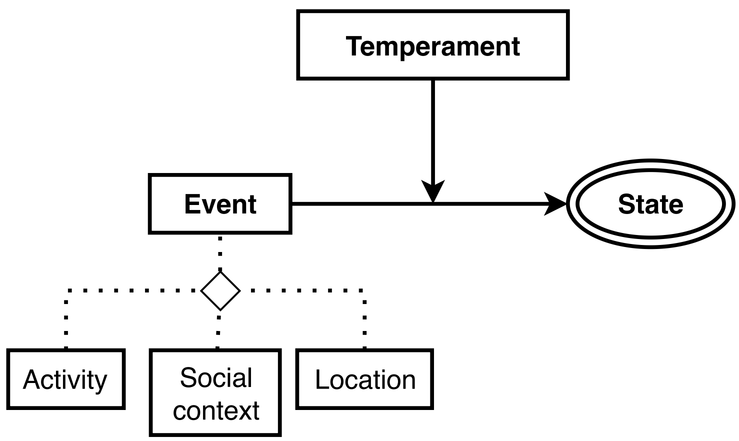

Figure 58: An elementary model of the components of momentary stress

To model this, and to differentiate between person-level tendencies in stressful experience, this study proposes observations nested within time and person. To explore person, I include multiple subjects, with a subject-level measure of trait anxiety, which represents individual sensitivity to stressors. And in order to situate these measurements in time, this study uses a longitudinal approach. In particular, to avoid the potential complications of single-item and retrospective assessments (e.g. Broderick et al., 2008), I ask subjects to take momentary measurements, repeated over a relatively short time period.

One established, but still evolving approach to obtaining momentary ratings is experience sampling, a form of “ambulatory assessment” (Ebner-Priemer & Trull, 2009; Fahrenberg et al., 2007) that allows for random, interval, or event-contingent triggering of observations as and where they happen (P. Wilhelm et al., 2012). This helps researchers assess patterns of experience over time, and in their natural environment (Holt, 2018). Although overlooked in the past as too demanding (Fleeson, 2007), modern experience sampling uses technology to take on most of the burden of recording and repeating (Harari et al., 2016).

In summary, both temperament and context are important factors in determining momentary state (Hamaker, 2012; Harari et al., 2016; Strelau, 2010). In addition, I will be looking for evidence of some kind of temporal influence, perhaps a cycle represented by weekday (Stone et al., 2012), even if it is no more complex than a simple weekend vs. weekday effect (Harvey et al., 2015; Ryan et al., 2010). Idiographic methods like experience sampling (Conner et al., 2009) offer opportunities to study person and situation simultaneously, and these new research strategies use mobile technology to increase the reach and scope of such methods (Fahrenberg et al., 2007; Vigo et al., 2017). However, these data can also be more challenging to explore (Fleeson, 2007).

Working with nested observations

With momentary assessment methodologies, measurements are often taken multiple times per day (Eid et al., 2012). That makes these kinds of data “inherently multilevel” (Nezlek, 2012); that is, observations within the same day are naturally clustered together as mood changes spread out over time, and the same can be said for observations by the same participant due to a person’s inherent characteristics. The basic idea is that measurements taken on one day, or by one person, are likely to be more similar to each other than they will be to another person or day. That is, moment-to-moment variation in stress will be at least partly explained by differences between individuals, and by changes over time (Shrout et al., 2006).

In this study, I am proposing a three-level structure. This structure (figure 59) suggests specifically that levels of momentary stress on different days within the same participant are correlated, and that levels within the same day for a particular subject are also correlated. An appropriate type of statistical model for data with a multilevel structure like this is a linear mixed-effects model (Rabe-Hesketh & Skrondal, 2012, pp. 85–86), and so I will be using these models in my analysis. This means that the predictors are estimated as ‘fixed effects’, set changes in the expected values of the response variable due to estimates of the effects of covariates; and that the clustering or nesting is represented by ‘random effects’, allowing the intercept to vary by participant and by day.

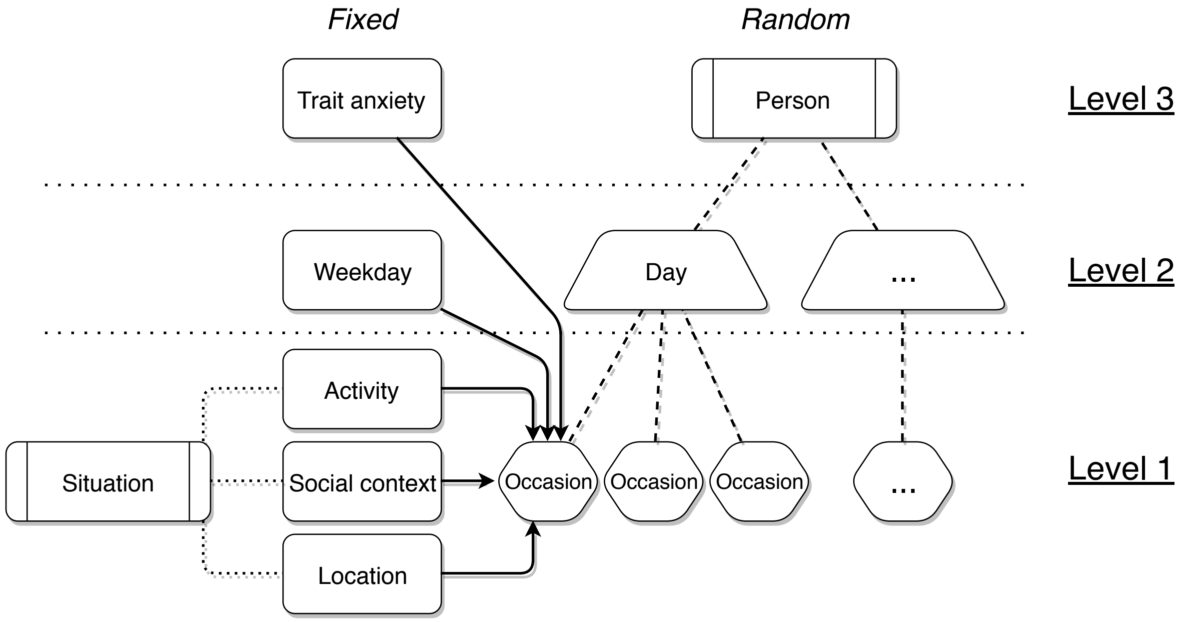

Figure 59: A mixed model of momentary stress

As this is a hierarchical model, some of the covariates occur at different levels. This means that for example, trait anxiety, measured at the start of the study, only varies at the participant level and is therefore a level three covariate. In fact, there are not too many predictors at levels two and three; there is also a single predictor at level two, day of the week, which varies between days but does not change within a particular day.

There are, however, several influences that occur at the occasion level in this model, and therefore change from measurement to measurement. These are the three situational factors in my elementary model of momentary states of stress seen in figure 58. This model proposes three essential elements of any situation: what you are doing (activity), who you are with (social context), and where you are (location). These are modelled as categorical or factor variables. A complete list of the possible categories for each is given in the results section further down.

Aims and questions

This study is intended to examine several ideas. I am testing recruitment for intensive longitudinal study via Mechanical Turk; I will be exploring the suitability of this data for a multilevel modelling analysis; and I will also be using this data to estimate the impact on momentary stress of influences at the trait and contextual level.

Whilst Mechanical Turk has been used extensively for longitudinal research, it is usually panel data with fewer measurements (Bolger & Laurenceau, 2013, p. 5), which often means web-based surveys instead of specialised software. Finding subjects should not be a problem - Apple have sold over a billion iPhones in the past decade (Apple Inc, 2016) - but it remains to be seen how this population of mobile phone users lines up with the population of Mechanical Turk workers. Here I hope that I will be able to recruit at least 20 workers with Apple devices to test the feasibility of the approach.

In addition, once the data is collected I expect to see reasonable evidence of clustering within person, and within time periods. The data should support examination of differences between, and similarities within, participant and day. This will allow estimation of fixed effects such as activity, social context, and trait anxiety as seen in figure 59, and random effects to represent clusters. So in addition to a general evaluation, I want to answer two specific theoretical questions.

Question one: How does structural dependence between and within levels explain variation in momentary stress between and within participants and over time? I will be looking at days nested within subjects, and occasions within days.

Question two: What are the fixed effects of trait, weekday, and situation (activity, social context, and location)? In earlier studies (see paper three) weekdays were significantly more stressful than weekends; and working, and being with work colleagues, was more stressful than relaxing, and being with friends and family. I am also expecting to see a positive effect of trait anxiety on momentary stress.

Method

Recruitment

Crowdsourcing is a way of distributing tasks of varying complexity to (mostly) anonymous workers across the internet (Benoit et al., 2016). It has become increasingly popular with social science researchers, not least because it offers a highly economical method of accessing a large, globally-located population. One of the most popular platforms is Amazon's Mechanical Turk (Litman et al., 2017). It was not designed as a research platform, but its flexibility means that with some work, it can be adapted for most tasks.

The unusual nature of this experiment, which required workers to install an application on their device, made recruitment slightly more uncertain, but these fears proved unfounded. A happy side effect of using custom software was that it was not really possible for 'spammers' (Benoit et al., 2014) to falsify results, given that much data capture was automated.

Participants

I recruited 24 workers by paying them five dollars for thirty minutes of their time, spread over the participation period (usually five days). There was no obvious bias towards either gender: my subjects were 52% male and 43% female, with 5% reporting ‘no answer’. This is roughly comparable to other, larger studies (see Mason & Suri, 2012), where 12.5% overall did not give their gender, but 55% of those who did identified as being female, and 45% of them as male. Ages ranged between 20 and 45 years old. Using data from 24 participants should be adequate to support the kind of modelling I use, with random effects, for which Rabe-Hesketh and Skrondal (2012) recommend at least 10 to 20 clusters.

Measures

I developed my own custom software to administer these measures, running on Apple’s iOS 9 or above (iOS is their touch-oriented mobile operating software; the acronym doesn’t stand for anything). I recommended an iPod Touch or iPhone, and participants all used iPhones. The software will run on an iPad, but these devices are not well-suited for experience sampling as they generally do not fit in a pocket.

Trait anxiety: The Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) (Spielberger et al., 1983), was administered by the software at the start and end of participation. It consists of twenty “anxiety-present” and “anxiety-absent” items (Marteau & Bekker, 1992), with ‘absent’ items reversed to give a total score of between 20 and 80. Not all subjects completed the final STAI-T, so I use the starting value alone in my analyses.

Momentary stress: The Six-item State-Trait Anxiety Inventory (STAI-6), an abbreviated version (Marteau & Bekker, 1992) of the Spielberger State-Trait Anxiety Inventory - State (STAI-S) (Spielberger et al., 1983), was used for the short daily readings. It has reliability and consistency comparable to the 20-item version (Nilsson et al., 2012; Tluczek et al., 2009), and contains three “anxiety-present” and three “anxiety-absent” items (Marteau & Bekker, 1992). A drop-down list presents four possible answers, rating agreement with the statement, and are scored from 1-4, with an overall score of 6-24, rescaled for ease of visual interpretation to between one and four overall.

Situation: Each observation also collects self-reported contextual information. In keeping with the simplified approach, there are three items: activity, company, and location. These are answered by selecting from a drop-down list. The set of answers are modelled on Kahneman’s (2004; 2006), coded from a large sample and with proven validity (e.g. Stone et al., 2006). All possible answers can be seen in the results section.

Time: Complete timestamps were added automatically to every observation, which were semi-random during a time specified by the subject. The period (e.g. 9am to 9pm) was divided by the software into 'slots' by number of measurements per day, then the notification would be generated at a randomised time within the slot. For example, five measurements between 9am and 9pm would result in five slots of 144 minutes each, with each measurement occurring at a random time during those 144 minutes.

Procedure

I built an application to run on the iPhone and iPod Touch using Apple's Xcode integrated development environment (IDE). This enables a graphical, touch-screen interface for ease of use. Figure 60 shows a sample screenshot with the wording of the Spielberger STAI obscured with dummy text to preserve copyright. Responses are from drop-down menus, which is quick and no typing is required.

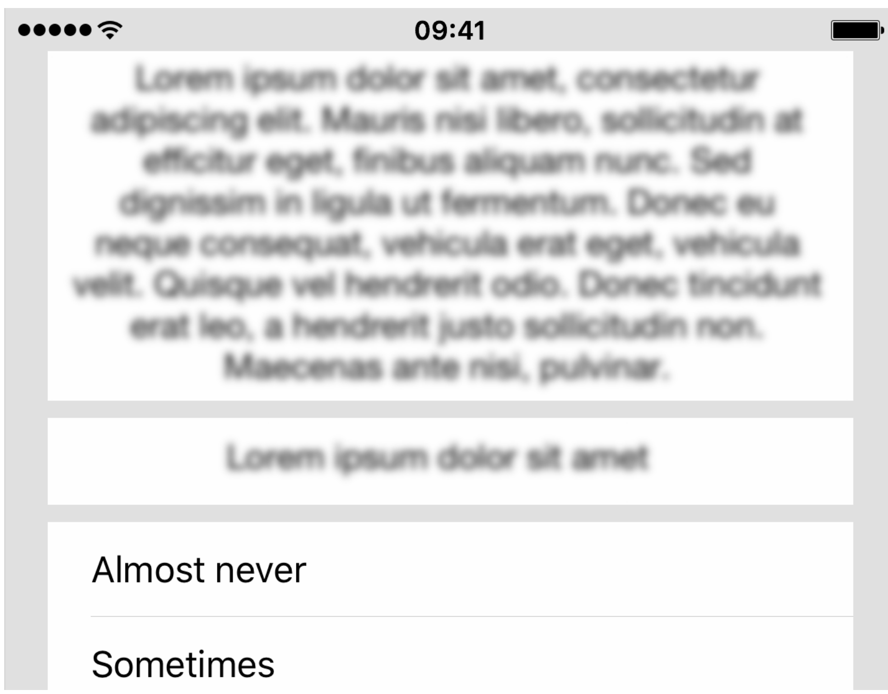

Figure 60: User interface

Measurements were taken with the application running locally on the device, at semi-random intervals, and the results were uploaded to a secure server using the device's wi-fi.

Conducting the study required a combination of services: a crowdsourcing platform for recruitment, and a distribution platform to actually get the application to users' devices (figure 61). As mentioned earlier, I used Mechanical Turk by Amazon for recruitment, paying a few dollars per worker. I used HockeyApp, by Microsoft, for distributing the application, which is free and requires little information from the participants.

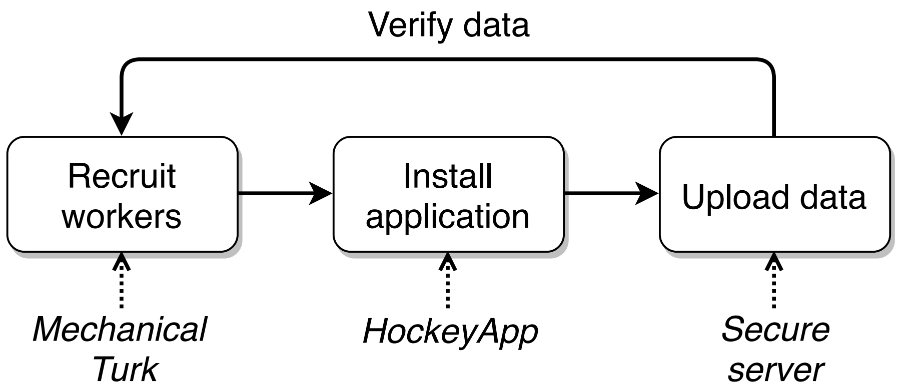

Figure 61: Participation procedure

Other than resolving the occasional minor installation problem via email, I had no contact with the participants other than through the data they uploaded. Whilst not unusual for a crowd-sourced study, this is a marked contrast to earlier recruitment efforts (see paper one and paper two) and may mark a future area for experience sampling study. Installed software, as opposed to web applications, offers several advantages in reliability, customisation, and usability, as the continued use of PDAs in other studies shows (e.g. Holt, 2018). It was also very cost-effective compared to the price of a device. Yet cost is not the only concern, and there are other benefits to having a dedicated device or devices to work with. In particular, supplying a device can allow you to work with otherwise hard-to-reach populations (as in paper one). It is worth considering which approach best suits the study in question.

Results

This section starts with exploratory indications for a multilevel/hierarchical structure by summarising the data by subject and day. We then look at evidence for variation within and between participants in the situational categories by examining frequencies.

Overall descriptives

Table 28 describes levels of momentary stress, clustered by participant.

Table 28: Summary of momentary stress by participant

		Mean	Std. Dev.	Min	Max	Observations
Stress	overall	1.87	0.73	1.00	4.00	Total = 683
	between		0.55	1.00	3.01	Clusters = 24
	within		0.51	0.53	3.79	Average/cluster = 28.46

There are 683 total observations in table 28, with 24 clusters representing individual subjects. On average there are 28.46 measurements per participant, showing some participants took more than the 25 required. Momentary stress varies slightly more between participants than it does within participants, but the standard deviations are roughly comparable, half a point on the 1-4 scale. This implies that analyses at the person and measurement level will prove useful.

Because this study includes an element of time, looking at these statistics grouped by day is also useful (table 29).

Table 29: Summary of momentary stress by day

		Mean	Std. Dev.	Min	Max	Observations
Stress	overall	1.87	0.73	1.00	4.00	Total = 683
	between		0.26	1.58	2.6	Clusters = 17
	within		0.72	0.91	4.16	Average/cluster = 40.18

Here there are 17 clusters, representing individual days of participation. Looking ahead to table 30 for confirmation shows that most subjects took part for five or six days, with a very small percentage going beyond requirements. Although almost three times as much of the variation occurs within days, compared to between days, there is still a reasonable amount between. This suggests that, as proposed earlier, at least part of the variation should be explained by the passage of time.

Figure 62 shows the similarities in averages for the first three days of study, grouped by subject.

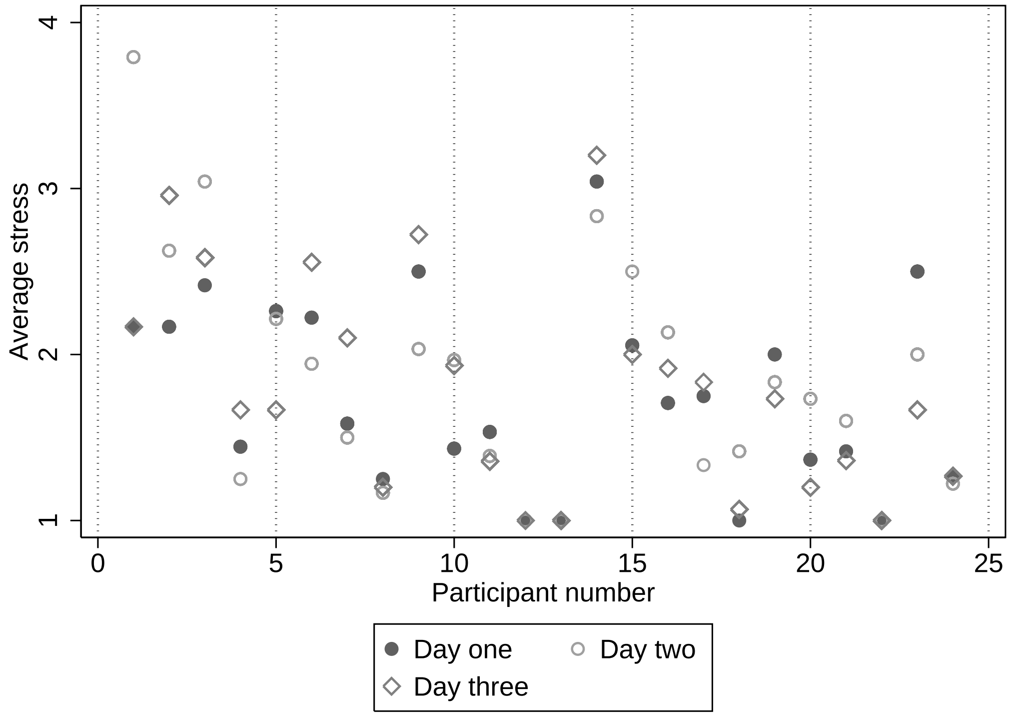

Figure 62: Average stress for first three days

We can see that measurements appear to cluster within participants more than they do across participants, and that days close together in time look like they tend to be more similar within participant, too.

Participation

Table 30: Summary of days participated

Elapsed day	Number of observations	Percent of observations	Number of participants
1	86	12.59	24
2	106	15.52	24
3	103	15.08	24
4	120	17.57	24
5	105	15.37	24
6	74	10.83	17
7	25	3.66	8
8	15	2.2	3
9	10	1.46	2
10	7	1.02	1
11	5	0.73	1

Elapsed day	Number of observations	Percent of observations	Number of participants
12	6	0.88	1
13	5	0.73	1
14	5	0.73	1
15	5	0.73	1
16	2	0.29	1
17	4	0.59	1
Total	683	100	158

Table 30 describes the pattern of participation among subjects. For instance, 12.59% of overall measurements were taken on day one of participation, and there are no differences between participants in the first five days as 100% of them took at least one measurement on those days. This shrinks increasingly rapidly as we pass six days.

