Essays on the Economic Psychology of Well-Being.

Sergio V. Pirla Lopez

TESI DOCTORAL UPF / Year 2022

THESIS SUPERVISOR
Daniel Navarro-Martinez & Jordi Quoidbach
Department of Economics and Business

Universitat Pompeu Fabra
Barcelona

Acknowledgements

I am deeply grateful to my two advisors – Daniel Navarro-Martinez and Jordi Quoidbach – for making this thesis possible. Their guidance, patience, and support during these years has allowed me to expand my intellectual interests, navigate challenging situations, and improve myself as a researcher. Getting to know and work closely with people I admire so much has been one of the best parts of my PhD journey. I feel extremely lucky to have you as advisors and I look forward to our future collaborations.

Many other professors shaped this thesis. I am especially grateful to Maxime Taquet for all the discussion and work that led to the first chapter of this thesis. I am also thankful to Robin Hogarth for fruitful discussions and feedback during these years. Other professors have directly impacted my work – either through their teaching or their feedback. For this, I am gratefult to Larbi Alaoui, Jose Apesteguia, Alex Frug, Gael LeMens and Gert Cornellissen. I am also thankful to the administrative assistant at UPF and ESADE Business School who have made my life easier during these years (Marta Araque, Laura Agusti, Sara Fernandez, Pablo Lopez-Aguilar, and Katica Boric). Thanks as well to my fellow PhD students at UPF and ESADE for our talks and discussions.

To my friends and family, thank you for your encouragement these years. To Olga, Javier and Isma – I cannot thank you enough for everything that you've done for me. Your example inspires me everyday to become a more caring, better person. Finally, thank you to my partner, Alma, for this would not have being possible without your love and support.

Abstract

This thesis consists of three chapters that improve our knowledge of how income shapes well-being. Chapter 1 ("Measuring Affect Dynamics: An Empirical Framework") presents an empirically-derived framework to conduct experience sampling studies of affect dynamics. Chapter 2 ("Happiness Without a Financial Safety Net: Low Income Predicts Emotional Volatility") examines how income shapes the emotional lives of 23,000 individuals whose happiness was tracked in real-time using a smartphone app. Lower income is associated with increased happiness volatility, a relationship that is partially explained by the experience of more frequent and intense periods of extreme unhappiness among low-income individuals. Chapter 3 ("Income, Boredom, and Mental Health") shows that low-income individuals experience more frequent boredom and that their affective experience of this emotion is more closely associated with depressed and anxious mood. Consequently, income moderates the relationship between boredom and the experience of clinical depression episodes.

Resumen:

Esta tesis consta de tres capítulos que contribuyen a mejorar nuestro conocimiento sobre la relación entre renta y bienestar. El Capítulo 1 ("Midiendo Dinámicas Afectivas: Un Marco Empírico") presenta un marco empírico para diseñar estudios de dinámicas afectivas usando el método de muestreo de experiencias. El Capítulo 2 ("Felicidad Sin Red de Seguridad Financiera: Ingresos Bajos Predicen Volatilidad Emocional") examina como los ingresos dan forma a las vidas emocionales de 23.000 personas cuya felicidad fue seguida en tiempo real usando una aplicación móvil. Los resultados muestran que una renta baja se asocia a una mayor volatilidad emocional. El Capítulo 3 ("Renta, Aburrimiento y Salud Mental") muestra que los individuos con rentas bajas experimentan aburrimiento más frecuentemente, y que, para estos individuos, la experiencia de esta emoción se asocia de forma más cercana a la experiencia de estados anímicos como la ansiedad o la depresión.

Preface

The relationship between income and well-being is a central topic in the social sciences, with implications in fields as diverse as management, economics, psychology, sociology, or political science. In our lives, understanding the relationship between income and well-being is key to comprehend people's daily behaviors and motivations, develop effective public policies, and design optimal incentive and compensation schemes.

The importance of the relationship between income and well-being has always attracted significant scientific attention – attention that was renewed at the onset of the new century when researchers across the social sciences shifted their focus towards subjective measures of well-being. These novel measures not only contributed to a better understanding of human welfare, but were pivotal in the development of new fields of study – most notably, the economics of happiness and the science of well-being (Cloninger, 2004; Diener et al., 2018; Diener, 2009; Eid and Larsen, 2008; Ferrer-i-Carbonell, 2013; Huppert et al., 2005; MacKerron, 2012; Van Praag and Ferrer-i-Carbonell, 2011). These fields provided the study of the relationship between income and well-being with theoretical structure and empirical rigor, but the standardization in methods and practices lead researchers to focus on a limited subset of subjective well-being measures, often neglecting the multidimensionality of human welfare. As a result, the last few years have witnessed an increased number of research and policy initiatives that call for more ample definitions of well-being (Diener et al., 2009; Diener and Tov, 2012; Forgeard et al., 2011; OECD, 2011; UN, 2012).

Despite this call for more multidimensional measures of well-being, progress in this area has been limited. To date, research on the relationship between income and subjective well-being has focused on static operationalizations of hedonic (i.e. happiness, positive affect, or life satisfaction, Diener et al., 2018; Kahneman et al., 1999) or eudaimonic (i.e. meaning or realizing one's potential, Ryan and Deci, 2001; Ryff, 1989; Vittersø, 2016) well-being. That is, well-being is typically operationalized as either happiness or meaning, and it is measured either once or on average. This relatively narrow conception of well-being is problematic. First, by focusing exclusively on hedonic or eudaimonic operationalizations of well-being, researchers might inadvertently ignore an important aspect of what constitutes a "good life". People not only want to live a happy or meaningful life, but a life that is filled with interesting experiences and novel or unique emotional moments - a "psychologically rich life" (Oishi et al., 2020; Oishi and Westgate, 2021). To fully capture the extent of the relationship between income and well-being, we need to move past the hedonic-eudaimonic dichotomy to also consider how income is related to monotony, uneventfulness, or boredom. Second, by focusing on static measures (i.e. measuring well-being once or on average), we might be obscuring important aspects of the relationship between income and well-being. Averaging across well-being reports can obscure the relationship between income and the experience of rare but extreme moments of suffering - moments that can have far reaching consequences through decision-making (Andrade and Ariely, 2009; Dunn and Schweitzer, 2005) and an impoverished physical and mental health (aan het Rot et al., 2012; Houben et al., 2015; Koval et al., 2013; Kuppens, 2015; Kuppens and Verduyn, 2017). In this thesis, we move past static measures of hedonic or eudaimonic well-being to consider how income relates to moment-to-moment happiness dynamics (Chapter 2) and boredom (Chapter 3).

The first chapter of this thesis ("Measuring Affect Dynamics: An Empirical Framework", joint with Maxime Taquet and Jordi Quoidbach) presents an empirical framework to conduct experience sampling studies of affect (or happiness) dynamics. Leveraging a dataset of 7,016 individuals that provided a minimum of 50 affect reports, we derive a set of general principles and tools to help researcher design well-powered and efficient experience sampling studies. Our results prove that for most dynamic measures of affect, a sample of 200 participants and 20 observations per person yields sufficient power to detect medium size associations. As the ideal sampling approach varies across affect dynamics measures, the chapter presents an R-package and an online app to help researchers conduct sample calculations and power analyses for studies of affect dynamics.

Using this framework, Chapter 2 ("Happiness Without a Financial Safety Net: Low Income Predicts Emotional Volatility" joint with Jordi Quoidbach) presents an investigation on the relationship between income and happiness fluctuations. Using over a million happiness reports from 23,000 individuals whose happiness was tracked in real-time using a smartphone app, we show that lower income is associated with increased happiness volatility – a relationship that replicates across multiple specifications of volatility and an additional sample of 25,000 individuals from 6 developing countries. Using a point and collective anomaly detection algorithm (Fisch et al., 2019), we move past classic psychometric approaches to fully identify what it means to live a financially deprived life. Our results suggest that financial scarcity is intimately linked to the experience of moments and periods of extreme unhappiness. Compared to high-income individuals, low-income earners experience more frequent and intense moments and periods of extreme unhappiness. The happiness gap between the highest and lowest earners during episodes of intense unhappiness was 1.5 to 3 times the size of the gap in average happiness between these two groups. Income is, nevertheless, unrelated to the experience of periods or moments of extreme happiness. Exploiting the exogeneity of monthly payments, we find that low-income people experience more moments and periods of anomalous happiness the last few days of the month, suggesting a causal relationship between income and happiness volatility. As fluctuations in happiness have been shown to impose a severe tax on a person's physical and mental health (aan het Rot et al., 2012; Houben et al., 2015; Koval et al., 2013; Kuppens, 2015; Kuppens and Verduyn, 2017), our results have important policy implications.

Finally, in Chapter 3 (" Income, Boredom, and Mental Health" joint with Daniel Navarro-Martinez and Jordi Quoidbach), we move past hedonic or eudaimonic characterizations of well-being to investigate how income relates to boredom – an emotional state linked to noise and decision errors, anti-social behavior, the development of addictions, and poor mental health outcomes. Using information on 65,000 individuals across 28 countries, we show that low-income individuals experience boredom more often. In addition, we take a network approach to show that the experience of boredom is more closely associated with depressed and anxious mood for low-income individuals. Consequentially, we demonstrate that income moderates the relationship between boredom and the experience of clinical depression episodes. On average, experiencing high levels of boredom is associated with an increase in the probability of suffering clinical depression of 10.9%. For each decrease of 1 SD in income, the difference in incidence of depression episodes across low and high-boredom individuals increases by 1.4%. Focusing on a subset of 1,907 individuals that experienced a depression episode in the last 12 months, we provide further evidence on the specific depression symptoms associated with experiencing high levels of boredom. Boredom is associated with a significant increase in the propensity to experience 6 depression symptoms including morbid thoughts, loss of appetite, sleeping problems, and feelings of hopelessness, anxiety, and restlessness. Together with previous literature on the decision-making consequences of boredom, our results portray this emotion as a potential poverty self-reinforcing mechanism. Our findings contribute to better understand the mental health consequences of boredom and open a venue to future research and policies that address the full extent of the emotional tax exerted by financial hardship.

Contents

List of Figures

List of Tables

Chapter 1

MEASURING AFFECT DYNAMICS: AN EMPIRICAL FRAMEWORK

Joint with Maxime Taquet and Jordi Quoidbach

Abstract

A fast-growing body of evidence from experience sampling studies suggests that affect dynamics are associated with well-being and health. But heterogeneity in experience sampling approaches impedes reproducibility and scientific progress. Leveraging a large dataset of 7016 individuals, each providing over 50 affect reports, we introduce an empirically-derived framework to help researchers design well-powered and efficient experience sampling studies. Our research reveals three general principles. First, a sample of 200 participants and 20 observations per person yields sufficient power to detect medium size associations for most affect dynamic measures. Second, for trait and time-independent variability measures of affect (e.g., S.D.), distant sampling study designs (i.e., a few daily measurements spread out over several weeks) leads to more accurate estimates than close sampling study designs (i.e., many daily measurements concentrated over a few days), whereas differences in accuracy across sampling methods were inconsistent and of little practical significance for temporally-dependent affect dynamic measures (i.e., RMSSD, autocorrelation coefficient, TKEO, and PAC). Third, across all affect dynamics measures, sampling exclusively on specific days or time windows leads to little to no improvement over sampling at random times. Because the ideal sampling approach varies for each affect dynamics measure, we

provide a companion R-package, an online calculator, and a series of benchmark effect sizes to help researchers address three fundamental how's of experience-sampling: How many participants to recruit? How often to solicit them? And for how long?

1.1. Introduction

With the advent of mobile phones, the experience sampling method (ESM; Csikszentmihalyi and Larson, 1984; also known as ecological momentary assessment; Stone and Shiffman, 1994) has quickly become a gold standard to study human emotion (Lucas et al., 2021; Stone et al., 1998). Rather than relying on retrospective reports ("How did you feel yesterday?") or cross-sectional surveys ("How do you feel in general?"), researchers in psychology, psychiatry, and behavioral science are now routinely capturing people's subjective experience in the moment through short mobile questionnaires. Experience-sampling not only alleviates recall and evaluative bias (Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996; Schimmack and Oishi, 2005), but also allows to uncover how dynamic aspects of people's emotional lives (e.g., fluctuation, inertia) play a crucial role in mental and physical health (for a meta-analysis, see Houben et al., 2015).

Since the first ESM studies in the 70's, countless articles have discussed the promises of the method to study emotion (Ellison et al., 2020; Fisher and To, 2012; Myin-Germeys et al., 2018; Schimmack, 2003; Scollon et al., 2009), and many technical solutions have blossomed (see Arslan et al., 2020 and Meers et al., 2020 for overviews). However, scientists have been astonishingly left to their own devices when it comes to conducting such research. Imagine, for example, that you want to assess how happy a person feels. How many moments of their daily life should you observe to capture their average happiness accurately? What about their propensity to experience mood swings? How spread in time or concentrated should your observations be? These questions are critical to the design of well-powered, cost-efficient ESM studies in affective sciences. However, an abysmal 2% of emotion ESM studies justify their sampling procedure (Trull and Ebner-Priemer, 2020), leading to important power, reproducibility, and suboptimal resource-allocation issues (e.g., Aguinis et al., 2013; Calamia, 2019; Kirtley et al., 2021) In what follows, we first provide a brief overview of the experience-sampling method in emotion research and the primary individual differences studied through this method. We then review the wide variety of sampling practices used to capture these individual differences. Finally, we stress the importance of relying on actual data to make critical decisions about how many participants to recruit and how often, when, and for how long to observe them.

1.1.1. Experience Sampling and Affective Sciences

Experience sampling involves repeated measurement of people's experiences, as it unfolds in real time in their everyday lives (Conner et al., 2009). It offers several advantages over traditional lab or survey-based emotion research.

First, by capturing emotions as they naturally occur in everyday life - rather than relying on artificial laboratory manipulation - ESM helps uncover how complex, intertwined, and diverse our affective reactions truly are (e.g., Dejonckheere et al., 2018; Kerr et al., 2021). For example, while theorists have debated the idea that people can experience two oppositely valenced emotions for decades, results from experience sampling suggest that this is a ubiquitous experience in everyday life: People report experiencing mixed emotions about a third of the time (Trampe et al., 2015).

Second, by capturing emotions in real time, ESM reduces recall and evaluative biases (e.g., Solhan et al., 2009; Stone et al., 1998). For example, people's retrospective ratings of how they felt during emotional experiences are overly influenced by these experiences' last and most intense moments (Fredrickson and Kahneman, 1993; Kahneman et al., 1993; Redelmeier and Kahneman, 1996). Similarly, global reports of affective states can be tainted by aspects of one's life that happen to be salient at the moment (see Schimmack and Oishi, 2005, for a meta-analysis) - for example, asking people questions about politics right before asking them how happy they feel overall substantially reduces happiness scores (Deaton and Stone, 2016).

Third, by capturing emotions on multiple occasions, ESM allows studying the role of changing contexts on people's emotions. For example, researchers have been able to quantify what type of daily activities (Choi et al., 2017; Taquet et al., 2016) or social interaction partners (Quoidbach et al., 2019) impact people's momentary happiness. For instance, Mueller et al. (2019) examined over 50,000 episodes of social interactions. They found that social (vs. task-oriented) conversations with close (vs. less close) others were associated with higher momentary happiness.

1.1.2. Experience Sampling and Affect Dynamics Measures

Beyond increased ecological validity and accuracy, a major contribution of ESM is that it allows to uncover how individual differences in affect dynamics, that is, trajectories, patterns, and regularities in people's emotion over time, play a critical role in mental health and psychopathology (Kuppens, 2015; Kuppens and Verduyn, 2017). Dozens of new affect dynamics measures have been introduced over the past decade, each designed to evaluate a unique aspect of people's emotional lives. Whereas the incremental validity of several of these indicators is currently debated (Dejonckheere et al., 2019; Lapate and Heller, 2020; Wendt et al., 2020), the most common measures of affect dynamics in the literature include trait affect, affect variability, affect instability, and affect inertia (see Table 1.1).

Trait affect represents people's propensity to experience negative or positive affect and is considered a relatively stable personality characteristic (e.g., Watson and Tellegen, 1985). It is typically captured as the individual mean of affective states. Affect variability represents whether people's affective state tends to change over time, regardless of when these changes occur. It is typically operationalized as the intraindividual standard deviation in affective states (Nesselroade and Salthouse, 2004) or a mean-corrected version of this intraindividual standard deviation that avoids confounding effects of the mean (Mestdagh et al., 2018). In contrast, affect instability is a function of temporal order and represents whether people's affective states tend to change abruptly from one moment to the next. Across different research domains, instability has been typically measured as the root mean square of successive differences (RMSSD; Jahng et al., 2008), the probability of acute change (PAC; Trull et al., 2008), or the Teager-Kaiser Energy Operator (TKEO; Solnik et al., 2010; Tsanas et al., 2016). Finally, affect inertia represents the degree to which people's affective states persist from one moment to the next. It is typically captured as an autoregressive correlation between an individual's current affective state and their previous affective state in time series (AR; e.g., Kuppens et al., 2010).

Accumulating empirical evidence shows that affect dynamics are associated with well-being and health. For example, research shows strong associations between average affect and depression (Golier et al., 2001; Thompson et al., 2012), post-traumatic stress disorder (Golier et al., 2001), borderline personality disorder (Zeigler–Hill and Abraham, 2006), and anxiety disorders (Bowen et al., 2006). Likewise, affect variability predicts lower subjective well-being (Gruber et al., 2013) and affective disorders (Bowen et al., 2004; Golier et al., 2001; McConville and Cooper, 1996). Affect instability is linked to poor mental health and several psychological disorders, including anxiety (Pfaltz et al., 2010), bipolar disorder (Jones et al., 2005), borderline personality disorder (Ebner-Priemer et al., 2007; Santangelo, Bohus, et al., 2014), major depressive disorder (aan het Rot et al., 2012), and bulimia nervosa (Anestis et al., 2010). Finally, affect inertia is related to low self-esteem, neuroticism, and trait rumination (see Trull et al., 2015 for a review).

Measure	Index	Formal Definition	Interpretation
Trait	Average (M)	$\sqrt{\frac{\sum x_i}{n}}$	Average affect.
Variability	Standard Deviation (SD)	$\sqrt{\frac{\sum(x_i - M)^2}{n}}$	Standard deviation of affect.
Variability	Relative Standard Deviation (Rel. SD)	$\frac{SD}{max(SD\|M)}$	Mean-corrected estimate of the standard deviation for bounded variables (Mestdagh et al. 2018).
Instability	Root Mean Square of Successive Differences (RMSSD)	$\sqrt{\frac{\sum(x_i - x_{(i+1)})^2}{n-1}}$	Average change across successive affect observations.
Instability	Teager-Kaiser energy operator (TKEO)	$\frac{\sum(x_i^2 - x_{(i-1)}x_{(i+1)})}{n-2}$	Measure of change across 3 affect reports. Useful in identifying mood spikes.
Instability	Probability of Acute Change (PAC)	$\frac{\sum I(x_{i+1} - x_i, d_{0.9})}{n-1}$	Likelihood of extreme affect changes.
Inertia	Auto-correlation Coefficient	$\sqrt{\frac{\sum(x_i - M)(x_{i+1} - M)}{\sum(x_i - M)^2}}$	Correlation between successive affect reports.

Table 1.1: Affect dynamics measures included in our study. In the formulas, x_i stands for the i^{th} current affect report of a given individual. Similarly, n represents the total number of observations collected for the individual. SD and M represent respectively the standard deviation and mean affect reported by a given individual. Finally, $I(x_{i+1} - x_i, d_{0.9})$ defines a binary variable taking a value of 1 if $(x_{i+1} - x_i)$ is greater than $d_{0.9}$ in absolute terms and 0 otherwise, where $d_{0.9}$ represents the 90th percentile in the distribution of absolute affect changes across all participants in the sample.

1.1.3. Affect Dynamics Measures: The Wild West of Sampling Approaches

The field of affect dynamics holds great promise. But the wide range of outcomes that have been related to affect dynamics measures is met by an even wider range of methodological approaches to study them. We examined the sampling characteristics of 423 ambulatory assessment studies of affect included in five major review articles (aan het Rot et al., 2012; Dunster et al., 2021; Ebner-Priemer and Trull, 2009; Houben et al., 2015; Myin-Germeys et al., 2009). Of these, 88 studies estimated at least one core affect dynamics measure. Our examination revealed a wide range of practices with samples ranging from 10 to 500 individuals and 14 to over 400 observations per individual (see Figure 1.1). Studies also crucially differed with regard to when and for how long they surveyed participants. Some studies favored close sampling - many questionnaires collected over a short period (e.g., ten questionnaires a day for a week; Delespaul and DeVries, 1987; Myin-Germeys et al., 2000; Peeters et al., 2010) - whereas others favored distant sampling - few questionnaires per day collected over a longer period (e.g., two questionnaires a day for two weeks; Chepenik et al., 2006; Links et al., 2003). Some studies systematically sent questionnaires on specific days (weekdays vs. weekends; Beal and Ghandour, 2011) or at specific times (e.g., morning, afternoon, or evenings; Gruber et al., 2013; Knowles et al., 2007; Links et al., 2003; Zeigler–Hill and Abraham, 2006), while other studies probed participants at random times (Havermans et al., 2007; Peeters et al., 2006; Trull et al., 2008).

The lack of a standardized approach has profound ramifications. First, it leads researchers to rely on heuristics, opportunities, or unfounded conventions to define their sample size, rather than rely on adequate power calculation. For example, a common design in the ESM literature (around 40% of the studies) is to collect observations ten times a day for six consecutive days, even if this approach is neither based on power considerations nor necessarily optimal (Myin-Germeys et al., 2018). The current lack of evidence to guide sampling decisions might result in underpowered studies, leading to missed opportunities to discover true effects and inflated effect sizes of discovered effects (Ioannidis, 2008). Combined with publication bias and the difficulty to publish null results, underpowered studies are a root cause of the dire claim that most research findings are false (Ioannidis, 2005). Whereas underpowered studies are of great concern, researchers should not find solace in overpowered studies. Recruiting more participants than is needed or running a study for longer than necessary puts an unnecessary burden on participants, increases the risk of attrition, and misallocates essential resources. It might also be unethical if the answer to the research question at hand can improve people's health or quality of life, and so should be sought with a degree of urgency.

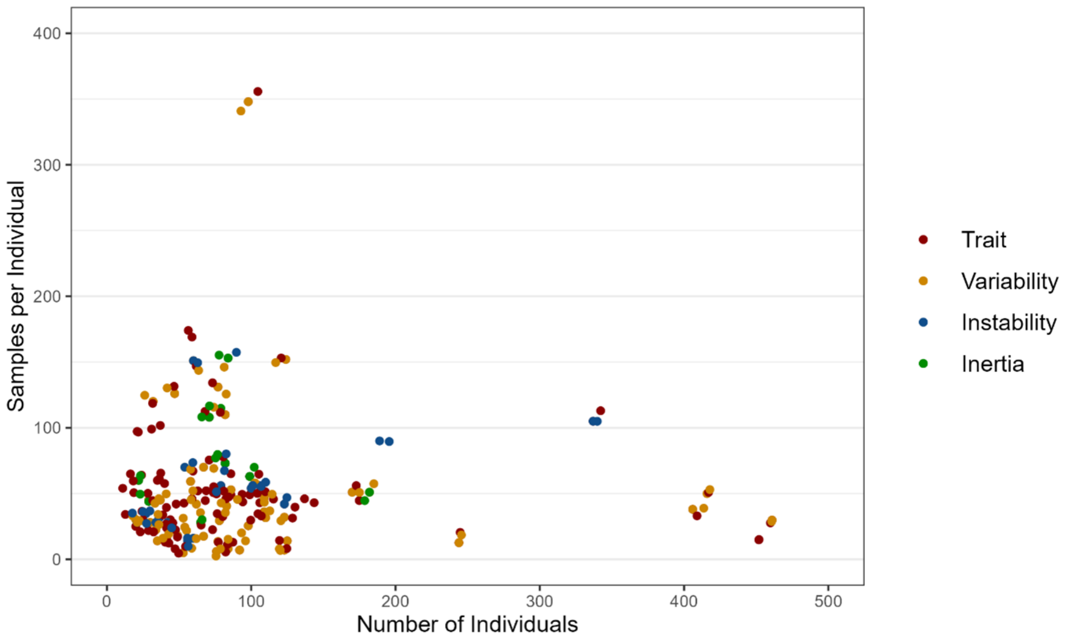


Figure 1.1: Distribution of the number of individuals sampled and the number of observations per individual in 88 emotion ESM studies.

1.1.4. Developing an Empirical Framework

The goals of affective scientists when conducting experience sampling studies are two-fold. First, they might be interested in precisely estimating an affect dynamic measure for a given group of individuals. Second, they might be interested in analyzing the relationship between an affect dynamic measure and another variable. In this paper, we consider both cases, presenting results that will be of use to those researchers concerned with estimation accuracy and those looking for guidance about power analysis.

A validated framework for study design would considerably advance the study of affect dynamics. But this framework needs to be determined on real affect data and not on simulations (Arend and Schäfer, 2019; Astivia et al., 2019; Lane and Hennes, 2018). In particular, while power analysis is a valid criterion to conduct inference under a set of plausible distributional assumptions of the data, defining a valid set of plausible distributional assumptions for affect dynamic studies is challenging. This is because the data generation process is complex and cannot be accurately captured by parametric models. Affect time series are stochastic processes that depend, in nonlinear ways, on various intertwined variables (e.g., time, weather, social interactions, cortisol level, physical wellness), many of which cannot be measured. Moreover, affect dynamics measures (e.g., the root mean squared successive differences) are themselves nonlinear summary statistics derived from these time series. Therefore, any valid framework to designing affect dynamics studies needs to link the probability distribution of these non-linear transformations of non-uniformly sampled stochastic time series to the sampling process. In practice, this is most readily achieved using real data and assessing power empirically.

To address these issues, we build on a large dataset of 7016 individuals, each providing over 50 affect reports at random moments using smartphones. We first analyze how many samples are needed to accurately estimate a person's affect dynamics in terms of trait affect (i.e., average), affect variability (i.e., within-person standard deviation), affect instability (i.e., RMSSD, TKEO, and PAC), and affect inertia (i.e., autocorrelation). We also investigate how strategic considerations in terms of timing between samples, time of the day, and days of the week change the number of samples needed to accurately estimate these affect dynamics measures. Second, we examine how the power to detect an association between the different measures and a given outcome varies as a function of sampling procedures. In doing so, we provide researchers with an easy-to-use companion R-package and an online calculator to address the three fundamental how's of experience sampling studies: How many participants to recruit? How often to solicit them? And for how long?

1.2. Method

1.2.1. Participants and Experience Sampling

We collected our data using "58 seconds", a free francophone smartphone application designed to assess different aspects of people's well-being by sending short questionnaires at random times of the day. Participants provided basic information on age, gender, and country of residence at sign-up (see Note 1 of Supplementary Materials). They were then asked to select which days of the week, within what time windows, and how many sample requests they wanted to receive (default = 4 questionnaires daily between 9 a.m. and 10 p.m. each day of the week). Taking into account each user's preferences and time constraints, the app sent questionnaire requests at random times throughout the day. By design, the minimum time between two consecutive notifications was set to 1 hour. We ensured random sampling through a notification system that did not require users to be connected to the internet. Each questionnaire consisted of 4-6 questions selected from an extensive battery of items. The sample and item pool has been extensively described in other publications (Quoidbach et al., 2019; Taquet et al., 2020). For the purpose of this study, we focused on participants who reported their current affective state (using a slider from 0-very unhappy to 100-very happy) at least 50 times. This subsample included 7016 individuals (M Age = 29.9, S.D. Age = 9.9; 74% female) who each provided an average of 111.6 (S.D. = 87.8) momentary affect reports.

1.2.2. Analytical Approach

Estimating Affect Dynamics Accurately

To analyze the number of reports required per individual to estimate each of the seven core affect dynamics measures reliably, we began by estimating their "true" value using the complete set of observations available for each individual. For example, if a participant provided 150 momentary affect reports, we computed the seven core affect dynamics measures for this participant (e.g., average happiness, within-person standard deviation, autoregressive coefficient) using all 150 observations. Then, we randomly selected a subset of N affect reports for each individual (with N varying from 3 to 30) and computed the affect dynamics measures using this smaller set of observations. We repeated this process 1000 times for each participant and for each value of N. We calculated an individual's root mean square error (RMSE) of the estimates (compared to the "true" measure based on the full sample) for each value of N. We averaged the RMSE across participants to examine how the accuracy of the estimates changed as one increased the number of reports used to compute the different affect dynamics measures. To provide intuitive benchmarks against which these RMSE values can be compared, we also report, for each affect dynamics measure, the standard deviation of the "true" value in our population. This allows readers to appraise how big or small an RMSE is. For instance, if we were measuring people's weight, an RMSE of 1 gram would be considered very small because the standard deviation of weights in the population is several kilograms. But if we were measuring insects' weights, an RMSE of 1 gram would be considerably larger. If for a given affect dynamics measure and number of affect reports per individual, our average RMSE equals one standard deviation in the true affect dynamics measure across individuals, we can expect the within-person estimation error to be equal in size to one between-person standard deviation in the true measure.

Optimizing Sampling Approaches

Could researchers reduce estimation errors of affect dynamics measures - and thus the number of reports required per individual - by probing participants at specific moments? To test whether sampling strategies can be optimized, we compared the accuracy of affect dynamics measures computed using reports selected at random times with affect dynamics measures computed (1) with temporally close or distant reports, (2) reports obtained at specific times, and (3) reports obtained on specific days (see details below).

To assess the accuracy of affect dynamics measures estimated using reports elicited at random times, we followed the procedure outlined in the previous section (see Estimating Affect Dynamics Accurately section). These baseline accuracy estimates were then compared to those obtained using alternative sampling strategies. To assess our results' robustness, for each condition and number of reports used, we bootstrapped over the individual-specific RMSE estimates to obtain the 95% confidence intervals for the average RMSE across individuals.

Random, close or distant sampling

Close sampling consists in collecting many reports over a short period of time. In this study, we consider close sampling to be the set of consecutive affect reports that were collected within the shortest possible time period for each individual (imposing a maximum of 24 hours between each affect report). In contrast, distant sampling consists in collecting reports less frequently but for a longer period of time. In this study, we consider distant sampling to be the individual's maximally distant reports. To determine an individual's maximally distant reports, we divided the temporal window in which each participant provided reports (from their first to their last) into N - 1 equally spaced time intervals (where N takes on values between 3 and 30, depending on the number of reports used in the computation). We then computed the different affect dynamics measures selecting reports that fell as close as possible to an equally spaced design. Note that by construction there is only a single set of reports for each individual that is considered close and distant sampling. Thus, for these sampling strategies, only one value of each affect dynamics measure was calculated per individual for each value of N (instead of resampling and estimating them 1000 times).

Random vs. specific times sampling

Specific times sampling differs from random sampling in that we estimated the affect dynamics measures using reports collected exclusively in the morning (from 6 am to 12 pm), afternoon (12 pm to 4 pm), evening (4 pm to 8 pm), or at night (8 pm to 6 am). For each of these conditions and number of affect reports from 3 to 30, we resampled and estimated the affect dynamics measures 1000 times. We introduced a bias-correction term in the estimates of affect dynamics measures to account for any baseline differences that might exist between specific sampling times (e.g., on average, affect tends to be more pleasant in the evening than in the morning). To debias the estimates, we first estimated a time window-specific bias by subtracting from the population average of affect dynamics measures based on all available affect records, the population average of the same measure estimated with affect reports from our time-window of interest. We then subtracted this bias from each of our estimates of affect dynamics measures averaged over 1000 bootstrap samples. For example, when analyzing the performance of the estimations of the average affect with reports collected at night, we first obtained a time window-specific bias. To calculate this bias term, we 1) estimated each individual's average affect using all reports available, 2) estimated each individual's average affect using all reports collected at night, 3) subtracted the population average of estimates in (1) from the population average of estimates in (2). The bias term is then added to each individual's average affect. This debiasing procedure allowed us to account for "time-window fixed effects", any bias across individuals that did not affect the relative ordering of individuals in terms of their affect dynamics measure of interest. Results obtained when excluding this bias-correction term can be found in Supplementary Note 2. For each time window, we excluded from our estimations participants that had not provided a minimum of 30 affect reports within that time window. This resulted in a final sample of 2806 individuals in the morning condition (i.e., 40% of the total sample), 2126 in the afternoon condition (i.e., 30.3% of the total sample), 2475 in the evening condition (i.e., 35.3% of the total sample), and 914 individuals in the night condition (i.e., 13% of the total sample).

Random vs. specific days sampling

Specific days sampling differs from random sampling in that we estimated the affect dynamics measures using reports collected exclusively during the weekends (weekend sampling) or during the week (weekday sampling). For each of these conditions, we resampled and estimated each affect dynamics measure 1000 times using a specific number of reports from 3 to 30. Again, we included a bias-correction procedure and omitted the data from participants that did not provide a minimum of 30 affect reports in each condition. This resulted in a final sample of 6982 individuals in the weekday condition (i.e., 40% of the total sample), and 2482 individuals in the weekend condition (i.e., 13% of the total sample).