Frequency of covariates

These tables give an overview of the distribution of situational variables. They are clustered by participant, which means subject numbers are represented in the 'Between' column.

Table 31: Summary of activities

Activity	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Commuting	37	5.42	15	62.5	8.77
Computer/e-mail/Internet	146	21.38	21	87.5	24.79
Eating	26	3.81	15	62.5	7.07
Exercising	19	2.78	9	37.5	9.34
Housework	62	9.08	15	62.5	12.53
Napping	20	2.93	10	41.67	8.15
On the phone	14	2.05	8	33.33	5.58
Pray/worship/meditate	2	0.29	2	8.33	6.17
Preparing food	25	3.66	12	50	6.67
Relaxing	106	15.52	17	70.83	21.21
Shopping	10	1.46	7	29.17	5.44
Socializing	17	2.49	7	29.17	7.47
Taking care of my children	25	3.66	9	37.5	8.69
Watching TV	64	9.37	16	66.67	12.82
Working	110	16.11	18	75	23.18

Activity	Overall		Between		Within
Total	683	100	181		

Taking ‘Commuting’ in table 31 as an example, the ‘Between’ column shows that 62.5% of participants, or 15 people, rated themselves as ‘Commuting’ at least once. The interpretation of the ‘Overall’ column is that participants rated themselves as ‘Commuting’ in 5.42% of total measurements, or on 37 occasions. And using the ‘Within’ column we can see that for those 15 participants who did commute, that is, rated themselves as ‘Commuting’ at least once, the number of occasions when they were ‘Commuting’ averaged 8.77% of their total occasions.

These statistics also demonstrate differences between individuals in activities. Staying with the commuting example, we saw that 15 people reported commuting at least once. Compare this to the nine who recorded ‘Taking care of my children’ once or more: two-thirds as many subjects, but within those two groups the number of occasions when they reported taking part in these activities was about the same (8.77% and 8.69% respectively). This implies that these two groups spend about the same amount of time on these activities, but that taking care of children is less common over total participants.

Table 32: Summary of social context

Company	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Alone	384	56.22	23	95.83	60.15
Associates	11	1.61	6	25	9.09
Boss	16	2.34	5	20.83	12.18
Family	192	28.11	18	75	33.88
Friends	44	6.44	13	54.17	12.74
Strangers	8	1.17	6	25	4.91
Workmates	28	4.1	6	25	16.05
Total	683	100	77		

In this table, table 32, participants rated themselves as ‘Alone’ in 56.22% of the measurements, and 95.83% of participants took a measurement when they were ‘Alone’ at least once. For those who took any, an average of 60.15% of their measurements were taken ‘Alone’. As an illustration of the differences between subjects, six participants (the ‘Between’ column) reported being with strangers at least once, and six with ‘Workmates'. However, those who recorded 'Workmates' did so comparatively often, around 16% of measurements compared to about 5% with strangers.

Table 33: Summary of locations

Location	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
At home	502	73.5	24	100	72.24
At work	94	13.76	14	58.33	24.48
Somewhere else	87	12.74	18	75	17.98
Total	683	100	56		

The percentages here in table 33, for example 'At home', show that participants said that they were at home 73.5% of the time, and that 100% of participants took at least one reading when they were at home. On the other hand, 75% reported being 'Somewhere else' at least once, but those 18 subjects reported this only 18% of the time recorded.

Table 34: Summary of weekdays

Weekday	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Sunday	100	14.64	22	91.67	16.44
Monday	98	14.35	19	79.17	18.07
Tuesday	91	13.32	19	79.17	17.18
Wednesday	108	15.81	23	95.83	16.22
Thursday	82	12.01	20	83.33	13.95
Friday	107	15.67	20	83.33	18.18
Saturday	97	14.2	22	91.67	16.04
Total	683	100	145		

Summarising weekday by participant in table 34 shows that most participants (around 80%) were represented on every weekday, which shows a good spread over the week given that just under one week was the average participation period.

Table 35: Summary of trait anxiety scores

Mean	Std. Dev.	Min	Max
43.25	13.08	20	66

Self-assessed trait anxiety at the start of the study, seen in table 35, is quite variable, but scores tend more towards the low end (20 points is the minimum possible and 80 points is the maximum.)

Mixed-effects models for the measurements of stress

I used Stata 11 (StataCorp, 2009c) to carry out a series of linear mixed effects analyses of the relationships between and within levels. To begin with, I evaluated whether partitioning variance by participant and day is justified by the data.

Using the likelihood ratio test, I first tested the hypothesis that between-subjects, or participants, variance is equal to zero (the null hypothesis) by comparing a model with a random intercept for participant to the null model without a random intercept. The test statistic was 399.67, giving a p value of <0.001, meaning we can reject the null hypothesis. I then tested the two-level model (random intercept at the participant level) against the three-level model (a random intercept for the between-days and within-participants variance). Here the test was 95.24, also giving a p value of <0.001 against the null hypothesis.

When I estimated the full model there were 24 participants with an average of 28.5 measurements each, and 158 days-within-participants (days multiplied by participants) with an average of 4.3 observations per day. The results are unstandardised.

Table 36: Estimates for fixed effects

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.232	0.242	0.337		-0.242	0.706
Trait anxiety	0.032***	0.005	<0.001		0.022	0.043
Activity						
Commuting	-0.111	0.123	0.367		-0.351	0.130
Computer/e-mail/Internet	-	-	-			
Eating	-0.083	0.094	0.379		-0.268	0.102
Exercising	0.014	0.106	0.894		-0.194	0.222
Housework	0.257***	0.070	<0.001		0.120	0.394
Napping	-0.193*	0.105	0.067		-0.399	0.013
On the phone	0.840***	0.122	<0.001		0.601	1.079
Pray/worship/meditate	-0.171	0.309	0.582		-0.777	0.436
Preparing food	0.105	0.096	0.274		-0.083	0.294
Relaxing†	-0.086	0.065	0.183		-0.213	0.041
Shopping	-0.243	0.172	0.158		-0.580	0.095
Socializing	0.062	0.117	0.597		-0.168	0.292
Taking care of my children	0.434***	0.101	<0.001		0.236	0.632
Watching TV	-0.123*	0.068	0.070		-0.256	0.010
Working	0.268***	0.067	<0.001		0.137	0.399
Social context						
Alone†	-	-	-			
Associates	0.144	0.141	0.308		-0.133	0.421
Boss	0.185	0.129	0.151		-0.067	0.438
Family	0.003	0.049	0.955		-0.093	0.099
Friends	-0.268***	0.097	0.006		-0.459	-0.077
Strangers	-0.135	0.169	0.425		-0.466	0.196
Workmates	0.008	0.113	0.947		-0.214	0.229
Location						
At home†	-	-	-			
At work	0.039	0.074	0.602		-0.106	0.183
Somewhere else	0.382***	0.100	<0.001		0.186	0.578
Weekday						
Sunday†	0.153	0.110	0.164		-0.063	0.370
Monday	0.111	0.112	0.324		-0.109	0.331
Tuesday	0.080	0.112	0.475		-0.139	0.299
Wednesday	-	-	-			
Thursday	0.142	0.113	0.208		-0.079	0.363
Friday	0.153	0.109	0.162		-0.061	0.367
Saturday	0.139	0.109	0.201		-0.074	0.353

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Consistent with predictions, in table 36 the level three covariate, trait anxiety, is estimated to have a significant association of 0.03 (95% confidence interval (CI) from 0.02 to 0.04, $p < 0.001$) additional points of momentary stress per additional point of tendency to anxiety. This is on a scale of 20 to 80, so the minimum estimated effect would be an increase of 0.6 (20 by 0.03) on the momentary stress scale of 1-4 points total, which is quite large, certainly comparable to any of the other covariates.

While there are significant estimates of the slope for activity, social context, and location (table 36) at level one, there are no significant differences by weekday at level two. I tested a separate model with weekday as a two-category variable (weekday vs. weekend), but that was still not significant at $p = 0.443$.

The intercept for momentary stress is estimated at 0.23. With other variables held constant, 'Working' estimates a 0.27 point increase (95% CI from 0.14 to 0.4, $p < 0.001$) in momentary stress on the scale of 1-4, compared with general use of a computer for email or internet. The reference category is chosen as the most commonly encountered value (see occurrence section for numbers, for example table 31 for activity). Being with friends is associated with a 0.27 point decrease (95% CI -0.46 to -0.08, $p < 0.001$) in momentary stress compared with being alone. And being 'Somewhere else' is estimated to increase momentary stress by 0.38 points (95% CI 0.19 to 0.58, $p < 0.001$), contrasted with being at home, and after controlling for other variables.

Table 37: Estimates for random effects

	Estimate	(*SE*)	95% CI	(*lower*)	(*upper*)
Standard deviation: participant	0.299	0.056		0.208	0.431
Standard deviation: day (within participants)	0.319	0.029		0.267	0.380
Standard deviation: occasion (within days and participants)	0.375	0.012		0.353	0.399

Examining the random part of the model, table 37, the between-participant standard deviation is estimated as 0.299 (95% CI: 0.208 to 0.431). Within participants, and between days, the standard deviation is estimated as 0.319 (95% CI: 0.267 to 0.380). This implies that the correlation is strongest for the same person on the same day, but that there is also a strong correlation between measurements for the same person on different days.

Fitting a box plot of predicted values, figure 63, makes the interpretation clearer. We can see that more variation occurs within participants, but that there is a fair amount between participants too, just under half as much. There are also some outliers at day and occasion, which could be worth examining in a follow-up, particularly day.

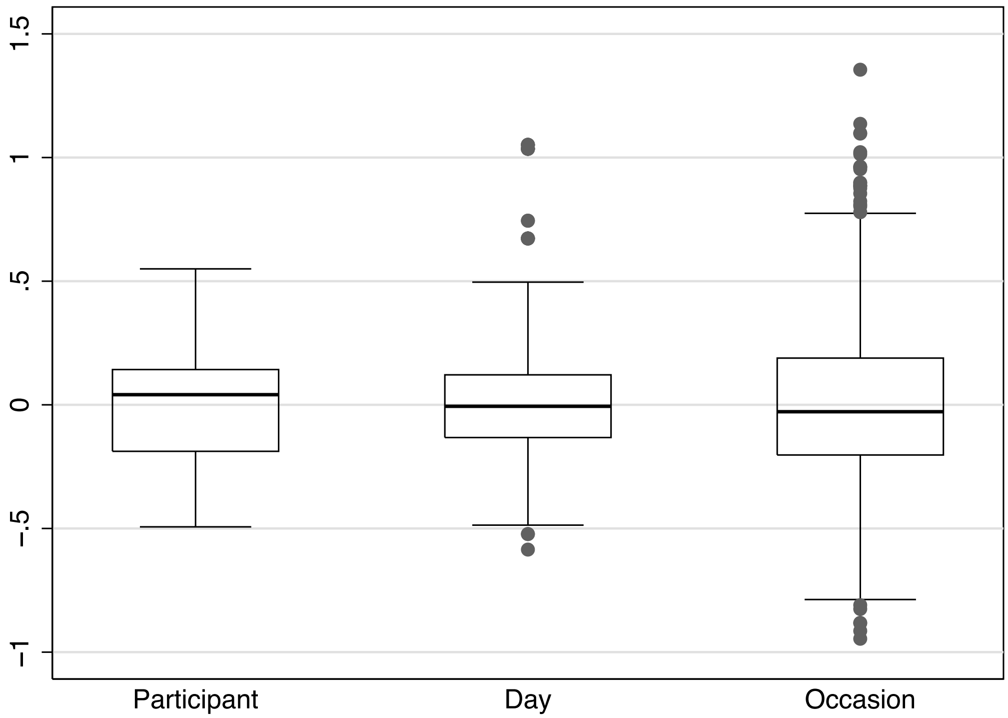

Figure 63: Box plots of predicted values for the intercepts at participant and day, and residuals at the occasion level

Finally, graphing the predicted values for a single participant in figure 64 shows that the model appears to fit the data reasonably closely. This suggests that my model is a useful one.

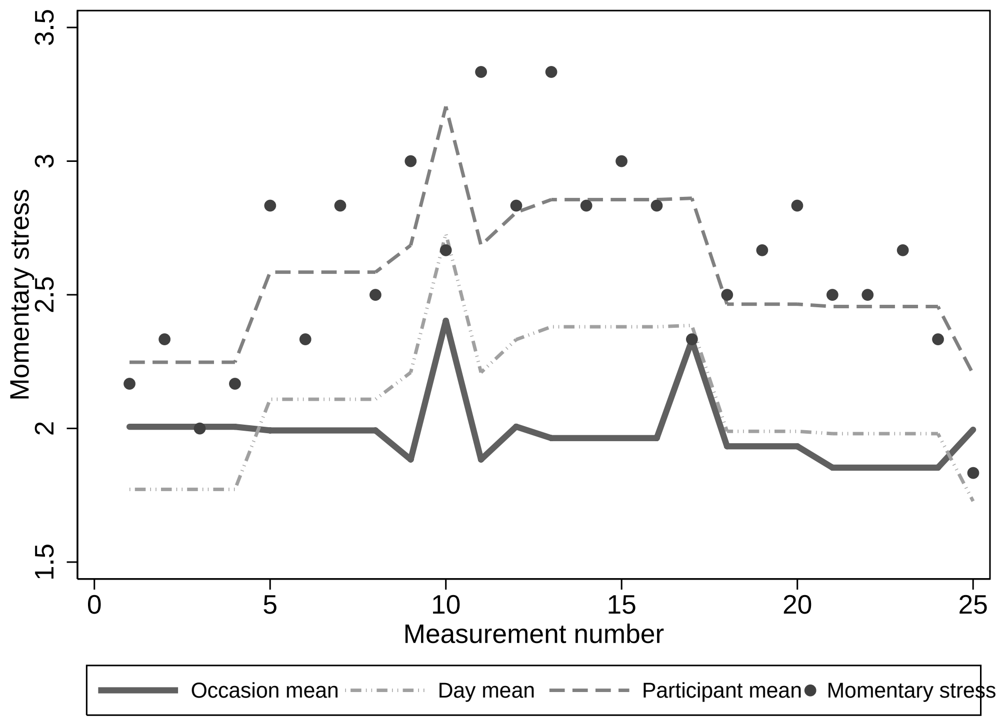

Figure 64: Predicted values for participant, day, and occasion for participant two

Discussion and conclusions

This study aimed to test a novel recruitment and measurement approach, and find evidence for a situational, hierarchical model of momentary stress. For the first part of the study, the investigation focussed on whether it was possible to recruit participants for intensive study over time, who would be required to use purpose-built software, through a crowdsourcing approach. The answer is a definite yes. Costs were very reasonable, participation rates were good, and it was straightforward to recruit the number of subjects I was aiming for. I suggest that future studies could recruit many more participants with very little additional effort.

The verdict for the theoretical questions was also a positive one. The first question was about the structuring of the data, which was expected to explain some of the between and within variations in momentary stress. There was good evidence for stability over days and participants (table 37), with notable correlations within subjects over days, and within days for the same subject. This matches previous findings about momentary affect, and stress in particular (Bolger & Amarel, 2007; Pottie & Ingram, 2008).

The second question was concerned with the effects of the different ecological, or contextual, variables at the occasion and subject levels. The results are in table 36. I had proposed a fixed effect of trait anxiety, or tendency to anxiety, and of three basic situational factors: activity, company, and location. The individual categories within situational factors were classified very broadly to make them suitable for a general population. I had also suggested a basic working day effect of increased stress against a weekend effect of reductions. There was little evidence for that, unlike in a previous study ([see paper three][A year of anxiety: An individual experience-sampling study]). There is, on the other hand, evidence that this kind of effect is strongly influenced by demographic and socio-cultural variables (Stone et al., 2012), so this was not completely unexpected. However, there were significant effects of trait anxiety at the person level, and of the three situation factors at the observation level.

Trait anxiety, in fact, has one of the largest predicted effect sizes, although this is camouflaged by it being unstandardised. This means that we need to bear in mind that the minimum score is 20 points, or 20 times the effect size, and scores can go as high as 80, theoretically. In this study the mean was 43.

Effect sizes for activity, location, and company were appreciable, too, and predicted values based on this model seem to fit the data well (figure 64). This shows that the 'elementary' model (figure 58 and figure 59) is not too abbreviated to make useful predictions about behaviour. This finding suggests that the model could be used for a more confirmatory study with a fair amount of confidence. It might also be reliable enough for something more unusual like an intervention application: for example, tailored feedback on regulation strategies during periods of high stress (Ebner-Priemer & Trull, 2009). A similar approach has been used with some success in activity trackers, which remind users to move about after periods of inactivity.

One limitation of this study was a lack of detailed information about the participants' demographic characteristics. I left this deliberately imprecise in order to encourage participation, as I was concerned that subjects would be unwilling to share too much information about themselves when they were supplying such intimate data about their daily experiences. Having established the feasibility of recruiting in reasonable numbers, it may be time to more closely examine individual characteristics. These were less important for this particular study, with no related hypotheses, but it would almost certainly prove necessary for studies of specific populations.

Overall, the results of this study suggest that even a streamlined approach to ecologically modelling momentary experience can be an effective one. And the effect sizes and significance imply that a very simple mixed effects modelling approach nonetheless has good explanatory power when we wish to analyse between and within sources of variation. I hope that more research can build on this.

Preface: Paper Five

This paper takes an extended look at the influence of trait anxiety as a moderator of momentary anxiety, considering situational factors such as location and activity as demands, or stressors. Hierarchical modelling is used to account for similarities within individuals, and within time periods, by the use of clustering.

The starting point is study four, which is used to inform sample size for a power analysis and later, as a population comparison. Using the same, proven recruitment procedure and data collection application, the required number of subjects were recruited using Amazon's Mechanical Turk, and the data were used to test the moderation effect of temperamental tendency toward anxiety on contextual elements of experience.

Temperament as a moderator of stressful situations: Expanding on crowdsourced momentary experiences of stress

Introduction

It is not controversial to state that stable individual patterns of behaviour can explain momentary individual differences in reaction to activity or circumstance (Kekäläinen, Sipilä, et al., 2021). In this series of studies, temperament has been interpreted in a specific way, one related to momentary anxiety, as a subject-level trait. This trait is defined as an inclination to increased or decreased amplitude of reactions to stress (Spielberger & Reheiser, 2009), which is one of the factors in vulnerability to daily stressors (Almeida, 2005). It is reasonable therefore to expect to see measurable differences in levels of momentary anxiety, for different individuals at different levels of trait anxiety (Rudaizky & Macleod, 2013), when observing their reaction to the "stimulating value of demands" (Strelau, 2010, p. 124).

This study tests this empirically, longitudinally, and in the moment. To begin with, additional data collection will further strengthen the findings of paper four, which starts from a similar perspective. It also allows for extension. In paper four results showed significant effects of trait and situation, and allowing variations at the intercept level for participant and day (random intercepts) made the model better fitted to both the data and the theoretical assumption of similarities within person and time period. This gave some evidence for the influence of temperament, as correlations were strongest within person and day, but invited further analytical exploration of the true effect.

Therefore, this paper explicitly fits an extended set of variables, in terms of interactions, to more fully model trait anxiety as a moderator of stressful experiences (figure 65). This approach allows analyses to build on the mixed effects model described in paper four.

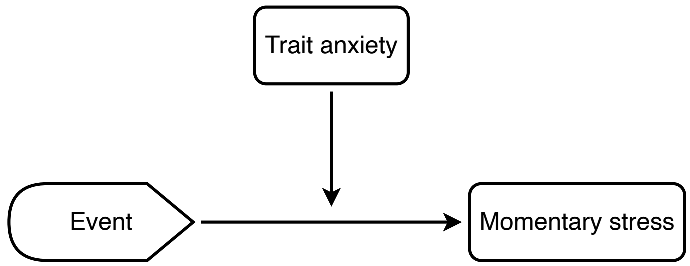

Figure 65: Simplified moderation model

To ensure the data sets were comparable, a very similar data-gathering procedure was carried out, with a slight update to the distribution method, but not the measurement tool or the assumed population. To this end, the software application (‘app’) used for data collection was identical, and the study recruited from the same population of Amazon Mechanical Turk (‘MTurk’) workers. Before data collection, a set of simulation-based power analyses (Lafit et al., 2021) were performed to determine the sample size. Following this determination, the study recruited 30 participants, who each took at least 25 measurements over five days.

After collecting sufficient data, several sets of mixed effects models were fitted using both study four and study five sets of data in turn. The study five data were tested without a moderator, for replication, and then separate extended moderation models for studies four and five were fitted: a model which included a moderating effect of trait anxiety.

The following section summarises this study’s aims. Pre-collection analyses are then discussed, alongside the recruitment procedure, and an examination of the characteristics of the participants recruited. One way in which the sample differs from the previous study is the impact of the COVID-19 (coronavirus) pandemic. The first data collection batches were run in March 2020 and the last in May 2021, just before the UK’s Prime Minister Boris Johnson declared the pandemic over in June 2021 (Mahase, 2021). Even before the pandemic, anxiety disorders ranked high among global health burdens (Santomauro et al., 2021), and these data sets happen to give some indication of changes in the frequency and distribution of situations and activities between the study periods. The analysis will touch briefly on this in a later section.

Aims and hypotheses

This study tests two simple hypotheses. The first is held in common with paper four, whilst the second extends the focus of this study to consider questions of moderation.

1. Moment-to-moment changes in anxiety are partly explained by changes in activity, social context, location, and weekday.
2. The direct effects of these situational variables are moderated by trait anxiety.

Expanding on the results of paper four, this study aims to find a sufficient sample size to test the new hypothesis, and to recruit an appropriate pool of participants. Like paper four, this study will be recruited through Mechanical Turk (MTurk). This is slightly different to most common tasks on MTurk, as momentary longitudinal studies require more intensive participation, and some observations were gathered. Also novel is the historical data, which gives the newer study an empirical basis for estimating the recruitment numbers, and so a power analysis will be carried out next.

The extended (study five) model will be fitted to the previous data (study four), to see how the moderation hypothesis holds up. Reciprocally, the paper four model will be fitted to the current data. This will allow comparison of the models, albeit in a limited way for verification. Although the extended (study five) model will be fitted to the previous data (study four), to see how the moderation hypothesis holds up, the main aim of this paper is to explore possible moderation effects using the new set of data. As such most of the replicatory analyses are found in the appendix at the end of the paper.

Method

The current methodological approach is built on the design from study four. So, for example, the same measurement items were used to ensure compatibility; the same software, to ensure the same processes were followed; and the analyses are built on a common model which is fitted jointly on each study.

The power analysis was a new and substantial addition to the process; the following sections go into more detail, some of it technical, on the procedure, followed by the recruitment and participation process, and a summary of the resulting participant details. There is also a brief discussion of some quirks of recruiting study participants through Amazon Mechanical Turk.

Power analysis

Traditionally, power calculations are difficult when there are multiple levels of correlation (Arnold et al., 2011), and are difficult to carry out with longitudinal data (Lafit et al., 2021). In the current study, there are two sources of correlation in the dependent variable: within participants and within days (see figure 68), and data are unbalanced.

It is suggested that such a clustered distribution is suitable for a simulation approach to power estimation (e.g. Arnold et al., 2011; Lafit et al., 2021). In this case it is possible to estimate plausible values for the hypothetical parameters of possible distributions using a training or pilot dataset, in this instance the existing data from study four (n=24), and use this structure to simulate response variables. The simulation repeats these steps for each potential sample size, and/or each specified value of the coefficient(s).

1. Sample (with replacement) the desired number (n) of subjects from the real dataset from study four (n=24) to generate a dataset of independent variables clustered by participant and day. This dataset includes all the observations for each sampled subject, to ensure that the values of the independent variables, and their intraclass correlations, are plausible. Each subject in the study four data has an average of 28 observations, so each of these re-sampled datasets include approximately 24x28 = 672 subject-occasion observations.
2. Sampling with replacement means that some subjects may be included several times, but will be treated as distinct for the purposes of simulation in step 2.
3. Using a random-effects model similar to the fitted model from study four, the response variable is then simulated based on the estimated regression coefficients and variance parameters from the fitted model. Most of the parameters will be set at their estimated values, but this function allows setting one or more coefficients to different values, to assess estimated power given these values.
4. Carry out significance tests of the coefficients of interest by fitting the chosen model to the simulated dataset.
5. Store the results.

This is done 1000 times. To determine the power, the proportion of these 1000 datasets for which the null hypothesis of zero is rejected for a coefficient with a given true value, at a given level of significance, must be found. This proportion gives the estimated power for that coefficient at that level of significance. The whole process is repeated for each different sample size and effect level, until the desired power level is reached.

R 3.6.1 (R Core Team, 2019) was used to run the analyses, via R Studio (RStudio Team, 2019), using packages lme4 (Bates et al., 2015), car (Fox & Weisberg, 2019), plyr (Wickham, 2011), and forcats (Wickham, 2019). The results suggested that, for an alpha of 0.05, the sample size should be between 25 (power approximately 0.78) and 35 (power 0.92). To ensure the study reached the recommended power of at least 80%/0.8 (Fritz et al., 2015) the required sample size for this study was set at 30.

Recruitment and participation procedure

As in study four, participants for this study were 'crowdsourced'; that is, a number of self-contained work requests were made available to a remote population (Benoit et al., 2016), who guarantee to carry them out to an agreed standard. Each request is a self-contained unit which the requester either passes or fails, and as they are isolated from each other, many requests can be carried out at the same time without having an effect on the results. On Amazon Mechanical Turk ('MTurk') these work requests are known as Human Intelligence Tasks ('HITs').

In this study the HITs were written and formatted identically to the HITs from paper four. This is easy to do on MTurk because past requests are stored and can be duplicated and re-set, making them available to a fresh population. Sets of requests usually go out in batches, larger or smaller depending on how much monitoring or interaction from the requester is required. Five subjects turned out to be the optimal number, because the verification process for longitudinal data (essentially, manually checking the submitted data) can be time-consuming. In addition, because of the complexity of the task, users sometimes made mistakes, for example by thinking it was a simple cross-sectional survey of the type often found on MTurk (Buhrmester et al., 2018).

How the app is installed to an individual device changed slightly between studies, to make it easier for the participants (see figure 66). Paper four used Microsoft's distribution platform, HockeyApp, which required users to create a HockeyApp account. In this study a switch was made to Apple's own distribution platform, TestFlight. This places a bit more of a burden on the developer in terms of validation and polish (e.g. the app must pass various checks and tests of quality before it can be loaded for distribution) but makes actually installing the app faster and more secure, with support from Apple themselves on the installation method and no requirement for a special account. By providing a restricted link on the HIT request, it was possible to limit recruitment to authorised participants without needing to link user email addresses to their anxiety data.

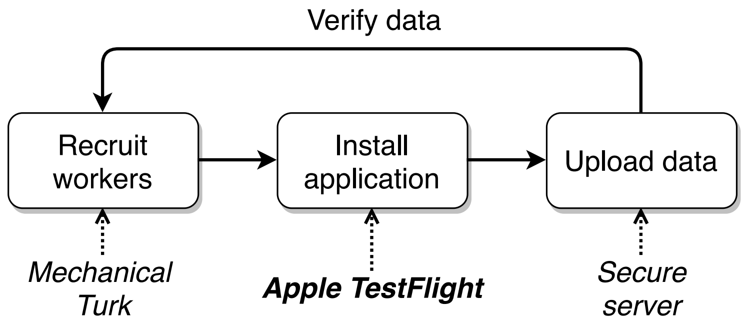

Figure 66: Recruitment procedure; changed distribution platform in bold

In the very early days this was absolutely impossible as quite a lot of collaboration was needed to produce, certify, and install a personalised beta version of the app, for each individual device. Much of this process is now automated on the distribution side, which increases security and privacy. (A fuller description of this process can be found in the previous paper, study four.)

Once the app is installed, participants go through an 'onboarding' process that explains what is required of them, has them choose the times when they do not wish to receive reminders, and walks them through the first survey, the Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) (Spielberger et al., 1983). For the remainder of the participation period their device will notify them automatically when a reading is due. Simply tapping on the notification will take them directly to the reading screen where they are asked for their current anxiety level, measured by the Six-item State-Trait Anxiety Inventory (STAI-6) (Marteau & Bekker, 1992), and their current situation, measured by selecting from a list of pre-determined possible activities, social contexts, and locations. They are asked to do this at least five times per day for five days. Following this regimen for 30 participants will give 750 total observations.

Issues peculiar to longitudinal studies using Mechanical Turk

Studies which require repeated measurements are more complex for the researcher on MTurk than cross-sectional ones. This is because individual workers can pick up and accept the assignment (Amazon's term for individual instances of a task) at any time during the task's available period, so they may start the study at their convenience. Then they have a certain number of days to complete the assignment; usually this is the duration of the study plus a buffer to cover any contingencies, so this interval can also vary. Once they mark the assignment as complete the requester has a certain period in which to check the data, and reject incorrect submissions, before the work is automatically approved.

So if Worker A1 starts the study on the 12th and takes five days to complete it, the requester needs to verify the submission before it is automatically approved on the 20th (the actual approval time can be picked by the requester). Worker A2 starts the study on the same day but takes eight days to complete it, so the requester has to remember to check their data before the 23rd. Amazon does not automatically notify requesters when these time periods are about to expire, and they do not allow requesters to reverse an approval, automatic or otherwise. Worker B did not read the instructions properly and they have just realised that they do not have enough time left to record five days of observations. And Worker C is not going to submit any data at all. This is not necessarily malicious, but does waste slots that more reliable participants could have filled.

Based on these two crowdsourcing studies, papers four and five, around 50% of workers are like Workers A1 and A2; 30% are like Worker B; and the remainder, Worker C. All of the issues can be attended to by keeping track of worker/task status, and making sure to reject submissions which do not pass your manual verification before the automatic approval deadline. In these studies a simple spreadsheet keeps track of the status of each task, updated daily.

Keeping track also allows workers who have made a mistake to complete the task past the automatic expiration time limit. Worker type B is distinguished here from type C by a readiness to complete the study, given extra time and support. This can be established through correspondence, and once their data has been received the rejection can be reversed and approval (and payment) can go through. It is fairly simple to manually change a rejected status to approved.

So longitudinal researchers should be aware that there is a hidden overhead for tasks that require participation over time. For a small team this can slow down recruitment, and this should be borne in mind when planning budgets and timetables.

Participant characteristics

The participation pool for MTurk tends to skew towards North Americans (Buhrmester et al., 2018), but is nonetheless more diverse than an average study based on undergraduate college students. Participants tend to be younger, better-educated, and more liberal than the general population (Chandler & Shapiro, 2016), although again, this is compared to North American residents.

The population of responders in study five appears to be similar but not identical to the respondents in study four. The participants were asked to fill in a very short form with basic demographic information. This study does not use them as predictors because the subjects were promised that these data would only be included in aggregate, anonymised form, and with no potentially identifying information. Answering the demographic questions was non-compulsory, although only one participant opted out (29 opted in).

The survey asked for age in years, general location (using the seven continents model), and gender, for which the options were male, female, or non-binary. Ages ranged from 22 to 61 years old, with an average age of 39 and a standard deviation of 10 years. This meant that participant age skewed slightly older than the previous study, in which the age range was 20 to 45 years old. 70.5% reported living in North America, 13.6% in Asia, 9.1% in South America, 4.5% in Europe, and 2.3% in Africa. The gender of respondents was split almost evenly between male (47.7%) and female (52.3%); the previous study reported similar proportions, 52% male and 43% female, although 5% reported 'no answer'.

Measures

The procedure described above is designed to collect primary, longitudinal data, with no time lag between events and self-reported observation. As in paper four, this study tried to operationalise the concept of situation by framing each observation as an event which includes context and time. The effect on anxiety is derived by measuring the reaction produced by the individual. So events are dynamic and changing, usually external; where are you, what are you doing, and so on. To represent internal influences, a fairly catholic conceptualisation of temperament as a regulator is intended to model individual differences in broad strokes.

Temperament from this perspective is a reasonably stable thing, a collection of components that tend to be inherited rather than acquired (Zuckerman, 2012). In this sense a temperament is its own element, separate from personality traits. So it may change slowly over time (Kagan, 2010, p. 11), but from a short- to medium-term perspective it does not vary very much. In this model, seen in figure 67, it is posited that the relationship between the effect of stressors in events and environment, and an individual's reaction in the moment, is one that is modulated by individual temperament. We expect to see differences in reactivity, or regulation (Allan et al., 2015), diminishing and strengthening the effect of situational factors on the state of anxiety.

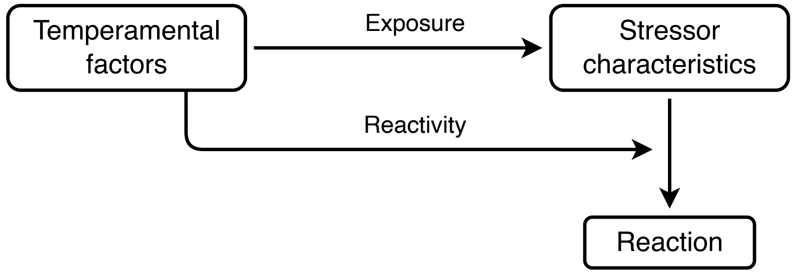

Figure 67: Stress processes. Adapted from Almeida (2005)

To be specific, the variables used to model events are exogenous: location, chronology, activity, and so on. Research suggests that differences in momentary states are explained at the person level by an approximately even split between the stable and the dynamic. For example, Hudson et al. (2017) found that about 50% of the variance in well-being could be accounted for by "constant, trait-level dynamics", and the other 50% by changeable processes at the state level.

Therefore, it is also suggested that situational factors have some, perhaps not overriding, but still measurable intrinsic quality about them. That is, more demanding contexts, events, and social situations, will generally be associated with increased stress, and therefore will increase an individual's experience of anxiety. The converse will be true for other stressors, which will reduce momentary anxiety. To take an example, location in these studies is not physical location, which in any case would not be very meaningful, but is rather conceptualised as a few broader constructs which represent domains of experience (Anusic et al., 2017; Fisher, 2000). In this model, 'at home' and 'at work' have a set of associated demands and resources (A. Bakker et al., 2003), and hence stressors. These have different effects on states of anxiety; work-related states do 'come home' (A. Bakker et al., 2005) and vice versa, but it is generally accepted that it is useful to consider these domains separately. Both trait effects and situational effects can account for similar amounts of variability (McAdams, 1995), but situational factors help explain differences in outcome within subjects, as well as between subjects.

To resolve concerns about reliability and specificity, the measures were all chosen for their ease of administration and established validity (Anusic et al., 2017; Lades et al., 2020). The investigation, as in the previous studies, focused on three essential explanatory factors and one dependent measurement.

The independent variables were:

1. Trait tendency to anxiety, as measured by the Spielberger State-Trait Anxiety

Inventory - Trait (STAI-T) (Spielberger et al., 1983), in the form of ten "anxiety-present" and ten "anxiety-absent" items (Marteau & Bekker, 1992), to give a total score of between 20 and 80.

2. Current situation, i.e. activity (15 choices), company (7 choices), and location (3 choices), modelled on proven studies by Kahneman and others (Anusic et al., 2017; Kahneman et al., 2004; Kahneman & Krueger, 2006; Stone et al., 2006). These were chosen for generalisability: in studies using the same measures of situation the results are "generally similar" (Anusic et al., 2017) some 15 years afterward. All possible choices can be seen in the results section.
3. Time, a system-generated timestamp complete with date and, adjusted for time zone.

The response variable measures:

1. Current experience of stress, measured by the Six-item State-Trait Anxiety Inventory (STAI-6) (Marteau & Bekker, 1992) using three "anxiety-present" and three "anxiety-absent" items giving a total score of 6-24.

In addition, proximity of events within time and person suggests that there will be differences between-subjects, and between time periods (Bolger & Laurenceau, 2013, pp. 27–30); that is, outcomes for the same individual will tend to be similar to each other. And the greatest similarities should be found in measurements within one person and close together in time, as in figure 68.

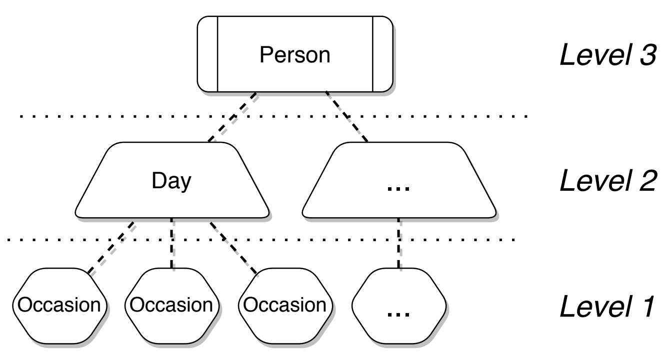

Figure 68: Clustering of events by subject and day

Results

This section begins with an inspection of the response variable, momentary anxiety, followed by brief analyses of various descriptive statistics, such as participation rates, and proportions of different categories within contextual factors (activity, social context, location and day of the week). Finally, a number of mixed-effects models for the stress outcome are fitted and examined, and discussion follows.

Participation rates

Experience sampling studies often have problems with participation and retention (e.g. Bolger & Laurenceau, 2013, p. 21). One aim of this series of studies is to try to address these issues, so it is important to examine whether recruited participants take part as much, less than, or even more than requested, as this will give an indication of how accessible they found it. Table 38 shows summary statistics for this paper's data collection. The requested length and numbers were five measurements per day, for five days.

Table 38: Study five: Participation summary

	Mean	Std. Dev.	Min	Max	Observations
Days participated	6.35	1.20	5	9	854
Daily measurements	5.1	1.34	1	8	”

These are clustered by participant (n=30); the number of measurements per participant is 28.47. As the table shows, the average number of daily measurements is around five. Much of the deviation can be explained by the minimum; from observing the data the participants would often start the study in the afternoon or evening, taking less than five observations that day, and then top up to 25 total by taking one or two extra on a sixth and final day (see 'Days participated' in the same table). This suggests that these numbers of days and measurements are not burdensome.

As further support for the method, the next table (table 39) summarises the participation for the previous study, study four.

Table 39: Study four: Updated participation summary

	Mean	Std. Dev.	Min	Max	Observations
Days participated	6.12	1.05	5	9	589
Daily measurements	4.92	1.55	1	9	”

There was originally a single outlier, who participated for 17 days. Removing this unusual number brought the mean days participated down to just under that of study five (otherwise 7.62). But the number of daily measurements did not change very much (5.1 to 4.9), because this 'super-participant' appears to have been extremely conscientious about following the prescribed schedule.

Clustering and variation

There were 854 observations in the study five dataset, and 30 clusters representing individual participants, giving the average number of measurements as 28.5. This is slightly more than the required total for completing the study. The intraclass correlation (table 40) for measurements within participants is quite sizeable, and the correlation for measurements within the same day for a participant is noticeably larger.

Table 40: Study five: Summary of variance (random effects only)

	Variance	Intraclass correlation
Participant	0.349	0.558
Day (within participants)	0.060	0.654
Occasion (within days and participants)	0.216	

Therefore the data appear to be clustered by participant, and more so by measurements taken on the same day of participation within participants. This structure supports the model's suggestion of a nested formation. This implies correlation within subject, and within day within subject (see also figure 68), which will be taken into account later in the fitted models.

In the earlier study (table 41), the initial analysis was rather similar. There were 683 observations, but in 24 clusters, so the average number of measurements was 28.46, almost identical. This is to be expected, as a minimum was required for successful task completion.

Table 41: Study four: Summary of variance (random effects only)

	Variance	Intraclass correlation
Participant	0.259	0.478
Day (within participants)	0.102	0.666
Occasion (within days and participants)	0.181	

The intraclass correlations in study four also support the idea that measurements within the same subject on the same day tend to be most similar.

Covariates and event frequency

This section presents a descriptive overview of the situational variables. During analysis data were clustered by participant, which means by-subject numbers are most often represented. These groupings are also used in the categorical plots, which were produced using the catplot package in Stata (Cox, 2003). Additional formatting was provided by the grstyle package (Jann, 2018).

Because the majority of readings for the current study were taken during the COVID-19 pandemic, some differences to paper four are expected. This means that two sets of results will be sometimes be presented. They will be clearly titled and any commentary on the previous study will be in a truncated form with references for further reading. For additional readability, many of the tables referred to in this, and following, sections have been placed in an appendix at the end of the paper.

Activity

The variable with the greatest number of possible values was self-reported activity. Answers to all these sections in the app were selected from a scrolling list of preset categories, as explained in the 'Measures' section above (and in more detail in the overall methodology section of this thesis). To begin with, a visual comparison of the two studies is in figure 69, which also shows all possible categories; the current study is represented by the lighter bar, and paper four, the darker bar.