Statistical Power as a Function of Sampling

In this section, we derive statistical power estimates for a two-tail t-test on the Pearson correlation coefficient between a given variable and an affect dynamic measure. That is, given two variables (one of them being an affect dynamic measure), we analyze power for a two-tail t-test examining the null hypothesis that

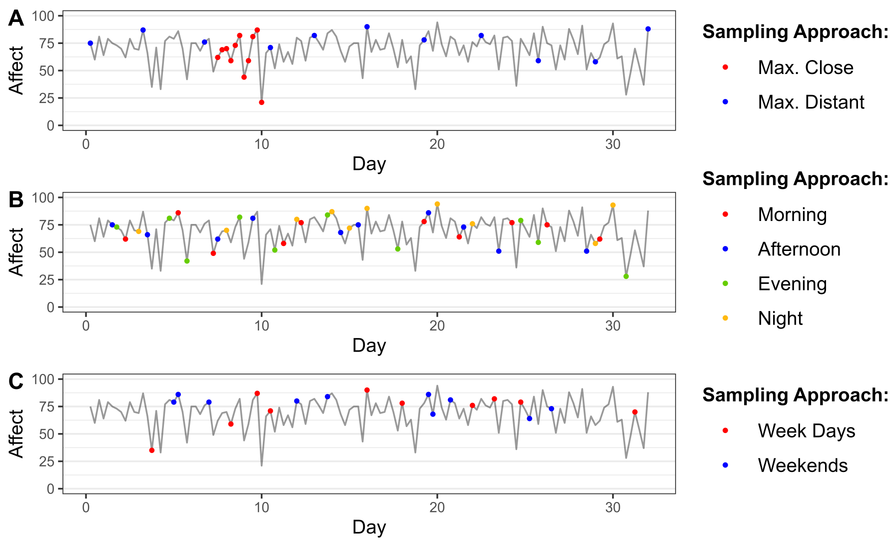


Figure 1.2: Graphical representation of the different sampling strategies tested.

the Pearson correlation between them is equal to zero, against the alternative hypothesis of a non-zero Pearson correlation coefficient. Throughout this paper, our tests employ a 0.05 significance level - but extensions of our analyses to different significance levels are included in our online calculator and R-package.

To conduct these analyses, we first estimated the seven affect dynamics measures for each individual using all the observations at our disposal. We then simulated random variables displaying a weak (Pearson's r = .10), medium (r = .30), and strong (r = .50) positive correlation with each affect dynamics measure by adding orthogonal random Gaussian noise (with a mean of 0 and standard deviation of 1) to projections of our variables of interest on vectors displaying the desired correlations. In doing so, we obtained variables displaying a weak, medium, and strong correlation with the affect dynamics measures derived from our full sample. We repeated this process to obtain a large enough set of simulated variables (2,500 simulated variables per effect size and affect dynamic measure). To evaluate how the power to detect these correlations changes when affect dynamics measures are computed from smaller numbers of participants and smaller number of observations per participant, we considered ten different numbers of participants (N Participants = 10, 20, 40, 80, 160, 320, 640, 1280, 2560, and 5120) and ten different numbers of observations per participants (N Observations = 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50), leading to 100 (= 10 x 10) sampling specifications in total.

For each combination of number of participants and number of observations per participant, we created 2,500 datasets by resampling from our original data. For each of these 2,500 datasets, we computed the seven affect dynamics measures for each participant. For each of these measures, we analyzed its correlation with a corresponding simulated variable (i.e., a simulated variable displaying the desired correlation with the full sample measure). We quantified power as the proportion of simulated datasets with a statistically significant positive correlation between the affect dynamics measures and the simulated variable.

Benchmarks for Plausible Effect Sizes

Like other power calculation tools, the sampling recommendations derived from our empirical framework require researchers to anticipate plausible effect sizes for the association they are interested in (or to set a minimum effect size that they want their study to detect). Such anticipated effect sizes can be informed by systematic literature review, preliminary data, and meta-analyses. But in practice, it may be challenging for affective scientists to come up with realistic effect size estimates as the field of affect dynamics is relatively new, and such estimates may not exist. Moreover, historical data may offer little guidance as past estimates tend to be overestimates given reporting and publication bias favoring significant results (Gelman and Carlin, 2014). Therefore, we provide a series of benchmarks based on 10 variables that we measured alongside affect in our experience-sampling project: (1) age, (2) gender, (3) average sleep time, (4) life satisfaction, (5) meaning in life, as well as the proportion of time spent with (6) friends, (7) family, (8) alone, (9) working, and (10) exercising (see Supplementary Note 5 for the complete list of variables and their operationalizations). Note that for life satisfaction and meaning in life, the associations we report are based on matched measures. For instance, we report the correlation between trait affect and trait life satisfaction, the correlation between affect instability and life satisfaction instability, and the correlation between affect inertia and life satisfaction inertia (vs. non-matching pairs).

We chose to report these 10 variables because they are commonly used demographic, well-being, and contextual measures in the experience sampling literature and cover a wide range of effect sizes-displaying correlations from $|r| = 0.002$ to $|r| = 0.856$ with our affect dynamic measures. By considering the magnitude of the relationships between these ten variables and the different affect dynamics measures, we hope to help researchers design optimized ESM studies based on plausible effect size estimates.

1.3. Results

1.3.1. Measuring Affect Dynamics Accurately

Figure 1.3 depicts changes in RMSE as we increase the number of observations per individual used to compute the seven affect dynamics measures. Our results show a large degree of heterogeneity between measures. We found that the number of observations needed to estimate our affect dynamics measures with a minimum accuracy of one between-subject standard deviation in the true measures ranges from 3 for trait affect to over 30 for the autocorrelation coefficient.

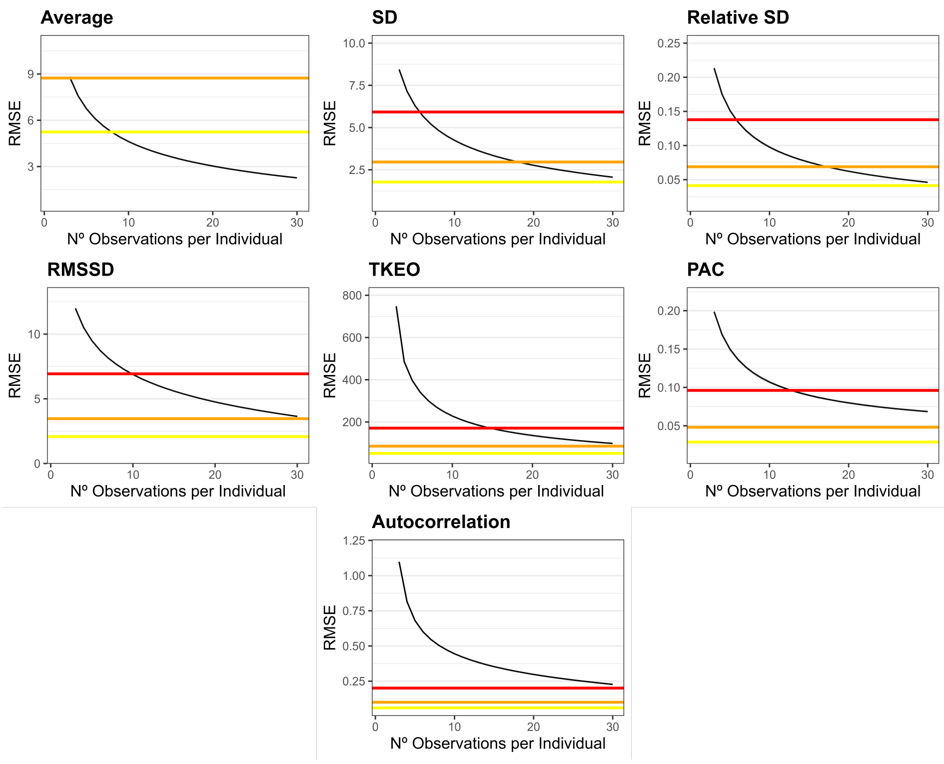


Figure 1.3: Average RMSE in the estimation of affect dynamics measures as a function of the number of observations per participant. The horizontal lines provide accuracy benchmarks depicting 1 (red), 0.5 (orange), and 0.3 (yellow) between-subjects standard deviations in the affect dynamics measure estimated on the full sample.

1.3.2. Optimizing Sampling Approaches

Random, close or distant sampling

Is it better to conduct short intense studies or longer less-demanding ones? As shown in Figure 1.4, the optimal measurement method depends on the affect dynamics measure of interest and the number of observations used to estimate it. We found large differences in the estimation error across sampling methods when calculating affect dynamics measures that are not temporally dependent (i.e., average affect, standard deviation, and relative standard deviation). Estimations of these three measures under close sampling were significantly less accurate than under random and distant sampling. For example, we can estimate a person's average affect more accurately with ten observations collected at random times over multiple days or weeks than with over 30 consecutive observations over shorter periods of time. In addition, when only a few observations can be collected, we found that distant sampling leads to more accurate estimations than both close and random sampling. Note that the difference between distant and random sampling is small and not statistically significant when at least 27 observations per individual are included in the estimation.

The differences in accuracy across sampling methods were substantially smaller and less consistent for temporally dependent affect dynamics measures (i.e., RMSSD, autocorrelation coefficient, TKEO, and PAC). For RMSSD and the autocorrelation coefficient, estimates obtained through close and distant sampling did not differ, though both of these strategies outperformed random sampling. For TKEO, close sampling largely outperformed both distant and random sampling, especially when the number of observations per participant is small. For the PAC, distant sampling outperformed close and random sampling, especially when the number of observations per participant is large.

Random vs. specific times sampling

Are there better moments than others to capture people's affective states? For non-temporally dependent measures (i.e., average affect, standard deviation, and relative standard deviation), random sampling tended to outperform estimates based solely on observations collected at specific times - with estimates based on night hours leading to the highest estimation error (see Note 2 of Supplementary Materials). Note that the differences were small and, in many cases, non-significant. For affect instability measures (i.e., RMSSD, TKEO, PAC) sampling exclusively at specific times outperforms random sampling, although the differences are small and non-significant across most numbers of samples. Sampling earlier in the day, either in the morning or in the afternoon yielded the best results. For affect inertia

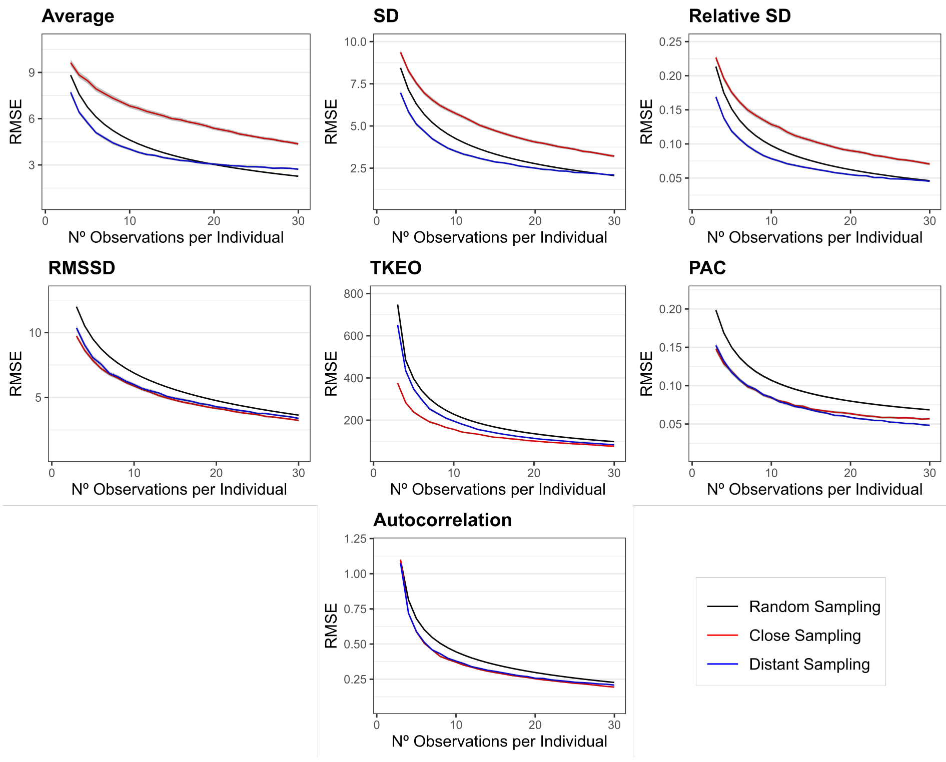


Figure 1.4: Average RMSE in the estimation of affect dynamics measures as a function of the number of observations per participant collected under random (black), close (red), or distant (blue) sampling. Gray areas around the lines represent the 95% confidence intervals for the average RMSE.

(i.e., autocorrelation coefficient), sampling exclusively at specific times performed better than random sampling-with estimates based on night hours providing the best performance. Detailed results for random vs. specific times sampling can be found in Supplementary Note 2.

Random vs. specific days sampling

Are there better days than others to capture people's affect dynamics? For non-temporally dependent measures (i.e., average affect, standard deviation, and relative standard deviation), random sampling tended to perform better than sampling on specific days - with estimates based on weekend observations yielding the highest estimation error. Again, these differences were small and, in many cases, non-significant. For measures of affect instability, we did not find differences between random sampling and sampling on specific days for TKEO and PAC, but we found small differences favoring sampling on the weekends for the estimation of the RMSSD. For affect inertia (i.e., autocorrelation coefficient), sampling exclusively on the weekends and sampling exclusively on the weekdays performed better than random sampling-with sampling on the weekends yielding the best performance. Detailed results for random vs. specific days sampling can be found in Supplementary Note 2.

1.3.3. Statistical Power as a Function of Sampling

Figure 1.5 displays the minimal combinations of number of individuals and observations per individual needed to achieve 80% power to detect an association of medium size (r = .30) using a two-tail t-test and an alpha of .05. The different curves are intended to provide a quick overview of how the number of individuals and samples per individuals can be traded off. Detailed information about (1) the method we used to estimate these curves, (2) the specific power achieved for all tested combinations of number of individuals and samples per individual, and (3) other effect sizes and power levels are presented in Supplementary Note 3 and in the online app (https://sergiopirla.shinyapps.io/powerADapp).

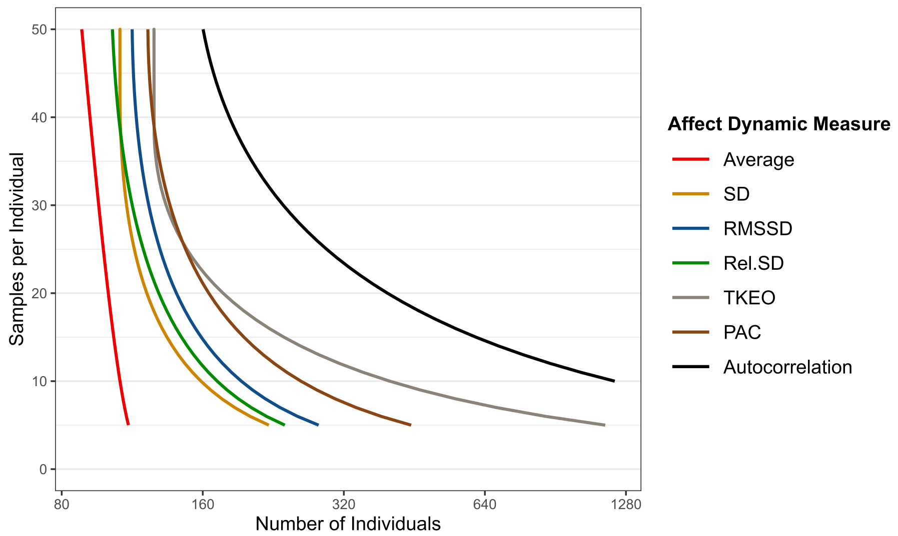


Figure 1.5: Minimum number of individuals and samples per individual required to achieve sufficient power ($\geq$ 80%) to detect a correlation of medium effect size (r = .30) with a two-tail t-test and an alpha of 0.05. The x-axis is in log-scale.

Adequate power could be achieved with a relatively small number of observations per individual. As a general rule, as long as a study includes at least 200 participants, sampling 20 observations per individual yields sufficient power for most affect dynamics measures. For average affect, standard deviation, and relative standard deviation, sufficient power was even achieved with 5 to 10 observations for 200 individuals. For measures of affect instability (i.e., RMSSD, PAC, and TKEO), 20 observations for 200 individuals were required. The only exception to the 200 $\times$ 20 rule arises with affect inertia (i.e., autocorrelation coefficient), for which over 40 observations for 200 individuals were required. It is important to note that these sample recommendations apply to studies with an expected medium-size association of interest (r = .30). However, as our plausible effect sizes benchmarks suggest, many affect dynamics measures display relatively weak associations with demographic, well-being, and time-allocation outcomes (see next section).

Overall, averaging across the range of all sampling combinations, affect dynamics measures, alpha levels, and effect sizes, increasing the number of individuals had a larger impact on power than increasing the number of observations per individual - with the exception of affect inertia which showed the opposite pattern (see Note 4 of Supplementary Materials).

1.3.4. Benchmarks for Plausible Effect Sizes

In power calculation, researchers are asked to anticipate the effect sizes of their associations of interest or to decide on a minimum effect size that they are willing to detect. How can one know in advance what plausible effect sizes might be? Figure 1.6 displays the magnitude of the associations between affect dynamics measures and ten outcomes: (1) age, (2) gender, (3) average sleep time, (4) life satisfaction, (5) meaning in life, as well as the proportion of time spent with (6) friends, (7) family, (8) alone, (9) working, and (10) exercising. These values can be used as broad benchmarks when attempting to postulate plausible effect sizes (see Supplementary Note 5 for additional information and results). For example, researchers interested in examining the relationship between average affect and the propensity to eat carrots could ask themselves whether they expect this relationship to be smaller or greater than the link between average affect and age (r = .06), time spent alone (r = -.24), or trait meaning in life (r = .84). Likewise, researchers interested in examining the relationship between affect instability and family history of bipolar disorder could ask themselves whether they expect the relationship to be smaller or greater than the link between affect instability and time spent with friends (r = .10), age (r = -.27), or life satisfaction instability (r = .34). In practice, researchers should not exclusively rely on these benchmark effect sizes to establish an expected effect size but consider information from different sources (including meta-analytic evidence, preliminary results, or past literature). These benchmarks thus provide a useful complementary source of information to help defining an expected effect size.

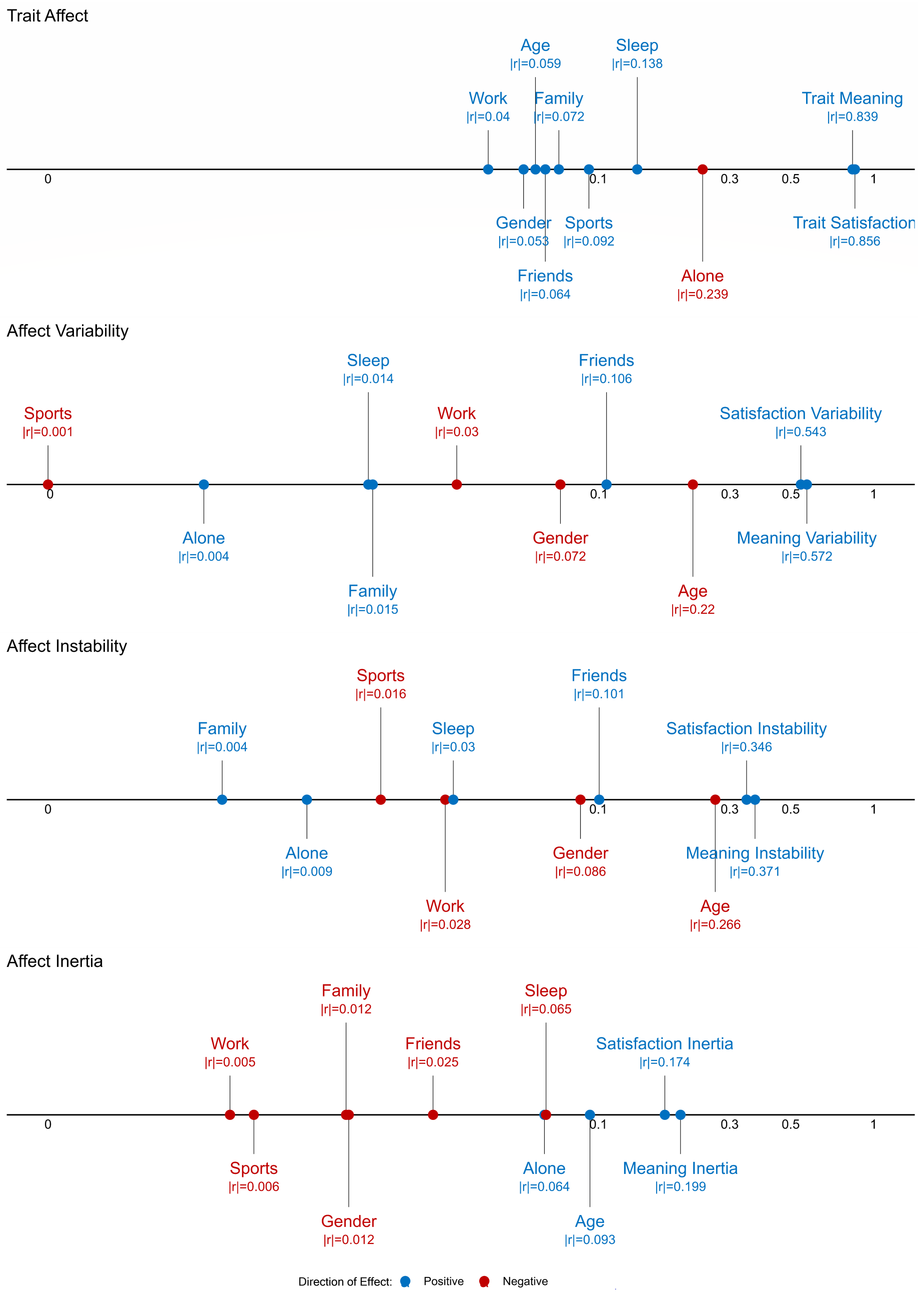


Figure 1.6: Correlations between affect dynamics measures and different outcome variables in our dataset. Positive and negative correlations are presented in blue and red, respectively.

1.3.5. R-package and Online Power Calculator

Building on our results and expanding our power calculations to all effect sizes, we developed an R-package ("powerAD") and a Shiny App [1] to help researchers make empirically-informed decisions about study design of affect dynamics studies. We refer to the package site [2] for more information on how to download, install and run its primary functions.

Our Shiny app is composed of two main panels. On the first panel ("Sampling Calculator"), users can estimate a set of valid sampling approaches for each affect dynamics measure given a specified statistical power, effect size, and alpha level. On the second panel ("Power Calculator"), users can estimate the statistical power achieved by a specific study based on its characteristics (sampling approach, affect dynamics measure, effect size, and alpha level). For example, Figure 1.7 shows the minimal combinations of number of individuals and number of observations per individual to obtain a statistical power of 80% to detect an r = .30 at the 5% significance level for the Taeger Kaiser Energy Operator (TKEO). Figure 1.8 provides the precise power estimate for the same r = .30 effect size and TKEO measure given a specific sample of 400 participants each surveyed 11 times. Finally, the app also provides a series of benchmark effect sizes for each affect dynamics measures to help researchers estimate plausible effect sizes.

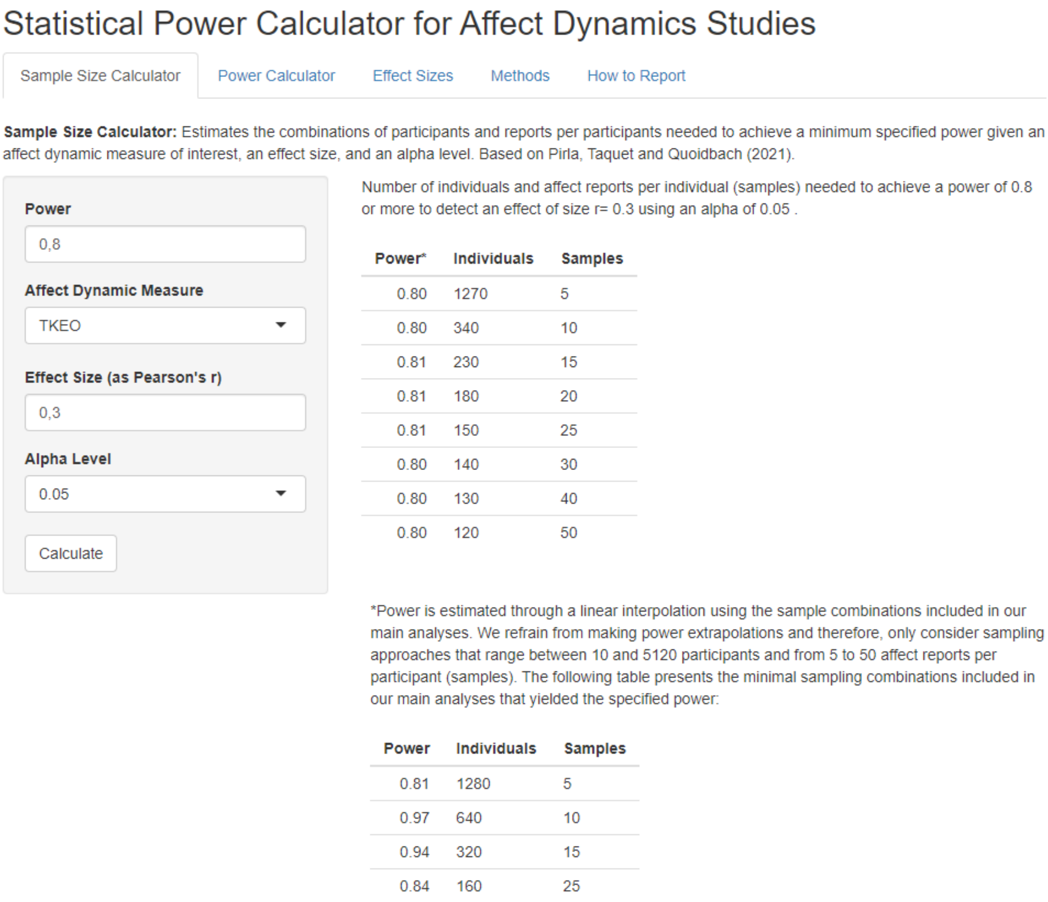
1.4. Discussion

This paper introduces an empirically-derived framework to help researchers design well-powered and efficient experience-sampling studies in the growing field of affect dynamics. To illustrate the value of this contribution, imagine that a group of researchers want to design an ESM study examining the association between affect variability and burnout risk. Using the online tool ("Effect sizes" tab) they anticipate that the effect size should be in the same ballpark as the relationship between affect variability and average life satisfaction (which, using our benchmarks, they observe to be r = .20). Using the "Sample Size Calculator" tab and setting the power to 0.80, the effect size to 0.20, and the alpha level to 0.05, they notice that they have a range of options to achieve this power. For instance, they could recruit 240 participants and collect 40 affect records from each or they could recruit 510 participants and collect 5 affect records from each. Because they are mindful that retention of participants can be an issue, they opt for the latter option.

[1] https://sergiopirla.shinyapps.io/powerADapp/

[2] https://sergiopirla.github.io/powerAD

Panel A:



Statistical Power Calculator for Affect Dynamics Studies

Sample Size Calculator | Power Calculator | Effect Sizes | Methods | How to Report

Sample Size Calculator: Estimates the combinations of participants and reports per participants needed to achieve a minimum specified power given an affect dynamic measure of interest, an effect size, and an alpha level. Based on Pirla, Taquet and Quoidbach (2021).

Power
0,8

Affect Dynamic Measure
TKEO

Effect Size (as Pearson's r)
0,3

Alpha Level
0.05

Calculate

Number of individuals and affect reports per individual (samples) needed to achieve a power of 0.8 or more to detect an effect of size r= 0.3 using an alpha of 0.05 .

Power*	Individuals	Samples
0.80	1270	5
0.80	340	10
0.81	230	15
0.81	180	20
0.81	150	25
0.80	140	30
0.80	130	40
0.80	120	50

*Power is estimated through a linear interpolation using the sample combinations included in our main analyses. We refrain from making power extrapolations and therefore, only consider sampling approaches that range between 10 and 5120 participants and from 5 to 50 affect reports per participant (samples). The following table presents the minimal sampling combinations included in our main analyses that yielded the specified power:

Power	Individuals	Samples
0.81	1280	5
0.97	640	10
0.94	320	15
0.84	160	25

Figure 1.7: Shiny App to calculate power in affect dynamics studies. Panel A: Sample size calculator.

Whereas the ideal sampling approach depends on the specific affect dynamics measure under consideration, three design principles emerge from our research. First, a sample of 200 participants each providing 20 observations (i.e., 200 $\times$ 20 rule) yields sufficient power to detect medium size associations for most affect dynamics measures. Second, the optimal sampling strategy depends on the affect dynamics measure of interest. For trait affect and affect variability, it is often better to run longer less-demanding studies (i.e., few daily measurements spread out over several weeks) than shorter intense ones (i.e., many daily affect measurements spread out over several days). For measures of instability and inertia, both short intense studies and longer less-demanding studies outperform random samples with little difference between the two designs. Third, little differences were

Panel B:

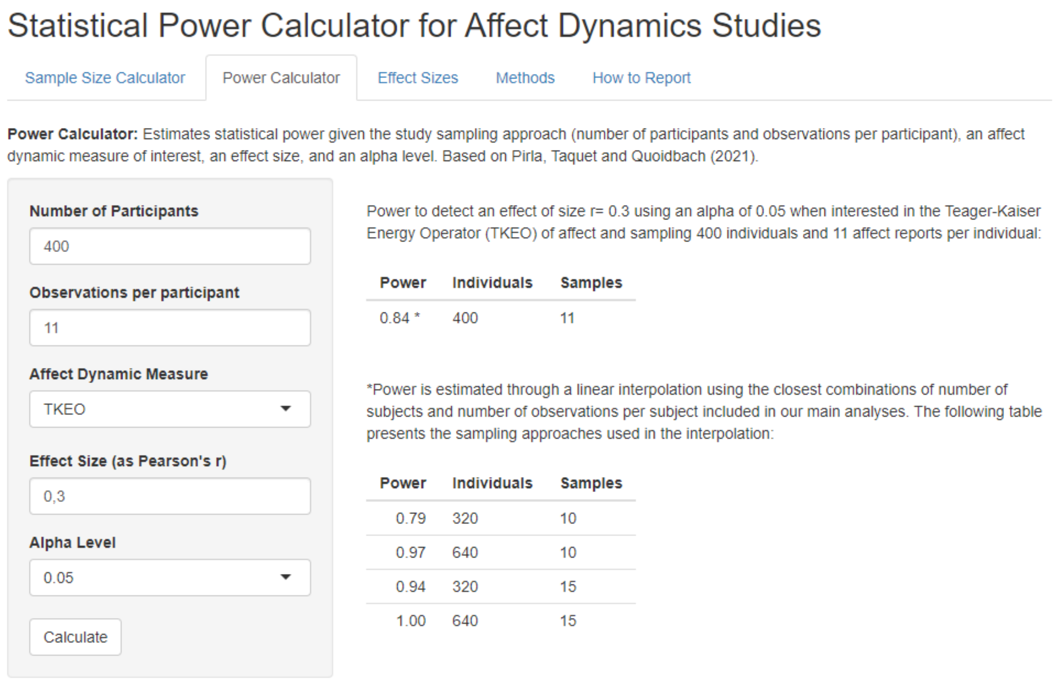


Figure 1.8: Shiny App to calculate power in affect dynamics studies. Panel B: Power calculator.

observed between random sampling and sampling at specific times or on specific days, so that the choice of sampling moments can be dictated by other considerations (such as the individual's preferences or practicalities related to the study at hand).

The present study provides a robust empirical framework to conduct ESM studies in affective science. But it is important for future research to address several limitations. First, our "true" values (i.e., those based on all the available measurements for an individual) were based on at least 50 observations per participant. It might be that more extensive data at the participant level (e.g., 1000 observations per individual) would lead to somewhat different inferences. Second, our recommendation about when researchers should survey participants is limited to relatively basic strategies (e.g., random moments vs. specific days or times). Future research is needed to examine whether advanced context-aware strategies (e.g., sending surveys in response to changes in participants' environmental or psychological circumstances) lead to substantial gains in accuracy and statistical power. Third, although we relied on an exceptionally large sample, our participants may not be representative of the general population. Future research is also needed to examine whether our recommendations need to be adjusted for specific groups of people (e.g., patients with depression, older adults). Fourth, our recommendations are based on accuracy and statistical power considerations. They do not take into account how different sampling strategies may affect burden, compliance, and careless responding in ESM research. Our data did not include information on non-answered notifications, limiting our ability to test the impact of our sampling recommendations on burden and compliance. While recent research suggests that sampling frequency has no impact on participant's burden, data quantity, and data quality (Eisele et al., 2022), further research is needed to examine whether other recommendations derived from our findings are similarly free of negative consequences. Finally, our framework focused on a general, unidimensional measure of affect (unhappy - happy) and the optimal sampling strategies to detect correlations. In future research, it is important to examine how different affect measurements impact estimation precision and statistical power. Further work should also explore how our recommendations apply to other affective states, including specific emotions, mixed-effects models, and non-linear relationships between affect dynamics measures and outcomes. We hope that the data and code provided will allow researchers to expand our framework, opening the door to fast and exciting advances in the study of human emotions.

1.5. Supplementary Materials

1.5.1. Note 1: Sample Summary Statistics

The following tables present the summary statistics for the sample of participants included in our analyses:

Sample Variable	Average	SD	Min	Max
Age	29.97	9.96	13	73
Number of affect reports per participant	111.55	87.79	51	1689

Table 1.2: Sample summary statistics, numeric variables. n=7016.

Sample Variable	**Proportion**
Gender (Female)	73.82%
Gender (Male)	26.18%
Country (France)	92.63%
Country (Switzerland)	4.83%
Country (Belgium)	0.59%
Country (Other)	0.37%

Table 1.3: Sample summary statistics, categorical variables. n=7016.

Sample Variable	**Proportion**
Morning (from 6 am to 12 pm)	31.15%
Afternoon (from 12 pm to 4 pm)	25.16%
Evening (from 4 pm to 8 pm)	27.72%
Night (from 8 pm to 6 am)	15.85%

Table 1.4: Temporal distribution of affect reports.