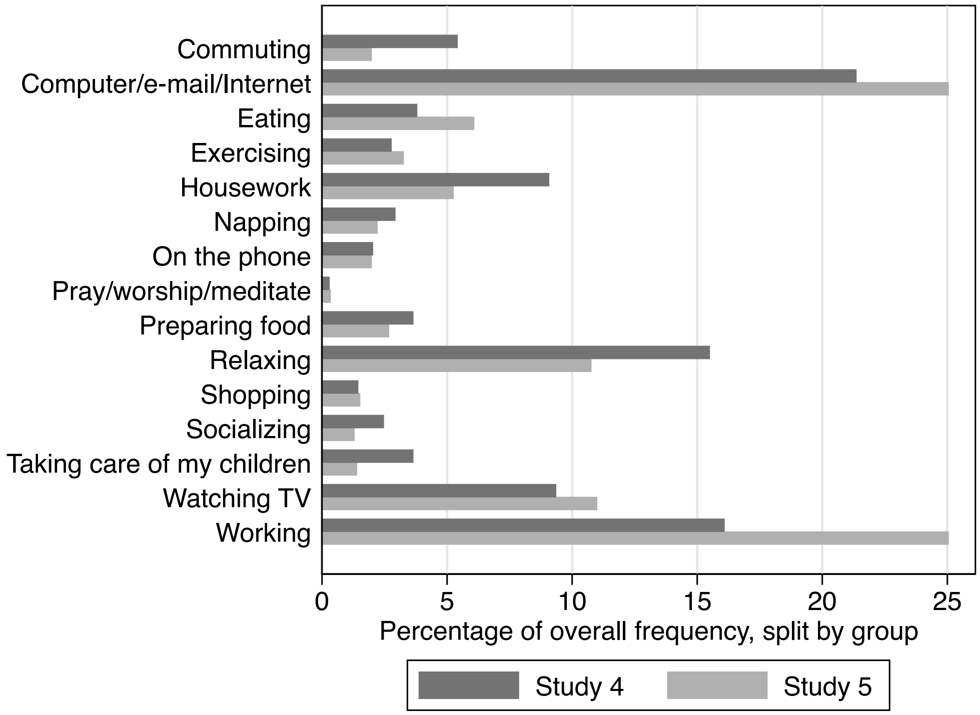

Figure 69: Studies four and five: Categorical plot of activities

As expected, there are clear differences in frequency; with working, relaxing, and eating, for example. For a closer examination, tables of clustered frequencies will be used to get an idea of how things have changed between study periods. In the first table, table 50, in the appendix, figure 69 is expanded upon with exact numbers for self-rated activity frequencies for study five participants. For clarity they are sorted by the total number of times each activity was recorded ('Overall'). The frequency in the 'Between' column denotes the number of subjects who reported this activity at least once (from n=30). And 'Within' enumerates how often each subject from 'Between' recorded the activity in question.

The most frequently reported activities, at the top of table 50, have the largest number of subjects recording them ('Between'). Although no single activity was recorded by every participant, using a computer or the internet, working, watching TV, relaxing, and eating were all reported at least once by at least 75% of the participants, which is useful for comparative purposes.

To give an example, the 'Between' column in table 50 shows that ten participants, or 1/3, reported 'Commuting' at least once. However, from the 'Overall' column it can be seen that just under 2% of all observations in the study, 17 occasions, were recorded as 'Commuting'. Furthermore, for the ten participants who commuted at least once, the ‘Within’ percentage is 6.5%, representing how often they reported ‘Commuting’. So even for those who did commute, the study suggests that they did not engage in it for too long.

These numbers also make it possible to compare differences between individuals’ activity frequencies. For instance, only six people reported ‘Socializing’ at least once; almost half as many compared to the ten who reported ‘Commuting’. Yet the ‘Within’ percentages are 6.5% for ‘Commuting’ and 6.9% for ‘Socializing’, suggesting that although socializing is less common overall (20%), those who did socialize spent just as much time on it as commuters did on travelling to work.

Compare this to the results in table 51. Here commuting, measured in number of subjects reporting at least one occasion (‘Between’) was nearly double by percentile frequency (63%, above; 33%, table 50). The global pandemic is known to have affected ‘physical activity’ behaviour (Kekäläinen, Sipilä, et al., 2021), largely independently from the influence of personality traits. It has been suggested that this kind of decrease in mobility, along with high infection rates, was a primary driver of global increases in mood disorders (Santomauro et al., 2021). Although the total subjects commuting has gone down a substantial amount, the ‘Within’ frequency is actually not too different (8.8 vs. 6.5). This suggests that while the number of participants reporting any commuting at all was greatly reduced, the amount of time spent by those who really had to commute has dropped only a little. The drop in participants reporting socializing (from 30% to 20%) is not so great as commuting, and the time socializers spent on it has dropped even less, just half a percent.

Social context

Second in terms of granularity, social context is another theoretically important determinant of stressful experience (A. Bakker et al., 2005). The visual illustration in figure 70, which shows all possible categories, seems to indicate that, compared to study four, participants in the current study spent more time alone, less time with family and friends, and relatively longer with workmates and associates, by frequency.

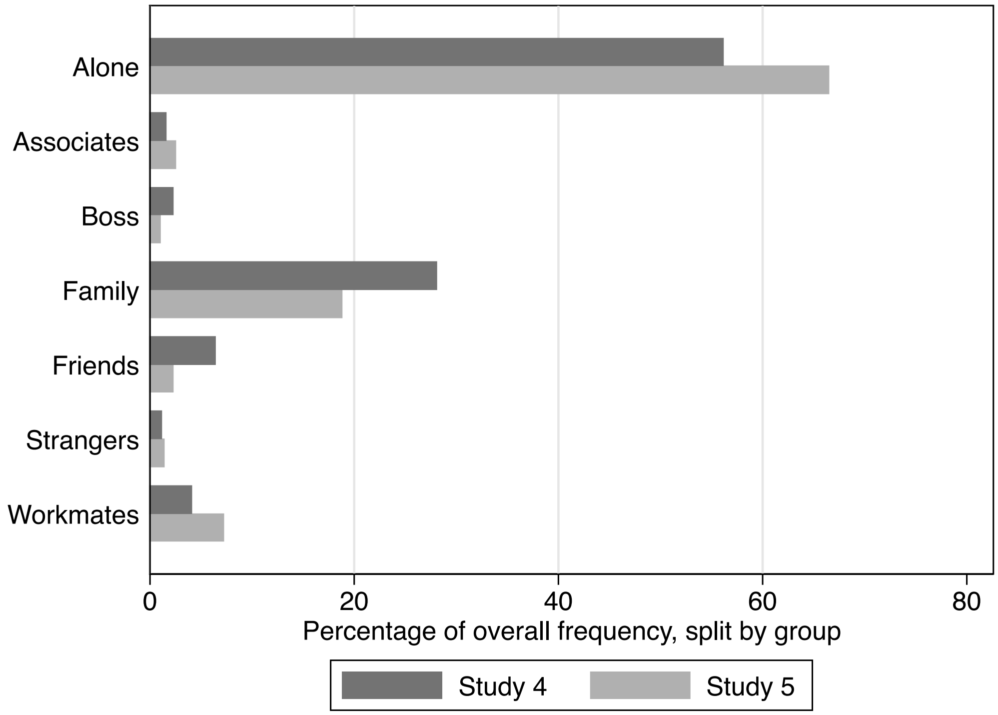

Figure 70: Studies four and five: Categorical plot of social contexts

These differences show up more clearly when decomposed into between and within by subject, in table 52 in the appendix. In table 52, ‘Alone’ and ‘Family’ are the top two overall, at 66.5% and 18.9% respectively. Taking into account that the most frequently reported activity was ‘Working’ or spending time on the computer (tables 52 and 53), around 50% of reports combined, it suggests that the bulk of measurements were taken during working hours, although not necessarily at an office (see below). This is not unexpected, as the participants were allowed to choose their own hours to receive notifications. To partly ameliorate this known problem with non-mandatory sampling techniques, the period of participation had to be a block of at least five hours per day, and the readings were randomised among those hours. This meant that it was easier to participate, but in common with similar studies (D. Bakker et al., 2016; Bolger & Laurenceau, 2013, pp. 21–22) there may be inconsistent recording of rare events.

As an example, participants rated themselves as ‘Alone’ in 66.5% of the measurements overall, and 100% of participants took at least one measurement when they were alone. Six participants reported being with ‘Workmates’ and ‘Strangers’ at least once, but those who spent time with workmates reported doing so for 34.2% of their measurements, about four times as much as the 7.9% for strangers. Both of these percentages are almost double those in the previous study, seen in the appendix, table 53, but not too far off in ratio (four to three times as frequently.)

In the previous study, 23 out of 24 participants reported being 'Alone' at least once during sampling. And comparing 'Workmates' and 'Strangers' is again illustrative. Of the six participants who recorded either of each context at least once, around 5% of their reported occasions were with strangers, while workmates were reported on 16% of occasions, three times as often.

The largest changes in overall frequency, that is, for all participants' reports, were an increase of around ten percent more time alone, and an almost equal reduction in time spent with family (see table 54, in the appendix). Of course, differences in population characteristics cannot be ruled out, but it is generally accepted that travel restrictions and periods of quarantining led to a reduced frequency of social contact (Kekäläinen, Hietavala, et al., 2021). It is true that around 12% more participants reported spending time with family, but those who did (within participants) reported 12% fewer occasions than in study four. Similarly, the overall frequency of time with friends was quite low in study five, just over 2%, with participants who reported spending time with friends dropping from 54% to just 26%. On the other hand, while 14% fewer total participants spent time with their boss, those who did actually recorded a higher frequency of occasions.

Location

Location is the broadest of the self-reported categorical variables, designed to capture the differing demands of the two most common environments (A. Bakker et al., 2003), work and home. Even under this sweeping classification, studies have found divergent relationships between experiences in different locales (Anusic et al., 2017). Here the categorical plot in figure 71 shows the percentage of reported frequencies,in all possible categories; home, work, and elsewhere. Working from home is still work; in fact, working from home during lockdown has been associated with additional negative mental health outcomes (Niederkrotenthaler et al., 2022).

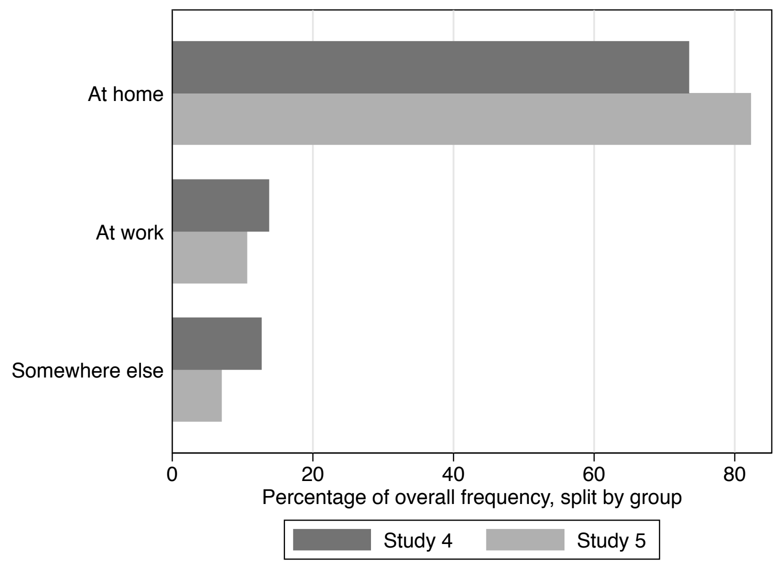

Figure 71: Studies four and five: Categorical plot of locations

Given these constraints, it is not surprising to see that the frequency of 'at home' measurements has gone up compared to study four, while the other two categories have gone down. (The full counts and percentages can be found in the appendix, table 55; for study four, table 56; and a summary of percentage changes in table 57.)

The proportion of participants who reported spending time at work in study five ('Between' in the tables), for instance, was 40% compared to 58% in study four. But while the total number of participants who were observed being at work has gone down, the proportion of time spent at work by those participants ('Within') has stayed almost the same, even increasing slightly (25.8%) versus study four (24.5%). As seen in table 52 earlier, for those who did spend time at work the percentage of observations recorded there actually increased. A possible interpretation is that those who could choose not to go into work did so, but that those who couldn't decline still had to attend about as much as before. Combining the two sets of reported frequencies (activity and location) gives figure 72.

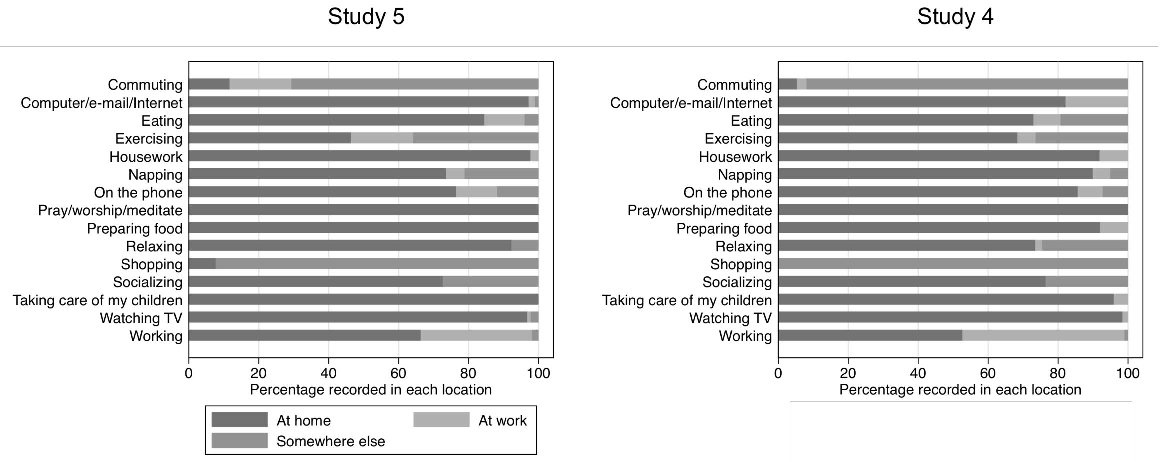

Figure 72: Studies four and five: Reported location of each activity

Almost all activities have increased frequency at the home location.

Weekday

Taking the previous results (see paper four) into account, weekday is not really expected to have a significant main effect. It should, however, be included as a control, as some variation is clustered within days (which are themselves within subjects).

Some studies suggest that it can be a proxy for various factors that produce a “weekly mood cycle” (Stone et al., 2012), or a so-called ‘weekend effect’ (Harvey et al., 2015). However, cultural and demographic differences in approaches to weekdays (for example, some 15% of participants were in Asia or Africa, and there are often differences by age) often make applying this pattern inappropriate.

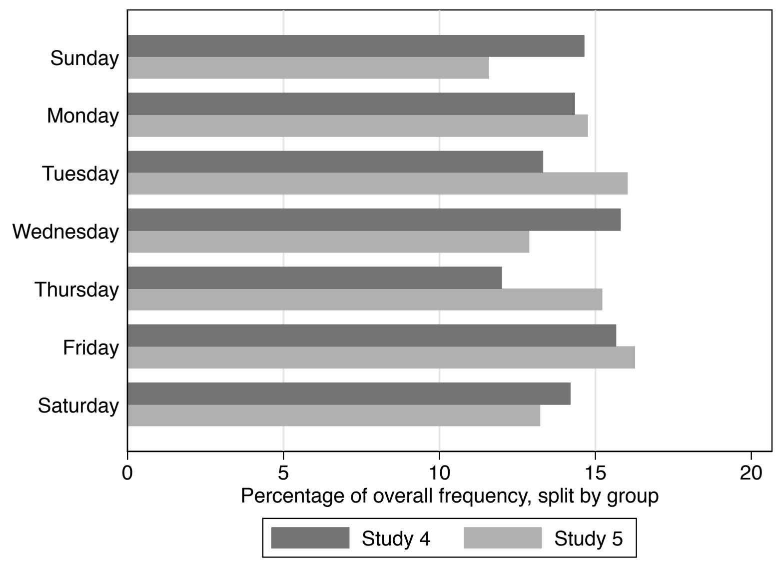

Figure 73: Studies four and five: Categorical plot of weekdays

As shown in figure 73, each day had at least 75% of participants take at least one observation; every day but Sunday had more than 80% (see also table 58 in the appendix). This suggests a fairly representative mix of observations by day, which had been a concern, given that the minimum participation time is mandated at five days, and the actual average was just over six.

Trait/temperament

As mentioned in the measures section, each participant began the study by completing the Spielberger State-Trait Anxiety Inventory - Trait (STAI-T) questionnaire in digital form, using drop-down selection lists within the iOS application. The scale runs from 20 (low tendency to anxiety) to 80 (highest). These scores are not expected to change over the time of the study, and are reliable: traits have been shown to have "longitudinal stability" (McAdams, 1995) whether self-, spouse-, or peer-rated.

While the mean was quite similar in each study (study four: 43.3; study five: 44), the maximum was several points higher (study four: 66; study five: 72). The standard deviation was 13.1 in study four and 16.1 in study five, the current study.

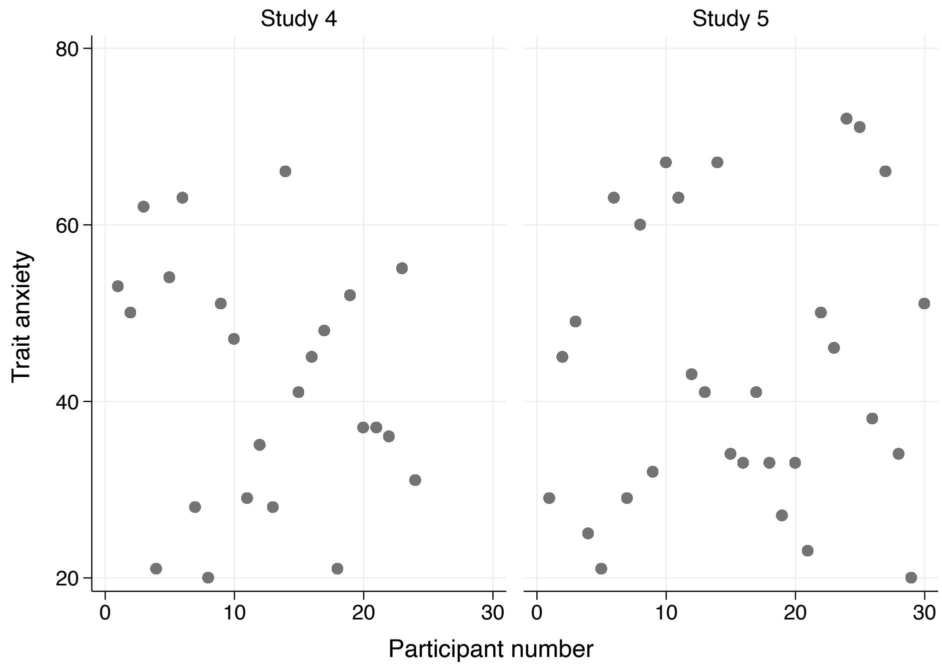

Figure 74: Both studies: Plot of trait anxiety

The scatterplot in figure 74 suggests a slight tendency towards higher STAI-T scores in study five; eight scored about 60 compared to three in study four, for example.

Mixed-effects models

The xtmixed command in Stata 11.2 (StataCorp, 2009c) was used to fit a series of mixed-effects models using maximum likelihood estimation (MLE) with random intercepts at participant and day (within participant) levels.

The first step was to fit the model without fixed effects (explanatory variables), to examine the intra-participant and intra-day correlations of the observations (see figure 68). Next the model is fitted without moderation by temperament. Finally, a moderating effect of temperament is added, by including interaction terms which allow the impact of events to vary between different subjects' by levels of trait anxiety.

Although each model has a large number of parameters, particularly with the addition of a moderator, a number of variables nevertheless proved significant, and various estimates are included. In later sections the esttab package (Jann, 2007) was used to help produce some of the larger tables of estimates.

Structure

Following the procedure of Rabe-Hesketh and Skrondal(2012, p. 452), to verify the necessity of partitioning the variance through random effects, a set of likelihood-ratio tests were carried out (table 42).

Table 42: Study five: Likelihood ratio test results, random effects only

	Test statistic	p
Participant (2-level) vs. null (1-level)	665.22	<0.001
Day and participant (3-level) vs. 2-level	53.99	<0.001

The first test is for the hypothesis that between-subjects (participants) variance is equal to zero, the null hypothesis. This compares a model with a random intercept at participant level to a null model with no random intercepts. Given the test statistic of 665.22, the p value is <0.001, meaning the null hypothesis can be rejected. Testing this two-level model (random intercept at the participant level) against the three-level model (a random intercept for the between-days and within-participants variance) gave a test statistic of 53.99, which also gives a p value of <0.001 against the null hypothesis. This suggests that variance between participants, and by days within participants, should be accounted for in the model structure.

Table 43: Study five: Likelihood ratio test results, fixed effects and moderation

	Test statistic	p
Fixed effects vs. random effects only	98.47	<0.001
Moderation vs. main effects only	45.49	0.02

Next, the days within participants (3-level) model with random effects only was tested against the same 3-level model, but with the fixed effects for situational variables. In this case the test statistic was 98.47, giving a p value of <0.001, suggesting that the null hypothesis should be rejected, and at least some of the situational variables should be retained. And finally, testing a model with interaction terms (moderated) and main effects against the model with main effects only was also significant, with a test statistic of 45.49 giving a p value of 0.02. Therefore the moderation model is preferred.

Replication

The second stage is to compare the situational model, from paper four, with the newer data, from the current study. As a model without a moderating effect of trait has already been fitted for the participants in study four, it will not be duplicated here; it can be seen in full in the appendix in table 61, and in the results section of paper four. This model can be compared with a three-level model with the same variables, and no moderation effect, to be fitted to the study five data.

The results of the fitted model without moderation are in table 60, in the appendix. We know that some stressors are worse than others (Almeida, 2005), and just as in study four, there is good evidence that negative situational factors, such as working or commuting, predict an increase in momentary anxiety. The direct effect of trait is also similar in size to the previous study (here, 0.029; study four, 0.032) and in significance (both $p<0.001$). In study five, compared to a person relaxing at home alone on a Sunday, changing the activity to working is predicted to add 0.292 points in momentary anxiety, on the STAI-6 scale of 1 to 4; conversely, someone working at home on a Sunday but changing the social context to 'with friends' is predicted to be just 0.054 points higher.

Intraclass correlation and variance

In table 44, produced by putting the estimates for the variance components (column two) into the expressions for intraclass correlation (Rabe-Hesketh & Skrondal, 2012, p. 394), it is evident that the intraclass correlation for measurements taken by the same participant on the same day is substantially higher than the correlation for participant alone.

Table 44: Study five: Summary of variance (fixed and random effects, without moderation)

	Variance	Intraclass correlation
Participant	0.121	0.323
Day (within participants)	0.056	0.472
Occasion (within days and participants)	0.198	

For reference, the same components in study four showed similar results. In table 45, the intraclass correlation for measurements taken by the same participant on the same day is also considerably higher than the correlation for participant alone.

Table 45: Study four: Summary of variance (fixed and random effects, without moderation)

	Variance	Intraclass correlation
Participant	0.090	0.270
Day (within participants)	0.102	0.577
Occasion (within days and participants)	0.141	

Adding a moderation term

Much research has endorsed the idea that there is a path in terms of reactivity (Almeida, 2005) between temperament and behaviour (Strelau, 2010, pp. 123–130), particularly in experiences of stress. Moderation occurs when the effect of one variable varies by the level of another (Preacher et al., 2016). In the theoretical model used in this study (seen in figure 75), it is proposed as an amplifier or dampener on the impact of events. This section explores a model which adds this effect to the model above.

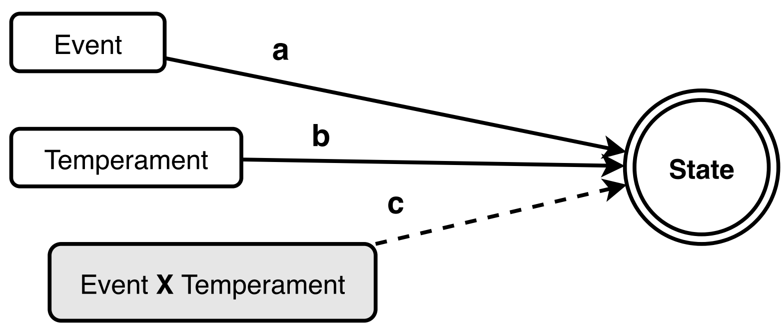

Figure 75: Suggested path diagram

A temperamental factor can be considered as a sort of dispositional baseline (Kraemer et al., 2002; McAdams, 1995), which interacts with an outcome of interest. Commonly, this type of factor is taken as fixed, or pre-measured (Fairclough, 2010, p. 105). In this study, trait anxiety was measured by the STAI-T (Spielberger & Reheiser, 2009), administered at the start of participation.

Table 46 presents the set of terms predicted by adding the effect of this variable as a moderator. Whilst coefficients are individually small, the associations implied by them can nevertheless be considerable given the values of the trait anxiety scale (its average was 44; see figure 74 and the related analyses) and the range of 1-4 in summed anxiety on the STAI-6.

Table 46: Study five: Estimates for fixed effects (moderation terms)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.651**	0.310	0.036		0.043	1.259
Trait anxiety	0.028***	0.007	<0.001		0.015	0.042
Activity X Trait anxiety						
Commuting	0.010	0.009	0.250		-0.007	0.028
Computer/e-mail/Internet	0.005	0.004	0.296		-0.004	0.013
Eating	0.005	0.005	0.378		-0.006	0.015
Exercising	0.010	0.006	0.133		-0.003	0.022
Housework	-0.012**	0.005	0.032		-0.022	-0.001
Napping	-0.004	0.008	0.621		-0.019	0.011
On the phone	-0.009	0.010	0.362		-0.028	0.010
Pray/worship/meditate	0.015	0.028	0.582		-0.039	0.070
Preparing food	0.005	0.007	0.514		-0.009	0.019
Relaxing†	-	-	-			
Shopping	0.001	0.010	0.943		-0.020	0.021
Socializing	0.033**	0.013	0.011		0.007	0.058
Taking care of my children	0.018	0.012	0.133		-0.005	0.041
Watching TV	0.001	0.005	0.840		-0.008	0.010
Working	0.002	0.005	0.617		-0.007	0.011
Social context X Trait anxiety						
Alone†	-	-	-			
Associates	-0.006	0.008	0.422		-0.021	0.009
Boss	-0.069***	0.022	0.001		-0.112	-0.027
Family	0.010**	0.004	0.016		0.002	0.018
Friends	-0.004	0.010	0.690		-0.023	0.015
Strangers	0.010	0.010	0.319		-0.009	0.028
Workmates	-0.008	0.008	0.340		-0.024	0.008
Location X Trait anxiety						
At home†	-	-	-			
At work	0.001	0.006	0.899		-0.011	0.012
Somewhere else	-0.006	0.007	0.441		-0.020	0.009
Weekday X Trait anxiety						
Sunday†	-	-	-			
Monday	-0.003	0.006	0.550		-0.015	0.008
Tuesday	0.002	0.006	0.672		-0.009	0.014
Wednesday	-0.002	0.006	0.699		-0.014	0.009
Thursday	0.000	0.006	0.940		-0.011	0.012
Friday	-0.002	0.006	0.792		-0.013	0.010
Saturday	-0.007	0.006	0.252		-0.018	0.005

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Table 47: Study five: Estimates for fixed effects (main effects)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Activity						
Commuting	-0.176	0.414	0.670		-0.988	0.635
Computer/e-mail/Internet	-0.021	0.190	0.911		-0.393	0.351
Eating	-0.136	0.258	0.599		-0.643	0.371
Exercising	-0.152	0.325	0.639		-0.789	0.485
Housework	0.575**	0.233	0.013		0.119	1.031
Napping	-0.016	0.393	0.967		-0.786	0.753
On the phone	0.521	0.462	0.260		-0.385	1.426
Pray/worship/meditate	-0.645	1.479	0.663		-3.544	2.254
Preparing food	-0.174	0.330	0.598		-0.820	0.473
Relaxing†	-	-	-			
Shopping	0.380	0.483	0.432		-0.566	1.325
Socializing	-1.314**	0.541	0.015		-2.375	-0.254
Taking care of my children	-0.588	0.486	0.226		-1.541	0.365
Watching TV	-0.095	0.202	0.640		-0.492	0.302
Working	0.166	0.193	0.387		-0.211	0.544
Social context						
Alone†	-	-	-			
Associates	0.374	0.353	0.288		-0.317	1.065
Boss	3.754***	1.152	0.001		1.496	6.012
Family	-0.366**	0.173	0.035		-0.706	-0.026
Friends	-0.052	0.436	0.905		-0.906	0.802
Strangers	-0.563	0.520	0.279		-1.581	0.456
Workmates	0.387	0.341	0.256		-0.281	1.055
Location						
At home†	-	-	-			
At work	-0.030	0.257	0.906		-0.534	0.473
Somewhere else	0.148	0.321	0.643		-0.480	0.777
Weekday						
Sunday†	-	-	-			
Monday	0.151	0.270	0.576		-0.378	0.679
Tuesday	-0.140	0.275	0.610		-0.678	0.398
Wednesday	0.056	0.285	0.844		-0.503	0.615
Thursday	-0.087	0.276	0.754		-0.627	0.454
Friday	0.065	0.269	0.809		-0.462	0.592
Saturday	0.143	0.270	0.596		-0.386	0.671

† *indicates reference category*
** $p \leq 0.1$*
*** $p \leq 0.05$*
**** $p \leq 0.01$*

To illustrate the moderation effects in this model, consider for example the activity of working, compared to the reference activity, relaxing. The estimated main effect of working is 0.166 (table 47), and the coefficient of the interaction between working and trait anxiety is 0.002 (table 46). Controlling for the other contextual variables (social context, location and weekday), the expected difference in momentary anxiety is then $0.166 + 0.002 * (traitanxietyoftheindividual)$. At the lowest (20), medium (50), and highest (80) levels of trait anxiety, this difference is 0.206, 0.266, and 0.326 respectively, so it is larger for respondents with higher levels of trait anxiety (note, however, that this interaction is not statistically significant).

Table 48 presents such comparisons in expected momentary anxiety between different context categories against their baseline categories, again evaluated at trait anxiety levels of 20, 50 and 80.

Table 48: Moderation estimates (main effect and interaction) for low (20), medium (50), and high (80) trait anxiety

	Low trait	Medium trait	High trait	*p*
Activity X Trait anxiety				
Commuting	0.028	0.335	0.642	0.250
Computer/e-mail/Internet	0.071	0.208	0.346	0.296
Eating	-0.041	0.101	0.242	0.378
Exercising	0.042	0.332	0.623	0.133
Housework	0.342	-0.006	-0.355	0.032**
Napping	-0.092	-0.205	-0.318	0.621
On the phone	0.343	0.077	-0.189	0.362
Pray/worship/meditate	-0.339	0.119	0.577	0.582
Preparing food	-0.080	0.061	0.202	0.514
Relaxing†	-	-	-	-
Shopping	0.394	0.416	0.439	0.943
Socializing	-0.661	0.318	1.298	0.011**
Taking care of my children	-0.230	0.308	0.846	0.133
Watching TV	-0.076	-0.047	-0.019	0.840
Working	0.212	0.279	0.347	0.617
Social context X Trait anxiety				
Alone†	-	-	-	-
Associates	0.252	0.070	-0.113	0.422
Boss	2.367	0.288	-1.791	0.001***
Family	-0.168	0.128	0.425	0.016**
Friends	-0.130	-0.247	-0.364	0.690
Strangers	-0.372	-0.087	0.198	0.319
Workmates	0.228	-0.010	-0.249	0.340
Location X Trait anxiety				
At home†	-	-	-	-
At work	-0.015	0.007	0.030	0.899
Somewhere else	0.037	-0.130	-0.297	0.441
Weekday X Trait anxiety				
Sunday†	-	-	-	-
Monday	0.082	-0.021	-0.124	0.550
Tuesday	-0.091	-0.018	0.055	0.672
Wednesday	0.010	-0.060	-0.130	0.699
Thursday	-0.078	-0.064	-0.051	0.940
Friday	0.035	-0.010	-0.056	0.792
Saturday	0.012	-0.184	-0.380	0.252

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Examining the four significant coefficients is enough to show how the moderator changes the slope of effects; for example, the predicted effects for socializing (compared to relaxing). The moderation effect predicts an increase of 0.033 per point of trait anxiety, and with the lowest possible score on the unadjusted STAI-T at 20 points, a minimum of 0.66 ($20 * 0.033$) points of additional momentary anxiety on the STAI-6 scale of 1-4. The combination of the two effects for a low trait subject is therefore predicted to be -0.654 in total. On the other hand, a subject with the maximum trait anxiety, 80 points, is predicted 2.64 ($80 * 0.033$) additional points. In this case, the addition of the moderation effect would make the total predicted effect 1.326, reversing the direction. So high trait anxiety has notably strengthened the anxious effect of the activity. This is a nearly two point difference from the low trait representative effect; the slope is much steeper for those with a high trait anxiety.

Similarly the coefficient for the main effect of being in the company of family suggests a reduction in anxiety, -0.366, but including the moderating effect of trait even at the medium level, 0.495 (roughly $50 * 0.01$), overturns this prediction and suggests the total effect is an increase of 0.129, compared to being alone.

Nonetheless, the model does not just predict strengthening effects. The predicted main effect for housework, 0.575, is reduced by higher trait scores, -0.012 points per one point increase in trait. This suggests that while subjects with low trait anxiety have a combined increase of 0.342 from the effect of housework alone, the direction is reversed again, this time to a predicted combination effect of -0.355 for those with high trait anxiety. Whilst this is not as large a dampening effect as the strengthening effect for socializing, it is still significant.

Spending time with one's boss has only two subjects recording readings ('Between' frequency, appendix table 52), perhaps due to pandemic restrictions. Unusually, the person with the higher trait documents lower relative values of momentary anxiety in this situation. These readings are probably related to the unique circumstances of the higher trait participant. They seem to mostly attend gym sessions with their boss, possibly at the company gym, as their most common jointly-located situation is 'exercising' 'at work'. So although the 'essential' main effect of being with one's boss is a very high positive anxiety, the interaction effect dampens this sharply compared to 'being alone'. (The other subject has only work-related interactions.)

Although not many are individually significant, the majority of activity interaction effects are positive, implying high trait anxiety is associated with higher momentary anxiety in many situations, as seen in the rightmost column in table 48. Overall, these results suggest that high trait anxiety does regulate different situations differently;

in some situations high trait is a dampener, and some an amplifier. It appears to emphasise some things, while reducing the impact of others. It is difficult to generalise, because anxiety triggers are complex and individual (Corr & Fajkowska, 2011). Some variable factors like the number of people present (rather than social category), excessive auditory or visual stimuli such as an open environment (Almeida, 2005; Krohne & Hock, 2011; Naragon-Gainey & Watson, 2011), and others, may have changeable influences across person even when broadly controlling for situation.

Table 49: Study five: Summary of variance (fixed and random effects, with moderation)

	Variance	Intraclass correlation
Participant	0.131	0.351
Day (within participants)	0.058	0.505
Occasion (within days and participants)	0.185	

Table 49 shows that the intraclass correlation for measurements taken by the same participant on the same day (with a moderation effect), in this model, remains notably higher than the correlation for participant alone.

Other estimates

Analyses not discussed here in detail are included in the appendix. Fitting the moderation model to the study four data is found in table 62 and table 63, with the variances and intraclass correlations in table 64.

Additionally, given the broad similarities (see above), the model was fitted to both sets of data combined, and these results are also available in the appendix. Combined moderation terms are in table 65, the main effects are in table 66, and variance and ICC are in table 67.

Significances, directions, and sizes are comparable; to stay with the same examples, socializing is predicted to reduce anxiety but is moderated positively by temperament, and housework vice versa. The effect of being with the boss is similar, proportionally, but the effect sizes are halved. Also interesting is the main effect of trait anxiety, which is still around 0.03 per point on the STAI-T (in fact it is 0.03 exactly, with a standard error of 0.005, p value <0.001).

Lastly, using the predicted effects, a number of fitted plots were produced. Figure 76 shows an example for four randomly-selected participants.

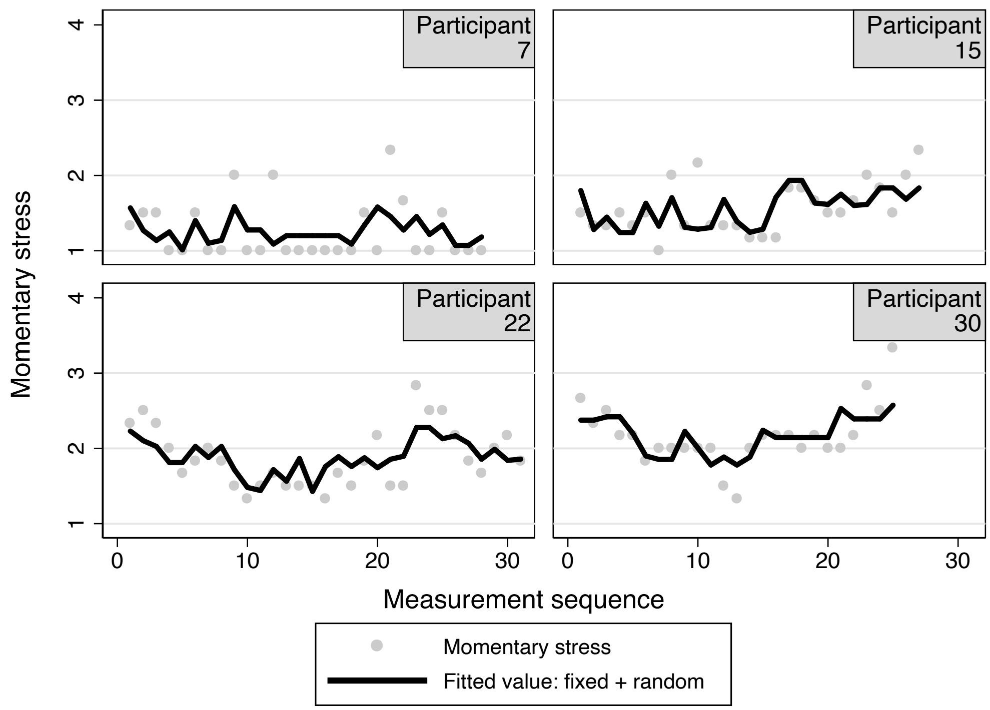

Figure 76: Study five: Plots of actual momentary stress readings with fitted lines of predicted values

The closeness of fit appears reasonable, as suggested in the statistical analyses preceding.

Discussion

The main purpose of this study was to examine the effect of trait anxiety as a moderator on state, or momentary, anxiety. Much research on anxiety has focussed on between-person trait-level structures (Rudaizky & Macleod, 2013), but newer methods of momentary data collection give an opportunity to examine within-person variations at the same time (Wilt et al., 2011). Consistent with previous findings (e.g. Strelau & Zawadzki, 2011), there was an association between subject levels of trait anxiety and within-subject anxious reaction to events. This captures essential attributes of the participants' responsiveness to alterations in stimulation (Rothbart & Posner, 2006).

Reactions were observed, with some caveats, to be stronger in more negative situations, as expected (Tull et al., 2007). Reactivity and self-regulation are both captured in the STAI-T trait measure (Rudaizky & Macleod, 2013), and this was found to moderate responses. However, as Fink (2016) notes, there are three essential factors when determining the magnitude of stress produced in an individual: heightened arousal, perceived aversion to the situation, and a feeling or lack of control. In our model, arousal is moderated by reactivity, while individual feelings of aversion or control are approximated by the situational variables. In this case, it is not surprising that the model paints with rather a broad brush.

While the time use comparisons are moderately informative given the worldwide events between this and the previous study, it may have caused some of the additional sampling variation that has reduced significance. In additional, these studies have explored little when it comes to socioeconomic factors, support networks, and other life-level events (Almeida, 2005), but must acknowledge that they are also important.

Finally, because anxiety is multi-dimensional (Corr & Fajkowska, 2011), there may be “lower order traits” (Naragon-Gainey & Watson, 2011) or aspects such as anxiety reactivity and anxiety perseveration that are not well-distinguished by the STAI-T (Rudaizky & Macleod, 2013). Therefore it is hard to go beyond a general reading of trajectories. Nevertheless the removal of the memory retrieval step in processing and encoding of anxious events (Krohne & Hock, 2011) again proves illuminative (Lavy & Littman-Ovadia, 2011), and so this study has contributed a little towards the study of trait in state.