1.5.2. Note 2: Optimizing Sampling Approaches

In this section we present in greater detail the results obtained when optimizing sampling based on specific times and days. To compare performance across time windows, we followed these steps: For a given individual and number of reports between 3 and 30, we 1) resampled without replacement and estimated affect dynamics measure 1000 times, 2) we applied a debiased step to each of these 1000 estimates, 3) we estimate an individual's Root Mean Square Error (RMSE) for each number of reports based on the full sample "true" estimates of affect dynamics measures, and 4) we average the RMSE across participants for each number of reports used in the computation of the affect dynamics measures. To debias the estimates, we first obtained a condition-specific bias by subtracting from the average of the population of true values of an affect dynamics measure, the average of the population of the same measure estimated with affect reports from our time-window of interest. We then subtracted this bias to each of our estimates of affect dynamics measures obtained from resampling 1000 times. For example, when analyzing the performance of the estimations of the TKEO with reports collected at night, we first obtained a condition-specific bias. To calculate this bias term, we 1) estimated each individual's TKEO in affect using all reports available, 2) estimated each individual's TKEO in affect using all reports collected at night, 3) averaged 1 and 2 across individuals, and 4) subtracted from the average of the population of TKEOs estimated with all reports, the average of the population of TKEOs estimated with the reports collected at night. This debiasing procedure allowed us to account for "condition fixed effects", any constant bias across individuals that did not affect the relative ordering of individuals in terms of their affect dynamics measure of interest. For completeness, we also present the results of the analyses without a debiasing step. Including the debiasing step did not substantially change our results. The results are presented in Figures 1.9-1.12. Figure 1.13 presents the results obtained when considering a sampling strategy that includes at least one observation from each time window. The estimations of performance in this last condition did not include a debiasing step. We followed the same procedure outlined here to compare performance when sampling across specific days. Figures 1.14-1.15 present the results obtained when sampling on specific days.

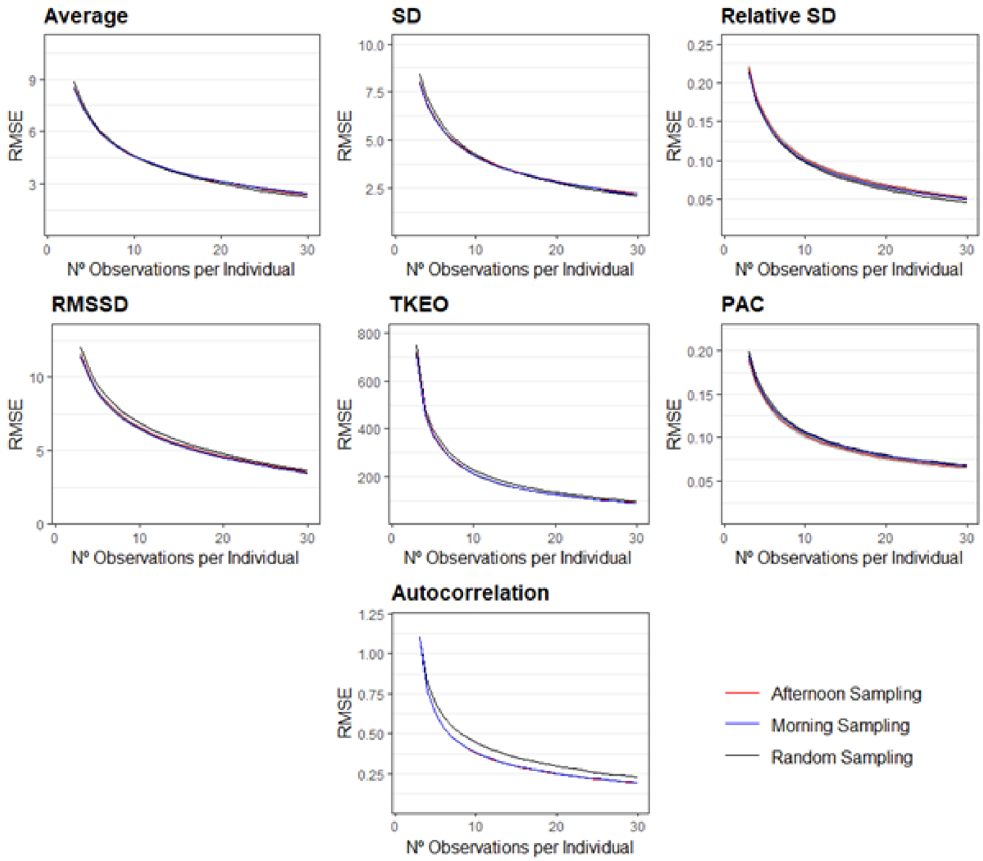


Figure 1.9: Between-subject mean RMSE for affect dynamics measures as a function of time of the day. Random times are depicted in black, morning (6 am to noon) in blue, and afternoon (noon to 4 pm) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated after a debiasing step.

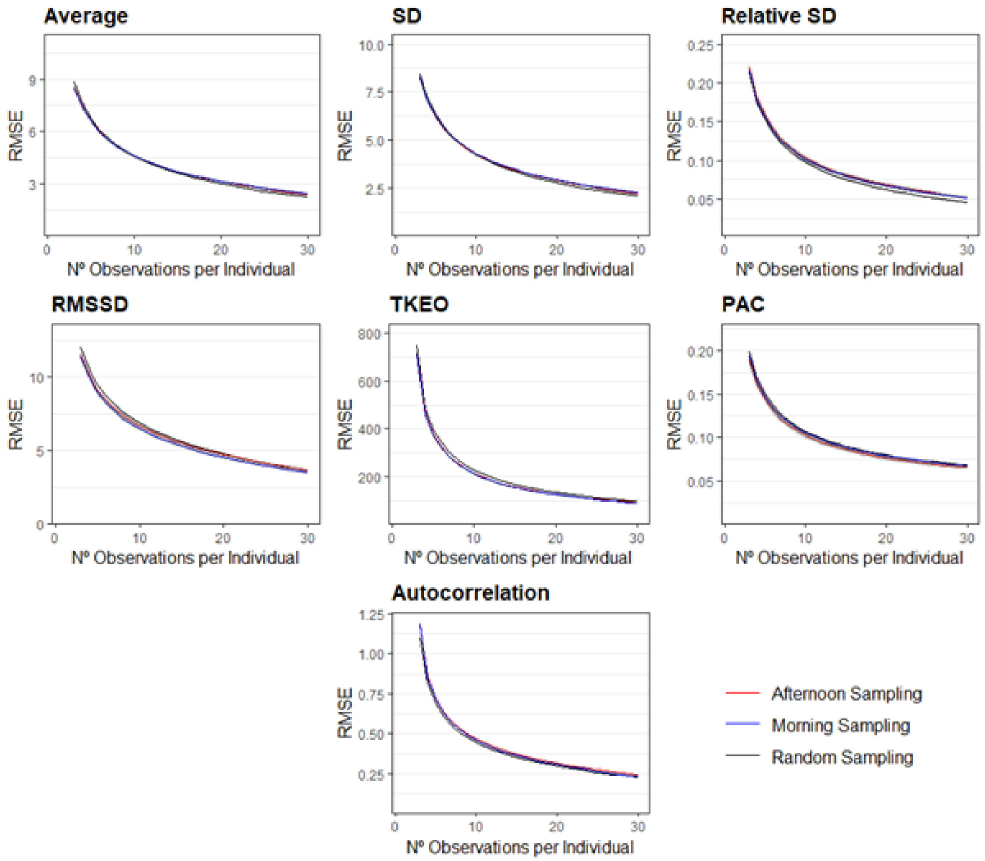


Figure 1.10: Between-subject mean RMSE for affect dynamics measures as a function of time of the day. Random times are depicted in black, morning (6 am to noon) in blue, and afternoon (noon to 4 pm) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated without a debiasing step.

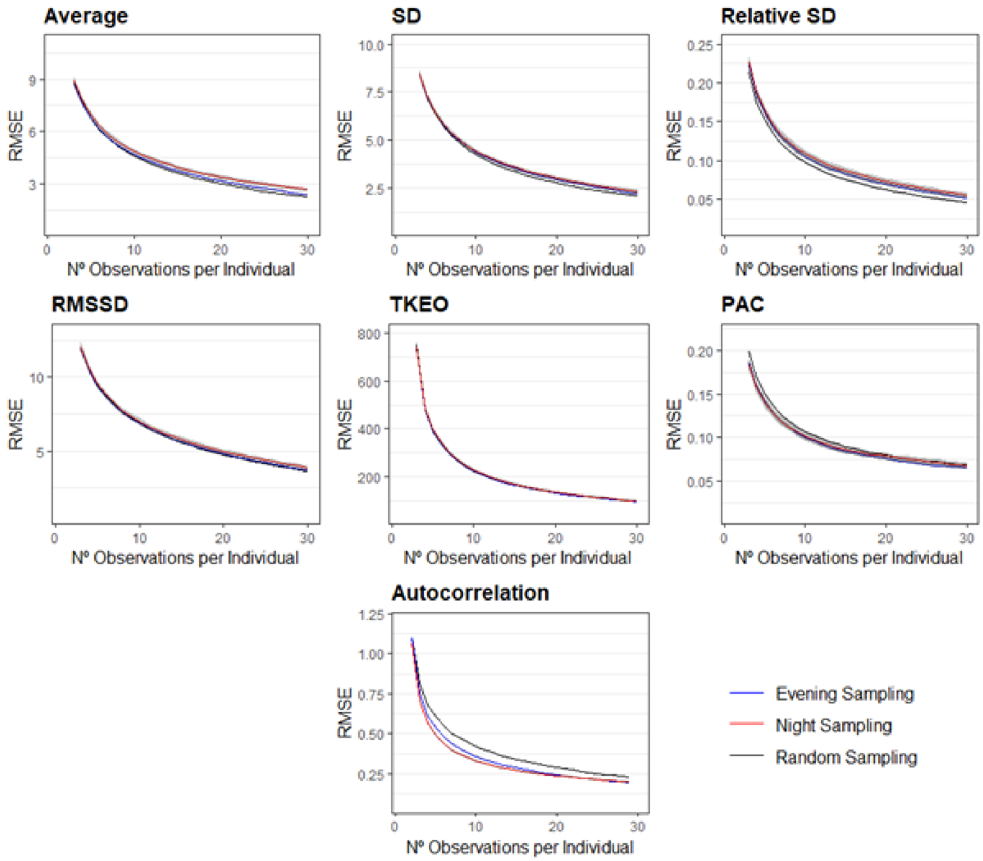


Figure 1.11: Between-subject mean RMSE for affect dynamics measures as a function of time of the day. Random times are depicted in black, evening (4 pm to 8 pm) in blue, and night (8 pm to 6 am) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated after a debiasing step.

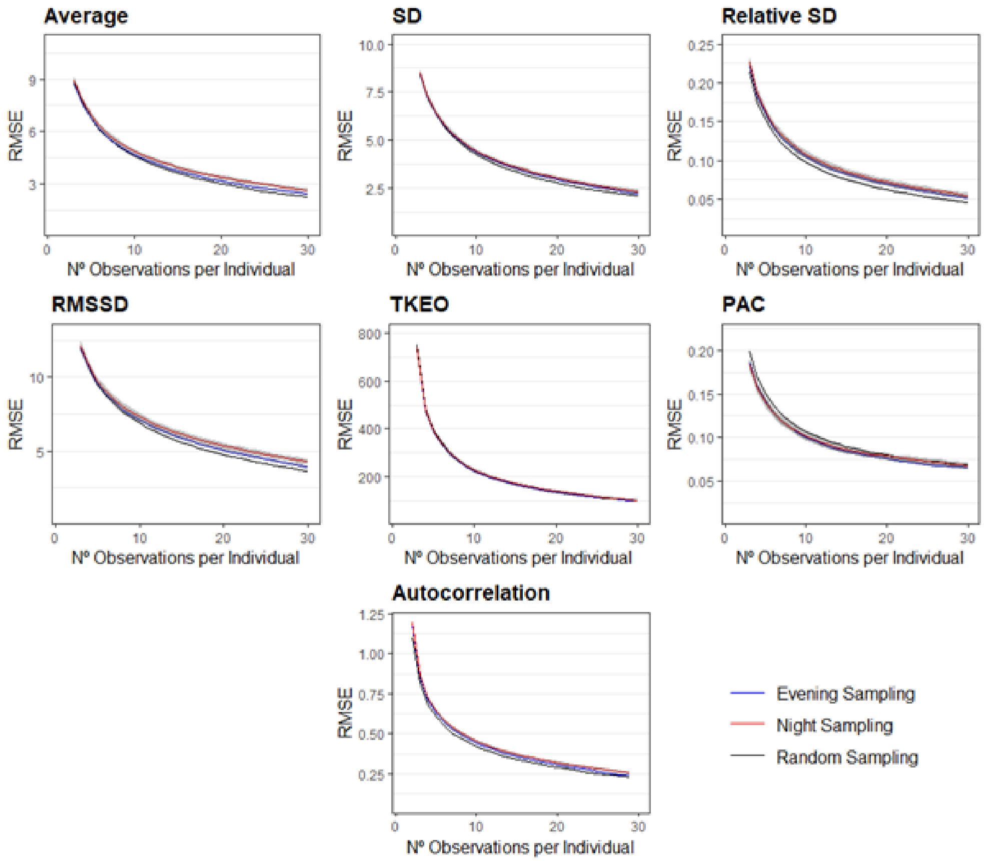


Figure 1.12: Between-subject mean RMSE for affect dynamics measures as a function of time of the day. Random times are depicted in black, evening (4 pm to 8 pm) in blue, and night (8 pm to 6 am) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated without a debiasing step.

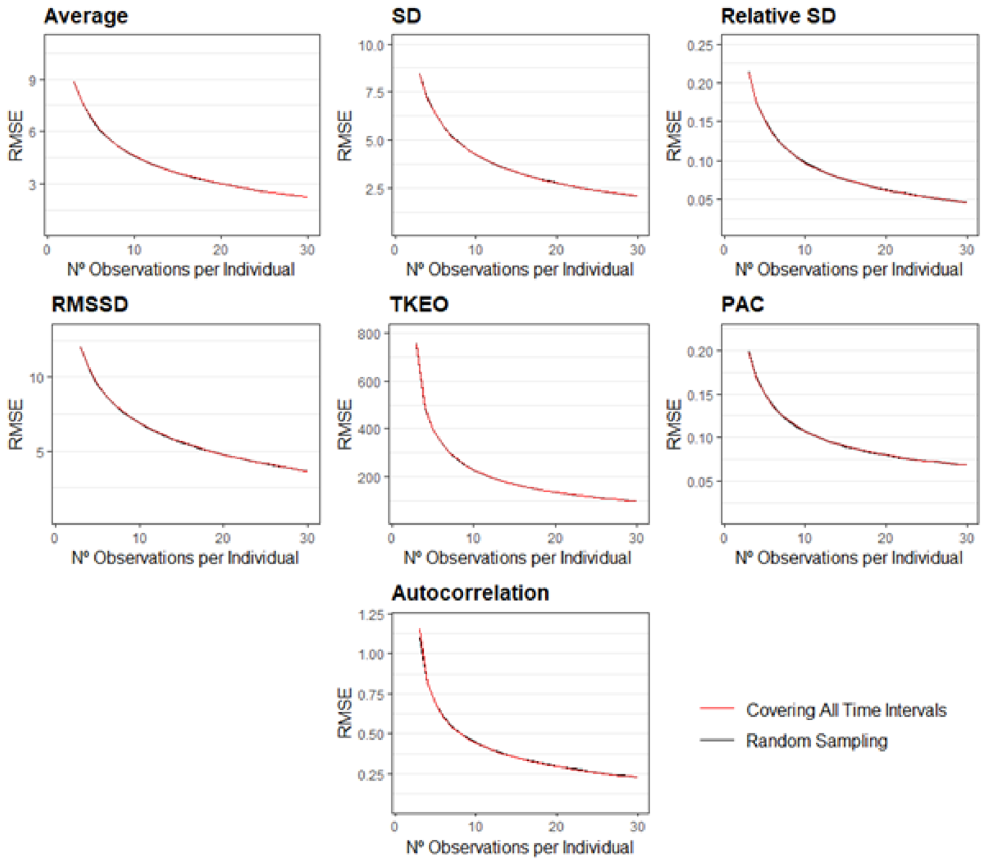


Figure 1.13: Between-subject mean RMSE for affect dynamics measures as a function of time of the day. Random times are depicted in black, and estimates obtained when sampling a minimum of 1 observation from each time interval (morning, afternoon, evening and night) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE.

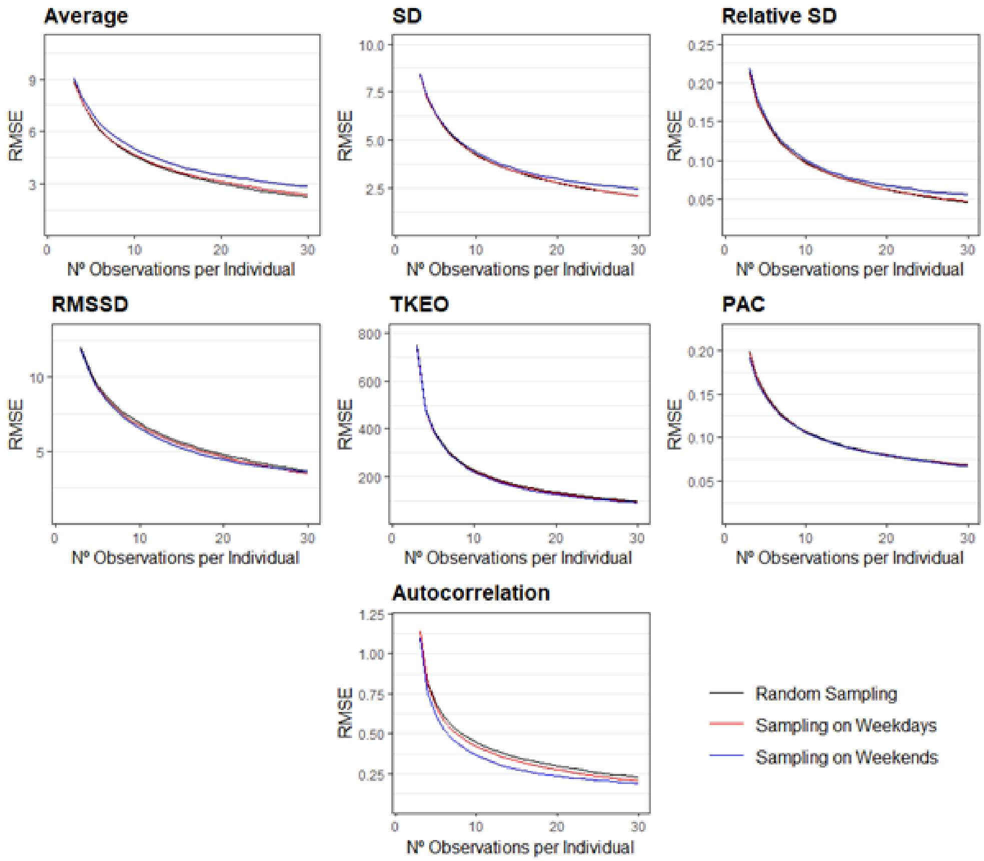


Figure 1.14: Between-subject mean RMSE for affect dynamics measures as a function of day of the week. Random days are depicted in black, weekends in blue, and weekdays (8 pm to 6 am) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated after a debiasing step.

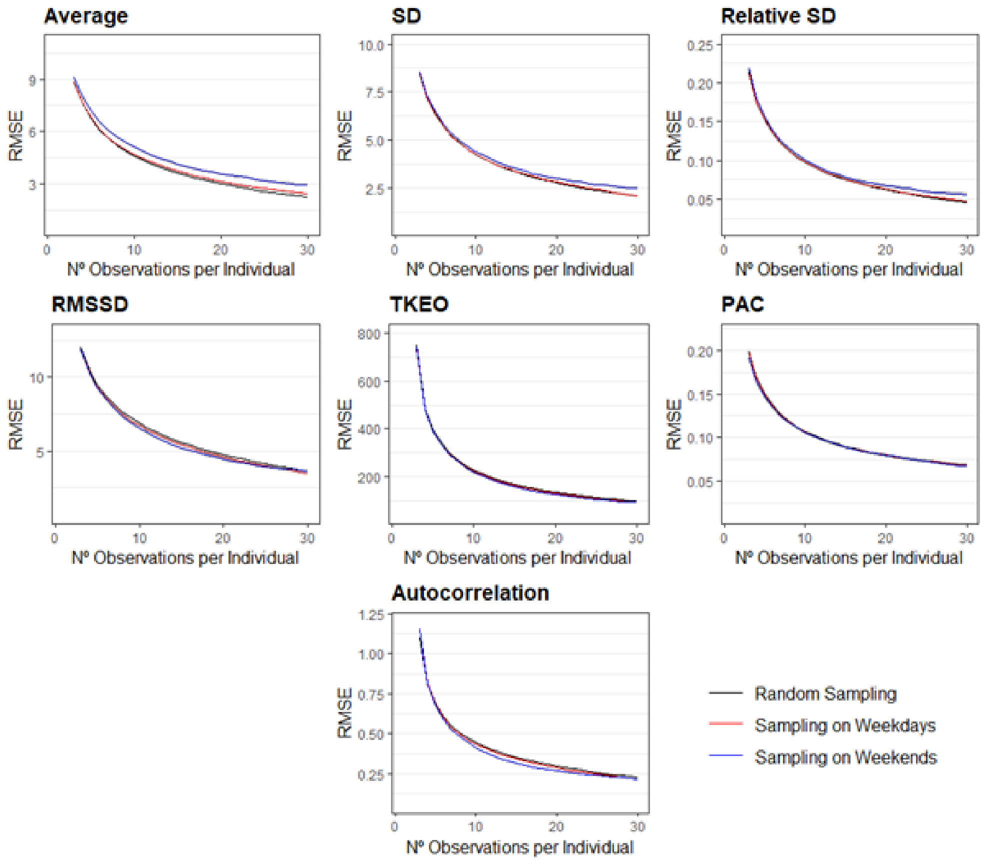


Figure 1.15: Between-subject mean RMSE for affect dynamics measures as a function of day of the week. Random times are depicted in black, weekends in blue, and weekdays (8 pm to 6 am) in red. Gray areas around the lines represent the 95% confidence intervals for the average RMSE. These estimates are calculated without a debiasing step.

1.5.3. Note 3: Statistical Power as a Function of Sampling

In this note, we first detail the procedure followed to obtain Figure 1.5 of the main text and then present empirical power for all sampling combinations included in our study when interested in weak (r = .10), medium-size (r = .30) or strong correlations (r = .50).

Figure 1.5 presents the minimum number of individuals and number of samples per individual for each affect dynamics measure to yield 80% power to detect a medium size relationship (r = .30) with an alpha of 0.05. To estimate the curves, we used the results of the empirical power estimations obtained in our main analyses (presented below in this Note). Using these results, for each affect dynamics measure, we first focused on those analyses that assumed an alpha of 0.05 and a medium size correlation (r = .30). Then, for each number of samples per participant in our analyses (that is, from 5 to 50 in increments of 5), we selected the minimum number of participants that yielded a power larger than or equal to 80%, and the maximum number of participants that yielded a power lower than or equal to 80%. We divided the space between these two numbers of participants into a sequence with increments of 10 participants and linearly interpolated power for each number of participants included in this sequence. Using these interpolated power values, we selected, for each number of samples per participant, the minimum number of individuals with an approximated power larger than or equal to 80%. This process yielded, for each affect dynamics measure, 10 combinations of individuals and number of samples per participants with an approximate power over 80%. As small variations existed in power (approximated power ranged between 80%-83%) and we relied on linear approximations, we encountered some non-monotonic regions - that is, regions where the number of participants needed to achieve a minimum power of 80% did not decrease with the number of samples per participant but displayed minor increases. We directly imposed weak monotonicity by replacing these regions by sequences in which the number of participants remained unchanged as the number of samples per participant was increased. In doing so, for each affect dynamics measure we obtained our 10 final sampling combinations with an approximate power of 80%. Using these 10 combinations, we regressed, for each affect dynamics measure, the logarithm of the number of participants on the number of samples per participant and its logarithm. This specification was selected based on fit - the average R-squared across affect dynamics measures was above 95%. We used the fitted number of participants for each number of samples per participant to draw the curves. Table 1.5 presents the final 10 sampling combinations used to run the regression models for each affect dynamics measure. We abstain from making power extrapolations, and therefore, for affect inertia (“Auto.”), we do not provide the number of individuals needed when sampling affect 5 times from each participant. This number would be well above 5120 participants.

Samples	Number of Individuals						
	Average	SD	Rel. SD	RMSSD	TKEO	PAC	Auto.
5	120	230	260	300	1200	460	-
10	110	150	160	200	360	260	1690
15	110	130	140	160	240	210	810
20	100	130	130	140	180	160	510
25	100	120	120	140	170	160	400
30	90	110	110	130	150	140	290
35	90	110	110	120	140	140	250
40	90	110	110	120	140	130	210
45	90	110	110	120	130	130	190
50	90	100	110	120	130	130	160

Table 1.5: Sampling combinations yielding an approximate power of 80% to detect a medium size association (r = .30) with an alpha of 0.05 for each affect dynamics measure. "Samples" represent the number of samples per participant needed and "Number of Individuals" the total number of participants.

Next, we focused on the results of estimating the empirical power for weak (r = .10) and strong correlations (r = .50). To estimate power, we followed the same procedure presented in the main text. We took a conservative approach and set the type 1 error rate at 0.001.

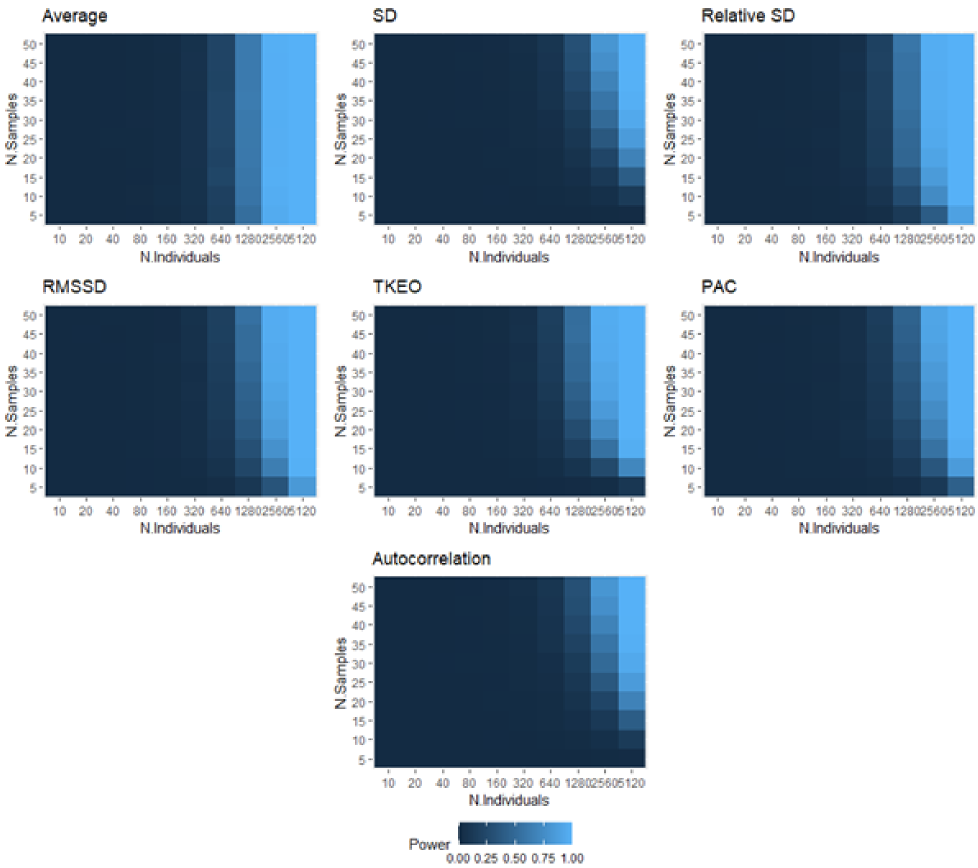


Figure 1.16: Power as a function of sampling strategy. Each panel represents the power to detect a small correlation (r = .10) between an affect dynamics measure and an outcome variable using a two-tailed t-test and an alpha of 0.001.

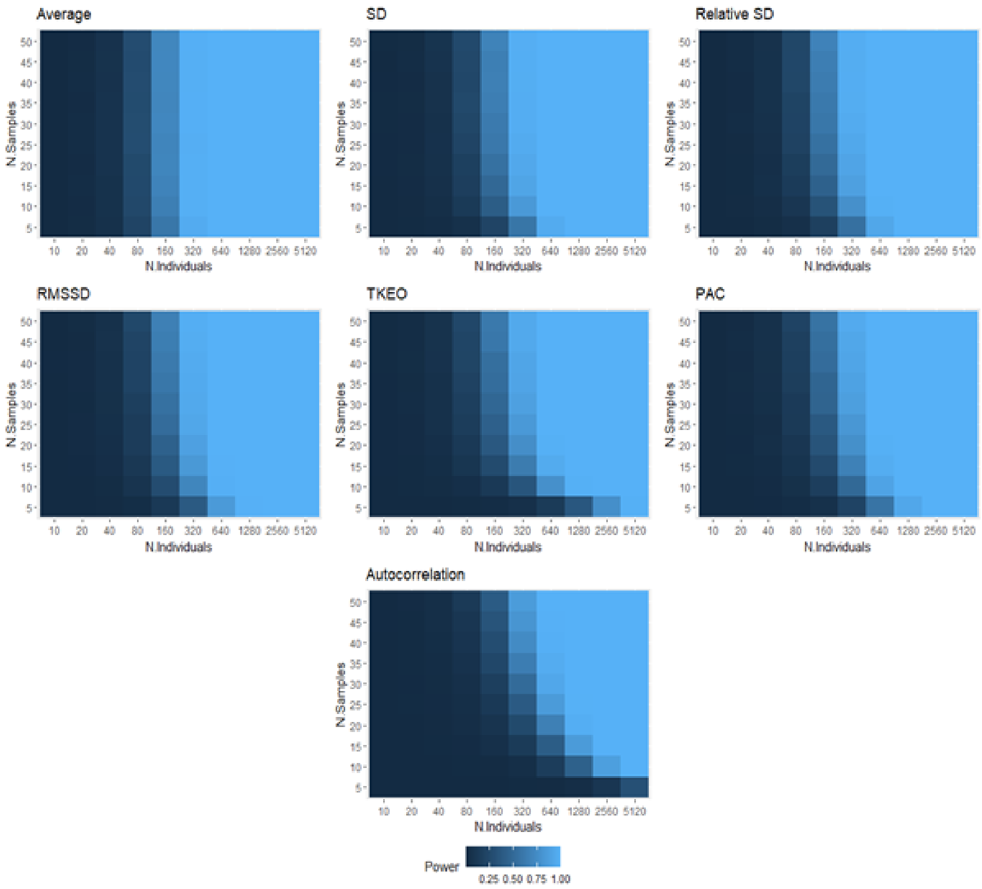


Figure 1.17: Power as a function of sampling strategy. Each panel represents the power to detect a small correlation (r = .30) between an affect dynamics measure and an outcome variable using a two-tailed t-test and an alpha of 0.001.

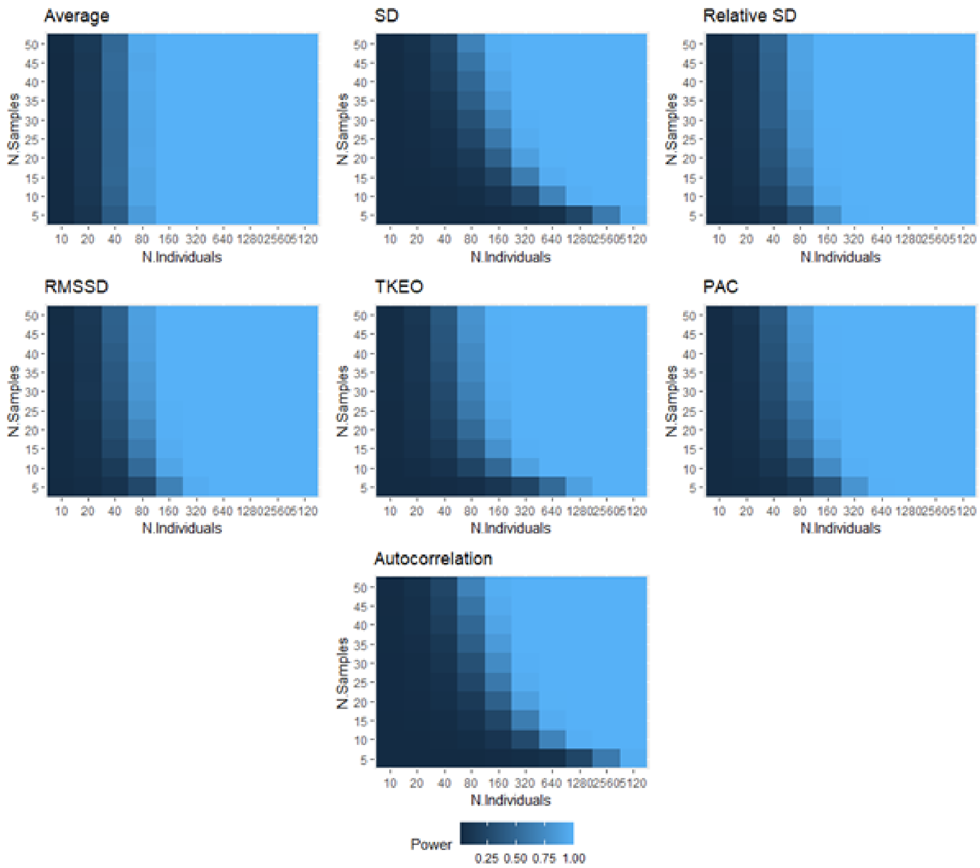


Figure 1.18: Power as a function of sampling strategy. Each panel represents the power to detect a small correlation (r = .50) between an affect dynamics measure and an outcome variable using a two-tailed t-test and an alpha of 0.001.