Appendix: Tables

Table 50: Study five: Summary of activities, sorted by overall frequency

Activity	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Computer/e-mail/Internet	214	25.06	25	83.33	29.38
Working	214	25.06	23	76.67	31.86
Watching TV	94	11.01	25	83.33	13.37
Relaxing	92	10.77	25	83.33	13.25
Eating	52	6.09	23	76.67	7.79
Housework	45	5.27	20	66.67	8.23
Exercising	28	3.28	11	36.67	9.33
Preparing food	23	2.69	14	46.67	5.69
Napping	19	2.22	10	33.33	7.3
Commuting	17	1.99	10	33.33	6.48
On the phone	17	1.99	11	36.67	5.71
Shopping	13	1.52	9	30	5.3
Taking care of my children	12	1.41	5	16.67	8.2
Socializing	11	1.29	6	20	6.85
Pray/worship/meditate	3	0.35	2	6.67	5.32
Total	854	100	219		

Table 51: Study four: Summary of activities, sorted by overall frequency

Activity	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Computer/e-mail/Internet	146	21.38	21	87.5	24.79
Working	110	16.11	18	75	23.18
Relaxing	106	15.52	17	70.83	21.21
Watching TV	64	9.37	16	66.67	12.82
Housework	62	9.08	15	62.5	12.53
Commuting	37	5.42	15	62.5	8.77
Eating	26	3.81	15	62.5	7.07
Preparing food	25	3.66	12	50	6.67
Taking care of my children	25	3.66	9	37.5	8.69
Napping	20	2.93	10	41.67	8.15
Exercising	19	2.78	9	37.5	9.34
Socializing	17	2.49	7	29.17	7.47
On the phone	14	2.05	8	33.33	5.58

Activity	Overall		Between		Within
Shopping	10	1.46	7	29.17	5.44
Pray/worship/meditate	2	0.29	2	8.33	6.17
Total	683	100	181		

Table 52: Study five: Summary of social context, sorted by overall frequency

Company	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Alone	568	66.51	30	100	66.19
Family	161	18.85	26	86.67	21.85
Workmates	62	7.26	6	20	34.16
Associates	22	2.58	9	30	9.21
Friends	20	2.34	8	26.67	9.38
Strangers	12	1.41	6	20	7.9
Boss	9	1.05	2	6.67	18
Total	854	100	87		

Table 53: Study four: Summary of social context, sorted by overall frequency

Company	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Alone	384	56.22	23	95.83	60.15
Family	192	28.11	18	75	33.88
Friends	44	6.44	13	54.17	12.74
Workmates	28	4.1	6	25	16.05
Boss	16	2.34	5	20.83	12.18
Associates	11	1.61	6	25	9.09
Strangers	8	1.17	6	25	4.91
Total	683	100	77		

Table 54: Studies four and five: Percentages and changes between studies, all social contexts

Company	Overall	*Change*	Between	*Change*	Within	*Change*
Alone						
- *Study five*	66.5%	10.3%	100%	4.2%	66.2%	6%
- *Study four*	56.2%		95.8%		60.2%	
Family						
- *Study five*	18.9%	-9.3%	86.7%	11.7%	21.9%	-12%
- *Study four*	28.1%		75%		33.9%	
Workmates						
- *Study five*	7.3%	3.2%	20%	-5%	34.2%	18.1%
- *Study four*	4.1%		25%		16.1%	
Associates						
- *Study five*	2.6%	1%	30%	5%	9.2%	0.1%
- *Study four*	1.6%		25%		9.1%	
Friends						
- *Study five*	2.3%	-4.1%	26.7%	-27.5%	9.4%	-3.4%
- *Study four*	6.4%		54.2%		12.7%	
Strangers						
- *Study five*	1.4%	0.2%	20%	-5%	7.9%	3%
- *Study four*	1.2%		25%		4.9%	
Boss						
- *Study five*	1.1%	-1.3%	6.7%	-14.2%	18%	5.8%
- *Study four*	2.3%		20.8%		12.2%	

Table 55: Study five: Summary of locations, sorted by overall frequency

Location	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
At home	703	82.32	30	100	82.26
At work	91	10.66	12	40	25.8
Somewhere else	60	7.03	18	60	12.36
Total	854	100	60		

Table 56: Study four: Summary of locations, sorted by overall frequency

Location	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
At home	502	73.5	24	100	72.24
At work	94	13.76	14	58.33	24.48
Somewhere else	87	12.74	18	75	17.98
Total	683	100	56		

Table 57: Studies four and five: Percentages and changes between studies, all locations

Location	Overall	*Change*	Between	*Change*	Within	*Change*
At home						
- *Study five*	82.3%	8.8%	100%	0%	82.3%	10%
- *Study four*	73.5%		100%		72.2%	
At work						
- *Study five*	10.7%	-3.1%	40%	-18.3%	25.8%	1.3%
- *Study four*	13.8%		58.3%		24.5%	
Somewhere else						
- *Study five*	7%	-5.7%	60%	-15%	12.4%	-5.6%
- *Study four*	12.7%		75%		18%	

Table 58: Study five: Summary of weekdays, sorted by overall frequency

Weekday	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Friday	139	16.28	28	93.33	17.19
Tuesday	137	16.04	28	93.33	17.4
Thursday	130	15.22	26	86.67	17.3
Monday	126	14.75	27	90	16.36
Saturday	113	13.23	25	83.33	16.19
Wednesday	110	12.88	25	83.33	15.29
Sunday	99	11.59	22	73.33	16.05

Weekday	Overall		Between		Within
Total	854	100	181		

Table 59: Study four: Summary of weekdays, sorted by overall frequency

Weekday	Overall		Between		Within
	Frequency	*Percent*	*Frequency*	*Percent*	*Percent*
Wednesday	108	15.81	23	95.83	16.22
Friday	107	15.67	20	83.33	18.18
Sunday	100	14.64	22	91.67	16.44
Monday	98	14.35	19	79.17	18.07
Saturday	97	14.2	22	91.67	16.04
Tuesday	91	13.32	19	79.17	17.18
Thursday	82	12.01	20	83.33	13.95
Total	683	100	145		

Table 60: Study five: Estimates for fixed effects, no moderation

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.587	0.215	<0.001		0.165	1.008
Trait anxiety	0.029***	0.004	<0.001		0.021	0.038
Activity						
Commuting	0.321**	0.148	0.030		0.032	0.610
Computer/e-mail/Internet	0.204***	0.066	0.002		0.075	0.333
Eating	0.091	0.084	0.279		-0.074	0.257
Exercising	0.271**	0.114	0.017		0.048	0.495
Housework	0.084	0.090	0.351		-0.093	0.261
Napping	-0.212*	0.126	0.094		-0.460	0.036
On the phone	0.086	0.130	0.511		-0.169	0.340
Pray/worship/meditate	0.061	0.294	0.834		-0.515	0.638
Preparing food	0.023	0.113	0.836		-0.199	0.246
Relaxing†	-	-	-			
Shopping	0.376**	0.167	0.025		0.048	0.703
Socializing	0.089	0.174	0.609		-0.252	0.429
Taking care of my children	0.093	0.155	0.547		-0.210	0.396
Watching TV	-0.061	0.073	0.410		-0.205	0.084
Working	0.292***	0.072	<0.001		0.151	0.434
Social context						
Alone†	-	-	-			
Associates	0.029	0.126	0.820		-0.218	0.275
Boss	0.113	0.190	0.554		-0.260	0.485
Family	0.069	0.054	0.197		-0.036	0.174
Friends	-0.238*	0.134	0.074		-0.500	0.024
Strangers	0.024	0.160	0.883		-0.290	0.337
Workmates	0.051	0.103	0.619		-0.151	0.253
Location						
At home†	-	-	-			
At work	0.061	0.091	0.501		-0.117	0.240
Somewhere else	-0.126	0.096	0.189		-0.314	0.062
Weekday						
Sunday†	-	-	-			
Monday	-0.007	0.094	0.944		-0.190	0.177
Tuesday	-0.061	0.095	0.523		-0.247	0.126
Wednesday	-0.046	0.099	0.639		-0.240	0.147
Thursday	-0.074	0.097	0.447		-0.263	0.116
Friday	-0.015	0.094	0.870		-0.199	0.168
Saturday	-0.175*	0.095	0.066		-0.361	0.012

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Table 61: Study four: Estimates for fixed effects, no moderation

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.232	0.242	0.337		-0.242	0.706
Trait anxiety	0.032***	0.005	<0.001		0.022	0.043
Activity						
Commuting	-0.111	0.123	0.367		-0.351	0.130
Computer/e-mail/Internet	-	-	-			
Eating	-0.083	0.094	0.379		-0.268	0.102
Exercising	0.014	0.106	0.894		-0.194	0.222
Housework	0.257***	0.070	<0.001		0.120	0.394
Napping	-0.193*	0.105	0.067		-0.399	0.013
On the phone	0.840***	0.122	<0.001		0.601	1.079
Pray/worship/meditate	-0.171	0.309	0.582		-0.777	0.436
Preparing food	0.105	0.096	0.274		-0.083	0.294
Relaxing†	-0.086	0.065	0.183		-0.213	0.041
Shopping	-0.243	0.172	0.158		-0.580	0.095
Socializing	0.062	0.117	0.597		-0.168	0.292
Taking care of my children	0.434***	0.101	<0.001		0.236	0.632
Watching TV	-0.123*	0.068	0.070		-0.256	0.010
Working	0.268***	0.067	<0.001		0.137	0.399
Social context						
Alone†	-	-	-			
Associates	0.144	0.141	0.308		-0.133	0.421
Boss	0.185	0.129	0.151		-0.067	0.438
Family	0.003	0.049	0.955		-0.093	0.099
Friends	-0.268***	0.097	0.006		-0.459	-0.077
Strangers	-0.135	0.169	0.425		-0.466	0.196
Workmates	0.008	0.113	0.947		-0.214	0.229
Location						
At home†	-	-	-			
At work	0.039	0.074	0.602		-0.106	0.183
Somewhere else	0.382***	0.100	<0.001		0.186	0.578
Weekday						
Sunday†	0.153	0.110	0.164		-0.063	0.370
Monday	0.111	0.112	0.324		-0.109	0.331
Tuesday	0.080	0.112	0.475		-0.139	0.299
Wednesday	-	-	-			
Thursday	0.142	0.113	0.208		-0.079	0.363
Friday	0.153	0.109	0.162		-0.061	0.367
Saturday	0.139	0.109	0.201		-0.074	0.353

† indicates reference category
** $p \leq 0.1$*
*** $p \leq 0.05$*
**** $p \leq 0.01$*

Table 62: Study four: Estimates for fixed effects (moderation terms)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.165	0.368	0.655		-0.557	0.886
Trait anxiety	0.035***	0.008	<0.001		0.019	0.051
Activity X Trait anxiety						
Commuting	0.027***	0.010	0.010		0.006	0.048
Computer/e-mail/Internet	-0.004	0.005	0.442		-0.014	0.006
Eating	-0.004	0.008	0.570		-0.020	0.011
Exercising	0.004	0.009	0.686		-0.014	0.021
Housework	-0.009	0.006	0.114		-0.021	0.002
Napping	-0.012	0.010	0.228		-0.033	0.008
On the phone	0.010	0.014	0.455		-0.017	0.038
Pray/worship/meditate	-0.014	0.024	0.576		-0.061	0.034
Preparing food	0.011	0.007	0.142		-0.004	0.026
Relaxing†	-	-	-			
Shopping	0.042*	0.024	0.089		-0.006	0.089
Socializing	0.005	0.007	0.497		-0.010	0.020
Taking care of my children	0.002	0.010	0.834		-0.017	0.022
Watching TV	-0.005	0.005	0.349		-0.015	0.005
Working	0.002	0.006	0.708		-0.009	0.013
Social context X Trait anxiety						
Alone†	-	-	-			
Associates	-0.003	0.014	0.806		-0.030	0.023
Boss	0.006	0.029	0.843		-0.051	0.063
Family	-0.002	0.004	0.611		-0.009	0.006
Friends	0.009	0.007	0.175		-0.004	0.022
Strangers	-0.026	0.026	0.312		-0.078	0.025
Workmates	0.028**	0.012	0.017		0.005	0.052
Location X Trait anxiety						
At home†	-	-	-			
At work	-0.005	0.007	0.511		-0.018	0.009
Somewhere else	-0.018*	0.009	0.054		-0.036	0.000
Weekday X Trait anxiety						
Sunday†	-	-	-			
Monday	0.004	0.009	0.615		-0.013	0.022
Tuesday	-0.001	0.009	0.893		-0.019	0.016
Wednesday	-0.003	0.009	0.724		-0.020	0.014
Thursday	-0.007	0.009	0.448		-0.025	0.011
Friday	-0.006	0.009	0.539		-0.023	0.012
Saturday	-0.004	0.008	0.670		-0.020	0.013

† *indicates reference category*
* $p \leq 0.1$
** $p \leq 0.05$
*** $p \leq 0.01$

Table 63: Study four: Estimates for fixed effects (main effects)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Activity						
Commuting	-1.151***	0.442	0.009		-2.017	-0.285
Computer/e-mail/Internet	0.254	0.231	0.271		-0.198	0.706
Eating	0.173	0.333	0.604		-0.479	0.825
Exercising	-0.036	0.431	0.933		-0.880	0.809
Housework	0.787***	0.283	0.005		0.233	1.341
Napping	0.442	0.447	0.323		-0.434	1.318
On the phone	0.478	0.620	0.441		-0.737	1.693
Pray/worship/meditate	0.444	0.941	0.637		-1.401	2.288
Preparing food	-0.293	0.348	0.400		-0.976	0.390
Relaxing†	-	-	-			
Shopping	-1.636**	0.814	0.044		-3.232	-0.041
Socializing	-0.099	0.355	0.780		-0.796	0.597
Taking care of my children	0.401	0.523	0.443		-0.623	1.425
Watching TV	0.165	0.238	0.488		-0.301	0.631
Working	0.247	0.260	0.342		-0.263	0.757
Social context						
Alone†	-	-	-			
Associates	0.326	0.482	0.499		-0.619	1.272
Boss	-0.009	1.380	0.995		-2.713	2.695
Family	0.103	0.177	0.561		-0.243	0.449
Friends	-0.667**	0.326	0.041		-1.307	-0.028
Strangers	0.716	0.865	0.408		-0.979	2.412
Workmates	-1.076**	0.474	0.023		-2.005	-0.146
Location						
At home†	-	-	-			
At work	0.220	0.309	0.476		-0.385	0.825
Somewhere else	1.124***	0.393	0.004		0.353	1.895
Weekday						
Sunday†	-	-	-			
Monday	-0.200	0.381	0.599		-0.947	0.547
Tuesday	-0.019	0.396	0.962		-0.795	0.757
Wednesday	0.001	0.394	0.999		-0.771	0.772
Thursday	0.325	0.425	0.445		-0.509	1.159
Friday	0.240	0.400	0.549		-0.544	1.024
Saturday	0.158	0.370	0.670		-0.567	0.883

† *indicates reference category*
** $p \leq 0.1$*
*** $p \leq 0.05$*
**** $p \leq 0.01$*

Table 64: Study four: Summary of variance (fixed and random effects, with moderation)

	Variance	Intraclass correlation
Participant	0.079	0.251
Day (within participants)	0.1	0.57
Occasion (within days and participants)	0.135	

Table 65: Combined data analysis: Estimates for fixed effects (moderation terms)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Intercept	0.517**	0.240	0.031		0.046	0.988
Trait anxiety	0.030***	0.005	<0.001		0.019	0.040
Activity X Trait anxiety						
Commuting	0.017***	0.007	0.008		0.005	0.030
Computer/e-mail/Internet	0.004	0.003	0.232		-0.003	0.011
Eating	0.004	0.004	0.288		-0.004	0.013
Exercising	0.012**	0.005	0.018		0.002	0.022
Housework	-0.007*	0.004	0.066		-0.015	0.000
Napping	-0.005	0.006	0.431		-0.016	0.007
On the phone	-0.007	0.008	0.394		-0.022	0.009
Pray/worship/meditate	0.014	0.015	0.369		-0.016	0.043
Preparing food	0.006	0.005	0.283		-0.005	0.016
Relaxing†	-	-	-			
Shopping	0.010	0.008	0.241		-0.006	0.026
Socializing	0.014**	0.007	0.038		0.001	0.027
Taking care of my children	0.015**	0.007	0.026		0.002	0.029
Watching TV	0.000	0.004	0.951		-0.007	0.007
Working	0.004	0.003	0.242		-0.003	0.011
Social context X Trait anxiety						
Alone†	-	-	-			
Associates	-0.004	0.006	0.528		-0.016	0.008
Boss	-0.034**	0.016	0.029		-0.065	-0.004
Family	0.005*	0.003	0.063		0.000	0.011
Friends	0.001	0.006	0.879		-0.010	0.012
Strangers	0.002	0.007	0.819		-0.012	0.015
Workmates	-0.001	0.007	0.865		-0.014	0.012
Location X Trait anxiety						
At home†	-	-	-			
At work	0.001	0.004	0.889		-0.008	0.009
Somewhere else	-0.009*	0.005	0.087		-0.020	0.001
Weekday X Trait anxiety						
Sunday†	-	-	-			
Monday	-0.001	0.005	0.865		-0.010	0.009
Tuesday	-0.001	0.005	0.887		-0.010	0.009
Wednesday	-0.004	0.005	0.403		-0.014	0.006
Thursday	-0.002	0.005	0.666		-0.012	0.008
Friday	-0.003	0.005	0.488		-0.013	0.006
Saturday	-0.006	0.005	0.229		-0.015	0.004

† indicates reference category
** $p \leq 0.1$*
*** $p \leq 0.05$*
**** $p \leq 0.01$*

Table 66: Combined data analysis: Estimates for fixed effects (main effects)

	Estimate	(*SE*)	*p*	95% CI	(*lower*)	(*upper*)
Activity						
Commuting	-0.572**	0.287	0.046		-1.135	-0.010
Computer/e-mail/Internet	-0.038	0.147	0.796		-0.327	0.251
Eating	-0.116	0.198	0.557		-0.504	0.271
Exercising	-0.345	0.253	0.172		-0.841	0.151
Housework	0.582***	0.182	0.001		0.226	0.939
Napping	0.061	0.280	0.827		-0.488	0.610
On the phone	0.755**	0.362	0.037		0.045	1.465
Pray/worship/meditate	-0.582	0.725	0.423		-2.003	0.840
Preparing food	-0.138	0.237	0.560		-0.603	0.327
Relaxing†	-	-	-			
Shopping	-0.255	0.352	0.469		-0.945	0.435
Socializing	-0.500*	0.302	0.098		-1.091	0.092
Taking care of my children	-0.375	0.335	0.263		-1.032	0.282
Watching TV	-0.042	0.156	0.786		-0.348	0.263
Working	0.135	0.152	0.377		-0.164	0.433
Social context						
Alone†	-	-	-			
Associates	0.222	0.257	0.387		-0.282	0.726
Boss	1.826**	0.773	0.018		0.310	3.342
Family	-0.212*	0.122	0.083		-0.453	0.028
Friends	-0.260	0.261	0.319		-0.770	0.251
Strangers	-0.124	0.339	0.714		-0.789	0.540
Workmates	0.054	0.273	0.842		-0.481	0.590
Location						
At home†	-	-	-			
At work	0.000	0.196	0.999		-0.383	0.384
Somewhere else	0.498**	0.235	0.034		0.036	0.959
Weekday						
Sunday†	-	-	-			
Monday	0.010	0.222	0.964		-0.425	0.445
Tuesday	-0.023	0.228	0.921		-0.470	0.425
Wednesday	0.084	0.233	0.720		-0.373	0.540
Thursday	0.068	0.235	0.772		-0.392	0.528
Friday	0.141	0.225	0.531		-0.301	0.583
Saturday	0.179	0.220	0.415		-0.251	0.610

† *indicates reference category*
** $p \leq 0.1$*
*** $p \leq 0.05$*
**** $p \leq 0.01$*

Table 67: Combined data analysis: Summary of variance
(fixed and random effects, with moderation)

	Variance	Intraclass correlation
Participant	0.119	0.321
Day (within participants)	0.078	0.532
Occasion (within days and participants)	0.174	

Summary and conclusions

Conclusion

Social scientists are often in the vanguard of new technology adoption, whether crowdsourcing data (Benoit et al., 2016) or testing portable heart-rate monitors (Ebner-Priemer & Trull, 2009), to broaden opportunities for inquiry. The growing population of mobile phone users has made device-contingent methodological approaches, which use built-in features to enhance data gathering, much more practical (Bolger & Laurenceau, 2013; Harari et al., 2016; Wheeler & Reis, 1991).

In this thesis I proposed that these advances have created a gap in methods research, especially for intensive observational techniques, which can benefit particularly from automation and streamlining to reduce the burden of participation (Barrett & Barrett, 2001). Addressing these issues will enable richer, more comprehensive measurements to be taken, and help to avoid the classical problem of low completion. The potential and pitfalls exemplify the fundamental conflict in researchers' desire to collect the most detailed data they can, yet maximise inclusion (Christensen et al., 2003; Furr, 2009; Shiffman et al., 2008).

My approach

In my studies I have argued that these new possibilities favour momentary observational methods. These ground measurements in time and in their environment (Reis, 2012), which allows improved modelling of temporal processes (Bolger et al., 2003), while at the same time reducing the hindsight biases inherent to reconstructed data (Shiffman et al., 2008; A. Stone & Shiffman, 2002).

My research brought together proposals about the importance of situation in the sense of environment and activity (R. A. Wilson & Clark, 2009), as well as situating observations in time. A core concern for these studies was this situatedness of experience (Gallagher, 2009), which is difficult to measure with retrospective methods. These often privilege cognitive influences (Bechtel, 2009); I incorporate person and trait, so my model does not ignore these effects, but I do suggest that what is happening around us is just as significant.

In research on stress and anxiety, it is common to separate stressful experience into stressors - what is happening - and reactivity - how you respond (Bolger & Zuckerman, 1995; Kagan & Snidman, 2009; Strelau, 2010). Often this is done by separating out the influence of personality and estimating its effect on reaction. My research goes one step further by uncoupling situational stressors from measurements of stress, and modelling them in their own right, which also helps delineate the impact of personal characteristics (Bolger & Zuckerman, 1995).

Few methods support measuring all these components. One that does is experience sampling, which collects data embedded in individuals' everyday surroundings (Smyth & Stone, 2003), and recorded in real time (Bolger et al., 2003). It does this by taking repeated measurements over time, at chosen or random intervals, often captured close together in time (Bolger & Laurenceau, 2013, p. 5). Traditionally seen as too demanding for widespread adoption, I have worked to address this concern by developing a supporting software framework. This takes away much of the burden of participation.

Each study took a distinct pathway, illustrating a different aspect of the method and its utility. Paper one was a collaborative undertaking with mental health specialists, whereas paper two was a broader study, remotely administered to a moderately-sized convenience sample. In paper three, the focus was on intensive data from a single subject, and paper four was another far-reaching project with a general population of Mechanical Turk users. Paper five built on this. I sum up each study, with a bit more detail, in the section below.

The studies

The individual studies detail the development, implementation, and evaluation of a mobile approach, using a number of different populations. This served two related aims: to test and evolve a situated momentary model of experience, and to demonstrate the feasibility of a technical solution to the methodological problems of experience sampling.