1.5.4. Note 4: Average Effect of Number of Participants and Samples on Power

In the following tables we present the results of regressing empirical power on the number of individuals and samples per participant for all affect dynamics measures. Note our sample size for each measure is 1500 observations as we include power obtained from our 100 sampling combinations, 3 effect sizes (small, medium and large) and 5 alpha levels (0.1, 0.05, 0.01, 0.005, 0.001).

	Power (All Measures)	Power (Average)	Power (SD)	Power (Rel.SD)
log(Individuals)	0.156***	0.153***	0.157***	0.157***
	(0.001)	(0.003)	(0.003)	(0.003)
log(Samples)	0.080***	0.008	0.045***	0.048***
	(0.004)	(0.010)	(0.010)	(0.010)
Constant	-0.525***	-0.211***	-0.379***	-0.394***
	(0.014)	(0.037)	(0.037)	(0.037)
Observations	10500	1500	1500	1500
R2	0.567	0.562	0.583	0.585

Table 1.6: OLS estimates of the effect of number of individuals and samples per participant on statistical power across effect sizes and significance levels. Standard errors are in parentheses. Statistical significance: *$p<0.1$; **$p<0.05$; ***$p<0.01$

	Power (RMSSD)	Power (TKEO)	Power (PAC)	Power (Auto.)
log(Individuals)	0.160***	0.158***	0.160***	0.147***
	(0.003)	(0.003)	(0.003)	(0.003)
log(Samples)	0.058***	0.128***	0.079***	0.196***
	(0.010)	(0.010)	(0.010)	(0.010)
Constant	-0.454***	-0.709***	-0.551***	-0.974***
	(0.036)	(0.036)	(0.037)	(0.036)
Observations	1500	1500	1500	1500
R2	0.596	0.612	0.596	0.598

Table 1.7: OLS estimates of the effect of number of individuals and samples per participant on statistical power across effect sizes and significance levels. Standard errors are in parentheses. Statistical significance: *p<0.1; **p<0.05; ***p<0.01

1.5.5. Note 5: Plausible Effect Sizes

We collected our data using "58 seconds", a free smartphone application. At sign-up, the participants answered a few questions regarding demographic information. Using such information, we coded the variable "Gender" to take a value of 1 for male participants and 0 for female participants. The numeric variable "Age" represents a participant's age in years. Data on life meaning, satisfaction, sleep and proportion of time spent with different groups of people or engaging in different activities was collected using a system of random notifications. Participants using the app received questionnaire prompts at random times of the day. These questionnaires consisted of 4 to 6 questions from a large battery of items. For meaning in life, the participants were asked to rate the following statement from 0 ("Not at all") to 100 ("Absolutely"): "Here and now I feel like I'm living a meaningful life". Similarly, for life satisfaction, participants provide a rating from 0 (Dissatisfied with my life) to 100 (Satisfied with my life) to the following statement: "Here and now, I feel ...". For life meaning as for life satisfaction, we used these numeric reports to estimate the seven dynamic measures included in this paper (Average, SD, Rel. SD, RMSSD, TKEO, PAC and Autocorrelation). Sleep was measured by asking participants the amount of sleep hours they had last night. Participants provided a numeric report ranging from 0 to 15, and we averaged across sleep reports to obtain an individual's average hours of sleep. Finally, some questionnaires included a list of activities and a list of groups of people. The participants facing these lists were asked to select all activities that they were doing before answering the questionnaire and select all the groups of people with whom they were when answering the questionnaire. For simplicity, we restricted our attention to the proportion of time an individual spent with family, friends, alone, studying or working, and exercising. To obtain proxies for the amount of time that an individual spent with these people or doing these activities, we estimated the proportion of times (out of all times that the participant was presented with each list) that the user reported being with the specific group of people or doing a specific activity.

For each participant, aside from these variables, we estimated the seven affect dynamics measures included in our main analyses using the full sample of affect reports at our disposal. To provide the reader with effect sizes to serve as reference, we estimated the Pearson's r coefficient between each affect dynamics measure and the demographic, well-being and time allocation variables described in the previous paragraph. The resulting coefficients are presented in Table 1.8-1.9.

Outcome	Average	SD	Rel. SD	RMSSD
Age	0.059	-0.248	-0.191	-0.282
Gender	0.053	-0.098	-0.047	-0.093
Average Meaning in life	0.839	-0.231	0.096	-0.201
SD Meaning in life	0.019	0.520	0.431	0.468
Rel.SD Meaning in life	0.144	0.414	0.623	0.365
RMSSD Meaning in life	0.014	0.472	0.387	0.469
TKEO Meaning in life	0.039	0.295	0.280	0.282
PAC Meaning in life	0.040	0.381	0.355	0.387
Auto. Meaning in life	0.017	0.016	0.026	-0.076
Average Life satisfaction	0.856	-0.197	0.109	-0.144
SD Life satisfaction	-0.076	0.508	0.393	0.439
Rel.SD Life satisfaction	0.141	0.363	0.579	0.323
RMSSD Life satisfaction	-0.087	0.471	0.353	0.447
TKEO Life satisfaction	0.003	0.251	0.227	0.231
PAC Life satisfaction	-0.034	0.375	0.325	0.366
Auto. Life satisfaction	0.054	-0.007	0.029	-0.080
Average sleep	0.138	0.001	0.028	0.036
Time spent alone	-0.239	0.033	-0.026	0.002
Time spent with family	0.072	0.001	0.030	0.008
Time spent with friends	0.064	0.100	0.113	0.105
Time spent at work/studying	0.040	-0.033	-0.027	-0.029
Time spent exercising	0.092	-0.018	0.016	-0.015

Table 1.8: Effect sizes (Pearson's r) of the correlation between different outcomes and measures of affect. Outcomes correspond to demographic variables, variables estimating the propensity of individuals to perform an activity or being in the presence of others or measures of life satisfaction, meaning and sleep. The correlations were estimated using our full sample of 7016 individuals each providing a minimum of 50 affect reports.

Outcome	TKEO	PAC	Auto.
Age	-0.247	-0.269	0.093
Gender	-0.081	-0.085	-0.012
Average Meaning in life	-0.194	-0.236	-0.029
SD Meaning in life	0.452	0.413	0.055
Rel.SD Meaning in life	0.373	0.318	0.059
RMSSD Meaning in life	0.448	0.422	-0.031
TKEO Meaning in life	0.284	0.249	0.005
PAC Meaning in life	0.383	0.360	-0.041
Auto. Meaning in life	-0.058	-0.082	0.199
Average Life satisfaction	-0.144	-0.181	-0.080
SD Life satisfaction	0.428	0.408	0.096
Rel.SD Life satisfaction	0.329	0.286	0.053
RMSSD Life satisfaction	0.431	0.420	0.013
TKEO Life satisfaction	0.241	0.208	0.021
PAC Life satisfaction	0.364	0.349	-0.014
Auto. Life satisfaction	-0.061	-0.079	0.174
Average sleep	0.023	0.030	-0.065
Time spent alone	0.015	0.009	0.064
Time spent with family	0.004	0.001	-0.012
Time spent with friends	0.091	0.106	-0.025
Time spent at work/studying	-0.031	-0.023	-0.005
Time spent exercising	-0.017	-0.017	-0.006

Table 1.9: Effect sizes (Pearson's r) of the correlation between different outcomes and measures of affect. Outcomes correspond to demographic variables, variables estimating the propensity of individuals to perform an activity or being in the presence of others or measures of life satisfaction, meaning and sleep. The correlations were estimated using our full sample of 7016 individuals each providing a minimum of 50 affect reports.

Chapter 2

HAPPINESS WITHOUT A FINANCIAL SAFETY NET: LOW INCOME PREDICTS EMOTIONAL VOLATILITY

Joint with Jordi Quoidbach

Abstract

Decades of research suggest that money buys very little happiness. However, previous studies have relied on static measures assessing people's well-being once or on average. We examine the "reel" of people's emotional lives through over 1 million reports from 23,000 individuals whose happiness was tracked in real-time using a smartphone app. Results show that lower income is associated with increased happiness volatility - a relationship that replicates across multiple operationalizations of volatility, statistical models, and a sample of individuals from six developing countries (N $>$ 25,000). An unsupervised anomaly detection algorithm further revealed that the greatest gap is between how frequent and intense the rich and the poor experience emotional downs, not ups. The happiness gap between the highest and lowest earners during episodes of intense unhappiness was 1.5 to 3 times the size of the gap in average happiness between these two groups. Finally, exploiting the exogeneity of monthly payments, we find that low-income earners experience more moments and periods of anomalous happiness the last few days of the month, suggesting a causal relationship between income and happiness volatility.

2.1. Introduction

Global poverty is rising for the first time in over 20 years due to the triple threat of COVID-19, conflict, and climate change (Lakner et al., 2021). How will this affect the well-being of the 120 million “new poor” around the world? If there’s a silver lining, it’s in the decades of scholarly research suggesting that money buys very little happiness (Aknin et al., 2009; Boyce et al., 2017; Frey and Stutzer, 2002; Kahneman et al., 2006). For example, the most recent meta-analyses suggest that income explains only 1% to 5% of how happy people feel overall (Jantsch and Veenhoven, 2019; Tan et al., 2020).

Yet, poverty has led to both individual and mass protests around the world. Have the hundreds of studies asking whether money buys happiness overlooked an important side of the question? One commonality shared by previous studies, whether they rely on global evaluative measures (e.g., “how happy do you feel in general?”) or momentary measures of affect in situ (e.g., “how happy do you feel right now?”), is that they solely capture people’s happiness once or on average. But a poor person reporting that they are only a percent less happy than a rich person overall doesn’t mean their day-to-day emotional experiences are the same. When you’re struggling to stay afloat, even regular events like paying your phone bill or rent can cause you to sink - not to mention drowning when catastrophes strike (Daly and Kelly, 2015; Morduch, 1994). The relationship between money and happiness may be less about general happiness than how much happiness fluctuates. And on an emotional rollercoaster, there could be moments of acute suffering or even extended periods of distress which could easily be missed when taking a snapshot of a person’s emotional life.

We shouldn’t overlook if the poor are afflicted by frequent emotional dips and crashes triggered by events that others cruise past. Hundreds of studies in psychology, psychiatry, and medicine show that emotional volatility is a key feature of bipolar, depressive, and anxiety disorders (aan het Rot et al., 2012; Anestis et al., 2010; Bowen et al., 2004; Ebner-Priemer et al., 2007; Golier et al., 2001; Houben et al., 2015; Jones et al., 2005; Koval et al., 2013; Kuppens, 2015; Kuppens and Verduyn, 2017; McConville and Cooper, 1996; Pfaltz et al., 2010; Santangelo, Reinhard, et al., 2014; Servaas et al., 2017; Snir et al., 2017; Thompson et al., 2012; Zeigler–Hill and Abraham, 2006) — all leading causes of disability worldwide and major contributors to the global burden of disease (Murray et al., 2012; Vos et al., 2016). Emotional volatility can also impose a severe tax on a person’s physical health (Hardy and Segerstrom, 2017; Jenkins et al., 2018; Koval et al., 2013) - for a review, see Houben et al. (2015). For example, emotional volatility is related to increases in cardiac conditions (Chan et al., 2016). Moreover, relatively rare moments of severe emotional distress can foster behaviors that have far-reaching consequences, from overeating, substance abuse, and gambling (Ciccarelli et al., 2017; Hull et al., 1986; Masheb and Grilo, 2006) to self-harm, aggression, and violence (Berkowitz, 1989; Gratz, 2003).

A number of policy initiatives call for better measures of well-being as a means of enhancing policies that improve people's lives (OECD, 2011; UN, 2012). Here, we respond to these calls by addressing the limitations of static measures in a comprehensive study of emotional dynamics among people from various income brackets. First, we examine the relationship between income and happiness volatility in a sample of over 23,000 people whose happiness was tracked in real-time for several weeks using a smartphone app. We then corroborated our results by comparing happiness volatility in an independent data set of 25,634 people from six developing countries obtained from the World Health Organization Study on Global Aging and Adult Health (WHO SAGE). Going beyond traditional psychometric approaches, we used unsupervised anomaly detection and clustering algorithms to capture the subtle - yet meaningful - ways income may shape our emotional lives. Finally, to provide suggestive evidence for a causal link between income and happiness dynamics, we examined how daily ups and downs change over the month as a function of income.

2.2. Methods

2.2.1. Participants and Experience Sampling.

Participants volunteered for the study by downloading 58 seconds, a free iPhone and Android mobile app designed to measure users' well-being through short questionnaires presented at random times throughout the day. Participants could customize which days of the week, within what time windows, and how many times they wished to receive questionnaire requests (default = 4 questionnaires a day; 7 days a week from 9:00 AM to 10:00 PM). The app then divided each participant's day into as many intervals as the number of requested questionnaires and chose a random time within each interval-setting a minimum of 1 hour between two questionnaires to avoid large artifactual autocorrelations. We ensured random sampling through a notification system that did not require an internet connection. The app generated notifications at new random times each day, independently for each participant.

Power analyses for affect dynamics time series (see Chapter 1) revealed that ten happiness observations per participant would ensure reliable estimates across different operationalizations of happiness volatility (see hereafter). Therefore, our study focuses on 23,471 users who completed at least ten happiness reports. On average, these participants each provided 50.8 happiness observations, for a total of 1,191,912 observations. In line with previous research, we excluded from our primary analyses individuals whose income figure could not be reliably measured because they selected the lowest ("no income") and highest ("over 7500 euros") categories, respectively, leaving a final sample of 17,278 individuals, each providing an average 52.2 happiness observations (901,816 observations in total). Note that results on the entire sample of 23,471 people, treating income as an ordinal variable, yielded identical results (see SM Note 4).

2.2.2. Happiness.

After accepting a questionnaire request, participants were presented with four to six questions drawn from an extensive battery of items - see Quoidbach et al. (2019). Here, our focal measure was a general happiness item ("How do you currently feel?"; answered on a slider from 0 "very unhappy" to 100 "very happy").

2.2.3. Income and demographics.

In addition to the repeated happiness item, participants were asked different demographic questions (once), including age, gender, country of residence, profession type, and monthly income after taxes (asked on a 13-point bracket scale, see Table 2.1 - 2.2 for detailed information on demographics and income distribution). The non-response rates were high for profession (44%) and income (75%). Because missingness of income data is typically related to key personal characteristics, including financial and health status, focusing only on complete-case analysis can introduce important biases (Schenker et al., 2006). To handle missing data on profession and income, we performed an imputation by random forests using the MissRanger R-package (Mayer, 2019). Note that the direction and statistical significance of all the results we report do not change if we only focus on the subset of complete data or use an alternative hot deck imputation method (see SM Note 4).

2.2.4. Measuring Happiness Volatility.

There are five main operationalizations of affect volatility in the literature (see SM Note 2 for formal definitions). First, one can focus on overall affect variability using the within-participant standard deviation (iSD). While iSD is the most widely-used metric (Rocke et al., 2009), research demonstrates that variability in a construct can be dependent on mean levels of the same construct, especially when measurements are bounded within scales (Mestdagh et al., 2018). That is, a person with a mean happiness level of 10 (or a mean of 90) cannot display as much variability as somebody with a mean of 50, since the scores of the latter individual are less constrained by the scale boundaries. To avoid confound with the mean, variability can be measured through a mean-adjusted version of the within-person standard deviation (Relative iSD) that takes into account the maximum possible variance given an observed mean and scale endpoints (Mestdagh et al., 2018). Second, one can focus on affect instability from one moment to the next, using the Root Mean Square of Successive Differences (RMSSD), the Probability of Acute Change (PAC), and the Traeger-Kaiser Energy Operator (TKEO).

While each of these measures is designed to capture unique dynamical aspects of our emotional life, recent research shows considerable interdependencies between them (Dejonckheere et al., 2019). Therefore, for parsimony, we mainly report results for happiness volatility using within-person standard deviations, the most common and straightforward metric. Note that results were virtually identical for all other operationalizations of happiness volatility (see SM Note 3).

2.2.5. Estimating the Income - Happiness Relationships.

Several studies suggest that the relationship between money and happiness is not linear. Therefore, we examined how income is related to both averages and volatility in happiness through Generalized Additive Models (GAMs; Hastie and Tibshirani, 1987) using the mgcv package for R (Wood, 2003). These models allow fitting data with smooths, or splines, which are functions that can take on a wide variety of shapes. GAMs provide more flexibility than polynomial transformations in the GLM framework (Wood and Augustin, 2002) and limit the risk of false positives through a parsimonious automatic model selection process (Mckeown and Sneddon, 2014). For completeness, we also examined the income - happiness relationships using two-lines tests, estimating separate regression lines for low and high values of income based on the Robin Hood algorithm (Simonsohn, 2018). The two-lines tests yielded identical conclusions to GAMs analyses (see SM Note 3).

2.2.6. Robustness and Specification Curves.

We performed a specification curve analysis to ensure the robustness of our main finding that income is associated with happiness variability (Simonsohn et al., 2020). In this specification curve, we consider five different operationalizations of happiness volatility and three methods to deal with missing data (imputation by random forest, imputation by hot deck, and removing data from individuals with missing data). We considered income either as a continuous or categorical ordinal variable, with specifications including and excluding data from individuals who selected the highest income response ("more than 7500 euros/month") and imputing their income as 9000 euros per month. For specifications including income as an ordinal variable, we also examined the impact of including or excluding individuals who reported no income. Finally, we considered the effect of adding or removing demographic control variables from the models (Age, Gender, Country). To make the comparison of coefficients possible across specifications, we standardized our dependent variables. In total, we included 180 different specifications.

2.2.7. Unsupervised Collective and Point Anomaly Detection.

Classic models of affect dynamics assume that people's emotional lives can be summarized through a series of parametric measures (e.g., mean, variance, probability of acute change). However, these summary statistics often mask the complexity of human emotional life. To paint a detailed and complex picture of how income relates to everyday happiness, we used a Collective and Point Anomaly Detection method (Fisch et al., 2019). That is, we used a nonparametric penalty-based approach to identify happiness reports that are anomalous given an individual's happiness time series. To ensure reliable estimates, we focused our analyses on a subsample of 5002 participants who provided a minimum of 50 happiness observations in our mobile app study (see SM Note 6 for details).

2.3. Results

To set the stage for our primary analyses, we first examined the relationship between the logarithm of people's monthly income after taxes and their average happiness. We found a small association $(r = .075, p < .0001)$. Consistent with other studies (Jebb et al., 2018; Kahneman and Deaton, 2010; Killingsworth, 2021), Generalized Additive Models (GAMs; Hastie and Tibshirani, 1987), suggest that the relationship between income and average happiness was weaker at higher levels of income $(edf = 3.754$, $p < .0001$, deviance explained $= 0.6\%)$. On average, doubling one's salary from 1000 to 2000 euros per month was associated with a 0.08 SD increase in mean happiness (1.4 points out of 100, Fig. 2.1; left panel). In contrast, doubling a person's salary from 3000 to 6000 euros per month was associated with a 0.07 SD increase in happiness (1.2 points). This pattern remained unchanged when controlling for age, gender, and country-specific fixed effects, as well as when using raw income in euros and categorical income ranks (see Specification Curve in SM Note 4).

Does money relate to happiness beyond how people feel on average? As shown in Fig. 2.1 (right panel), the logarithm of monthly income after taxes was negatively related to happiness volatility ($r = -.149$, $p < .0001$), and GAM analyses suggest that the magnitude of this relationship decreased at higher levels of income ($edf = 2.933$, $p < .0001$, deviance explained $= 2.28\%$), reaching a plateau at about 3,300 euros per month. On average, doubling one's salary from 1000 to 2000 euros per month was associated with a reduction in happiness volatility of 0.2 SD. In contrast, doubling a person's salary from 3000 to 6000 euros per month was virtually unrelated to happiness volatility changes (a decrease of 0.015 SD). Again, this pattern remained unchanged when controlling for age, gender, country-specific fixed effects, raw income, and categorical income ranks. The relationship also held for all major operationalizations of happiness volatility, including standard deviations, probability of acute changes (PAC), root mean successive square differences (RMSSD), and Teager-Kaiser energy operator (TKEO) - see Specification Curve in SM Note 4. Importantly, the relationship between income and happiness volatility remains significant when controlling for average happiness ($\beta = -.138$, $p < .0001$; $edf = 2.889$, $p < .0001$) and when measuring happiness volatility using relative standard deviations ($r = -.119$, $p < .0001$; $edf = 2.776$, $p < .0001$) - a variability measure designed to account for the confounding of mean and standard deviation in bounded variables (Mestdagh et al., 2018). That is, the impact of income on happiness volatility cannot be explained by its effect on average happiness. Additional analyses show that the relationship between income and happiness volatility is not limited to our relatively wealthy European sample. We examined data from the World Health Organization SAGE study in which a sample of over 25,000 individuals from China, Ghana, India, Mexico, Russia, and South Africa reported their happiness an average of 4.8 times using a short Day Reconstruction survey and provided a measure of permanent income estimated from the household ownership of various country-specific durable goods (see SM Note 5). Regression analyses accounting for country-specific fixed effects revealed that permanent income predicted lower happiness volatility ($\beta = -.057$, $p < .0001$) - a relationship robust to all major operationalizations of happiness volatility, and the inclusion of average happiness and different demographic controls (see SM Note 5).

Our findings show a robust association between financial hardship and people's propensity to experience volatile levels of happiness. But what does a volatile emotional life look like exactly? Because increased volatility was apparent across all major psychometric operationalizations and given the sizable statistical overlap between these metrics (Dejonckheere et al., 2019), it is difficult to fully appraise the shape of people's ups and downs. Therefore, going beyond classic affect dynamics measurements, we employed an unsupervised Collective and Point Anomaly (CAPA; Fisch et al., 2019) machine learning algorithm to identify and

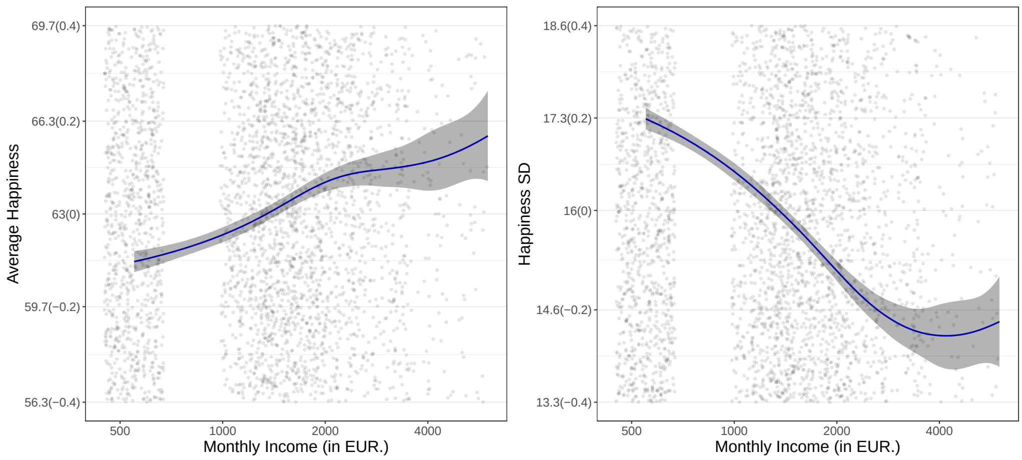


Figure 2.1: Average happiness (Left Panel) and happiness fluctuation (Right Panel) as a function of monthly income in the mobile application study (France, Belgium, and Switzerland). Shadow areas represent 95% confidence intervals. Standardized scores in parentheses.

quantify how happiness changes as a function of income. Specifically, we first take a within-person approach and identify, for each participant, the presence of "anomalous" moments (i.e., observations) and periods (i.e., sequences of observations) in their happiness time series. We then take a between-person approach and examine whether income predicts people's propensity to experience anomalous happiness-related moments and periods, as well as the magnitude of these anomalies.

Fig. 2.2 provides a schematic representation of the CAPA results for prototypical respondents in the lowest and highest income group, respectively. Over 22% of participants experienced at least one anomalous happiness moment throughout the study. Most were instances in which individuals reported being a lot less happy than usual (88%), and some were instances in which individuals reported being a lot happier (12%). Income did not predict the frequency of extreme unhappiness moments ($\beta_{\text{log income}} = .006$, $t = .42$, $p = .67$), which happened on average every 330 happiness observations (i.e., approximately once every 3 months in our dataset). However, it was significantly related to their severity ($\beta_{\text{log income}} = .12$, $t = 3.36$, $p < .001$). For example, participants in the lowest income group ($<$ 1100 euros/month) had extreme unhappy moments that were rated 7 points lower in happiness (an average of 23 vs. 30) than participants in the highest income group ($\sim$ 6000 euros/month) - a difference over 50% larger than the gap in average happiness between the two groups (4.5 points). In contrast, income did not predict the magnitude of extreme moments of happiness

$(\beta_{\text{log income}} = -.06,\, t = -.58,\, p = .56)$.

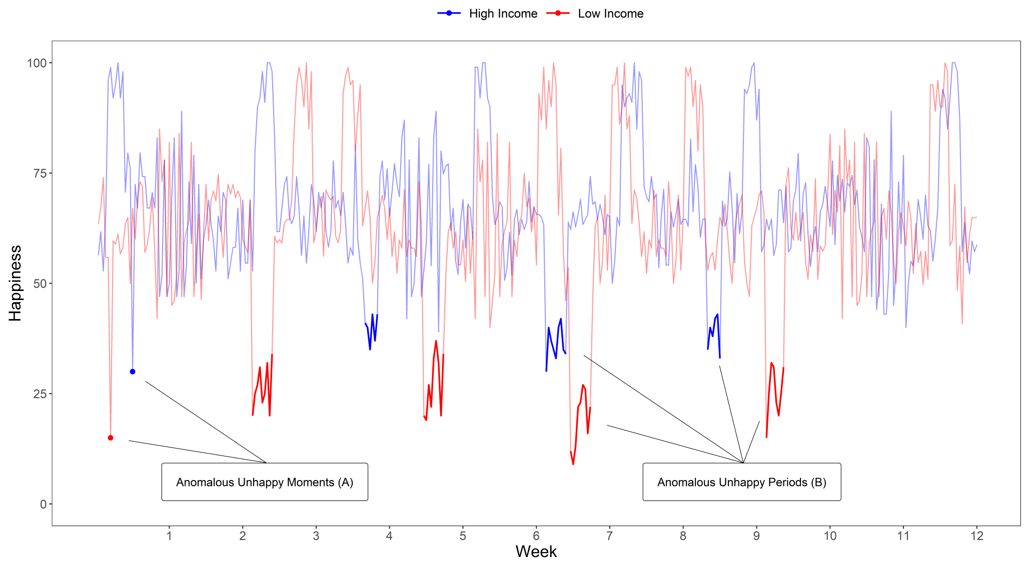


Figure 2.2: Representative happiness dynamics for low and high-income individuals over three months, based on the average sample parameters. Low-income individuals experience harsher moments of extreme unhappiness (A), as well as more frequent and severe periods of prolonged unhappiness (B).

Anomalous periods were far more common than anomalous moments, with over 94% of people experiencing at least one sequence of happiness states that significantly differed from their typical sequences (e.g., a strange couple of days). K-means clustering (see SM Note 6) suggested that these sequences fell into three categories: 1) unusually prolonged periods of unhappiness (23% of anomalous sequences), 2) unusually prolonged periods of happiness (48%), and 3) unusual sequences of high happiness volatility (28%). We estimated the frequency, duration, and intensity of each of these categories of anomalous sequences and examined how they related to income.

On average, anomalous periods of prolonged unhappiness last two days and occur once every 120 happiness observations (i.e., approximately once a month in our dataset). Income significantly predicts both the frequency $(\beta_{\text{log income}} = -.04,\, t = -2.89,\, p = .004)$ and intensity (i.e., mean happiness) of these periods $(\beta_{\text{log income}} = .10,\, t = 3.51,\, p < .001)$, but not their duration $(\beta_{\text{log income}} = .02,\, t = 1.65,\, p = .09)$. For example, the highest income group experienced 30% fewer prolonged unhappiness periods than the lowest income group. These painful periods were also less extreme overall (mean happiness: 43 vs. 30)-a difference almost three times the size of the gap in average happiness between the two groups (4.5 points).

Anomalous periods of prolonged happiness happened approximately twice a month and typically lasted two days. Periods of unusual volatility occur once every 25 days on average and last for approximately four days. The relationship between income and these types of anomalous sequences was substantially weaker and above a .05 significance cut-off: frequency of prolonged happiness periods ($\beta_{\text{log income}} = -.02$, $t = -1.47$, $p = .14$), average intensity of prolonged happiness periods ($\beta_{\text{log income}} = .04$, $t = 1.95$, $p = .051$), frequency of unusual volatility periods ($\beta_{\text{log income}} = .02$, $t = 1.09$, $p = .27$), duration of unusual volatility periods ($\beta_{\text{log income}} = -.0015$, $t = -0.88$, $p = .37$).

Given the non-linear associations between income and happiness, we performed CAPA analyses separately within the lower-income (less than 3300) and higher-income (more than 3300 euros per month) brackets. Results were in line with the notion of an income plateau: all the CAPA results mentioned above replicated when considering income variation from low to middle income. In contrast, income only predicted the intensity (average happiness) of anomalous sequences of both prolonged unhappiness ($\beta_{\text{income}} = .21$, $t = 3.92$, $p < .001$) and prolonged happiness ($\beta_{\text{income}} = .11$, $t = 3.01$, $p = .003$), when considering income variation from middle to high income.

Taken together, these findings show that people with relatively low income have more volatile emotional lives, as reflected by (1) the experience of more extreme "rock bottom" moments and (2) more frequent and intense periods of lasting unhappiness. Results from the CAPA analyses are robust to alternative, more conventional, ways to identify extreme observations. For example, income significantly relates to people's propensity to experience happiness moments that are in the bottom 1%, 5% or 10% of the distribution of happiness observations across and within individuals (see SM Note 7). However, the observational nature of our data precludes causal inferences. To provide suggestive evidence for a causal link between income and happiness volatility, we examined how people's propensity to experience anomalous happiness states changes over the month as a function of income. We reasoned that lower-income individuals might experience more frequent anomalous moments and periods of happiness at times of heightened financial strain (i.e., in the last few days of the months when most Europeans are waiting for their monthly salary). As shown in Fig. 2.3, income was associated with fewer anomalous affective experiences overall ($\beta_{\text{income group}} = -.021$, $t = -10.56$, $p < .0001$), and its effect grows larger at the end of the month ($edf_{\text{income group x time}} = 7.45$, $p < .0001$). For example, while in the first three weeks of the month, individuals with income 1 S.D. above the mean report 6% fewer anomalous happiness observations than individuals with income 1

S.D. below the mean (95% C.I for the relative difference [4.36%, 7.42%]), this difference roughly doubles in the last week (11.33% relative difference; 95% C.I [9.42%, 12.29%]), and triples in last few days of the month (17% relative difference; 95% C.I [15.19%, 18.55%]).

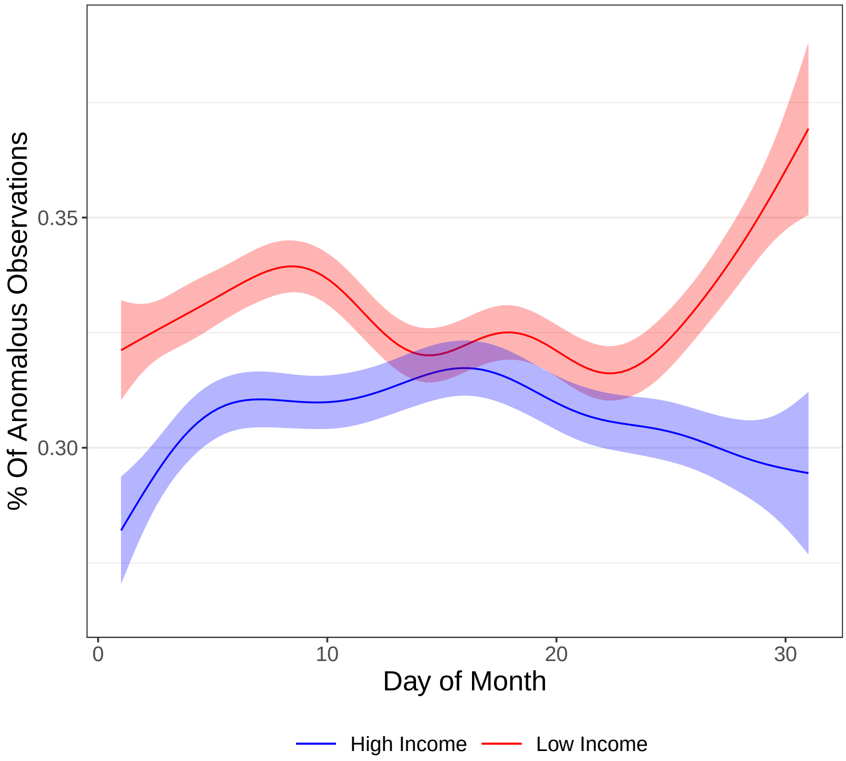


Figure 2.3: Proportion of anomalous happiness-related observations for individuals with income 1 SD above the mean (> 2900 euros; in blue) and 1 SD below the mean (< 1100 euros; in red). Differences between the two groups grow larger towards the end of the month. As in France, salaries and wages are only paid once at the end of the month, results are consistent with the notion that income has a causal effect on happiness volatility.

2.4. Discussion

The lay notion that "money buys happiness" has been challenged by decades of empirical research revealing that money has a surprisingly small impact on happiness, especially in wealthier countries (Boyce et al., 2017; Frey and Stutzer, 2002; Kahneman et al., 2006). Accordingly, scholars have recommended that policies should consider alternative ways to increase people's happiness - for example, by focusing not only on economic gains but also income redistribution (Kang and Rhee, 2021; Ono and Lee, 2016).

While these discussions are important, they must be informed by understanding how income shapes our emotional lives. By looking beyond static snapshots of people's happiness, we found that we may have underestimated the impact of income on happiness. Across multiple countries, measurement choices, and model specifications we found a robust negative relationship between financial hardship and people's propensity to experience volatile levels of happiness. This relationship was far from trivial. To put it into perspective, the difference in emotional volatility between the lowest and highest-income group in our European sample (Δ = 0.4 SD) was similar in size to the difference between patients with bipolar disorder and healthy controls (Δ = 0.32 SD; Stanislaus et al., 2020), and about half the size of the difference between people with and without borderline personality disorder (Δ = 0.7 SD; Ebner-Priemer et al., 2007). And the greatest gap between people of lower and higher income is between how frequent and intense they experience emotional downs, not ups.

The overall impact of relatively rare but extreme episodes of unhappiness on people's lives can live longer than the emotional experience itself. Intense affective states have been repeatedly shown to guide people's choices, even when those emotions are incidental to the decision setting. When we feel miserable, we may eat and procrastinate more (Grunberg and Straub, 1992; Tice et al., 2001) and help and trust less (Dunn and Schweitzer, 2005; Manucia et al., 1984). Moreover, the decisions we make based on fleeting emotions can become the basis for future decisions after those emotions have passed. For example, when we make a poor decision out of anger, we tend to repeat that mistake even after cooling off (Andrade and Ariely, 2009).

In line with many other studies (Jebb et al., 2018; Kahneman and Deaton, 2010), we also found strong evidence that the relationship between money and happiness diminishes at higher levels of income. It's important to note that there were also no gains in emotional stability beyond 3,300 euros per month-if anything, the data showed a trend toward decreasing stability.

It's important to note the limitations of the present study. First, while the satiation figure above is consistent with previous research on satiation points in Western Europe (Jebb et al., 2018), we note that determining satiation from categorical income brackets makes it difficult to estimate a precise cutoff. Second, while we found robust associations between income and happiness volatility in a sample of individuals from six developing countries, we could only apply unsupervised anomaly detection techniques to our non-representative experience-sampling dataset from Europe. The relationship between income and the frequency, intensity, or duration of happiness anomalies could differ in the general population. Furthermore, although our data show a stronger connection between income and experienced happiness when people face challenges related to lacking money at the end of the month, it would be valuable to substantiate this “money crunch” hypothesis beyond the time-of-month analysis. Future research should explore, for example, whether money shortages spill over into interpersonal conflict in the family, which may be a more proximal driver of unhappiness. This conjecture is compatible with the observation that income is associated with fluctuations in marital satisfaction rather than overall marital satisfaction (Jackson et al., 2017).

Money may not buy happiness, but our research strongly suggests that an impoverished life is an emotionally volatile life punctuated with rare - but extreme - moments of distress. While future research is needed to fully assess the personal and social repercussions of income-induced emotional volatility, seemingly rare episodes of misery may only be the beginning - even after the emotional distress itself fades, the suffering is likely to continue in a cascade of repeated poor decisions that set the conditions for social alienation and emotional relapse. Harnessing this knowledge could be useful for public policy. With more and more governments focusing on measuring and increasing happiness (Trudel-Fitzgerald et al., 2019; Verma, 2017), policymakers and researchers may take into consideration not only whether we can increase general happiness but whether we can buy emotional stability.

2.5. Supplemental Materials

2.5.1. Note 1: Summary Statistics

Table 2.1 and 2.2 present the summary statistics of our “58 seconds” data. In Table 2.3 we present the distribution of income responses before ($N_{Raw} = 6,010$) and after our main imputation by Random Forests ($N_{RF} = 23,471$). Table 2.4 and Table 2.5 present the summary statistics for the World Health Organization SAGE dataset ($N = 25,739$ including 105 observations with missing values for income).

Numeric Variable	Average	SD	Median
Age	27.905	9.215	26
Income (in EUR/month)	1,243.134	1,350.856	1,200
Average Happiness	62.555	16.657	63.176
Happiness SD	16.429	6.727	15.484
Happiness Rel. SD	0.371	0.152	0.351
Happiness RMSSD	19.621	8.759	18.137
Happiness TKEO	249.772	240.404	185.226
Happiness PAC	0.125	0.125	0.091

Table 2.1: Summary Statistics - continuous variables in the “58 seconds” dataset.

Categorical Variable	Category	N	Sample Proportion
Gender	Female	16,065	68.4%
Gender	Male	7,406	31.6%
Country	France	22,040	93.9%
Country	Switzerland	866	3.6%
Country	Austria	339	1.4%
Country	Belgium	135	0.06%
Country	Other	1	0%

Table 2.2: Summary Statistics - categorical variables in the “58 seconds” dataset.

	Raw Responses		RF Imputation	
Monthly Income	Obs.	Rel. Freq.	Obs.	Rel. Freq.
No income	1630	27.1%	5988	25.5%
Less than 1100 eur.	1121	18.7%	4843	20.6%
Between 1100 and 1300 eur.	628	10.4%	2691	11.5%
Between 1300 and 1400 eur.	236	3.9%	997	4.2%
Between 1400 and 1500 eur.	279	4.6%	1054	4.5%
Between 1500 and 1700 eur.	425	7.1%	1686	7.1%
Between 1700 and 1900 eur.	327	5.4%	1329	5.6%
Between 1900 and 2100 eur.	285	4.7%	1168	4.9%
Between 2100 and 2500 eur.	345	5.7%	1336	5.6%
Between 2500 and 3300 eur.	345	5.7%	1216	5.1%
Between 3300 and 4500 eur.	191	3.2%	575	2.4%
Between 4500 and 7500 eur.	129	2.1%	383	1.6%
More than 7500 eur.	69	1.1%	205	0.8%

Table 2.3: Distribution of income responses in the “58 seconds” dataset.

Numeric variables	Average	SD	Median
Age	43.015	14.737	44
Income (ladder)	0.335	0.561	0.295
Average Happiness	1.324	0.621	1.400
Happiness SD	0.331	0.322	0.289
Happiness Rel. SD	0.313	0.327	0.242
Happiness RMSSD	0.439	0.439	0.377
Happiness TKEO	0.129	0.323	0
Happiness PAC	0.084	0.189	0

Table 2.4: Summary Statistics - numeric variables in the WHO SAGE dataset.

Categorical Variable	Category	N	Sample Proportion
Gender	Female	14,567	56.7%
Gender	Male	11,130	43.3%
Country	China	9,131	35.4%
Country	Ghana	3,162	12.3%
Country	India	8,223	31.9%
Country	Mexico	646	2.5%
Country	Russia	2,082	8.1%
Country	South Africa	2,495	9.7%

Table 2.5: Summary Statistics - categorical variables in the WHO SAGE dataset.

2.5.2. Note 2: Formal Definitions of Happiness Fluctuations

Table S6 presents the formal definitions of the different operationalizations of happiness fluctuations included in our studies. In our formulas, x_i stands for the i^{th} happiness observation of a given individual. M, SD, and n represent an individual average happiness, standard deviation in happiness and total number of happiness reports. Finally, $I(x_{i+1} - x_i, d_{0.9})$ defines a binary variable taking a value of 1 if the absolute value of $(x_{i+1} - x_i)$ is greater than $d_{0.9}$ and 0 otherwise, where $d_{0.9}$ represents the 90^{th} percentile in the distribution of absolute happiness changes across all participants.

Index	Formal Definition	Interpretation
Standard Deviation (SD)	$\sqrt{\frac{\sum(x_i - M)^2}{n}}$	Variability in happiness reports.
Relative Standard Deviation (Rel. SD)	$\frac{SD}{max(SD\|M)}$	Mean-corrected estimate of the standard deviation for bounded variables (Mestdagh et al. 2018).
Root Mean Square of Successive Differences (RMSSD)	$\sqrt{\frac{\sum(x_i - x_{(i+1)})^2}{n-1}}$	Average change across successive happiness observations.
Teager–Kaiser energy operator (TKEO)	$\frac{\sum(x_i^2 - x_{(i-1)}x_{(i+1)})}{n-2}$	Average change across 3 happiness reports. Useful in identifying happiness spikes.
Probability of Acute Change (PAC)	$\frac{\sum I(x_{i+1} - x_i, d_{0.9})}{n-1}$	Likelihood of extreme change in happiness.

Table 2.6: Operationalizations of Happiness Fluctuations.

2.5.3. Note 3: Main Results - Robustness Checks

In this section, we show the robustness of our results to multiple specifications. These robustness checks are performed using our "58 seconds" data (with missing income observations imputed by a random forest approach). As in the main body of the paper, all our models exclude the data from participants that reported having no income or an income of over 7,500 euros per month (N=17,278). Further robustness checks are presented in the next section (SM Note 4). In order to accommodate the non-linear nature of the relationship between income and happiness volatility, our main results are estimated using Generalized Additive Models (GAMs). All our models use income as the main explanatory variable. Each model uses a different measure of average happiness or happiness volatility as the dependent variable. Table 2.7 presents the extension of our main results to all operationalizations of happiness fluctuations. Table 2.8 shows that these relations are significant when controlling for demographic variables. Table 2.9 shows that these relations remain significant when, in addition to demographic variables, we control for the effect of income on average happiness. As the GAM coefficients are not directly interpretable, in tables 2.10-2.12, we present the same analyses using Linear Regressions (OLS) and the logarithm of income as the main explanatory variable. To allow for comparisons across specifications, we report the standardized regression coefficients (with standard errors in parenthesis) for all numeric variables (log income, age, average happiness). For binary variables (gender), we report the coefficients obtained when regressing this raw variables on the standardized dependent variable. We also present visually the results of the GAM models (with income as a unique explanatory variable) in Figure 2.4.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(3)	(4)
Income Statistics:						
Edf	3.754	2.933	2.776	2.991	2.915	2.889
Ref.df	4.474	3.554	3.375	3.620	3.534	3.505
F	22.752	111.261	74.861	102.915	85.979	90.811
P Value	<.00001	<.00001	<.00001	<.00001	<.00001	<.00001

Table 2.7: GAM results.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(3)	(4)
Income Statistics:						
Edf	3.750	2.483	2.284	2.243	2.255	2.168
Ref.df	4.472	3.042	2.808	2.761	2.775	2.672
F	14.097	16.405	8.162	11.876	12.502	11.544
P Value	<.00001	<.00001	.00004	<.00001	<.00001	<.00001

Table 2.8: GAM results (controlling for age, gender and country of residence).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(3)	(4)
Income Statistics:						
Edf	-	2.401	2.430	2.081	2.153	1.966
Ref.df	-	2.945	2.979	2.569	2.655	2.434
F	-	12.14	12.93	9.12	9.725	8.095
P Value	-	$<.00001$	$<.00001$	.00003	.00001	.00012

Table 2.9: GAM results (controlling for age, gender, country of residence and average happiness).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Log Income	0.075^{***}	-0.148^{***}	-0.119^{***}	-0.145^{***}	-0.131^{***}	-0.135^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	-0.000	0.000	0.000	0.000	0.000	-0.000
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Observations	17,278	17,278	17,278	17,278	17,278	17,278
R^2	0.006	0.022	0.014	0.021	0.017	0.018
Adjusted R^2	0.006	0.022	0.014	0.021	0.017	0.018
Residual Std. Error	0.997	0.989	0.993	0.989	0.991	0.991
F Statistic	98.041^{***}	389.327^{***}	250.016^{***}	370.598^{***}	303.050^{***}	318.299^{***}
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.10: Linear regression results.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Log Income	0.061^{***}	-0.063^{***}	-0.048^{***}	-0.054^{***}	-0.056^{***}	-0.053^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Age	0.010	-0.204^{***}	-0.183^{***}	-0.228^{***}	-0.188^{***}	-0.203^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Male	0.140^{***}	-0.131^{***}	-0.058^{***}	-0.080^{***}	-0.071^{***}	-0.063^{***}
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Constant	-0.102	0.112	0.071	0.064	0.088	-0.009
	(0.117)	(0.114)	(0.115)	(0.114)	(0.115)	(0.115)
Observations	17,278	17,278	17,278	17,278	17,278	17,278
R^2	0.012	0.061	0.044	0.066	0.048	0.054
Adjusted R^2	0.011	0.060	0.043	0.066	0.048	0.054
Residual Std. Error	0.994	0.969	0.978	0.966	0.976	0.973
F Statistic	25.449^{***}	139.792^{***}	98.342^{***}	153.576^{***}	109.280^{***}	123.612^{***}
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.11: Linear regression results (controlling for age, gender and country of residence).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Log Income	0.061^{***}	-0.055^{***}	-0.056^{***}	-0.047^{***}	-0.050^{***}	-0.044^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Average Happiness		-0.134^{***}	0.138^{***}	-0.113^{***}	-0.096^{***}	-0.147^{***}
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age	0.010	-0.203^{***}	-0.184^{***}	-0.227^{***}	-0.187^{***}	-0.201^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Male	0.140^{***}	-0.112^{***}	-0.077^{***}	-0.064^{***}	-0.058^{***}	-0.043^{***}
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Constant	−0.102	0.098	0.085	0.052	0.078	−0.024
	(0.117)	(0.113)	(0.114)	(0.113)	(0.115)	(0.114)
Observations	17,278	17,278	17,278	17,278	17,278	17,278
R^2	0.012	0.079	0.062	0.079	0.057	0.076
Adjusted R^2	0.011	0.078	0.062	0.079	0.057	0.075
Residual Std. Error	0.994	0.960	0.969	0.960	0.971	0.962
F Statistic	25.449^{***}	163.890^{***}	127.779^{***}	164.903^{***}	116.594^{***}	156.693^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.12: Linear regression results (controlling for age, gender, country of residence and average happiness).

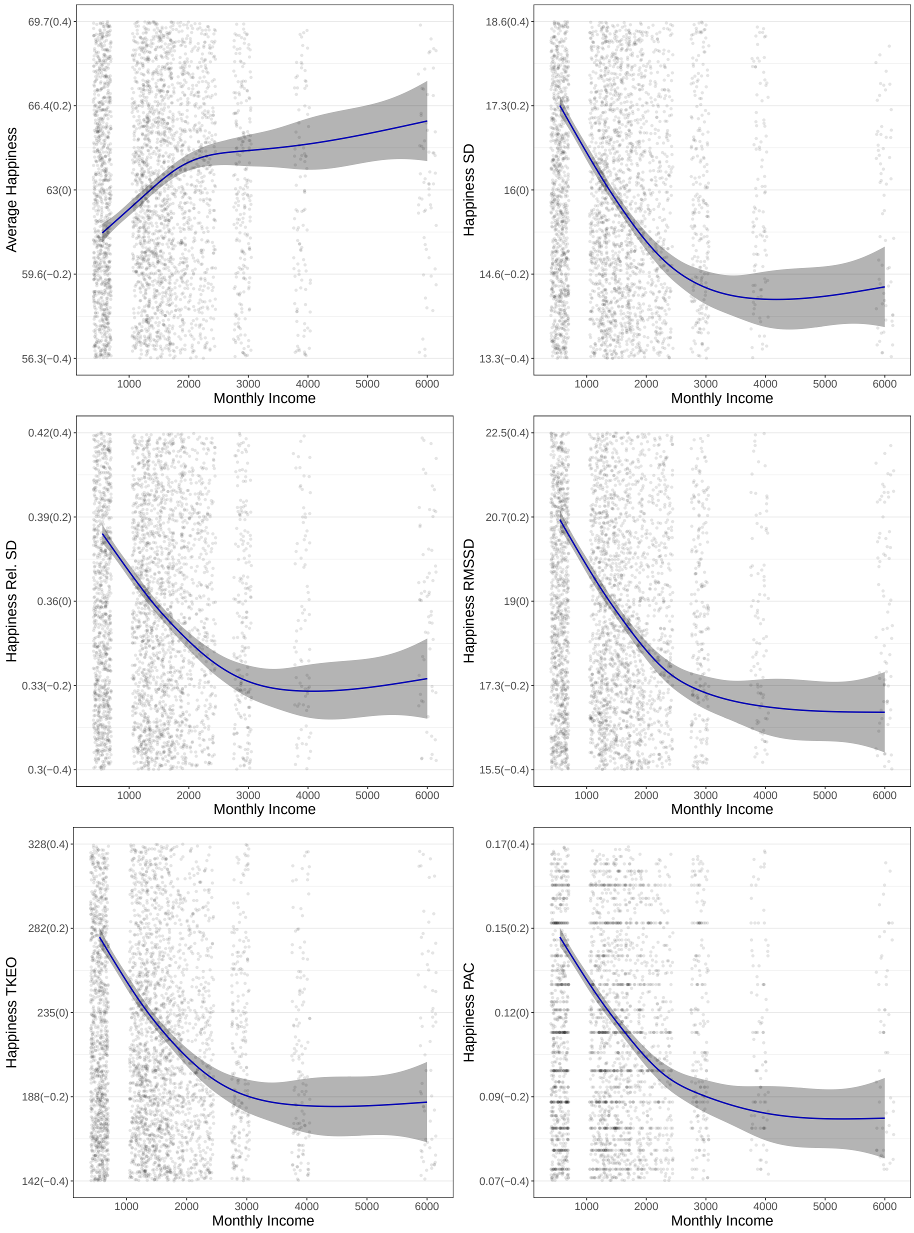


Figure 2.4: GAM Fit with 95% C.I. (in gray).

In the main manuscript, we claim that the effect of income on happiness volatility satiated at a monthly income of 3,000 euros. To estimate this satiation point, we used the confidence intervals of the derivatives of the GAM splines (estimated in a model with income as the unique explanatory variable). Figure 2.5 presents the 95% confidence intervals of the splines' derivatives for each of our GAM models. To identify the precise satiation points, we found the lowest income level that corresponded to a spline derivative containing a slope of 0 in its 95% confidence interval. The specific satiation points were found at 2,111.05 euros per month for average happiness, 3,069.6 euros for happiness SD, 3,042.21 euros for the relative SD, 2,987.44 euros for RMSSD, 2,960.05 euros for TKEO and 2,768.34 euros for PAC. As stated in the main body of the paper, we also analyzed our data using separate linear regressions for low and high-income individuals. To separate our sample based on income, we made use of the Robin Hood algorithm (Simonsohn, 2018). Following the algorithm results, we estimated two regression lines, one for low-income individuals and one for high income individuals. The Robin Hood algorithm suggested using a cutoff income of 2,000 euros per month for the analysis of average happiness and 2,300 euros for all analyses of happiness volatility. Note that this amount is not the satiation point, but the point that would maximize the probability of finding a u-shaped relationship. Table 2.13 presents the linear regression results for low-income individuals and Table S14 presents the results for high-income individuals. As in our previous analyses, controlling for age, gender or country of residence does not significantly affect our results. These coefficients are estimated without including the individuals with a cut-off income in the low or high-income samples. Including these individuals to either the low-income group or the high-income group does not change our results. All reported coefficients are standardized.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.063***	−0.116***	−0.091***	−0.109***	−0.101***	−0.101***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	0.000	0.000	0.000	0.000	0.000	0.000
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Observations	12,600	13,768	13,768	13,768	13,768	13,768
R^2	0.004	0.013	0.008	0.012	0.010	0.010
Adjusted R^2	0.004	0.013	0.008	0.012	0.010	0.010
Residual Std. Error	0.998	0.993	0.996	0.994	0.995	0.995
F Statistic	50.300***	186.148***	115.387***	166.222***	140.913***	140.809***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2.13: OLS estimates for low-income individuals.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.027	−0.003	0.005	−0.028	−0.008	−0.036*
	(0.017)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Constant	0.000	−0.000	−0.000	0.000	−0.000	−0.000
	(0.017)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Observations	3,510	2,174	2,174	2,174	2,174	2,174
R^2	0.001	0.00001	0.00003	0.001	0.0001	0.001
Adjusted R^2	0.0005	−0.0005	−0.0004	0.0003	−0.0004	0.001
Residual Std. Error	1.000	1.000	1.000	1.000	1.000	1.000
F Statistic	2.594	0.016	0.060	1.754	0.153	2.857*
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2.14: OLS estimates for high-income individuals.

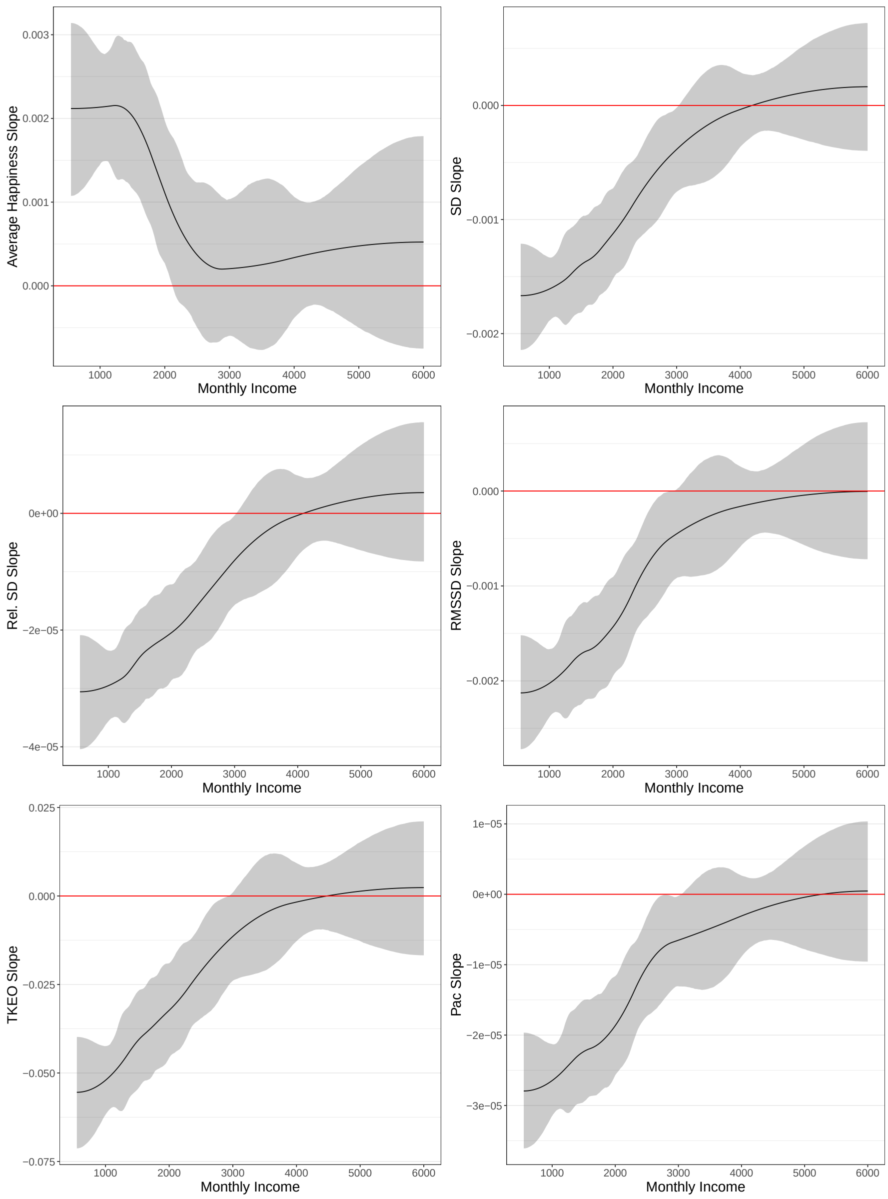


Figure 2.5: Splines' derivatives with 95% C.I. (in gray).

2.5.4. Note 4: Specification Curves

We run two separate specification curves for average happiness and happiness volatility. We start by looking at average happiness and consider the following specification choices:

- Income: We considered income either as an ordered categorical variable (i.e, income category) or a numeric variable (i.e, Euros per month). When income is numeric, we regressed average happiness on the logarithm of income.

- Missing: Treatment given to missing values. We either removed the data from individuals with missing observations, imputed the missing observations by Random Forest (RF) or imputed the missing observations by means of a Hot-Deck Algorithm (HD).

- Lowest: We considered specifications that included the lowest income group (individuals reporting an income of zero) and specifications that excluded them from the analyses. Given our logarithmic relationship between income and happiness, we only included this group in specifications considering income as an ordered categorical variable.

- Highest: We considered specifications that included the highest income group (individuals reporting an income of over 7500 euros per month) and specifications that excluded them from the analyses. If included and income is numeric, we assume that the middle point of the interval would be 9000 euros per month.

- Controls: We included specifications including and excluding demographic control variables (age, gender and country of residence).

In total we considered 36 specifications. Figure 2.6 presents the specification curve results. The coefficient of income is statistically significant (at the 95% confidence level) for all specifications. The effect of income on average happiness is larger for specifications using income as a numeric variable and without imputed data. We also see that excluding observations that reported an income of zero lead to larger coefficient sizes. As argued by Kahneman and Deaton, 2010, zero-income reports suffer from important reliability issues. Unsurprisingly, the effect of income on average happiness was jointly significant across all specifications ($p < 0.002$ for each of the three significance criteria outlined in Simonsohn et al. (2020)).

Our second specification curve (Figure 2.7) analyzes the robustness of the relationship between happiness volatility and income. In addition to the previous analytical choices, we consider 5 different dependent variables (see Table 2.6), yielding a total of 180 specifications. Of these 180 specifications, income was significantly related to happiness fluctuations in 167 specifications. In all specifications, the coefficient of income was negative. The 13 specifications displaying a non-significant relationship between income and happiness volatility were estimated with data imputed by the Hot-Deck algorithm. Again, the effect of income on happiness fluctuations was jointly significant across all specifications ($p < 0.002$ for each of the three significance criteria outlined in Simonsohn et al. (2020)).

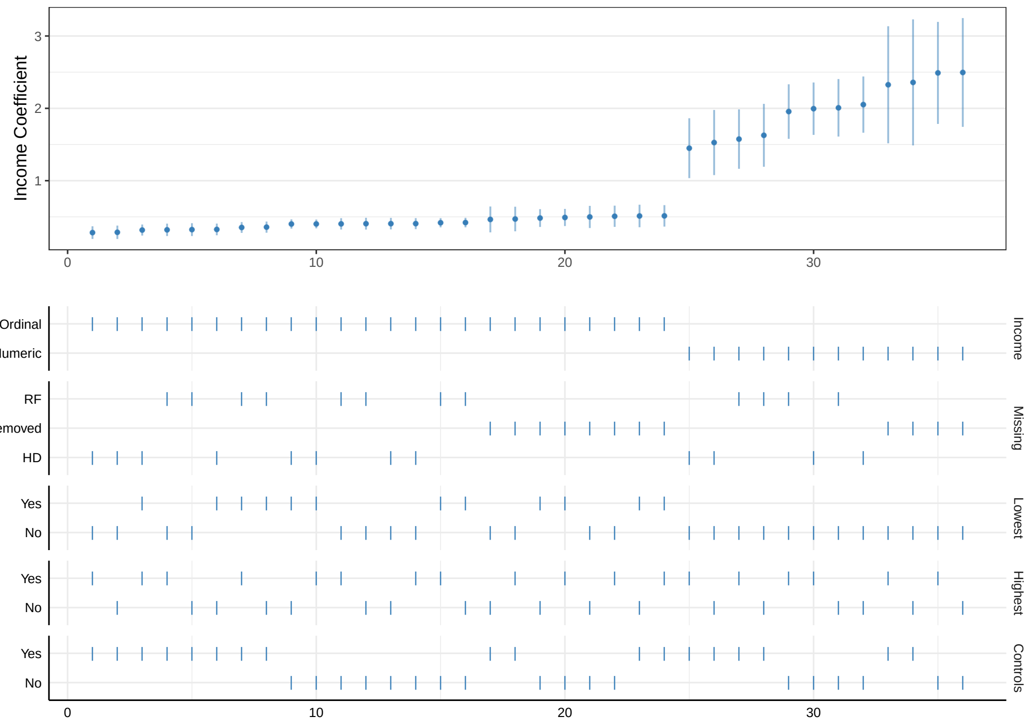


Figure 2.6: Specification curve 1 - income and average happiness.

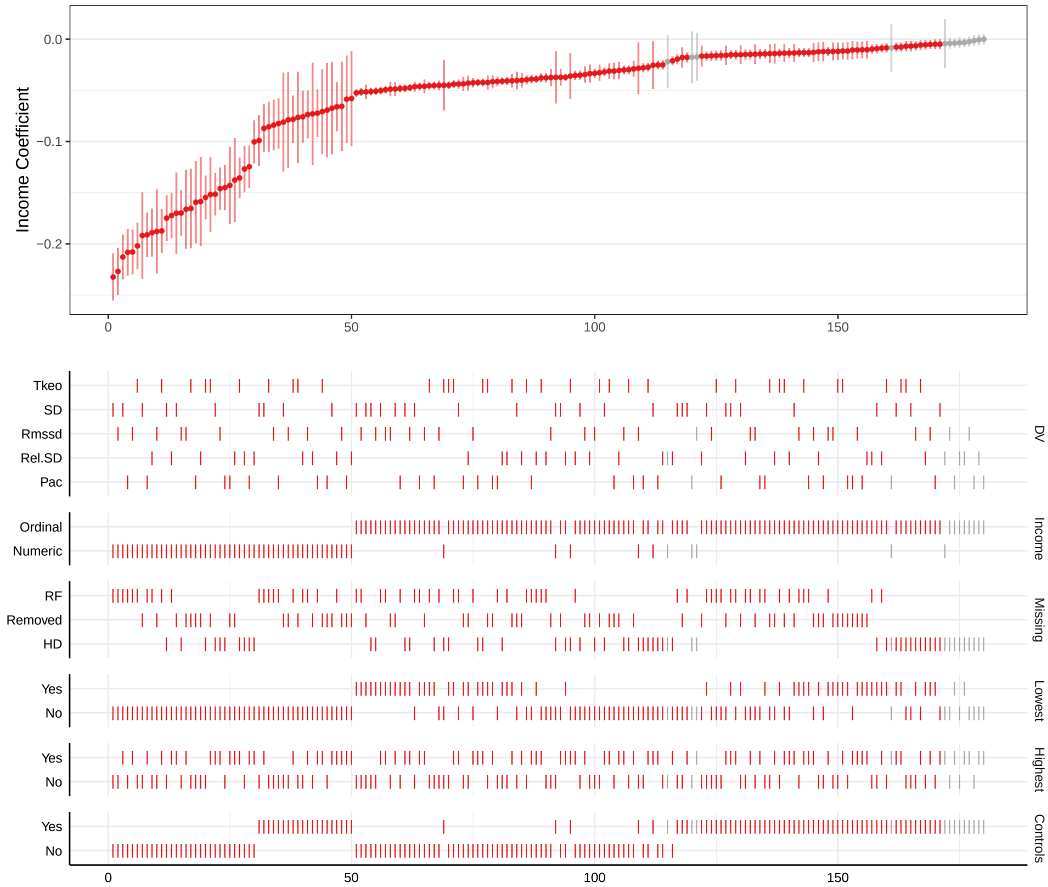


Figure 2.7: Specification curve 2 - income and happiness volatility.

2.5.5. Note 5: World Health Organization (WHO SAGE) Data and Results

Participants and Day Reconstruction Method

In the World Health Organization Study on Global Ageing and Adult Health (WHO SAGE), nationally representative samples of people in China, Ghana, India, Mexico, Russia, and South Africa completed a modified version of the Day Reconstruction Method - see Ayuso-Mateos et al., 2013; Kowal et al., 2012 for detailed descriptions of the study. Participants were asked to report, in chronological order, what they did and how they felt across different episodes of their previous day. As most operationalizations of happiness volatility require at least three measurements, we excluded from our analyses participants who provided less than three episodes. Hence, our final sample consists of 25,634 participants.

Happiness

For each episode, respondents were asked to rate the extent to which they felt 7 emotions on a scale from 1 ("Not at all") and 3 ("Very much"). We then calculated a total composite happiness score by subtracting the mean of the negative emotions (worried, rushed, irritated or angry, depressed, tense or stressed) from the mean of the positive emotions (calm or relaxed, enjoying), resulting in a continuous score from -2 to 2 - see Taquet et al. (2020) for a similar approach.

Income and demographics

The WHO SAGE study measures income from the household ownership of durable goods, access to services, and housing characteristics—an approach that provides more reliable estimates of income than direct self-report questions in developing countries (Ferguson et al., 2003). Each respondent provides information on a country-specific list of 21 items (durable goods, services, and housing characteristics; e.g., "do you own a refrigerator?"). Then, households are arranged in an "asset ladder" using an item random-effects model and Bayesian post estimation. We use this continuous "asset ladder" variable as a measure of income (see WHO SAGE documentation). This dataset also contains information on the respondents' age and gender.

Results

Using the WHO data we can show that the relationship between income and happiness volatility is not unique to Western industrialized and rich societies. Income in this dataset is not self-reported but estimated from the respondents' ownership of durable goods, housing characteristics and access to services (see Ferguson et al. (2003) for a complete description of the estimation of income based on these variables). As the income ladders are country specific, we control for the participants' country of residence in all the presented specifications. We abstain from estimating country-level random effects due to the small number of countries presented in our dataset.

Employing non-linear methods (GAMs), we show that income significantly predicts happiness fluctuations (Table S15), even when controlling for demographic information (Table S16) and average happiness (Table 2.17). To provide the reader with a more intuitive presentation of our results, we repeat these analyses using linear regressions (Table 2.18-2.20). As in our previous results, the reported regression coefficients are standardized for all numeric variables (income, age). We control for country of residence in all our analyses (coefficients not reported for brevity).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income Statistics:						
Edf	5.247	1	1	3.837	1.738	1.879
Ref.df	6.414	1	1	4.798	2.190	2.377
F	102.890	189.698	6.004	36.154	13.434	33.248
P Value	< 0.00001	< 0.00001	0.014	< 0.00001	< 0.00001	< 0.00001

Table 2.15: WHO SAGE data - GAM results.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income Statistics:						
Edf	5.258	1	1	3.745	1	1
Ref.df	6.427	1	1	4.689	1	1
F	104.868	207.884	9.236	40.792	32.481	87.059
P Value	< 0.00001	< 0.00001	0.002	< 0.00001	< 0.00001	< 0.00001

Table 2.16: WHO SAGE data - GAM results (controlling for age and gender).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income Statistics:						
Edf	-	3.819	2.464	2.996	2.340	2.554
Ref.df	-	4.778	3.118	3.779	2.964	3.231
F	-	9.954	8.268	8.088	2.727	4.580
P Value	-	< 0.00001	0.00001	0.00001	0.046	0.003

Table 2.17: WHO SAGE data - GAM results (controlling for age, gender and average happiness).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.173***	−0.100***	−0.018**	−0.095***	−0.040***	−0.065***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Constant	0.204***	−0.317***	−0.230***	−0.298***	−0.135***	−0.217***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	25,634	25,634	25,634	25,634	25,634	25,634
R^2	0.134	0.072	0.035	0.069	0.014	0.037
Adjusted R^2	0.134	0.072	0.035	0.069	0.014	0.037
Residual Std. Error	0.931	0.963	0.982	0.965	0.993	0.982
F Statistic	662.199***	331.160***	156.584***	316.401***	62.016***	165.859***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2.18: WHO SAGE data - Linear regression results.