Following the proposals laid out in the introduction, the first paper carried out a real-world assessment, with practical collaboration from two volunteers from a mental health charity. In this paper I described the process of design and the ideas behind it, the testing and deployment procedure, and explored evaluative feedback from prototyping and post-participation discussion sessions. I also made an examination of the data to verify its suitability when compared to the information my theoretical model requires.

Papers two and three dealt with refinements to the method and further tests of the model. In the second study, I tested the software and model in a new context. Instead of the tailored, one-to-one approach of the first paper, I recruited a larger, less-invested population; and rather than provide devices, I installed the software on user equipment using remote management. This was my first step in testing general applicability, and assessing the fitness of the software for protracted, unsupported use (Lahlou, 2010b).

The third paper was perhaps the most ambitious, using one year's worth of measurements of my own stress, heart rate, and sleep to evaluate the theoretical model, and assess the validity of the observations. The resulting considerable dataset provided a number of inferences about theory, and an opportunity for a general verification of my theoretical approach.

The fourth and fifth pieces are inferential papers, which gathered crowdsourced data to explore the theoretical proposals through analyses using hierarchical linear models. I found positive evidence of stability over participants and time, with strong correlations within subjects and days. There were also significant effects for all of the situational variables, and a significant interaction with trait anxiety.

Limitations

The most obvious limitation of this thesis is its reliance on self-reported data. Experience-sampling methods are designed to reduce recall and retrospective biases (Smyth & Stone, 2003), but the data are still weaker than direct observation in many cases (Furr, 2009). However, with experiences of emotion, it is difficult to make an assessment any other way. I chose to address verification of these measurements through subject evaluations (see paper one), comparisons with other studies and theory (found in all the studies), and correlation with physical measures (see paper three). While the studies showed good evidence for the validity of the observations, it would have been advantageous to carry out some other kind of corroboratory study to address this issue in depth. Gros et al. (2010), for example, asked pairs of friends to rate themselves and each other on trait anxiety and compared self-other agreement. I could also have tried comparing experience-sampling data to reconstructed data, such as the day reconstruction method (Kahneman et al., 2004), which tries to control for recall bias.

I exclusively used the State-Trait Anxiety Inventory (Spielberger et al., 1983) for assessments of both trait and state anxiety. This was for two main reasons: firstly, this meant that observations would be comparable between data sets; and secondly, the scale has a good reputation, with good consistency, reliability, and validity (Bados et al., 2010; Tluczek et al., 2009). It worked well from this perspective. But some research (e.g. Bados et al., 2010; Gros et al., 2010) suggests that it has poor divergent validity when it comes to distinguishing symptoms of depression from those of high anxiety. A complementary study could engage this concern head on. In addition, it is a copyrighted instrument so it is not possible to discuss individual questions or show those aspects of the user interface, and each use must be approved and licensed. Adopting something in the public domain would give more flexibility, but I felt that the trade-off was worth making to achieve reliable, consistent measurements.

It could also be said that the broadness of the situational categories was a limitation, as it was necessary to compromise between specificity and generalisability to carry out the research with different populations. To compensate for this I drew my generalised categories from larger studies (e.g. Kahneman et al., 2004; Stone et al., 2006), to increase correspondence with likely situations, and the resulting analyses showed effects of these factors were commensurate with theoretical predictions. Nevertheless, I recommend that studies which target distinct populations or behaviours think about making an independent study of expected and applicable events.

Summary of conclusions and contributions

This section sums up my contributions, and lays out my major conclusions. First I review my general findings, and then in the second half of the discussion I describe some specific outcomes.

Overall conclusions

My original research question was whether the method and model that I was designing was practical. The initial study showed that my software design, which implemented experience-sampling methods and measured a situational model of stress, was viable on a small scale. It also illustrated some of the advantages of an intensive design process, as a contribution to methods research.

This investigation was extended with a further test of the design in paper two, which looked at the robustness of the methodology with naive users and in a remote environment. The success of this study supported the positive findings in the first paper, and satisfied my question about the external potential of my work. On top of this I had found exploratory evidence for the importance of trait and situational influences in both papers, mainly by visual delineation of overall differences in levels.

At this point I had performed an initial evaluation of my theoretical model and proven the feasibility of my design. There was confirmation for my approach in both areas. This enabled me to make some methodological recommendations, most notably that researchers should use an iterative perspective when developing, and should not be intimidated by programming, which is not too different from scripting in statistical packages. These studies laid a good foundation for the next stage, the theoretical analyses.

Method assessment would continue to be an important part of the remaining studies, but the focus now shifted to my psychological questions. The next investigation looked at how the model matched with physical measures, and if my results corresponded with other findings. This was in the interest of situating my research in the wider field of the literature. I was able to further validate foundational research about the influence of resting heart rate and day of the week effects on levels of stress, whilst corroborating my proposals about situational factors, and placing them in the context of the field of experiential research.

The fourth study built on these conclusions and broke ground on a technique for recruiting intensive longitudinal data subjects through Amazon's Mechanical Turk. Using multiple subjects enabled me to assess the influence of trait on momentary levels of stress, and reinforce the importance of momentary situational factors with significant effects in all three categories. The evidence was poor for the 'weekday effect' (Stone et al., 2012) found in the third paper, but using hierarchical linear models allowed me to make a stronger determination of the position of person and time in my model. There was good evidence for similarities within days and subjects, i.e. that levels are likely to be closer together for the same day and person. I also found a strong effect of trait anxiety, the first time I was able to assess this rigorously.

Finally the moderation effect was tested in an additional study.

Specific findings

While papers one and two were exploratory, the inferential studies, papers three, four and five, carried out more formal statistical analyses. The inferential papers found evidence of notable situational effects. Of the three primary factors (what, who, and where), activity and social context were significant in both studies. For example, a subject rating themselves as 'working' was associated with a predicted increase in momentary stress in both data sets, and so was being in the company of work associates or your boss. Location was only analysed in paper four, and found being at home was associated with a decrease in momentary stress compared to being at work or somewhere else. An increase in resting heart rate (paper three only) was associated with a predicted increase in momentary stress, and so was the day being a weekday rather than a Sunday. Weekday was not significant in paper four, and the sleep measures in paper three had small effect sizes and likewise large p-values.

There were variations in frequency observed between the different populations; for example, 'Taking care of my children' was the most frequent activity in paper three, whilst 'Computer/email/internet' was the most common in paper four. Being alone was most common in study four, whereas study three found the company of 'Family' much more common. Overall, these studies present good evidence for my general situational model of stress, which is based on activity, company, and location.

Paper four was the only study to make an estimate of the effect of trait anxiety, as paper three was single-subject. We saw some variation and suggestions of clustering in papers one and two, but these investigations were exploratory in nature. This observation was verified in paper four's analysis, which predicted an effect of 0.03 points (95% confidence interval (CI) from 0.02 to 0.04, $p < 0.001$) of additional momentary stress on the scale of 1 to 4. The mean trait anxiety in the sample population was 43, which predicts 1.29 points (95% CI: 0.86; 1.72) of momentary stress.

he fourth study found positive indications of variation between participants, and between days within participants. These effects were estimated using a linear mixed-effects model (Rabe-Hesketh & Skrondal, 2012, pp. 85–86). The 'random effects' part of the model, which allows the intercept to vary by subject and by day, estimated between-participant standard deviation as 0.299 (95% CI: 0.208 to 0.431). The standard deviation within participants, and between days, is estimated as 0.319 (95% CI: 0.267 to 0.380). This suggests that there is a strong correlation between measurements for the same person on different days, but that the correlation for the same person on the same day is the strongest.

Finally, the fifth study extended the scope of paper four's conclusions. There was an association between subject levels of trait anxiety and within-subject anxious reaction to events, which captured the participants' responsiveness to alterations in stimulation.

Summary

In summary, the findings from these studies contribute to research in two different areas. Firstly, the methodological, where I have developed an additional technique for taking fine-grained measurements in real time, which assesses levels of momentary affect, and records situation and time. In my research I have concentrated on experiences of momentary stress, but this method is flexible enough to support almost any research area, from the highly quantitative to the deeply qualitative.

Secondly, I have investigated the nature of experience, outlining more clearly the links between ecological and psychological factors in processes of experience construction, and investigating temperament as a moderator on two data sets. Again, my findings are on stress and anxiety, but I feel that my theoretical conclusions could be extended to other states such as depression, work satisfaction, or fear of crime.

Future directions

Throughout this thesis, each individual study has proven general feasibility and made suggestions about areas for future study. In this penultimate section I collect and further develop some of these suggestions.

In general, the trajectory I would like to see following my work is a practical one: improved measurements and increased access to experience promoting improved and more timely interventions. These could even be momentary interventions, given in response to real-time measurements (e.g. Ebner-Priemer & Trull, 2009). For example, a software application could assess how well acceptance and commitment therapy (ACT) and cognitive behavioural therapy (CBT) work in treating anxiety disorders like post-traumatic stress disorder (PTSD). These treatments are common (Arch et al., 2012; Losada et al., 2015), but are usually evaluated via retrospective measures. It would be a natural fit with event-contingent observations, perhaps pre-and post treatment as it happens in daily life: taking a measurement, trying the therapy, taking an additional measurement. The software could even prompt the subject with instructions when an intervention is deemed necessary.

Another area of increasing urgency is late-life anxiety (Balsamo et al., 2015), or "the silent geriatric giant" (Cassidy & Rector, 2008). Behavioural therapies are of particular interest to an older population that cannot necessarily tolerate medication. A software approach might present interface challenges, especially when subjects are frail, but may also prove significant in reducing frailty. This could be by better tracking daily experiences of high stress and providing guidance on reducing occurrences, or in supplying professionals with evidence supporting the provision of therapy, which is especially pertinent in today's cost-conscious medical environment.

Anxiety in the workplace is of equal significance (e.g. Beehr, 2014), bearing in mind the numbers referred to earlier in this thesis: 12.5 million working days missed on sick leave due to work-related stress in the most recent statistics (Health and Safety Executive, 2017). Investigating burnout using a lens such as coping (Pottie & Ingram, 2008) or demand-control (Van Yperen & Snijders, 2000) would almost certainly prove significant, as momentary evidence, especially intensively-sampled measurements, is sparse.

On the theoretical side, as mentioned earlier, complementary studies comparing different sources of data, perhaps even recalled data versus momentary, could prove illuminating. At the least, such study would provide material for delineating some of the processes of recall construction. It might also be possible to extend the analysis of traits, and investigate more specifically what characteristics cause some to be more resilient to stressors over time (Cicchetti, 2010).

Another process of change that could benefit from intensive longitudinal examination is the occurrence of co-morbidity in disorders of anxiety and depression. This is known to predict how long sufferers may experience them (Spinhoven et al., 2011), and unpicking this process would be a definite step forward.

Lastly, as noted in previous studies, investigation of a specific research question or research population would almost certainly benefit from more specialised categories for situational variables. There are several routes to exploring these kinds of data, both quantitative and qualitative. Researchers could easily adapt this methodology to carry out an entirely qualitative study, asking for written, descriptive answers or even short recordings. A study could use something as simple as a text message reminding a participant to take a voice recording on the built-in microphone, or design a self-contained instrument as I have done over the course of the thesis.

Final summary

This thesis investigated a number of issues in methods and experience of stress. It validated a mobile approach to gathering momentary and situational data, demonstrating that such technologies can reduce the cost of intensive measurement methods and increase their reach. It also found good evidence for a situated model of stressful experience.

I think that I have made a contribution that is both practical, with implications for health and professionals, and theoretical, with significance for future studies into processes of experience construction and the effects of situation. I hope that the scientific and general population will benefit from my experiences.

References

Adam, E. K., Hawkley, L. C., Kudielka, B. M., & Cacioppo, J. T. (2006). Day-to-day dynamics of experience - cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences*, *103*(45), 17058–17063.

Adams, F., & Aizawa, K. (2009). Why the mind is still in the head. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 78–95). Cambridge University Press New York, NY.

Åkerstedt, T., Orsini, N., Petersen, H., Axelsson, J., Lekander, M., & Kecklund, G. (2012). Predicting sleep quality from stress and prior sleep – a study of day-to-day covariation across six weeks. *Sleep Medicine*, *13*(6), 674–679.

Allan, N. P., Norr, A. M., Macatee, R. J., Gajewska, A., & Schmidt, N. B. (2015). Interactive effects of anxiety sensitivity and emotion regulation on anxiety symptoms. *Journal of Psychopathology and Behavioral Assessment*, *37*(4), 663–672. https://doi.org/10.1007/s10862-015-9484-z

Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*, *14*(2), 64–68. http://www.jstor.org/stable/20182990

Anusic, I., Lucas, R. E., & Donnellan, M. B. (2017). The validity of the day reconstruction method in the German socio-economic panel study. *Social Indicators Research*, *130*(1), 213–232. https://doi.org/10.1007/s11205-015-1172-6

Apple Inc. (2016). *Apple celebrates one billion iPhones*. https://www.apple.com/newsroom/2016/07/apple-celebrates-one-billion-iphones/

Apple Inc. (2018a). *Apple Reports Fourth Quarter Results*. https://www.apple.com/newsroom/2018/11/apple-reports-fourth-quarter-results/

Apple Inc. (2018b). *Xcode Overview: Building a User Interface*. https://developer.apple.com/library/archive/documentation/ToolsLanguages/Conceptual/Xcode_Overview/UsingInterfaceBuilder.html#//apple_ref/doc/uid/TP40010215-CH42-SW1

Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology*, *80*(5), 750.

Arnold, B. F., Hogan, D. R., Colford, J. M., & Hubbard, A. E. (2011). Simulation methods to estimate design power: An overview for applied research. *BMC Medical Research Methodology*, *11*(1), 94. https://doi.org/10.1186/1471-2288-11-94

Augustine, A. A., & Larsen, R. J. (2012). Emotion research. In M. R. Mehl & T. S.

Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 497–510). Guilford Press.

Bados, A., Gómez-Benito, J., & Balaguer, G. (2010). The State-Trait Anxiety Inventory, Trait Version: Does It Really Measure Anxiety? *Journal of Personality Assessment*, *92*(6), 560–567. https://doi.org/10.1080/00223891.2010.513295

Bakker, A., Demerouti, E., & Schaufeli, W. (2003). Dual processes at work in a call centre: An application of the job demands-resources model. *European Journal of Work and Organizational Psychology*, *12*(4), 393–417.

Bakker, A., Demerouti, E., & Schaufeli, W. B. (2005). The crossover of burnout and work engagement among working couples. *Human Relations*, *58*(5), 661–689. https://doi.org/10.1177/0018726705055967

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health*, *3*(1), e7. https://doi.org/10.2196/mental.4984

Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., & Saggino, A. (2015). Measuring anxiety in the elderly: Psychometric properties of the State Trait Inventory of Cognitive and Somatic Anxiety (STICSA) in an elderly Italian sample. *International Psychogeriatrics*, *27*(6), 999–1008.

Barrett, L. F. (2004). Feelings or Words? Understanding the Content in Self-Report Ratings of Experienced Emotion. *Journal of Personality and Social Psychology*, *87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, *19*(2), 175–185.

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

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403.

Bechtel, W. (2009). Explanation: Mechanism, modularity, and situated cognition. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 155–170). Cambridge University Press New York, NY.

Beehr, T. A. (2014). *Psychological stress in the workplace (psychology revivals)*. Routledge.

Benoit, K., Conway, D., Lauderdale, B. E., Laver, M., & Mikhaylov, S. (2016). Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review*, *110*(2), 278–295.

Benoit, K., Conway, D., Laver, M., & Mikhaylov, S. (2014). Crowd-sourced data coding for the social sciences. *Unpublished Manuscript*.

Bolger, N., & Amarel, D. (2007). Effects of social support visibility on adjustment to stress: Experimental evidence. *Journal of Personality and Social Psychology*, *92*(3), 458.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579. https://doi.org/Article

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research* (p. 256). Guilford Press.

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology*, *69*(5), 890.

Brewer, M. B., & Crano, W. D. (2014). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 47–79). Cambridge University Press.

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain*, *139*(1), 146–157. https://doi.org/10.1016/j.pain.2008.03.024

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2), 149–154. https://doi.org/10.1177/1745691617706516

Busemeyer, J. J., Jessup, R. K., & Dimperio, E. (2009). The dynamic interactions between situations and decisions. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 307–321). Cambridge University Press New York, NY.

Cable, J., Henley, A., & Holland, K. (2002). Pot of Gold or Winner's Curse? An Event Study of the Auctions of 3G Mobile Telephone Licences in the UK. *Fiscal Studies*, *23*(4), 447–462. http://www.jstor.org/stable/24438306

Campbell, D. T. (1988). *Methodology and epistemology for social science : Selected papers*. University of Chicago Press.

Carlsson, C., Carlsson, J., Hyvonen, K., Puhakainen, J., & Walden, P. (2006). Adoption of mobile devices/services – searching for answers with the UTAUT. *Proceedings of the 39th annual Hawaii international conference on system sciences (HICSS'06)*, *6*, 132a–132a.

Cassidy, K.-L., & Rector, N. A. (2008). The silent geriatric giant: Anxiety disorders in late life. *Geriatrics and Aging*, *11*(3), 150–156.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, *12*(1), 53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623

Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003).

A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, *4*(1), 53–78.

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, *4*(1), 5–6.

Cicchetti, D. (2010). Resilience under conditions of extreme stress: A multilevel perspective. *World Psychiatry*, *9*(3), 145–154.

Cikszentmihalyi, M. (2012). Foreword. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. Guilford Press.

Conner, T. S. (2015). *Experience sampling and ecological momentary assessment with mobile phones*. https://www.otago.ac.nz/psychology/otago047475.pdf

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*(3), 292–313.

Corr, P. J., & Fajkowska, M. (2011). Introduction to special issue on anxiety. *Personality and Individual Differences*, *50*(7), 885–888. https://doi.org/https://doi.org/10.1016/j.paid.2011.01.031

Costa, G. (1996). *Occupational stress and stress prevention in air traffic control*. International Labour Office.

Cox, N. J. (2003). *CATPLOT: Stata module for plots of frequencies, fractions or percents of categorical data*. Statistical Software Components, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s431501.html

Crawford, K., Lingel, J., & Karppi, T. (2015). Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies*, *18*(4-5), 479–496.

Dahlgren, A., Kecklund, G., Theorell, T., & Åkerstedt, T. (2009). Day-to-day variation in saliva cortisol: Relation with sleep, stress and self-rated health. *Biological Psychology*, *82*(2), 149–155.

Doane, L. D., & Adam, E. K. (2010). Loneliness and cortisol: Momentary, day-to-day, and trait associations. *Psychoneuroendocrinology*, *35*(3), 430–441.

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057–1066.

Eagle, N., Pentland, A., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15274–15278. https://doi.org/10.1073/pnas.0900282106

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment: An innovative and promising approach for clinical psychology. *European Psychologist*, *14*(2), 109–119.

Eid, M., Courvoisier, D. S., & Lischetzke, T. (2012). Structural equation modeling of ambulatory assessment data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 384–406). Guilford Press.

Ellis, D. A. (2020). *Smartphones within psychological science*. Cambridge University Press.

Fahrenberg, J., & Myrtek, M. (1996). *Ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Hogrefe & Huber Publishers.

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment – monitoring behavior in daily life settings. *European Journal of Psychological Assessment*, *23*(4), 206–213. https://doi.org/10.1027/1015-5759.23.4.206

Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials* (Second). Chapman; Hall/CRC. https://doi.org/https://doi.org/10.1201/9781420061185

Fink, G. (2016). Stress, definitions, mechanisms, and effects outlined: Lessons from anxiety. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior* (pp. 3–11). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800951-2.00001-7

Fisher, C. D. (2000). Mood and emotions while working: Missing pieces of job satisfaction? *Journal of Organizational Behavior*, *21*(2), 185–202.

Fitbit Customer Support. (2018). *What should I know about my heart rate data?* https://help.fitbit.com/articles/en_US/Help_article/1565

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83–87.

Fleeson, W. (2007). Using experience sampling and multilevel modeling to study person-situation interactionist approaches to positive psychology. In A. D. Ong & M. H. M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 501–514). Oxford University Press.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808

Fritz, M. S., Cox, M. G., & MacKinnon, D. P. (2015). Increasing statistical power in mediation models without increasing sample size. *Evaluation & the Health Professions*, *38*(3), 343–366. https://doi.org/10.1177/0163278713514250

Funder, D. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*(1), 21–34. https://doi.org/10.1016/j.jrp.2005.08.003

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369–401.

Gallagher, S. (2009). Philosophical antecedents of situated cognition. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 35–51). Cambridge University Press New York, NY.

Gaskell, G. D., Wright, D. B., & O'Muircheartaigh, C. A. (2000). Telescoping of landmark events: Implications for survey research. *The Public Opinion Quarterly*, *64*(1), 77–89.

Genter, R. (2014). Understanding the POW Experience: Stress Research and the Implementation of the 1955 US Armed Forces Code of Conduct. *Journal of the History of the Behavioral Sciences*.