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.172***	−0.103***	−0.021***	−0.098***	−0.042***	−0.068***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
Age	0.016**	−0.045***	−0.035***	−0.042***	−0.039***	−0.040***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
Male	0.070***	−0.029**	0.022*	−0.022*	−0.020	−0.018
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Constant	0.166***	−0.293***	−0.234***	−0.278***	−0.117***	−0.200***
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	25,594	25,594	25,594	25,594	25,594	25,594
R^2	0.136	0.074	0.036	0.071	0.016	0.039
Adjusted R^2	0.135	0.074	0.036	0.070	0.015	0.039
Residual Std. Error	0.930	0.962	0.982	0.964	0.993	0.982
F Statistic	502.328***	255.387***	120.607***	242.757***	51.366***	129.460***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2.19: WHO SAGE data - Linear regression results (controlling for age and gender).

	Dependent variable:					
	Average	SD	Rel.SD	RMSSD	TKEO	PAC
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.172***	−0.040***	−0.038***	−0.038***	−0.014*	−0.023***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
Average Happiness		−0.363***	0.100***	−0.347***	−0.161***	−0.257***
		(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
Age	0.016**	−0.039***	−0.036***	−0.036***	−0.036***	−0.036***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
Male	0.070***	−0.004	0.015	0.002	−0.009	−0.001
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Constant	0.166***	−0.233***	−0.250***	−0.221***	−0.091***	−0.158***
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Observations	25,594	25,594	25,594	25,594	25,594	25,594
R^2	0.136	0.188	0.045	0.175	0.038	0.096
Adjusted R^2	0.135	0.188	0.045	0.175	0.038	0.095
Residual Std. Error	0.930	0.901	0.977	0.909	0.982	0.952
F Statistic	502.328***	657.993***	133.954***	602.233***	112.931***	300.843***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2.20: WHO SAGE data - Linear regression results (controlling for age, gender and average happiness).

2.5.6. Note 6: Point and Collective Anomalies

To jointly identify anomalous happiness moments and sequences, we used a Collective and Point Anomaly Detection (CAPA, Fisch et al., 2019) machine learning algorithm. Figure 2.8 exemplifies the results of this procedure for 3 participants in our “58 seconds” dataset. We restricted our analyses to individuals with a MAD (Mean Absolute Deviation) in happiness larger than 0 (as the CAPA procedure relies on a robust normalization that requires a positive MAD). The final sample consisted of 5002 participants and a minimum of 50 happiness observations per participant.

Anomalous moments

Table 2.21 presents the results of the frequency analyses. To compute the frequency with which an individual experiences anomalous happiness reports, we estimated the proportion of an individuals' happiness reports that were identified as anomalous momentary reports. All numeric variables are standardized before estimating the regression. To avoid potential confounds, we control for age, gender and country of residence (coefficients not reported for brevity) in all our subsequent analyses. To ensure that our results were not driven by individuals with extremely high or low average happiness, we dropped from our our analyses of anomalous moments those individuals with an average happiness of over 90 or below 10. This resulted in dropping approximately 2% of the individuals in our sample. Table 2.22 presents the estimated OLS coefficients of income on the happiness levels of positive (above and individual's average happiness) and negative (below average) anomalous happiness observations. All models include clustered standard error at the individual level. All numeric variables are standardized before estimating the regression.

	Dependent variable:	
	Frequency (Neg.)	Frequency (Pos.)
	(1)	(2)
Log Income	0.006	−0.005
	(0.016)	(0.016)
Age	0.004	−0.004
	(0.015)	(0.015)
Male	0.031	−0.015
	(0.032)	(0.032)
Constant	0.147	−0.035
	(0.224)	(0.224)
Observations	4,896	4,896
R^2	0.001	0.0003
Adjusted R^2	−0.001	−0.001
Residual Std. Error	1.000	1.000
F Statistic	0.393	0.189
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2.21: Frequency of negative and positive anomalous moments of happiness. The coefficients of numeric variables are standardized.

	Dependent variable:	
	Happiness (Neg.)	Happiness (Pos.)
	(1)	(2)
Log Income	0.121^{***}	-0.056
	(0.036)	(0.096)
Age	0.007	-0.177
	(0.037)	(0.14)
Male	0.084	0.064
	(0.086)	(0.201)
Constant	-0.857^{***}	1.137^{***}
	(0.222)	(0.201)
Observations	1471	206
R^2	0.023	0.049
Adjusted R^2	0.019	0.019
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2.22: Happiness of negative and positive anomalous moments. Coefficients of numeric variables are standardized. Standard errors are clustered at the participant level.

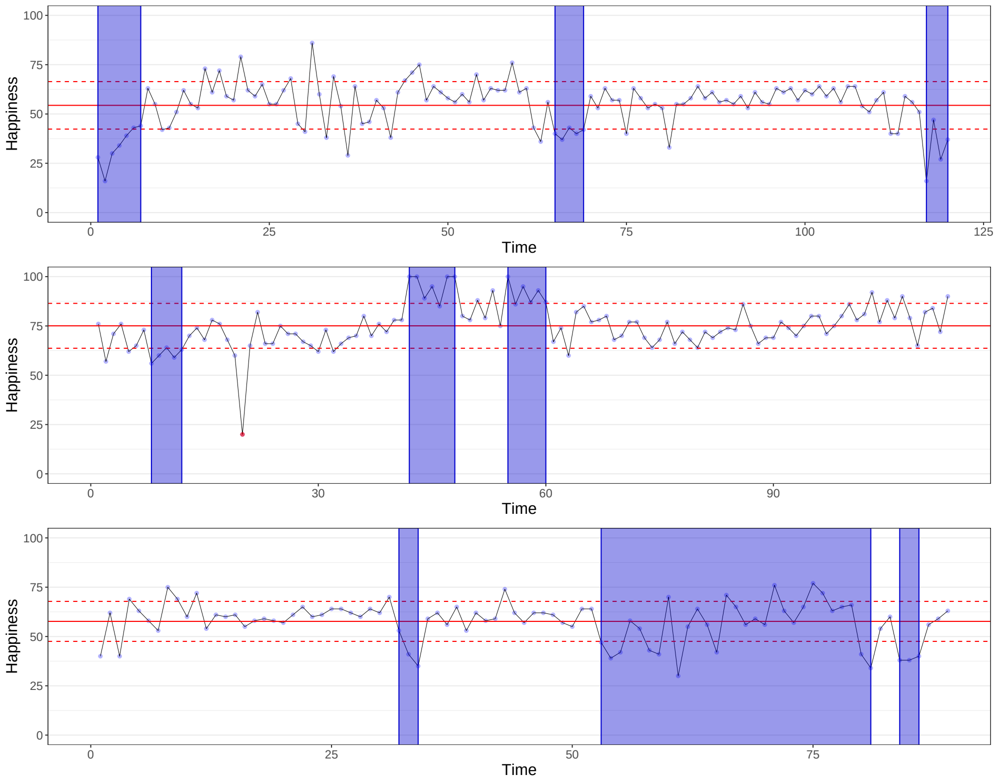


Figure 2.8: Results of PACA algorithm for 3 individuals. Intervals in blue represent anomalous periods of happiness. Points in red indicate anomalous moments of happiness. The horizontal solid line represents the individual's average happiness and the dotted lines the average happiness +1/-1 standard deviation.

Anomalous Sequences:

We categorized the anomalous periods experienced by the participants using a k-means clustering algorithm. Before performing the clustering algorithm, we centered the mean and SD in happiness of each sequence. In order to do so, we subtracted from a sequence's average and SD in happiness, the mean and SD across all observations of happiness of the individual experiencing the anomalous sequence. In doing so, our clustering procedure takes into account the sequence characteristics as compared to the typical mood of the individual experiencing it. We also standardized these centered variables to ensure the robustness of our procedure. Before running the clustering algorithm, we determined the optimal number of clusters using the average silhouette method. The silhouette method suggested an optimal partition consisting of 3 clusters (see Figure 2.9). Then, we performed the k-means algorithm and classified the sequences into three clusters. Figure 2.10 and Table 2.23 present the resulting clusters based on the centered variables and the summary statistics of the anomalous period of happiness contained in each cluster.

We included three metrics of interest as dependent variables in the following analyses. First, we considered the frequency of each type of anomalous period. We measured frequency as the number of anomalous sequences divided by an individual's total number of happiness reports. We measured a sequence intensity using its average happiness. Finally, we measured a sequence duration as its length (in number of happiness reports). Considering length in terms of actual time (hours) does not affect our results. Tables 2.24-2.26 presents the results of our analyses of frequency (Table 2.24), intensity (Table 2.25) and duration (Table 2.26) of anomalous happiness periods. All models control for country of residence (coefficients not reported for brevity).

	Average Happiness	Std. Deviation in Happiness	Median Duration (in hours)
Cluster 1	52.876	16.055	116.38
Cluster 2	84.356	2.410	50.46
Cluster 3	34.468	4.266	47.15

Table 2.23: Median duration, average happiness and standard deviation (SD) of happiness for each cluster of anomalous sequences. Average and SD in happiness are first estimated at the sequence level. Then the resulting mean happiness and happiness SD are averaged across sequences within each cluster.

	Dependent variable:		
	Frequency (Volatility)	Frequency (Pos.)	Frequency (Neg.)
	(1)	(2)	(3)
Log Income	0.017	−0.023	−0.045***
	(0.015)	(0.015)	(0.015)
Age	0.045***	−0.035**	−0.049***
	(0.015)	(0.015)	(0.015)
Male	0.083***	−0.024	0.004
	(0.032)	(0.032)	(0.032)
Constant	−0.050	−0.358	−0.037
	(0.223)	(0.224)	(0.223)
Observations	5,002	5,002	5,002
R^2	0.006	0.004	0.007
Adjusted R^2	0.005	0.002	0.005
Residual Std. Error	0.998	0.999	0.997
F Statistic	4.234***	2.586**	4.803***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.24: Frequency of anomalous periods of increased volatility (Model 1), sustained happiness (Model 2) and sustained unhappiness (Model 3). Coefficients of numeric variables are standardized.

	Dependent variable:		
	Happiness (Volatility)	Happiness (Pos.)	Happiness (Neg.)
	(1)	(2)	(3)
Log Income	0.031	0.046^{*}	0.101^{***}
	(0.024)	(0.024)	(0.028)
Age	0.017	-0.05^{**}	0.021
	(0.024)	(0.024)	(0.029)
Male	0.206^{***}	0.06	-0.076
	(0.052)	(0.048)	(0.063)
Constant	0.046	-0.685^{***}	-0.294
	(0.295)	(0.165)	(0.383)
Observations	5,084	8,739	4,280
R^2	0.014	0.006	0.013
Adjusted R^2	0.012	0.005	0.011
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.25: Intensity (happiness) of anomalous periods of increased volatility (Model 1), sustained happiness (Model 2) and sustained unhappiness (Model 3). Coefficients of numeric variables are standardized. All standard errors are clustered at the participant level.

	Dependent variable:		
	Duration (Volatility)	Duration (Pos.)	Duration (Neg.)
	(1)	(2)	(3)
Log Income	−0.015	−0.009	0.022*
	(0.017)	(0.013)	(0.013)
Age	0.010	0.082***	0.027
	(0.015)	(0.018)	(0.018)
Male	0.014	0.001	−0.001
	(0.035)	(0.028)	(0.036)
Constant	−0.256***	−0.104	−0.233***
	(0.039)	(0.108)	(0.045)
Observations	5,084	8,739	4,280
R^2	0.001	0.006	0.003
Adjusted R^2	−0.001	0.005	0.001
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.26: Duration of anomalous periods of increased volatility (Model 1), sustained happiness (Model 2) and sustained unhappiness (Model 3). Coefficients of numeric variables are standardized. All standard errors are clustered at the participant level.

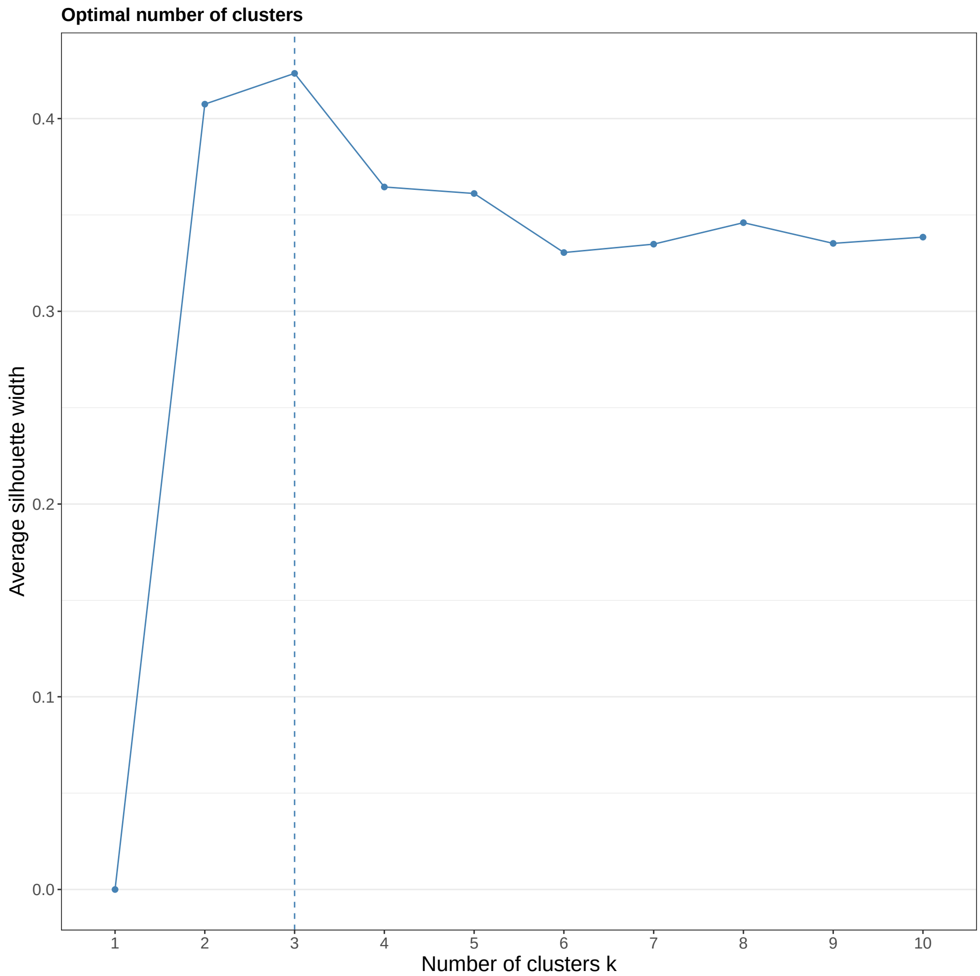


Figure 2.9: Optimal Number of clusters.

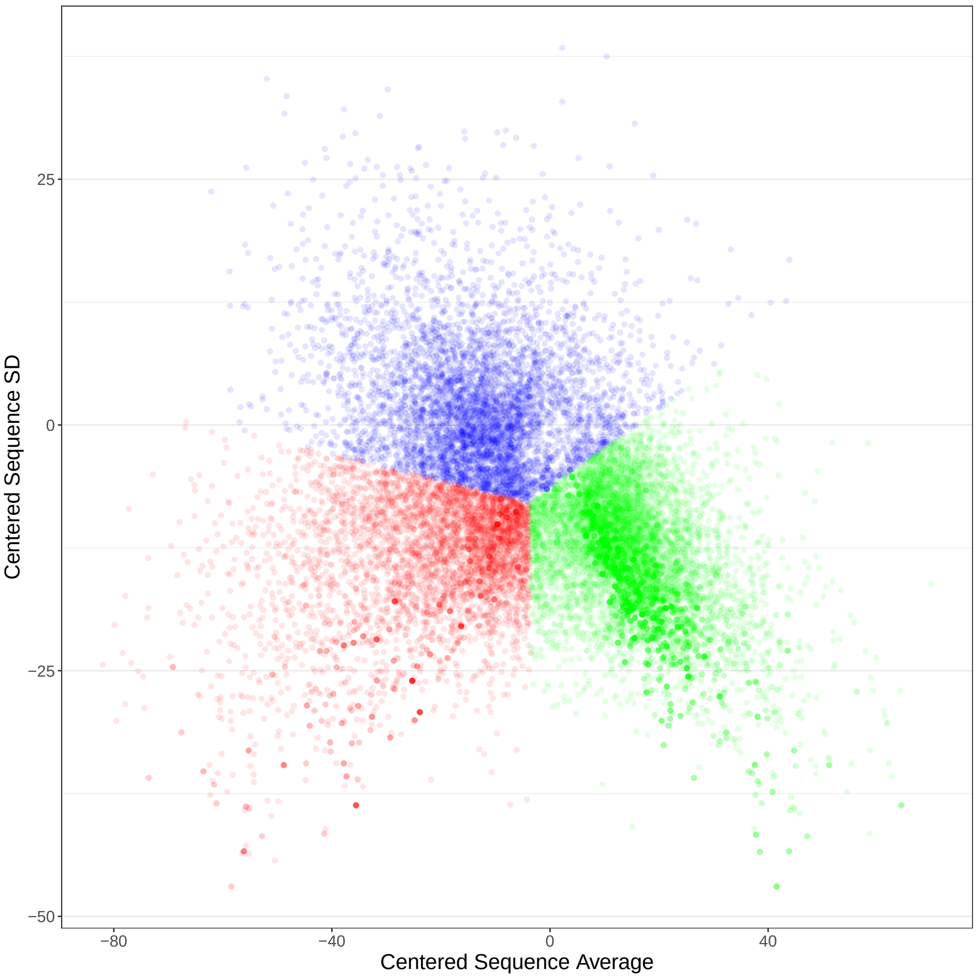


Figure 2.10: Clusters of anomalous sequences. Results of k-means clustering over our 18,103 anomalous sequences. Cluster 1 (blue) represents sequences with a higher emotional variability than an individual's typical variability. Cluster 2 (green) represents sequences that are on average happier than an individual typical mood. Cluster 3 (red) represents sequences that are lower on happiness than an individual's typical mood.

2.5.7. Note 7: Robustness of CAPA Results

Results from the CAPA analyses are robust to alternative, more conventional, ways to identify extreme observations. In this section, we show that focusing on the bottom 1%, 5%, and 10% of all happiness observations - either across individuals or within subjects - yields similar conclusions. In order to do so, we focus on our larger sample size of 17,278 individuals.

To ensure the robustness of our CAPA results, for each individual, we estimated the proportion of his or her happiness moments that fall within the bottom 1%, 5%, and 10% of all happiness observations across individuals. For those individuals that had at least 1 happiness report in the bottom 5% or 10% of all happiness observations across individuals, we also estimated the intensity (average happiness) of those extreme moments. We do not estimate the intensity of observations in the bottom 1% as these observations always take a value of 0. Table 2.27 presents the results (OLS coefficients) of regressing the proportion of extreme happiness observations on the logarithm of income (controlling for demographic characteristics). Table 2.28 presents the results of regressing the intensity (average happiness) of these extreme observations on income (controlling for demographic characteristics).

To further ensure the robustness of these results, we take a within-individual approach. For each participant in our sample, we estimated the intensity (average happiness) of the bottom 1%, 5%, and 10% of his or her happiness observations. Table 2.29 presents the results of regressing the intensity of these observation on income (controlling for demographic characteristics). For brevity, we don't report "country of residence" coefficients.

	Dependent variable:		
	Proportion 1%	Proportion 5%	Proportion 10%
	(1)	(2)	(3)
Log Income	-0.046^{***}	-0.049^{***}	-0.059^{***}
	(0.008)	(0.008)	(0.008)
Age	-0.057^{***}	-0.051^{***}	-0.050^{***}
	(0.008)	(0.008)	(0.008)
Male	-0.014	-0.054^{***}	-0.072^{***}
	(0.016)	(0.016)	(0.016)
Constant	0.264^{**}	0.135	0.078
	(0.118)	(0.118)	(0.117)
Observations	17,278	17,278	17,278
R^2	0.008	0.009	0.011
Adjusted R^2	0.008	0.008	0.010
Residual Std. Error	0.996	0.996	0.995
F Statistic	18.422^{***}	18.583^{***}	23.821^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.27: Proportion of observations per individual that belong to the bottom 1%, 5%, and 10% of all happiness observations across individuals. Coefficients of numeric variables are standardized.

	Dependent variable:	
	Happiness (Obs. in 5%)	Happiness (Obs. in 10%)
	(1)	(2)
Log Income	0.026^{**}	0.040^{***}
	(0.012)	(0.010)
Age	0.086^{***}	0.085^{***}
	(0.012)	(0.010)
Male	-0.065^{**}	0.016
	(0.025)	(0.020)
Constant	-0.267	-0.062
	(0.182)	(0.144)
Observations	7,941	11,671
R^2	0.011	0.012
Adjusted R^2	0.010	0.011
Residual Std. Error	0.995	0.994
F Statistic	10.823^{***}	17.449^{***}
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.28: Intensity (average happiness) of observations belonging to the bottom 5% and 10% of all happiness observations across individuals. Coefficients of numeric variables are standardized.

	Dependent variable:		
	Happiness (1%)	Happiness (5%)	Happiness (10%)
	(1)	(2)	(3)
Log Income	0.066***	0.072***	0.076***
	(0.008)	(0.008)	(0.008)
Age	0.081***	0.111***	0.109***
	(0.008)	(0.008)	(0.008)
Male	0.242***	0.212***	0.206***
	(0.016)	(0.016)	(0.016)
Constant	−0.144	−0.139	−0.136
	(0.116)	(0.116)	(0.116)
Observations	17,278	17,278	17,278
R^2	0.031	0.037	0.037
Adjusted R^2	0.030	0.037	0.037
Residual Std. Error	0.985	0.981	0.982
F Statistic	68.750***	83.600***	83.107***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.29: Intensity (average happiness) of observations belonging to the bottom 1%, 5% and 10% of happiness observations within individuals. Coefficients of numeric variables are standardized.

2.5.8. Note 8: Temporal Variability of Anomalous Happiness Observations and Periods

In the main text (Figure 2.4) we presented the GAM Smooth function of the prevalence of anomalous happiness reports (happiness reports belonging to either an anomalous happiness moment or an anomalous happiness period) across the month. For completeness, Figure 2.11 presents a similar analysis focusing exclusively on anomalous negative observations (either anomalous negative moments or period of sustained unhappiness). For low-income individuals, anomalous negative observations are more common during the first week of the month (i.e., when facing large expenditures such as rent or loan repayments) and during the last few days of the month (i.e., when waiting for their monthly salary). For high income individuals, the opposite picture arises, suggesting that the relationship between income and the negative anomalous moments and periods of unhappiness is causal.

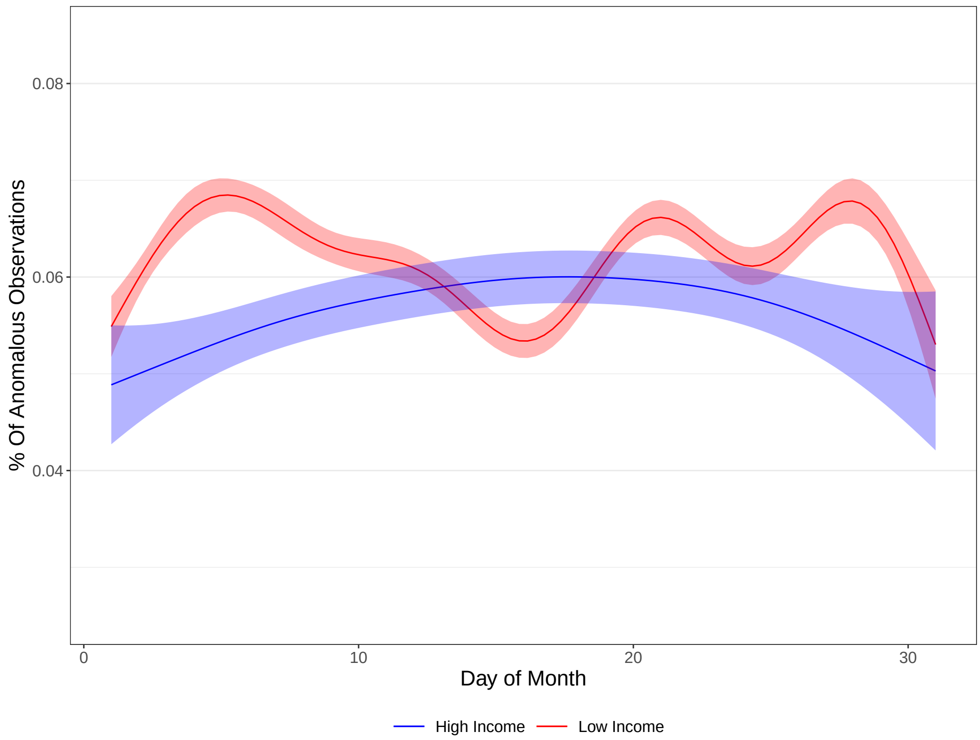


Figure 2.11: Differences in the proportion of anomalous negative observations between high (more than 3000 euros per month) and low-income (less than 3000 euros per month) individuals.

Chapter 3

INCOME, BOREDOM, AND MENTAL HEALTH

Joint with Daniel Navarro-Martinez and Jordi Quoidbach

Abstract

Using information on 65,000 individuals across 28 countries, we show that low-income individuals experience boredom more often - an emotional state linked to anti-social behavior, poor decision-making, and the development of addictions. Taking a network approach, we show that the experience of boredom is more closely associated with depressed and anxious mood for low-income individuals. As a result, income moderates the relationship between boredom and the experience of clinical depression episodes. On average, experiencing high levels of boredom is associated with an increase in the probability of suffering clinical depression of 10.9%. For each decrease of 1 SD in income, the difference in incidence of depression episodes across low and high-boredom individuals increases by 1.4%. Evaluating 13 different depression symptoms, we find a particularly robust association between boredom and the experience of morbid thoughts. Our results portray boredom as a potential poverty self-reinforcing mechanism, contribute to better understand the mental health consequences of this emotion, and open a venue to future research and policies that address the full extent of the emotional tax exerted by financial hardship.

3.1. Introduction

Understanding the relationship between income and well-being is key to comprehend people's daily behaviors and motivations, identify the full extent of a financially deprived life, and develop public programs and policies that actively improve people's lives. While decades of research seek to clarify this relationship, past work has either focused on hedonic (i.e. happiness) or eudaimonic (i.e. meaning) operationalizations of well-being. Yet, people not only want to live a happy or meaningful life, but a life that is filled with interesting experiences and novel or unique emotional moments - a "psychologically rich life" (Oishi et al., 2020; Oishi and Westgate, 2021). In other words, people are not only motivated to pursue happy or meaningful experiences in their everyday lives, but actively seek to avoid a life that is monotonous, uneventful, and boring (Oishi et al., 2020; Oishi and Westgate, 2021). Hence, in order to fully comprehend the relationship between income and well-being, we need to move past happiness or meaning, and also consider how income relates to boredom.

Beyond being an important part of a person's well-being, boredom shapes our everyday lives through a plethora of behavioral, motivational, and mental health consequences. Research on the consequences of boredom has demonstrated that this emotion leads to impulsivity and present bias (Moynihan et al., 2017; Watt and Vodanovich, 1992), noise and decision errors (Wolff et al., 2022; Yakobi and Danckert, 2021), aggression and antisocial behavior (Dahlen et al., 2004; Pfattheicher, Lazarevic, Nielsen, et al., 2021; Pfattheicher, Lazarevic, Westgate, et al., 2021; Rupp and Vodanovich, 1997; Yucel and Westgate, 2021), and the development of addictions (Blaszczynski et al., 1990; Iso-Ahola and Crowley, 1991; Sommers and Vodanovich, 2000). This line of research has also shown that bored individuals tend to engage more often in risky behaviors such as gambling, unprotected sex, or risky driving (Biolcati et al., 2018; Blaszczynski et al., 1990; Dahlen et al., 2005), but do not show reduced risk aversion in economic choices (Pirla and Navarro-Martinez, 2021). In addition to poor decision-making, boredom has been linked to an array of mental health issues (Farmer and Sundberg, 1986; Goldberg and Danckert, 2013; Masland et al., 2020; Sommers and Vodanovich, 2000; Todman, 2007; Vodanovich et al., 1991), including the development of depression and anxiety disorders.

Several reasons indicate that low-income individuals are more likely to suffer from the negative consequences of boredom. First, low-income individuals tend to hold more repetitive and monotonous jobs. Second, low-paying jobs are also characterized by a lack or individual agency, an important determinant of boredom (Raffaelli et al., 2018; Struk et al., 2021). Third, compared to high-income earners, low-income individuals have a limited ability to outsource boring tasks (i.e.

pay someone to clean the kitchen). Finally, laboratory evidence suggests that consumers can predict the experience of boredom in advance (Dal Mas and Wittmann, 2017). Accordingly, these consumers adjust their willingness to pay for stimulation when expected to get bored (Dal Mas and Wittmann, 2017). Given budget constraints, low-income individuals have a smaller choice set when it comes to entertainment and stimulation that high-income individuals. Altogether, low-income individuals encounter boredom more often in their daily lives.

Yet, past work has only identified a very small or non-existent relationship between income and boredom (Chin et al., 2017; Robinson, 1975). In this paper, we contribute to past literature by showing a robust association between income and boredom. Using data from 40,819 individuals across 22 European countries, we show that low-income individuals experience boredom more often than high-income earners. Beyond experiencing boredom more often, the affective experience of boredom is more closely associated with other negative emotions (especially depressed and anxious mood) for low-income individuals. The relationship between income and boredom is not unique to our European sample. Employing an additional dataset of over 25,000 individuals from 6 developing countries, we are able to replicate our main findings. As this dataset contains information on mental health outcomes, we further demonstrate the practical consequences of our results by showing that income moderates the relationship between boredom and depression. On average, experiencing high levels of boredom is associated with an increase in the probability of suffering clinical depression of 10.9%. For each decrease of 1 SD in income, the difference in incidence of depression episodes across low and high-boredom individuals increases by 1.4%. Finally, focusing on a reduced subset of 1,907 individuals that experienced a depression episode in the last 12 months, we outline the specific symptoms associated with boredom. High-boredom individuals experience 6 depression symptoms significantly more often than low-boredom individuals. Specifically, high-boredom individuals are more prone to suffer from loss of appetite, sleeping problems, and feelings of hopelessness, anxiety, and restlessness. Across symptoms, boredom is particularly linked to morbid thoughts. On average – and controlling for other negative emotions and demographic characteristics – high boredom levels are associated with a relative increase in the prevalence of morbid thoughts of 30% (30% for low-boredom individuals vs. 39% for high-boredom individuals). Our results portray boredom as a potential poverty self-reinforcing mechanism, contribute to better understand the mental health consequences of this emotion, and open a venue to future research and policies that address the full extent of the emotional tax exerted by financial hardship.

3.2. Methods

3.2.1. European Social Survey

Participants:

The European Social Survey (ESS) is a publicly available multi-country survey. In its 2006 wave, the ESS included information on 47,099 participants from 25 European countries. As information on income is missing from 3 countries, our final sample size consists of 40,819 participants from 22 countries. The ESS has been used to study a wide array of topics ranging from immigration attitudes (Card et al., 2005) to the determinants of physical health (Verbakel et al., 2017).

Income:

The ESS contains a self-reported measure of income. Participants disclosed their household income on a 12-point bracket scale. Missing responses (approximately 25% of the income observations) were imputed using a random forest approach and over 50 demographic, social, affective, and labor market variables (see Supplemental Materials Note 1 for details). Note that our main results remain unchanged when using only non-missing income reports (see Supplemental Materials Note 2).

Boredom and other affective states:

Affective states were measured on a 4-point Likert scale representing the prevalence of an emotion during the past week (from 1 – "None or almost none of the time" to 4 – "All or almost all of the time". In this study, we use information on boredom (exclusively collected in the ESS 2006 wave) and 7 additional negative emotions measured on the same scale ("Depressed", "Effortful", "Lonely", "Sad", "Anxious", "Tired", and "Could not get going").

3.2.2. World Health Organization Study on Global Aging and Adult Health

Participants:

The World Health Organization Study on Global Aging and Adult Health (WHO SAGE) gathers information on 25,739 participants from 6 low and middle-income countries (China, Ghana, India, Mexico, Russia, and South Africa). This dataset has been previously used to study, for instance, the relationship between income and emotional volatility (see Chapter 2).

Income:

In the WHO SAGE study, income is estimated indirectly using information on an individual's access to services and ownership of durable goods. More specifically, each participant provides information on a country-specific list of 21 items (e.g., do you own a refrigerator?). Using this information, households are positioned on an "asset ladder" and a measure of permanent income is obtained using a post-Bayesian estimator. In developing countries, this measure of permanent income is more reliable than self-reported questions (Ferguson et al., 2003).

Boredom and other affective states:

Boredom and other affective states (including worried, angry, or depressed mood) were measured on a binary scale, representing whether the participant experienced (or not) a specific feeling for much of the past day.

Depression Episodes:

The WHO SAGE dataset includes a module on depression and depressive episodes. Participants are asked whether in the last 12 months they experienced a feeling of depression, loss of interest, or increased tiredness during most of the day for a period of more than two weeks. Using this information, we constructed a binary measure indicating whether an individual experienced a depression episode in the last 12 months. A total of 1,907 individuals experienced depression episodes. For this subset of participants, the WHO SAGE dataset contains information on the prevalence of 13 different depression symptoms (measured on a binary scale). The list of symptoms and consequences includes physical and physiological issues (slowing down in your moving around, loss of appetite, decreased interest in sex), cognitive problems (slowing down in your thinking, difficulties concentrating), sleeping problems (falling asleep, waking up too early), affective issues (feeling anxious, restless, hopeless, or losing confidence), morbid thoughts (thinking or wishing to be dead), and suicide attempts.

3.2.3. Linear Models

Cultural, institutional, and social differences might impact baseline levels of reported boredom across countries. Similarly, cross-counties differences might affect the strength of the relationship between income and boredom. To account for this structure of random and fixed country-level effects, we use Linear Mixed Models (LMM, Bates, 2007). That is, our LMM estimates accommodate the existing variability in baseline differences in boredom and allow the effect of income on boredom to vary across countries. We also use LMM to account for the country-level structure of our WHO dataset in both our moderation analysis and our assessment of the impact of boredom on specific depression symptoms.

3.2.4. Non-linear Models

We confirm our Linear Mixed-Models effects using non-linear Generalized Additive Models (GAM, Hastie and Tibshirani, 1987; Wood, 2003; Wood and Augustin, 2002). This approach allows us to model the participants' levels of boredom as a linear combination of non-linear effects. In doing so, we are able to precisely estimate the shape of the relationship between income and boredom and its satiation points. As in our LMM models, our GAM specifications include country-specific random and fixed effects.

3.2.5. Lasso-Regularized Partial Correlation Network

Using data from the ESS, we estimate a Lasso-regularized Partial Correlation Network of negative emotions for low and high-income individuals. In order to test differences across income groups, we fist split our sample by selecting the top and bottom 10% of income earners for each country in our dataset. In doing so, we respect the country composition of our ESS dataset and ensure that the size of the different income groups is similar - a condition needed to perform valid permutation tests (Epskamp et al., 2018; Epskamp and Fried, 2018). This leads to a sample of 7,508 individuals in the low-income group and 8,210 in the high-income group. For each group, we estimate a lasso-regularized partial correlation network. That is, for each pair of emotions, we estimate their Lasso correlation controlling for all other negative emotions. We test differences across groups using a permutation test (Epskamp et al., 2018; Epskamp and Fried, 2018).

3.3. Results

To provide a first approximation at whether earning a higher income is associated with living a less boring life, we used data from a sample of over 40,819 individuals included in the European Social Survey (EES 2006). Consistent with previous literature on the impact of income on emotional states (Brown and Gathergood, 2020; Ferrer-i-Carbonell, 2005; Quispe-Torreblanca et al., 2021), we took the logarithm of income as our main explanatory variable. Using mixed-models (to account for country random and fixed-effects), we find that a higher income is associated with experiencing boredom less often ($\beta_{log(\text{Income})}$)) = - .135, t = - 9.303, p $<$ .00001) . This relationship is robust to including additional control variables such as demographic information (age, gender, and education), employment status, and job characteristics (job interestingness and agency at work) – see Supplemental Materials Note 2. The logarithmic relationship between income and boredom (confirmed with non-linear Generalized Additive Models, see Supplemental Materials Note 2) suggests that this relationship is stronger for lower levels of income. For example, we find a difference of 0.3 SD in boredom frequency between those with a monthly income of less than 1,000 euros and those earning between 1,000 and 2,000 euros per month. On the other hand, those making between 2,000 and 3,000 euros per month only experience a reduction in boredom frequency of 0.13 SD compared with those making between 1,000 and 2,000 euros per month. Using non-linear methods (see Figure 3.1), we estimate the satiation point of the relationship between income and boredom at an average of 30,000 euros in yearly income.

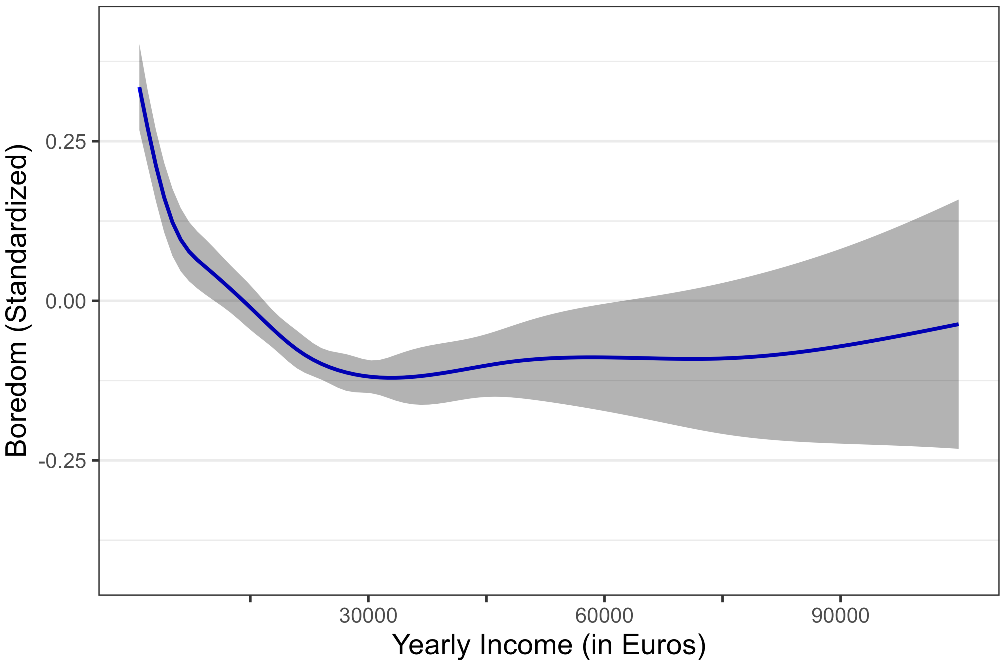


Figure 3.1: Predicted boredom frequency across income groups. Data: EES Survey (N = 40,819). Predicted boredom frequency is obtained using non-linear methods (Generalized Additive Models) accounting for country-level fixed and random effects.

Looking at the data from each country independently, we find a similar pattern of results. The zero-order correlation between the logarithm of income and the frequency of experienced boredom is negative and significant for 16 of the 22 countries included in the European Social Survey (see Figure 3.2). The median correlation coefficient across countries is r= - .09, with the correlation ranging from r = - .13 (Finland) to r = .08 (Latvia). We only find a positive and significant relationship between income and boredom for the Latvian sample.

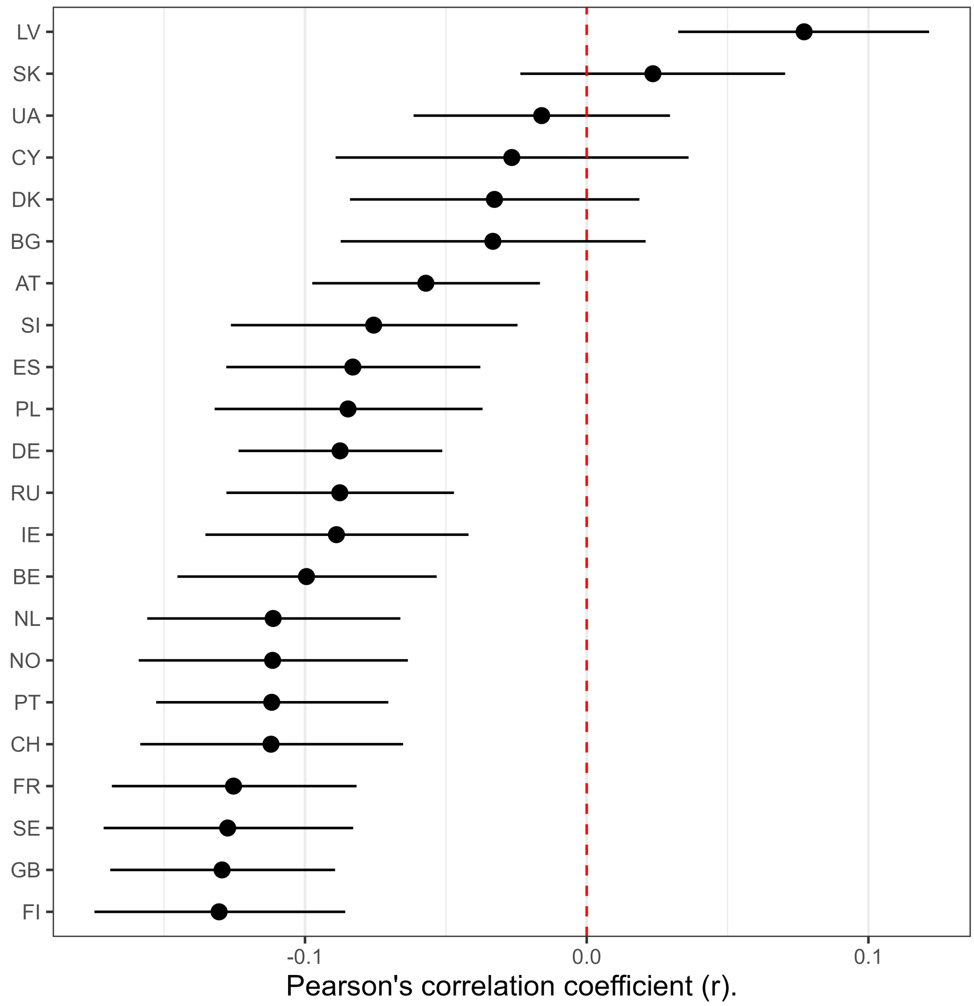


Figure 3.2: Pearson's correlation coefficient (with 95% confidence intervals) for the zero-order relationship between the logarithm of income and boredom across 22 countries.

Beyond experiencing boredom more often, low-income individuals might encounter more difficulties when regulating the intensity of this emotion. As intense boredom has been shown to lead to both depressed and anxious mood (Elhai et al., 2018; Goldberg and Danckert, 2013; Sommers and Vodanovich, 2000), this failure to regulate boredom would imply that this emotion is more closely associated with experiencing depressed and anxious mood for low-income individuals. To investigate this hypothesis, we take a complex systems approach to analyze the effect of income on the structure of affect across individuals. Using data from 8 different negative emotional states (depressed, effortful, lonely, sad, could not get going, anxious, tired, bored), we estimated a Lasso-regularized partial correlation network of negative emotions for individuals in the bottom and top 10% of the income distribution of each country (see Figure 3.3). Running a permutation test, we find that boredom is more central to the network of negative emotions for low-income individuals than for high income earners (p = .038). That is, for low-income individuals, experiencing boredom is more closely associated with their propensity to experience other negative emotions. Looking at the specific relationship between emotions, we find that the difference in centrality of boredom across income groups is driven by the association of boredom with anxious and depressed mood. In fact, boredom is significantly more closely associated with depressed (Diff = .057, p = .034) and anxious mood (Diff = .066, p = .016) for low-income individuals. The lasso-regularized partial correlations between boredom and depressed or anxious mood are three times larger for low-income individuals ($r_{\text{depressed}}$ = .093, r_{anxious} = .095) than for high-income earners ($r_{\text{depressed}}$ = .036, r_{anxious} = .029).

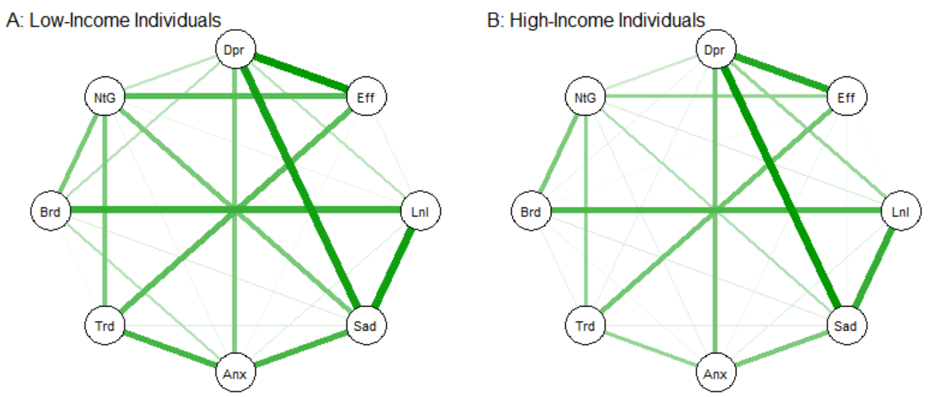


Figure 3.3: Networks of negative emotions for low (A) and high-income individuals (B). Each node represents an affective state. Affective states include depressed (Dpr), effortful (Eff), lonely (Lnl), sad (Sad), anxious (Anx), tired (Trd), bored (Brd) and "could not get going" (NtG). The thickness of the links between nodes represents how closely associated two emotions are.

Our results not only prove that boredom is more often experienced among low-income individuals, but also show that boredom is more closely associated with depressed or anxious mood for these individuals. These results suggest that the relationship between boredom and poor mental health might be moderated by income. While our sample of European individuals does not contain information on mental health outcomes, we can further investigate the relationship between income, boredom, and mental health outcomes using a sample of 25,000 individuals from 6 developing countries (WHO SAGE dataset). Using this sample of individuals, we are able to replicate our main relationship of interest – low-income individuals are more likely to experience high levels of boredom (β_{Income} = - .031, t = -15.598, p $<$ .00001) - a relationship that holds when controlling for demographic characteristics (such as age or gender) and similarly valenced emotions such as anxiety, worry, or depressed mood (β_{Income} = - .017, t = -9.314, p $<$ .00001). To put these numbers in context, on average, those with an income of 1 SD above their national average experience high levels of boredom with a probability of 9.78%. On the other hand, those individuals with an income of 1SD below the national average experience high boredom levels with a probability of 13.2%. Hence, compared with those making 1SD above the national average, those with an income of 1SD below the mean experience a 35% relative increase in their propensity to experience high levels of boredom. Within countries, we find a significant correlation between income and boredom for Russia (r = -0.12, t = -5.744, p $<$ .00001), India (r = -0.12, t = -10.709, p $<$ .00001), China (r = -0.11, t = -10.925, p $<$ .00001) and South Africa (r = -0.05, t = -2.25, p = 0.0245). We also find negative but non-significant correlations for Mexico (r = -0.05, t = -1.235, p = 0.217) and Ghana (r = -0.03, t = -1.531, p = 0.126).

For participants in the WHO SAGE dataset, apart from data on income and boredom, we have information on whether they experienced a depressive episode in the last 12 months (i.e. a period lasting a minimum of two weeks during which the individual experienced depressed mood, loss of interest, or lack of energy during most of the day). Using Linear mixed-models (and controlling for age and gender) we analyze how income, boredom, and their interaction are related to the experience of depression episodes. As expected, low-income individuals are more prone to experience depression episodes (β_{Income} = - .004, t = - 2.710, p = .0067). On average, those with an income 1SD above the national average experienced a depression episode in the last 12 months with a probability of 5.2%. On the other hand, those individuals with an income 1SD below the national average experience a depression episode with a probability of 6%. Boredom is also associated with the experience of depression episodes (β_{Boredom} = .109, t = 23.823, p $<$ .00001). Crucially, this relationship is significantly moderated by income ($\beta_{\text{Boredom x Income}}$ = - .014, t = -3.118, p = .0018). For example, for high income earners (1SD above the national average), experiencing high levels of boredom is associated with an increase in an individual's propensity to experience depression episodes of 9.5 percentage points ($P_{\text{Low Boredom}}$ = 4.4%, $P_{\text{High Boredom}}$ = 13.9%). For low income earners (1SD below the national average), experiencing high levels of boredom is associated with an increase in an individual's propensity to experience depression episodes of 12.3 percentage points ($P_{\text{Low Boredom}}$ = 5.2%, $P_{\text{High Boredom}}$ = 17.5%). That is, the relationship between boredom and the experience of depression episodes is 30% stronger for low-income individuals than for high-income earners.

We can further analyze the effect of boredom on mental health by analyzing its relationship with specific depression symptoms. Our dataset includes information on the prevalence of depressive symptoms across 1907 individuals who reported experiencing periods of sustained depressed mood, loss of interest, or increased tiredness. Figure 3.4 presents the coefficient of income on each of the 13 symptoms of depression included in our dataset. All the models include demographic controls (age, gender, and income), country-level fixed effects, and affective controls (depressed, worried, and stressed mood). Experiencing boredom is significantly associated with 6 of the 13 symptoms of depression included in our dataset. The relationship between boredom and depressive symptoms is particularly robust when looking at an individual's feelings of hopelessness or restlessness, and when considering whether an individual thought or wished to be dead. Controlling for the previously mentioned demographic and affective characteristics, depressed individuals reporting high levels of boredom are 9.3% more likely to think about death or wish they were dead (29.8% to 39%, over a 30% relative increase in the prevalence of morbid thoughts).

3.4. Discussion

Decades of research seek to understand what it means to live in financial scarcity. Our results suggests that, by ignoring the relationship between income and boredom, past work has overlooked an important aspect of what constitutes a financially deprived life. Across a set of 28 countries and over 65,000 individuals, we find a robust association between income and boredom. Compared with high-income earners, low-income individuals not only experience boredom more often, but their experience of boredom is more closely associated to other negative emotions such as depressed or anxious mood. Finally, while experiencing high levels of boredom is predictive of poor mental health outcomes for both high and low-income individuals, we find a stronger association between boredom and depression for low-income earners. Across depression symptoms, our results point at a particularly robust association between boredom and the experience of morbid thoughts.

Our results carry important theoretical and practical implications. First, our results suggest that - through worsened decision-making - boredom might act as a

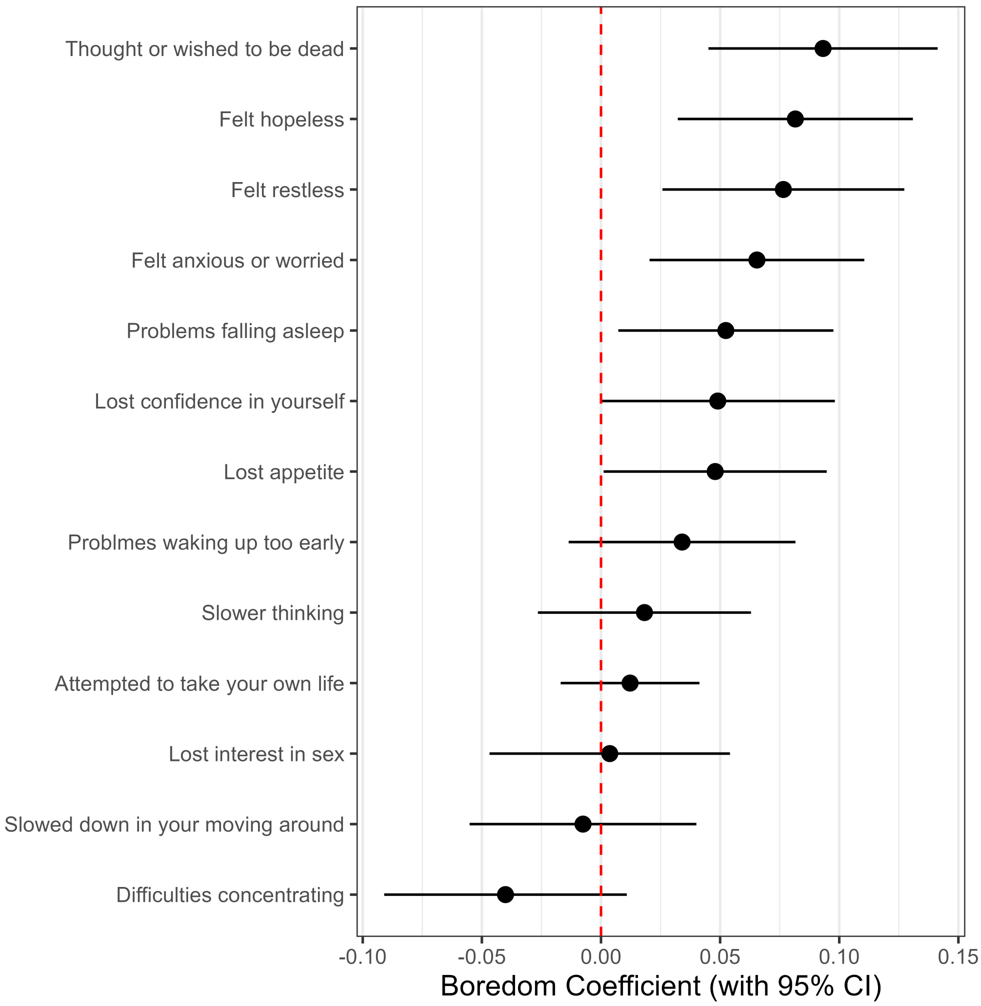


Figure 3.4: Association between boredom and depression symptoms controlling for country-level fixed effects, demographic information (age, gender, and income), and affective variables (depressed, worried, and stressed mood). Point estimates represent the increase in the prevalence of each symptom for high (vs. low) boredom individuals.

poverty self-reinforcement mechanism. Bored individuals are more prone to make impulsive choices (Moynihan et al., 2017; Watt and Vodanovich, 1992), fall prey to decision errors (Wolff et al., 2022; Yakobi and Danckert, 2021), behave antisocially (Dahlen et al., 2004; Pfattheicher, Lazarevic, Nielsen, et al., 2021; Pfattheicher, Lazarevic, Westgate, et al., 2021; Rupp and Vodanovich, 1997; Yucel and Westgate, 2021), and develop addictions (Blaszczynski et al., 1990; Iso-Ahola and Crowley, 1991; Sommers and Vodanovich, 2000). As boredom is more commonly experienced among low-income individuals, our results suggest that this emotion contributes to the perpetuation of financial scarcity. Our findings, therefore, contribute to the literature on psychologically driven poverty traps (Haushofer, 2019; Haushofer and Fehr, 2014; Ridley et al., 2020), illustrating a promising new area of research and opening a venue to the design of public policies that consider the full extent of the emotional tax exerted by financial scarcity. Second, our results improve our understanding of the mental health consequences of boredom. While past literature has established a link between boredom and poor mental health outcomes (Farmer and Sundberg, 1986; Goldberg and Danckert, 2013; Sommers and Vodanovich, 2000; Vodanovich et al., 1991), the mechanisms linking this emotion to the experience of psychopathologies remain unclear. By considering the relationship between boredom and each depression symptom independently, our results illuminate several potential mechanisms linking the experience of boredom to depression disorders. Third, given the robust association between boredom and the experience of depression symptoms, our results can be of practical relevance in forecasting and treating depression episodes – especially among low-income individuals and chronically bored populations - i.e., after a traumatic accident, see Goldberg and Danckert (2013). In fact, as presented in the Supplemental Materials Note 3, the relative importance of boredom in explaining the presence of depression episodes is statistically indistinguishable in size from that of stress and significantly larger than that of income, age, or gender. Finally, our results show that boredom is especially problematic among low-income individuals. Future research aimed at understanding the situational and contextual determinants of boredom among low-income individuals would pave the way to a better understanding of the relationship between financial scarcity and mental health outcomes.

Our findings, nevertheless, present a number of limitations. First, the observational nature of our data does not allow us to make causal inferences. Second, our measures of income and boredom represent imperfect estimates. Boredom is measured either as a 4-point or a binary scale. Similarly, income in our ESS dataset is measured on a 12-point interval scale. Although imperfect measures should not lead to biased estimates, it can lead to mitigated effect sizes. Third, we find substantial heterogeneity in the effect of income on boredom across countries. Although some of these differences might be due to sampling variability or measurement error, further work needs to investigate the cultural, social, and institutional factors shaping this relationship. Fourth, while our network approach allows us to analyze how boredom is - on average - associated with other negative emotions, our cross-sectional data does not allow us to evaluate the dynamic associations between boredom and other affective states. Further work needs to clarify the dynamic associations between boredom and other negative emotions, and its relationship to both income and mental-health outcomes. Fifth, recent work has identified different boredom profiles streaming from distinct combinations of attention and meaning deficits (Westgate, 2020; Westgate and Wilson, 2018). Our network approach suggest that the affective experience of boredom is different across income groups, but further research needs to clarify the origins – in terms of attention and meaning – of these differences and its impact on mental health. Finally, while our main findings replicate across 28 countries and over 65,000 individuals, our investigation on the relationship between boredom and depression symptoms is performed on a limited sample of 1,907 participants that experience a depression episode in the past 12 months. Future research needs to clarify the generalizability of the associations found on this limited subset of participants.

Recent work shows that people not only want to live a happy or meaningful life, but an exciting one (Oishi et al., 2020; Oishi and Westgate, 2021). Our result show that boredom plays a major role in explaining the relationship between income, well-being, and mental health.

3.5. Supplemental Materials

3.5.1. Note 1: Summary Statistics

Main Summary Statistics

Imputation Procedure

Approximately 25% of the income observations are missing. Removing these observations from our analyses could bias our results (Schenker et al., 2006). To gain a more robust understanding of the relationship between our variables of interest, we impute the missing income observations using a random forest approach. In Note 2, we show that our main results are robust to removing our missing observations.

To impute income, we selected a set of 57 variables from the European Social Survey. These variables included demographic information (e.g., age, gender, education), employment status and characteristics (e.g., job status, past unemployment spells, job satisfaction), and well-being and emotion measurements (e.g.,

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Age	41,626	47.321	18.573	14.17	32.25	61.58	101.33
Income	31,237	28,178	27,093	900	9,000	48,000	150,000
Boredom	41,281	1.471	0.681	1	1	2	4
Depressed Mood	41.416	1.518	0.703	1	1	2	4
Sadness	41,475	1.616	0.71	1	1	2	4
Anxious Mood	41.428	1.655	0.745	1	1	2	4
Effortfulness	41,518	1.754	0.804	1	1	2	4
Loneliness	41,492	1.483	0.749	1	1	2	4
Could not get going	41.016	1.615	0.732	1	1	2	4
Tiredness	41,618	2.048	0.749	1	2	2	4

Table 3.1: Summary Statistics: Main Numeric Variables

Variable	N	Percent
Gender	41,825	
... Female	22,801	54.5%
... Male	19,024	45.5%
Education	41,713	
... Less than lower secondary education (ISCED 0-1)	5,661	13.6%
... Lower secondary education completed (ISCED 2)	7,744	18.6%
... Post-secondary non-tertiary education completed (ISCED 4)	981	2.4%
... Tertiary education completed (ISCED 5-6)	11,287	27.1%
... Upper secondary education completed (ISCED 3)	16,040	38.5%

Table 3.2: Summary Statistics: Main Categorical Variables

life satisfaction, boredom, sadness). Using these variables, we imputed the missing income observations using a random forest approach. We implemented our imputation procedure using the missRanger package in R (Mayer, 2019). Table 3.3 presents an overview of the income distribution before and after imputation.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Income (before imputation)	31,237	28,178	27,093	900	9,000	48,000	150,000
Income (after imputation)	41,925	26,676	25,757	900	9,000	33,000	150,000

Table 3.3: Income summaries before and after imputation.

3.5.2. Note 2: Robustness Checks (Main Results, ESS)

As stated in the main text, we use Linear Mixed models with country-level random and fixed effects to derive our main results. In this note, we show that our results are robust to including job characteristics and employment status as additional control variables. We also show that our main results hold when we focus on the smaller subset of individuals with non-missing income responses. Finally, we show that our results are robust to using non-linear methods (Generalized Additive Models).

Additional Control Variables

Table 3.4 presents our main robustness checks. The coefficients of all numeric variables are standardized. Our results show that the relationship between income and boredom is significant even when controlling for employment status or job characteristics such as job interestingness (Job Int) or agency at work (Job Agency). As shown in Model 4 of the same table, the relationship between income and boredom is significantly stronger for the unemployed.

Excluding Imputed Observations

A similar pattern of results emerges when we focus exclusively on complete observations. Table 3.5 presents the results of these analysis. If anything, the relationship between income and boredom is stronger when focusing exclusively on complete observations.

Non-Linear Methods

In this section, we confirm our main results using non-linear methods. More specifically, we use Generalized Additive Models (GAMs) to estimate the same set of models but allowing for a non-linear relationship between all our explanatory numeric variables and boredom. As in our previous sections, all our models include country-level fixed and random effects. Table 3.6 presents the main results. Note that the coefficients obtained using generalized additive models are hardly interpretable. Yet, one can readily see that the relationship between income and boredom is significant in all models. To gain a better understanding of the nature of this relation, we plot the non-linear marginal effects of income on boredom for all our models in Figure 3.5. The variables included in each model are those of the corresponding model in the previous section. Hence, model 5 of table 3.6 (or plot E of Figure 3.5) represents the marginal effects of income on boredom when controlling for job characteristics (as in model 5 of table 3.5).

3.5.3. Note 3: Robustness Checks (WHO SAGE Dataset)

In this section, we provide a more detailed overview of the analyses conducted using the WHO SAGE dataset. Note that boredom (as the other emotion measures) is a binary variable. Our measure of permanent income was scaled to represent a within-country standardized measure of income. The income coefficients represent, therefore, the marginal increase in the probability to report large levels of boredom for a 1 SD increase in within-country income.

Robustness Check - Main Results

We present the main robustness checks in Table 3.7. Our models include country-level fixed effects. Given the small number of countries, we abstained from including random effects in our specifications.

Moderation Analysis

Table 3.8 presents our moderation analyses. Again, our specifications include country-level fixed effects. The dependent variable is a binary indicator representing whether an individual experience a depression episode in the last 12 months. Hence, the regression coefficients represent the marginal effect of each variable in the probability to experience depression episodes.

	Dependent variable:				
	Boredom				
	(1)	(2)	(3)	(4)	(5)
Income	-0.135^{***}	-0.117^{***}	-0.089^{***}	-0.118^{***}	-0.058^{***}
	(0.015)	(0.013)	(0.012)	(0.013)	(0.018)
Education		-0.119^{***}	-0.091^{***}	-0.091^{***}	-0.048^{***}
		(0.012)	(0.010)	(0.010)	(0.008)
Age		-0.070^{***}	-0.099^{***}	-0.100^{***}	-0.088^{***}
		(0.005)	(0.005)	(0.005)	(0.007)
Female		-0.025^{***}	-0.043^{***}	-0.042^{***}	-0.043^{***}
		(0.010)	(0.010)	(0.010)	(0.013)
Income*Employed				0.058^{***}	
				(0.010)	
Employed			-0.214^{***}	-0.213^{***}	
			(0.011)	(0.011)	
Job Int.					-0.125^{***}
					(0.007)
Job Agency					-0.058^{***}
					(0.007)
Constant	0.006	0.013	0.134^{***}	0.129^{***}	0.046
	(0.049)	(0.049)	(0.050)	(0.050)	(0.052)
Observations	40,819	40,348	40,246	40,246	21,479
Log Likelihood	−56,052.980	−55,088.810	−54,755.890	−54,743.840	−29,232.270
Akaike Inf. Crit.	112,118.000	110,201.600	109,537.800	109,515.700	58,492.540
Bayesian Inf. Crit.	112,169.700	110,304.900	109,649.600	109,636.100	58,604.190
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 3.4: Robustness Checks - Main Results ESS. Standard errors are in parenthesis.

	Dependent variable:				
	Boredom				
	(1)	(2)	(3)	(4)	(5)
Income	-0.177^{***}	-0.161^{***}	-0.129^{***}	-0.158^{***}	-0.067^{**}
	(0.021)	(0.021)	(0.020)	(0.021)	(0.025)
Education		-0.107^{***}	-0.083^{***}	-0.083^{***}	-0.051^{***}
		(0.011)	(0.009)	(0.008)	(0.009)
Age		-0.089^{***}	-0.124^{***}	-0.125^{***}	-0.100^{***}
		(0.006)	(0.006)	(0.006)	(0.007)
Female		-0.044^{***}	-0.060^{***}	-0.059^{***}	-0.060^{***}
		(0.011)	(0.011)	(0.011)	(0.015)
Income*Employed				0.058^{***}	
				(0.012)	
Employed			-0.214^{***}	-0.212^{***}	
			(0.013)	(0.013)	
Job Int.					-0.136^{***}
					(0.008)
Job Agency					-0.064^{***}
					(0.008)
Constant	0.003	0.020	0.147^{***}	0.143^{***}	0.057
	(0.048)	(0.045)	(0.047)	(0.047)	(0.051)
Observations	30,489	30,264	30,199	30,199	17,000
Log Likelihood	−41,821.970	−41,287.910	−41,063.650	−41,055.850	−23,118.160
Akaike Inf. Crit.	83,655.940	82,599.820	82,153.300	82,139.700	46,264.320
Bayesian Inf. Crit.	83,705.890	82,699.640	82,261.400	82,256.110	46,372.690
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 3.5: Robustness Checks - Complete Observations Only. Standard errors are in parenthesis.