Geyer, K., Ellis, D. A., & Piwek, L. (2019). A simple location-tracking app for psychological research. *Behavior Research Methods*, *51*(6), 2840–2846. https://doi.org/10.3758/s13428-018-1164-y

Goldschmidt, A. B., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Lavender, J. M., Peterson, C. B., Crow, S. J., Cao, L., & Mitchell, J. E. (2014). Ecological momentary assessment of stressful events and negative affect in bulimia nervosa. *Journal of Consulting and Clinical Psychology*, *82*(1), 30.

Grammenos, D., & Barker, S. A. (2014). On the transmethylation hypothesis: Stress, n, n-dimethyltryptamine, and positive symptoms of psychosis. *Journal of Neural Transmission*, 1–7.

Gray, E., Jackson, J., & Farrall, S. (2008). Reassessing the Fear of Crime. *European Journal of Criminology*, *5*(3), 363–380. https://doi.org/10.1177/1477370808090834

Gros, D. F., Simms, L. J., & Antony, M. M. (2010). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads. *Behavior Therapy*, *41*(3), 277–284.

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 144–159). Guilford Press.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 43–61). Guilford Press.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*(6), 838–854.

Harvey, B., Milyavskaya, M., Hope, N., Powers, T. A., Saffran, M., & Koestner, R. (2015). Affect variation across days of the week: Influences of perfectionism and academic motivation. *Motivation and Emotion*, *39*(4), 521–530.

Hasler, B. P., Mehl, M. R., Bootzin, R. R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, *42*(6), 1537–1546.

Health and Safety Executive. (2014). *Stress-related and psychological disorders in Great Britain 2014*. UK Government.

Health and Safety Executive. (2017). *Work-related stress, depression or anxiety statistics in Great Britain 2017*. UK Government.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage.

Hirotsu, C., Tufik, S., & Andersen, M. L. (2015). Interactions between sleep, stress, and metabolism: From physiological to pathological conditions. *Sleep Science*, *8*(3), 143–152.

Hofmann, W., & Patel, P. V. (2015). SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*, *33*(2), 235–253.

Hogarth, R. M., Portell, M., & Cuxart, A. (2007). What risks do people perceive in everyday life? A perspective gained from the experience sampling method (ESM). *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *27*(6), 1427–1439. https://doi.org/10.1111/j.1539-6924.2007.00978.x

Holt, N. J. (2018). Using the experience-sampling method to examine the psychological mechanisms by which participatory art improves wellbeing. *Perspectives in Public Health*, *138*(1), 55–65.

Hudson, N. W., Lucas, R. E., & Donnellan, M. B. (2017). Day-to-day affect is surprisingly stable: A 2-year longitudinal study of well-being. *Social Psychological and Personality Science*, *8*(1), 45–54. https://doi.org/10.1177/1948550616662129

Hunter, R. G., Gagnidze, K., McEwen, B. S., & Pfaff, D. W. (2014). Stress and the dynamic genome: Steroids, epigenetics, and the transposome. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1411260111

Hunter, R. G., & McEwen, B. S. (2013). Stress and anxiety across the lifespan: Structural plasticity and epigenetic regulation. *Epigenomics*, *5*(2), 177–194.

Hurlburt, R. T. (1993). *Sampling inner experience in disturbed affect*. Springer Science & Business Media.

Hurtado-Parrado, C., & López-López, W. (2015). Single-case research methods: History and suitability for a psychological science in need of alternatives. *Integrative*

Psychological and Behavioral Science, 1–27.

Jann, B. (2007). Making regression tables simplified. *The Stata Journal*, *7*(2), 227–244.

Jann, B. (2018). Customizing Stata graphs made easy (part 2). *The Stata Journal*, *18*(4), 786–802.

Jensen, M. (2014). Catalytic models developed through social work. *Integrative Psychological and Behavioral Science*, 1–17.

Jensen, M. T., Suadicani, P., Hein, H. O., & Gyntelberg, F. (2013). Elevated resting heart rate, physical fitness and all-cause mortality: A 16-year follow-up in the Copenhagen Male Study. *Heart*, *99*(12), 882–887.

Jones, F., & Bright, J. (2001). *Stress: Myth, Research, and Theory*. Prentice Hall.

Jose, P. E. (2013). *Doing statistical mediation and moderation* (p. 336). Guilford Publications. https://books.google.co.uk/books?id=ZxSoOsWXNloC

Kagan, J. (2010). *The Temperamental Thread: How Genes, Culture, Time, and Luck Make Us Who We Are*. Dana Pr.

Kagan, J., & Snidman, N. (2009). *The Long Shadow of Temperament*. Harvard University Press.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405. http://www.jstor.org.gate2.library.lse.ac.uk/stable/40062570

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives*, *20*(1), 3–24.

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.

Kastrenakes, J. (2021). Apple says there are now over 1 billion active iPhones. In *The Verge*. https://www.theverge.com/2021/1/27/22253162/iphone-users-total-number-billion-apple-tim-cook-q1-2021

Katz, J. (2004). Everyday lives and extraordinary research methods. *Social Science Information*, *43*(4), 609–619. https://doi.org/10.1177/0539018404047707

Kekäläinen, T., Hietavala, E.-M., Hakamäki, M., Sipilä, S., Laakkonen, E. K., & Kokko, K. (2021). Personality traits and changes in health behaviors and depressive symptoms during the COVID-19 pandemic: A longitudinal analysis from pre-pandemic to onset and end of the initial emergency conditions in Finland. *International Journal of Environmental Research and Public Health*, *18*(15). https://doi.org/10.3390/ijerph18157732

Kekäläinen, T., Sipilä, S., Saajanaho, M., & Kokko, K. (2021). The role of personality traits in leisure time physical activity during COVID-19 pandemic. *Personality and Individual Differences*, *182*, 111080. https://doi.org/https://doi.org/10.1016/j.paid.

2021.111080

Kenney, M., & Pon, B. (2011). Structuring the smartphone industry: is the mobile internet OS platform the key? *Journal of Industry, Competition and Trade*, *11*(3), 239–261.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, *59*(10), 877–883. https://doi.org/10.1001/archpsyc.59.10.877

Krohne, H. W., & Hock, M. (2011). Anxiety, coping strategies, and the processing of threatening information: Investigations with cognitive-experimental paradigms. *Personality and Individual Differences*, *50*(7), 916–925. https://doi.org/https://doi.org/10.1016/j.paid.2010.08.001

Kubiak, T., & Krog, K. (2012). Computerized sampling of experiences and behavior. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 124–143). Guilford Press.

Lades, L. K., Laffan, K., Daly, M., & Delaney, L. (2020). Daily emotional well-being during the COVID-19 pandemic. *British Journal of Health Psychology*, *25*(4), 902–911.

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the number of participants in intensive longitudinal studies: A user-friendly Shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–24.

Lahlou, S. (2010a). Augmented environments and design. In S. Lahlou (Ed.), *Designing user friendly augmented work environments* (pp. 1–29). Springer London. https://doi.org/10.1007/978-1-84800-098-8

Lahlou, S. (2010b). Experimental reality: Principles for the design of augmented environments. In S. Lahlou (Ed.), *Designing user friendly augmented work environments* (pp. 113–157). Springer London. https://doi.org/10.1007/978-1-84800-098-8

Lahlou, S. (2011). How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social Science Information*, *50*(3-4), 607–655.

Lahlou, S., Le Bellu, S., & Boesen-Mariani, S. (2015). Subjective evidence based ethnography: Method and applications. *Integrative Psychological and Behavioral Science*, 1–23. https://doi.org/10.1007/s12124-014-9288-9

Lavy, S., & Littman-Ovadia, H. (2011). All you need is love? Strengths mediate the negative associations between attachment orientations and life satisfaction. *Personality and Individual Differences*, *50*(7), 1050–1055. https://doi.org/https://doi.org/10.1016/j.paid.2011.01.023

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Losada, A., Márquez-González, M., Romero-Moreno, R., Mausbach, B. T., López, J., Fernández-Fernández, V., & Nogales-González, C. (2015). Cognitive–behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for dementia family caregivers with significant depressive symptoms: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *83*(4), 760.

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2021). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology*, *120*(3), 816–835. https://doi.org/10.1037/pspp0000289

Lupton, D. (2014). Self-tracking cultures: Towards a sociology of personal informatics. *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, 77–86.

MacKay, C. J., Cousins, R., Kelly, P. J., Lee, S., & McCaig, R. H. (2004). Management Standards and work-related stress in the UK: Policy background and science. *Work & Stress*, *18*(2), 91–112.

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, *23*(5), 992–1000.

Mackley, J. R. K. (2008). European 3G auctions: Using a comparative event study to search for a winner's curse. *Utilities Policy*, *16*(4), 275–283. https://doi.org/https://doi.org/10.1016/j.jup.2008.01.005

Mahase, E. (2021). Covid-19: Experts condemn UK "freedom day" as dangerous and unethical. *BMJ (Clinical Research Ed.)*, *374*, n1829. https://doi.org/10.1136/bmj.n1829

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, *63*(3), 365–396.

McCarty, R. (2016). The alarm phase and the general adaptation syndrome: Two aspects of Selye's inconsistent legacy. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior* (pp. 13–19). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800951-2.00002-9

Mehl, M. R., & Conner, T. S. (Eds.). (2012). *Handbook of Research Methods for Studying Daily Life*. Guilford Press.

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.

Mezick, E. J., Matthews, K. A., Hall, M., Kamarck, T. W., Buysse, D. J., Owens, J. F., & Reis, S. E. (2009). Intra-individual variability in sleep duration and fragmentation: Associations with stress. *Psychoneuroendocrinology*, *34*(9), 1346–1354.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237.

Morrissey, S., & Campbell, T. (2011). *iOS forensic analysis: For iPhone, iPad, and iPod touch*. Apress.

Naragon-Gainey, K., & Watson, D. (2011). Clarifying the dispositional basis of social anxiety: A hierarchical perspective. *Personality and Individual Differences*, *50*(7), 926–934. https://doi.org/https://doi.org/10.1016/j.paid.2010.07.012

Nezlek, J. B. (2012). Multilevel modeling analyses of diary-style data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 357–383). Guilford Press.

Niederkrotenthaler, T., Laido, Z., Kirchner, S., Braun, M., Metzler, H., Waldhör, T., Strauss, M. J., Garcia, D., & Till, B. (2022). Mental health over nine months during the SARS-CoV2 pandemic: Representative cross-sectional survey in twelve waves between april and december 2020 in austria. *Journal of Affective Disorders*, *296*, 49–58. https://doi.org/https://doi.org/10.1016/j.jad.2021.08.153

Nilsson, S., Buchholz, M., & Thunberg, G. (2012). Assessing children's anxiety using the modified short state-trait anxiety inventory and talking mats: A pilot study. *Nursing Research and Practice*.

Norton, K. S. (2001). Applying cross-functional evolutionary methodologies to web development. In S. Murugesan & Y. Deshpande (Eds.), *Web engineering: Managing diversity and complexity of web application development* (pp. 48–57). Springer Science & Business Media.

Nota, J. A., & Coles, M. E. (2014). Duration and timing of sleep are associated with repetitive negative thinking. *Cognitive Therapy and Research*, 1–9.

Office for National Statistics. (2014). *Measuring national well-being, personal well-being in the UK, 2013/14*. UK Government.

Office for National Statistics. (2017a). *Estimates of life satisfaction, worthwhile, happiness and anxiety in the UK for the years ending March 2012 to December 2017*. UK Government.

Office for National Statistics. (2017b). *Personal well-being in the UK: January to December 2017*. UK Government.

Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P., Lewis, S. W., & Dunn, G. (2011). Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta Psychiatrica Scandinavica*, *123*(1), 12–20.

Pasternak, H. (2018). *6 Things Your Resting Heart Rate Can Tell You About Your Health*. https://blog.fitbit.com/resting-heart-rate/

Perrez, M., Schoebi, D., & Wilhelm, P. (2000). How to assess social regulation of stress and emotions in daily family life? A computer-assisted family self-monitoring system (FASEM-C). *Clinical Psychology & Psychotherapy*, *7*(4), 326–339. https://doi.org/https://doi.org/10.1002/1099-0879(200010)7:4%3C326::AID-CPP266%3E3.0.CO;2-0

Population Reference Bureau. (2020). *2020 world population data sheet*. https://interactives.prb.org/2020-wpds/

Pottie, C. G., & Ingram, K. M. (2008). Daily stress, coping, and well-being in parents of children with autism: A multilevel modeling approach. *Journal of Family Psychology*, *22*(6), 855.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*(2), 189–205.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata: Vols. 1: Continuous responses* (Third edition). Stata Press.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods Research*, *37*(3), 426–454.

Reis, H. T. (2012). Why researchers should think ”real-world”: A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 3–21). Guilford Press.

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 3–10). Cambridge University Press New York, NY.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934.

Rothbart, M. K., & Posner, M. I. (2006). Temperament, attention, and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, volume 2: Developmental neuroscience* (2nd ed., pp. 465–501). John Wiley & Sons.

RStudio Team. (2019). *RStudio: Integrated Development Environment for R*. RStudio, Inc.

Rudaizky, D., & Macleod, C. (2013). Anxiety reactivity and anxiety perseveration represent dissociable dimensions of anxiety vulnerability: A replication and extension. *Australian Journal of Psychology*, *65*(4), 232–235.

Ryan, R. M., Bernstein, J. H., & Brown, K. W. (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. *Journal of Social and Clinical Psychology*, *29*(1), 95–122.

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic [Doi: 10.1016/S0140-6736(21)02143-7]. *The Lancet, Online First*. https://doi.org/10.1016/S0140-6736(21)02143-7

Sapolsky, R. (2004). *Why zebras don't get ulcers: The acclaimed guide to stress, stress-related diseases, and coping* (2nd ed.). New York: Henry Holt; Company.

Schwartz, N. (2012). Why researchers should think "real-time": A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 22–42). Guilford Press.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*(1), 5–34.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature*, *138*(3479), 32.

Selye, H. (1976). The stress of life (rev. edn.). *New York: McGrawHill*.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, *109*(5), 872–888. https://doi.org/10.1037/pspp0000036

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. http://www.ncbi.nlm.nih.gov/pubmed/18509902

Shrout, P. E., Herman, C. M., & Bolger, N. (2006). The costs and benefits of practical and emotional support on adjustment: A daily diary study of couples experiencing acute stress. *Personal Relationships*, *13*(1), 115–134.

Smith, E. R., & Conrey, F. R. (2009). The social context of cognition. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 454–466).

Cambridge University Press New York, NY.

Smith, T. W., & Anderson, N. B. (1986). Models of personality and disease: An interactional approach to type a behavior and cardiovascular risk. *Journal of Personality and Social Psychology*, *50*(6), 1166.

Smyth, J. M., & Stone, A. A. (2003). Ecological Momentary Assessment Research in Behavioral medicine. *Journal of Happiness Studies*, *4*(1), 35–52. https://doi.org/10.1023/A:1023657221954

Solan, M. (2018). *Your resting heart rate can reflect your current — and future — health*. https://www.health.harvard.edu/blog/resting-heart-rate-can-reflect-current-future-health-201606179806

Sorrel, C. (2010). *Inevitable iPhone 4 Flashlight Apps Flicker Into View*. https://www.wired.com/2010/06/inevitable-iphone-4-flashlight-apps-flicker-into-view/

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1983). *Manual for the State-Trait Anxiety Inventory*.

Spielberger, C. D., & Reheiser, E. C. (2009). Assessment of emotions: Anxiety, anger, depression, and curiosity. *Applied Psychology: Health and Well-Being*, *1*(3), 271–302.

Spinhoven, P., Balkom, A. van, & Nolen, W. A. (2011). Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Clinical Psychiatry*, *72*, 341–348.

StataCorp. (2009a). *Stata 11 base reference manual*. Stata Press.

StataCorp. (2009b). *Stata 11 base reference manual: Postestimation tools for regress with time series*. Stata Press.

StataCorp. (2009c). *Stata statistical software: Release 11*. StataCorp LP.

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

Stone, A. A., Schneider, S., & Harter, J. K. (2012). Day-of-week mood patterns in the United States: On the existence of "Blue Monday," "Thank God it's Friday" and weekend effects. *The Journal of Positive Psychology*, *7*(4), 306–314. https://doi.org/10.1080/17439760.2012.691980

Stone, A. A., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: Diurnal rhythms of a working day examined with the day reconstruction method. *Emotion*, *6*(1), 139.

Stone, A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals Of Behavioral Medicine*, *24*(3), 236–243.

Strelau, J. (2010). *Temperament as a regulator of behavior: After fifty years of research*. Eliot Werner Publications.

Strelau, J., & Zawadzki, B. (2011). Fearfulness and anxiety in research on temperament: Temperamental traits are related to anxiety disorders. *Personality and Individual Differences*, *50*(7), 907–915. https://doi.org/https://doi.org/10.1016/j.paid.2010.07.008

Swan, M. (2009). Emerging patient-driven health care models: An examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research and Public Health*, *6*(2), 492–525.

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85–99.

Szabo, S., Tache, Y., & Somogyi, A. (2012). The legacy of Hans Selye and the origins of stress research: A retrospective 75 years after his landmark brief letter to the editor of Nature. *Stress*, *15*(5), 472–478.

Tluczek, A., Henriques, J. B., & Brown, R. L. (2009). Support for the reliability and validity of a six-item state anxiety scale derived from the State-Trait Anxiety Inventory. *Journal of Nursing Measurement*, *17*(1), 19.

Tull, M. T., Barrett, H. M., McMillan, E. S., & Roemer, L. (2007). A preliminary investigation of the relationship between emotion regulation difficulties and posttraumatic stress symptoms. *Behavior Therapy*, *38*(3), 303–313. https://doi.org/https://doi.org/10.1016/j.beth.2006.10.001

Van der Doef, M., & Maes, S. (1999). The job demand-control (-support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, *13*(2), 87–114.

Van Dijk, L., Withagen, R., & Bongers, R. M. (2015). Information without content: A Gibsonian reply to enactivists' worries. *Cognition*, *134*(0), 210–214. https://doi.org/http://dx.doi.org/10.1016/j.cognition.2014.10.012

Van Yperen, N. W., & Snijders, T. A. (2000). A multilevel analysis of the demands–control model: Is stress at work determined by factors at the group level or the individual level? *Journal of Occupational Health Psychology*, *5*(1), 182.

Verdoux, H., Gindre, C., Sorbara, F., Tournier, M., & Swendsen, J. D. (2003). Effects of cannabis and psychosis vulnerability in daily life: an experience sampling test study. *Psychological Medicine*, *33*(1), 23–32. http://www.ncbi.nlm.nih.gov/pubmed/12537033

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Mechelen, I. V. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion*, *9*(1), 83–91. https://doi.org/10.1037/a0014610

Verduyn, P., Mechelen, I. V., Tuerlinckx, F., Meers, K., & Coillie, H. V. (2009). Intensity profiles of emotional experience over time. *Cognition & Emotion*, *23*(7), 1427–1443. https://doi.org/10.1080/02699930902949031

Vigo, M., Hassan, L., Vance, W., Jay, C., Brass, A., & Cruickshank, S. (2017). Britain breathing: Using the experience sampling method to collect the seasonal allergy symptoms of a country. *Journal of the American Medical Informatics Association*, *25*(1), 88–92.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, *59*(3), 339–354.

Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C., Derom, C., Thiery, E., Delespaul, P., & Van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*(1), 191–195.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. http://www.jstatsoft.org/v40/i01/

Wickham, H. (2019). *Forcats: Tools for working with categorical variables (factors)*. https://CRAN.R-project.org/package=forcats

Wilhelm, F. H., Grossman, P., & Müller, M. I. (2012). Bridging the gap between the laboratory and the real world: Integrative ambulatory psychophysiology. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 210–234). Guilford Press.

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 62–86). Guilford Press.

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. *European Journal of Psychological Assessment*, *23*(4), 258–267.

Wilhelm, P., Schoebi, D., & Perrez, M. (2004). Frequency estimates of emotions in everyday life from a diary method's perspective: a comment on Scherer et al.'s survey-study "Emotions in everyday life". *Social Science Information*, *43*(4), 647–665. https://doi.org/10.1177/0539018404047712

Wilson, R. A., & Clark, A. (2009). How to situate cognition: Letting nature take its course. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). Cambridge University Press New York, NY.

Wilson, T., & Gilbert, D. (2005). Affective forecasting - Knowing what to want. *Current Directions In Psychological Science*, *14*(3), 131–134.

Wilt, J., Oehlberg, K., & Revelle, W. (2011). Anxiety in personality. *Personality and Individual Differences*, *50*(7), 987–993. https://doi.org/https://doi.org/10.1016/j.paid.2010.11.014

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality*, *34*(2), 178–197. https://doi.org/10.1006/jrpe.1999.2275

Zuckerman, M. (2012). Models of Adult Temperament. In M. Zentner & R. L. Shiner (Eds.), *Handbook of Temperament* (pp. 41–66). Guilford Press.

Zunhammer, M., Eichhammer, P., & Busch, V. (2014). Sleep quality during exam stress: The role of alcohol, caffeine and nicotine. *PloS One*, *9*(10), e109490.