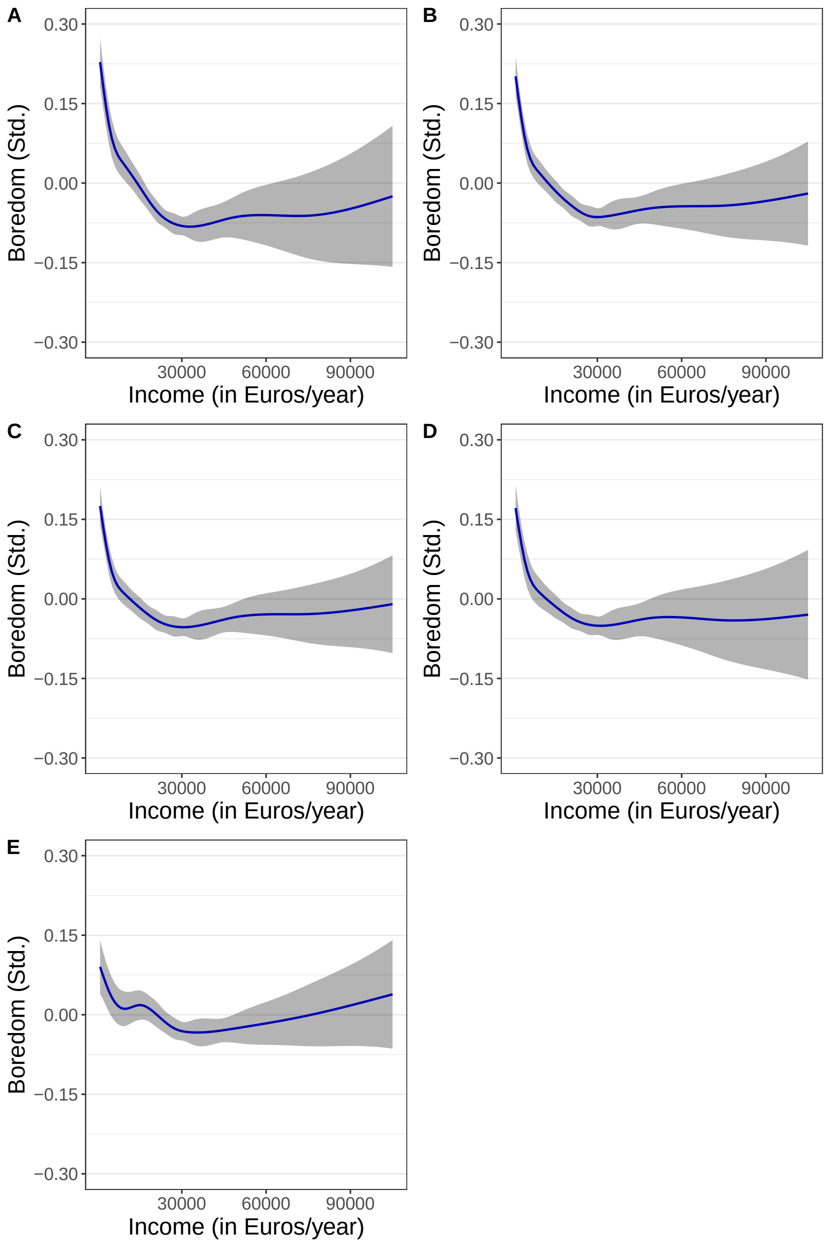


Figure 3.5: Plots of non-linear marginal effects of income on boredom for the five different models presented in this section. Shaded areas represent the 95% confidence interval of the effect.

	Dependent variable:				
	Boredom				
	(1)	(2)	(3)	(4)	(5)
Income Statistics:					
Edf	8.531	8.5066	8.3973	8.3284	6.9617
Ref.df	8.917	8.908	8.866	8.836	7.810
F	27.24	21.307	16.58	14.01	4.406
P Value	$<$0.00001	$<$0.00001	$<$0.00001	$<$0.00001	0.00006

Table 3.6: Robustness Checks - Generalized Additive Models

Relative Importance of Boredom in Predicting Depression Episodes

In the discussion section, we claim that "the relative importance of boredom in explaining the presence of depression episodes is statistically indistinguishable in size from that of stress and significantly larger than that of income, age, or gender". In this section, we present the analyses that lead to this conclusion.

To analyze the relative importance of different variables in explaining the presence of depression episodes, we use a relative importance approach (Genizi, 1993; Groemping, 2007; Groemping and Matthias, 2021). In short, this method allow us to estimate the average marginal variance in the dependent variable explained by each independent regressor across model permutations. We implement our analyses using the "relaimpo" package in R (Groemping and Matthias, 2021). We construct the confidence intervals of the relative importance of each independent variable by taking 1000 bootstrap samples. Table 3.9 presents the relative importance of each regressor (proportion of variance in variable "depression" explained on average) and its 95% confidence intervals.

	Dependent variable:		
	Boredom		
	(1)	(2)	(3)
Income	-0.031^{***}	-0.029^{***}	-0.017^{***}
	(0.002)	(0.002)	(0.002)
Male		-0.022^{***}	-0.010^{**}
		(0.004)	(0.004)
Age		0.017^{***}	0.014^{***}
		(0.002)	(0.002)
Stressed Mood			0.085^{***}
			(0.006)
Depressed Mood			0.247^{***}
			(0.008)
Worried Mood			0.148^{***}
			(0.007)
Constant	0.109^{***}	0.117^{***}	0.063^{***}
	(0.016)	(0.018)	(0.010)
Observations	25,595	25,555	25,523
Log Likelihood	−6,714.045	−6,684.681	−4,896.586
Akaike Inf. Crit.	13,436.090	13,381.360	9,811.172
Bayesian Inf. Crit.	13,468.690	13,430.250	9,884.498
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3.7: WHO SAGE Dataset - Robustness Checks. Standard errors are in parenthesis.

	Dependent variable:	
	Depression	
	(1)	(2)
Boredom	0.112***	0.109***
	(0.004)	(0.005)
Income	−0.006***	−0.004***
	(0.001)	(0.001)
Male	−0.011***	−0.011***
	(0.003)	(0.003)
Age	0.016***	0.016***
	(0.001)	(0.001)
Boredom*Income		−0.014***
		(0.004)
Constant	0.052***	0.052***
	(0.015)	(0.015)
Observations	25,443	25,443
Log Likelihood	1,980.479	1,980.843
Akaike Inf. Crit.	−3,946.958	−3,945.686
Bayesian Inf. Crit.	−3,889.948	−3,880.533
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3.8: WHO SAGE Dataset - Moderation Analysis. Standard errors are in parenthesis.

Ind. Variable	Variance Explained	95% CI Lower Limit	95% CI Upper Limit
Worried Mood	0.0277	0.0225	0.0329
Depressed Mood	0.0242	0.0189	0.0299
Boredom	0.0142	0.0105	0.0186
Stressed Mood	0.0130	0.0101	0.0169
Age	0.0027	0.0017	0.0038
Income	0.0006	0.0003	0.0012
Gender	0.0004	0.0001	0.0009

Table 3.9: Proportion of variance in Depression variable explained by each independent regressor (with 95% confidence intervals).

Bibliography

aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical psychology review*, *32*(6), 510–523.

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, *39*(6), 1490–1528.

Aknin, L. B., Norton, M. I., & Dunn, E. W. (2009). From wealth to well-being? money matters, but less than people think. *The Journal of Positive Psychology*, *4*(6), 523–527. https://doi.org/10.1080/17439760903271421

Andrade, E. B., & Ariely, D. (2009). The enduring impact of transient emotions on decision making. *Organizational Behavior and Human Decision Processes*, *109*(1), 1–8. https://doi.org/10.1016/j.obhdp.2009.02.003

Anestis, M. D., Selby, E. A., Crosby, R. D., Wonderlich, S. A., Engel, S. G., & Joiner, T. E. (2010). A comparison of retrospective self-report versus ecological momentary assessment measures of affective lability in the examination of its relationship with bulimic symptomatology. *Behaviour research and therapy*, *48*(7), 607–613.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on monte carlo simulation. *Psychological methods*, *24*, 1–19.

Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using r. *Behavior Research Methods*, *52*(1), 376–387.

Astivia, O. L. O., Gadermann, A., & Guhn, M. (2019). The relationship between statistical power and predictor distribution in multilevel logistic regression: A simulation-based approach. *BMC medical research methodology*, *19*(1), 1–20.

Ayuso-Mateos, J. L., Miret, M., Caballero, F. F., Olaya, B., Haro, J. M., Kowal, P., & Chatterji, S. (2013). Multi-country evaluation of affective experience: Validation of an abbreviated version of the day reconstruction method in seven countries. *PLOS ONE*, *8*(4), 1–8. https://doi.org/10.1371/journal.pone.0061534

Bates, D. (2007). Linear mixed model implementation in lme4. *Manuscript, University of Wisconsin*, *15*.

Beal, D. J., & Ghandour, L. (2011). Stability, change, and the stability of change in daily workplace affect. *Journal of Organizational Behavior*, *32*(4), 526–546.

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, *106*(1), 59–73. https://doi.org/10.1037/0033-2909.106.1.59

Biolcati, R., Mancini, G., & Trombini, E. (2018). Proneness to boredom and risk behaviors during adolescents' free time. *Psychological reports*, *121*(2), 303–323.

Blaszczynski, A., McConaghy, N., & Frankova, A. (1990). Boredom proneness in pathological gambling. *Psychological reports*, *67*(1), 35–42.

Bowen, R., Baetz, M., Hawkes, J., & Bowen, A. (2006). Mood variability in anxiety disorders. *Journal of affective disorders*, *91*(2-3), 165–170.

Bowen, R., Clark, M., & Baetz, M. (2004). Mood swings in patients with anxiety disorders compared with normal controls. *Journal of affective disorders*, *78*(3), 185–192.

Boyce, C. J., Daly, M., Hounkpatin, H. O., & Wood, A. M. (2017). Money may buy happiness, but often so little that it doesn't matter. *Psychological Science*, *28*(4), 544–546. https://doi.org/10.1177/0956797616672271

Brown, G. D., & Gathergood, J. (2020). Consumption changes, not income changes, predict changes in subjective well-being. *Social Psychological and Personality Science*, *11*(1), 64–73.

Calamia, M. (2019). Practical considerations for evaluating reliability in ambulatory assessment studies. *Psychological Assessment*, *31*(3), 285–291.

Card, D., Dustmann, C., & Preston, I. (2005). Understanding attitudes to immigration: The migration and minority module of the first european social survey.

Chan, D. K. C., Zhang, X., Fung, H. H., & Hagger, M. S. (2016). Affect, affective variability, and physical health: Results from a population-based investigation in china. *International Journal of Behavioral Medicine*, *23*(4), 438–446. https://doi.org/10.1007/s12529-015-9510-2

Chepenik, L. G., Ten Have, T., Oslin, D., Datto, C., Zubritsky, C., & Katz, I. R. (2006). A daily diary study of late-life depression. *The American journal of geriatric psychiatry*, *14*(3), 270–279.

Chin, A., Markey, A., Bhargava, S., Kassam, K. S., & Loewenstein, G. (2017). Bored in the usa: Experience sampling and boredom in everyday life. *Emotion*, *17*(2), 359.

Choi, J., Catapano, R., & Choi, I. (2017). Taking stock of happiness and meaning in everyday life: An experience sampling approach. *Social Psychological and Personality Science*, *8*(6), 641–651.

Ciccarelli, M., Griffiths, M. D., Nigro, G., & Cosenza, M. (2017). Decision making, cognitive distortions and emotional distress: A comparison between pathological gamblers and healthy controls. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 204–210. https://doi.org/10.1016/j.jbtep.2016.08.012

Cloninger, C. R. (2004). *Feeling good: The science of well-being*. Oxford University Press.

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.

Csikszentmihalyi, M., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. Basic Books, New York.

Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M. (2004). Boredom proneness in anger and aggression: Effects of impulsiveness and sensation seeking. *Personality and Individual Differences*, *37*(8), 1615–1627.

Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M. (2005). Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accident analysis & prevention*, *37*(2), 341–348.

Dal Mas, D. E., & Wittmann, B. C. (2017). Avoiding boredom: Caudate and insula activity reflects boredom-elicited purchase bias. *Cortex*, *92*, 57–69.

Daly, M., & Kelly, G. (2015). *Families and poverty: Everyday life on a low income*. Policy Press.

Deaton, A., & Stone, A. A. (2016). Understanding context effects for a measure of life evaluation: How responses matter. *Oxford Economic Papers*, *68*(4), 861–870.

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., Brose, A., Bastian, B., & Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of personality and social psychology*, *114*(2), 323–341.

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature human behaviour*, *3*(5), 478–491.

Delespaul, P., & DeVries, M. W. (1987). The daily life of ambulatory chronic mental patients. *Journal of Nervous and Mental Disease*.

Diener, E., Lucas, R. E., Oishi, S., Hall, N., & Donnellan, M. B. (2018). Advances and open questions in the science of subjective well-being. *Collabra: Psychology*, *4*(1).

Diener, E., Scollon, C. N., & Lucas, R. E. (2009). The evolving concept of subjective well-being: The multifaceted nature of happiness.

Diener, E., & Tov, W. (2012). National accounts of well-being. *Handbook of social indicators and quality of life research* (pp. 137–157). Springer.

Diener, E. (2009). *The science of well-being: The collected works of ed diener* (Vol. 37). Springer.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Dunster, G. P., Swendsen, J., & Merikangas, K. R. (2021). Real-time mobile monitoring of bipolar disorder: A review of evidence and future directions. *Neuropsychopharmacology*, *46*(1), 197–208.

Ebner-Priemer, U. W., Kuo, J., Kleindienst, N., Welch, S. S., Reisch, T., Reinhard, I., Lieb, K., Linehan, M. M., & Bohus, M. (2007). State affective instability in borderline personality disorder assessed by ambulatory monitoring. *Psychological medicine*, *37*(7), 961–970.

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., & Larsen, R. J. (2008). *The science of subjective well-being*. Guilford Press.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136–151.

Elhai, J. D., Vasquez, J. K., Lustgarten, S. D., Levine, J. C., & Hall, B. J. (2018). Proneness to boredom mediates relationships between problematic smartphone use with depression and anxiety severity. *Social Science Computer Review*, *36*(6), 707–720.

Ellison, W. D., Trahan, A. C., Pinzon, J. C., Gillespie, M. E., Simmons, L. M., & King, K. Y. (2020). For whom, and for what, is experience sampling more accurate than retrospective report? *Personality and Individual Differences*, *163*, 110071.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior research methods*, *50*(1), 195–212.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods*, *23*(4), 617.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness–the development and correlates of a new scale. *Journal of personality assessment*, *50*(1), 4–17.

Ferguson, B. D., Tandon, A., Gakidou, E., & Murray, C. (2003). Estimating permanent income using indicator variables.

Ferrer-i-Carbonell, A. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of public economics*, *89*(5-6), 997–1019.

Ferrer-i-Carbonell, A. (2013). Happiness economics. *SERIEs*, *4*(1), 35–60.

Fisch, A. T. M., Eckley, I. A., & Fearnhead, P. (2019). A linear time method for the detection of point and collective anomalies [arXiv: 1806.01947]. *arXiv:1806.01947 [cs, stat]*. http://arxiv.org/abs/1806.01947

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational behavior*, *33*(7), 865–877.

Forgeard, M. J., Jayawickreme, E., Kern, M. L., & Seligman, M. E. (2011). Doing the right thing: Measuring wellbeing for public policy. *International journal of wellbeing*, *1*(1).

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, *65*(1), 45–55.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, *40*(2), 402–435. https://doi.org/10.1257/jel.40.2.402

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651.

Genizi, A. (1993). Decomposition of r 2 in multiple regression with correlated regressors. *Statistica Sinica*, 407–420.

Goldberg, Y., & Danckert, J. (2013). Traumatic brain injury, boredom and depression. *Behavioral sciences*, *3*(3), 434–444.

Golier, J. A., Yehuda, R., Schmeidler, J., & Siever, L. J. (2001). Variability and severity of depression and anxiety in post traumatic stress disorder and major depressive disorder. *Depression and Anxiety*, *13*(2), 97–100.

Gratz, K. (2003). Risk factors for and functions of deliberate self-harm: An empirical and conceptual review. *Clinical Psychology: Science and Practice*, *10*, 192–205. https://doi.org/10.1093/clipsy.bpg022

Groemping, U. (2007). Relative importance for linear regression in r: The package relaimpo. *Journal of statistical software*, *17*, 1–27.

Groemping, U., & Matthias, L. (2021). Package 'relaimpo'. *Relative Importance of Regressors in Linear Models; R Fundation for Statstical Computing: Vienna, Austria*.

Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion*, *13*(1), 1–6.

Grunberg, N. E., & Straub, R. O. (1992). The role of gender and taste class in the effects of stress on eating. *Health Psychology*, *11*(2), 97–100. https://doi.org/10.1037/0278-6133.11.2.97

Hardy, J., & Segerstrom, S. C. (2017). Intra-individual variability and psychological flexibility: Affect and health in a national us sample. *Journal of Research in Personality*, *69*, 13–21. https://doi.org/10.1016/j.jrp.2016.04.002

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*(398), 371–386. http://www.jstor.org/stable/2289439

Haushofer, J. (2019). *Is there a psychological poverty trap?* (Tech. rep.). Working Paper.

Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *science*, *344*(6186), 862–867.

Havermans, R., Nicolson, N. A., & Devries, M. W. (2007). Daily hassles, uplifts, and time use in individuals with bipolar disorder in remission. *The Journal of Nervous and Mental Disease*, *195*(9), 745–751.

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological bulletin*, *141*(4), 901–930.

Hull, J. G., Young, R. D., & Jouriles, E. (1986). Applications of the self-awareness model of alcohol consumption: Predicting patterns of use and abuse. *Journal of Personality and Social Psychology*, *51*(4), 790–796. https://doi.org/10.1037/0022-3514.51.4.790

Huppert, F. A., Baylis, N., & Keverne, B. (2005). *The science of well-being*. Oxford University Press, USA.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, 640–648.

Iso-Ahola, S. E., & Crowley, E. D. (1991). Adolescent substance abuse and leisure boredom. *Journal of leisure research*, *23*(3), 260–271.

Jackson, G. L., Krull, J. L., Bradbury, T. N., & Karney, B. R. (2017). Household income and trajectories of marital satisfaction in early marriage. *Journal of Marriage and Family*, *79*(3), 690–704. https://doi.org/https://doi.org/10.1111/jomf.12394

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological methods*, *13*(4), 354–375.

Jantsch, A., & Veenhoven, R. (2019). Private wealth and happiness - a research synthesis using an online findings-archive. https://doi.org/10.1007/978-3-030-05535-6_2

Jebb, A. T., Tay, L., Diener, E., & Oishi, S. (2018). Happiness, income satiation and turning points around the world. *Nature Human Behaviour*, *2*(1), 33–38. https://doi.org/10.1038/s41562-017-0277-0

Jenkins, B. N., Hunter, J. F., Cross, M. P., Acevedo, A. M., & Pressman, S. D. (2018). When is affect variability bad for health? the association between affect variability and immune response to the influenza vaccination. *Journal of Psychosomatic Research*, *104*, 41–47. https://doi.org/10.1016/j.jpsychores.2017.11.002

Jones, S. H., Hare, D. J., & Evershed, K. (2005). Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder. *Bipolar disorders*, *7*(2), 176–186.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489–16493. https://doi.org/10.1073/pnas.1011492107

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

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.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? a focusing illusion. *Science*, *312*(5782), 1908–1910. https://doi.org/10.1126/science.1129688

Kang, H., & Rhee, D.-E. (2021). Does income (re)distribution matter for subjective well-being? evidence from cross-country panel data. *Social Science Quarterly*, *102*(2), 706–721. https://doi.org/10.1111/ssqu.12943

Kerr, M. L., Rasmussen, H. F., Buttitta, K. V., Smiley, P. A., & Borelli, J. L. (2021). Exploring the complexity of mothers' real-time emotions while caregiving. *Emotion*, *21*(3), 545–556.

Killingsworth, M. A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences*, *118*(4). https://doi.org/10.1073/pnas.2016976118

Kirtley, O. J., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2021). Making the black box transparent: A template and tutorial for registration of studies using experience-sampling methods. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920924686.

Knowles, R., Tai, S., Jones, S. H., Highfield, J., Morriss, R., & Bentall, R. P. (2007). Stability of self-esteem in bipolar disorder: Comparisons among remitted bipolar patients, remitted unipolar patients and healthy controls 1. *Bipolar Disorders*, *9*(5), 490–495.

Koval, P., Ogrinz, B., Kuppens, P., Van den Bergh, O., Tuerlinckx, F., & Sütterlin, S. (2013). Affective instability in daily life is predicted by resting heart rate variability. *PLOS ONE*, *8*(11), null. https://doi.org/10.1371/journal.pone.0081536

Kowal, P., Chatterji, S., Naidoo, N., Biritwum, R., Fan, W., Lopez Ridaura, R., Maximova, T., Arokiasamy, P., Phaswana-Mafuya, N., Williams, S., Snodgrass, J. J., Minicuci, N., D'Este, C., Peltzer, K., Boerma, J. T., & the SAGE Collaborators. (2012). Data Resource Profile: The World Health Organization Study on global AGEing and adult health (SAGE). *International Journal of Epidemiology*, *41*(6), 1639–1649. https://doi.org/10.1093/ije/dys210

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological science*, *21*(7), 984–991.

Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26.

Lakner, C., Yonzan, N., Gerszon Mahler, D., Castaneda Aguilar, R. A., & Wu, H. (2021). Updated estimates of the impact of covid-19 on global poverty: Looking back at 2020 and the outlook for 2021. https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty-looking-back-2020-and-outlook-2021

Lane, S. P., & Hennes, E. P. (2018). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships*, *35*(1), 7–31.

Lapate, R. C., & Heller, A. S. (2020). Context matters for affective chronometry. *Nature human behaviour*, *4*(7), 688–689.

Links, P. S., Heisel, M. J., & Garland, A. (2003). Affective instability in personality disorders. *American Journal of Psychiatry*, *160*(2), 394–395.

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.

MacKerron, G. (2012). Happiness economics from 35 000 feet. *Journal of Economic Surveys*, *26*(4), 705–735.

Manucia, G. K., Baumann, D. J., & Cialdini, R. B. (1984). Mood influences on helping: Direct effects or side effects? *Journal of Personality and Social Psychology*, *46*(2), 357–364. https://doi.org/10.1037/0022-3514.46.2.357

Masheb, R. M., & Grilo, C. M. (2006). Emotional overeating and its associations with eating disorder psychopathology among overweight patients with binge eating disorder. *International Journal of Eating Disorders*, *39*(2), 141–146. https://doi.org/10.1002/eat.20221

Masland, S. R., Shah, T. V., & Choi-Kain, L. W. (2020). Boredom in borderline personality disorder: A lost criterion reconsidered. *Psychopathology*, *53*(5-6), 239–253.

Mayer, M. (2019). *Missranger: Fast imputation of missing values* [R package version 2.1.0]. https://CRAN.R-project.org/package=missRanger

McConville, C., & Cooper, C. (1996). Mood variability and the intensity of depressive states. *Current Psychology*, *14*(4), 329–338.

Mckeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, *19*(1), 155–174. https://doi.org/10.1037/a0034282

Meers, K., Dejonckheere, E., Kalokerinos, E. K., Rummens, K., & Kuppens, P. (2020). Mobileq: A free user-friendly application for collecting experience sampling data. *Behavior Research Methods*, *52*(4), 1510–1515.

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707.

Morduch, J. (1994). Poverty and vulnerability. *The American Economic Review*, *84*(2), 221–225. http://www.jstor.org/stable/2117833

Moynihan, A. B., Igou, E. R., & van Tilburg, W. A. (2017). Boredom increases impulsiveness. *Social Psychology*.

Mueller, S., Ram, N., Conroy, D. E., Pincus, A. L., Gerstorf, D., & Wagner, J. (2019). Happy like a fish in water? the role of personality–situation fit for momentary happiness in social interactions across the adult lifespan. *European Journal of Personality*, *33*(3), 298–316.

Murray, C. J., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., Naghavi, M., Salomon, J. A., Shibuya, K., Vos, T., et al. (2012). Gbd 2010: Design, definitions, and metrics. *The Lancet*, *380*(9859), 2063–2066.

Myin-Germeys, I., Delespaul, P. A., & DeVries, M. W. (2000). Schizophrenia patients are more emotionally active than is assumed based on their behavior. *Schizophrenia Bulletin*, *26*(4), 847–854.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & Van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological medicine*, *39*(9), 1533–1547.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *59*(2), P49–P55.

OECD. (2011). How's life? : Measuring well-being. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/127381468326701628/hows-life-measuring-well-being

Oishi, S., Choi, H., Koo, M., Galinha, I., Ishii, K., Komiya, A., Luhmann, M., Scollon, C., Shin, J.-e., Lee, H., et al. (2020). Happiness, meaning, and psychological richness. *Affective Science*, *1*(2), 107–115.

Oishi, S., & Westgate, E. C. (2021). A psychologically rich life: Beyond happiness and meaning. *Psychological review*.

Ono, H., & Lee, K. S. (2016). *Redistributing happiness: How social policies shape life satisfaction*. Praeger, an imprint of ABC-CLIO, LLC.

Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, *6*(3), 383–391.

Peeters, F., Berkhof, J., Rottenberg, J., & Nicolson, N. A. (2010). Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. *Behaviour research and therapy*, *48*(8), 754–760.

Pfaltz, M. C., Michael, T., Grossman, P., Margraf, J., & Wilhelm, F. H. (2010). Instability of physical anxiety symptoms in daily life of patients with panic disorder and patients with posttraumatic stress disorder. *Journal of Anxiety Disorders*, *24*(7), 792–798.

Pfattheicher, S., Lazarevic, L. B., Nielsen, Y. A., Westgate, E., Krstic, K., & Schindler, S. (2021). I enjoy hurting my classmates: On the relation of boredom and sadism in schools.

Pfattheicher, S., Lazarevic, L. B., Westgate, E. C., & Schindler, S. (2021). On the relation of boredom and sadistic aggression. *Journal of Personality and Social Psychology*, *121*(3), 573.

Pirla, S., & Navarro-Martinez, D. (2021). Does boredom affect risk preferences? *Available at SSRN 3928897*.

Quispe-Torreblanca, E. G., Brown, G. D., Boyce, C. J., Wood, A. M., & De Neve, J.-E. (2021). Inequality and social rank: Income increases buy more life satisfaction in more equal countries. *Personality and Social Psychology Bulletin*, *47*(4), 519–539.

Quoidbach, J., Taquet, M., Desseilles, M., de Montjoye, Y.-A., & Gross, J. J. (2019). Happiness and social behavior. *Psychological science*, *30*(8), 1111–1122.

Raffaelli, Q., Mills, C., & Christoff, K. (2018). The knowns and unknowns of boredom: A review of the literature. *Experimental brain research*, *236*(9), 2451–2462.

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

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, *370*(6522), eaay0214.

Robinson, W. P. (1975). Boredom at school. *British Journal of Educational Psychology*, *45*(2), 141–152.

Rocke, C., Li, S.-C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging*, *24*(4), 863–878. https://doi.org/10.1037/a0016276

Rupp, D. E., & Vodanovich, S. J. (1997). The role of boredom proneness in self-reported anger and aggression. *Journal of Social Behavior and Personality*, *12*(4), 925.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology*, *52*(1), 141–166.

Ryff, C. D. (1989). Happiness is everything, or is it? explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, *57*(6), 1069.

Santangelo, P., Bohus, M., & Ebner-Priemer, U. W. (2014). Ecological momentary assessment in borderline personality disorder: A review of recent findings and methodological challenges. *Journal of personality disorders*, *28*(4), 555–576.

Santangelo, P., Reinhard, I., Mussgay, L., Steil, R., Sawitzki, G., Klein, C., Trull, T. J., Bohus, M., & Ebner-Priemer, U. W. (2014). Specificity of affective instability in patients with borderline personality disorder compared to posttraumatic stress disorder, bulimia nervosa, and healthy controls. *Journal of Abnormal Psychology*, *123*(1), 258–272. https://doi.org/10.1037/a0035619

Schenker, N., Raghunathan, T. E., Chiu, P.-L., Makuc, D. M., Zhang, G., & Cohen, A. J. (2006). Multiple imputation of missing income data in the national health interview survey. *Journal of the American Statistical Association*, *101*(475), 924–933. https://doi.org/10.1198/016214505000001375

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies*, *4*(1), 79–106.

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of personality and social psychology*, *89*(3), 395–406.

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

Servaas, M. N., Riese, H., Renken, R. J., Wichers, M., Bastiaansen, J. A., Figueroa, C. A., Geugies, H., Mocking, R. J., Geerligs, L., Marsman, J.-B. C., Aleman, A., Schene, A. H., Schoevers, R. A., & Ruhé, H. G. (2017). Associations Between Daily Affective Instability and Connectomics in Functional Subnetworks in Remitted Patients with Recurrent Major Depressive Disorder. *Neuropsychopharmacology*, *42*(13), 2583–2592. https://doi.org/10.1038/npp.2017.65

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, *1*(4), 538–555. https://doi.org/10.1177/2515245918805755

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Snir, A., Bar-Kalifa, E., Berenson, K. R., Downey, G., & Rafaeli, E. (2017). Affective instability as a clinical feature of avoidant personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *8*(4), 389–395. https://doi.org/10.1037/per0000202

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

Solnik, S., Rider, P., Steinweg, K., DeVita, P., & Hortobágyi, T. (2010). Teager–kaiser energy operator signal conditioning improves emg onset detection. *European journal of applied physiology*, *110*(3), 489–498.

Sommers, J., & Vodanovich, S. J. (2000). Boredom proneness: Its relationship to psychological-and physical-health symptoms. *Journal of clinical psychology*, *56*(1), 149–155.

Stanislaus, S., Faurholt-Jepsen, M., Vinberg, M., Coello, K., Kjærstad, H. L., Melbye, S., Sletved, K. S., Christensen, E. M., Frost, M., Bardram, J. E., & Kessing, L. V. (2020). Mood instability in patients with newly diagnosed bipolar disorder, unaffected relatives, and healthy control individuals measured daily using smartphones. *Journal of Affective Disorders*, *271*, 336–344. https://doi.org/https://doi.org/10.1016/j.jad.2020.03.049

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S., & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of personality and social psychology*, *74*(6), 1670–1680.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (ema) in behavorial medicine. *Annals of behavioral medicine*, *16*, 199–202.

Struk, A. A., Scholer, A. A., & Danckert, J. (2021). Perceptions of control influence feelings of boredom. *Frontiers in Psychology*, *12*, 2698.

Tan, J. J. X., Kraus, M. W., Carpenter, N. C., & Adler, N. E. (2020). The association between objective and subjective socioeconomic status and subjective well-being: A meta-analytic review. *Psychological Bulletin*, *146*(11), 970–1020. https://doi.org/10.1037/bul0000258

Taquet, M., Quoidbach, J., De Montjoye, Y.-A., Desseilles, M., & Gross, J. J. (2016). Hedonism and the choice of everyday activities. *Proceedings of the National Academy of Sciences*, *113*(35), 9769–9773.

Taquet, M., Quoidbach, J., Gross, J. J., Saunders, K. E., & Goodwin, G. M. (2020). Mood homeostasis, low mood, and history of depression in 2 large population samples. *JAMA psychiatry*, *77*(9), 944–951.

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of abnormal psychology*, *121*(4), 819–829.

Tice, D. M., Bratslavsky, E., & Baumeister, R. F. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! *Journal of Personality and Social Psychology*, *80*(1), 53–67. https://doi.org/10.1037/0022-3514.80.1.53

Todman, M. (2007). Psychopathology and boredom: A neglected association. *Psychological science: Research, theory, and future directions*, 145–160.

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PloS one*, *10*(12), e0145450.

Trudel-Fitzgerald, C., Millstein, R. A., Hippel, C. V., Howe, C. J., Tomasso, L. P., Wagner, G. R., & Vanderweele, T. J. (2019). Psychological well-being as part of the public health debate? insight into dimensions, interventions, and policy. *BMC Public Health*, *19*(1). https://doi.org/10.1186/s12889-019-8029-x

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, *129*(1), 56–63.

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, *7*(4), 355–361.

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of abnormal psychology*, *117*(3), 647–661.

Tsanas, A., Saunders, K., Bilderbeck, A., Palmius, N., Osipov, M., Clifford, G., Goodwin, G., & De Vos, M. (2016). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of affective disorders*, *205*, 225–233.

UN. (2012). Happiness : Towards a holistic approach to development. https://digitallibrary.un.org/record/715187?ln=en

Van Praag, B. M., & Ferrer-i-Carbonell, A. (2011). *Happiness economics: A new road to measuring and comparing happiness*. Now Publishers Inc.

Verbakel, E., Tamlagsrønning, S., Winstone, L., Fjær, E. L., & Eikemo, T. A. (2017). Informal care in europe: Findings from the european social survey (2014) special module on the social determinants of health. *The European Journal of Public Health*, *27*(suppl_1), 90–95.

Verma, R. (2017). Gross national happiness: Meaning, measure and degrowth in a living development alternative. *Journal of Political Ecology*, *24*(1). https://doi.org/10.2458/v24i1.20885

Vittersø, J. (2016). *Handbook of eudaimonic well-being*. Springer.

Vodanovich, S. J., Verner, K. M., & Gilbride, T. V. (1991). Boredom proneness: Its relationship to positive and negative affect. *Psychological reports*, *69*(3_suppl), 1139–1146.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z., et al. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the global burden of disease study 2015. *The Lancet*, *388*(10053), 1545–1602.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219–235.

Watt, J. D., & Vodanovich, S. J. (1992). Relationship between boredom proneness and impulsivity. *Psychological reports*, *70*(3), 688–690.

Wendt, L. P., Wright, A. G., Pilkonis, P. A., Woods, W. C., Denissen, J. J., Kühnel, A., & Zimmermann, J. (2020). Indicators of affect dynamics: Structure, reliability, and personality correlates. *European Journal of Personality*, *34*(6), 1060–1072.

Westgate, E. C. (2020). Why boredom is interesting. *Current Directions in Psychological Science*, *29*(1), 33–40.

Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The mac model of boredom and cognitive engagement. *Psychological Review*, *125*(5), 689.

Wolff, W., Bieleke, M., & Keller, L. (2022). Boredom proneness predicts self-assessed decision errors in sports but is unrelated to risk taking in general. *International Journal of Environmental Research and Public Health*, *19*(6), 3479.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114.

Wood, S. N., & Augustin, N. (2002). Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, *157*, 157–177.

Yakobi, O., & Danckert, J. (2021). Boredom proneness is associated with noisy decision-making, not risk-taking. *Experimental Brain Research*, *239*(6), 1807–1825.

Yucel, M., & Westgate, E. (2021). From electric shocks to the electoral college: How boredom steers moral behavior.

Zeigler–Hill, V., & Abraham, J. (2006). Borderline personality features: Instability of self–esteem and affect. *Journal of Social and Clinical Psychology*, *25*(6), 668–687.