

# UC Riverside

## UC Riverside Electronic Theses and Dissertations

### Title

Experience Sampling and Day Evaluations

### Permalink

<https://escholarship.org/uc/item/65385673>

### Author

Miller, Travis J

### Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
RIVERSIDE

Experience Sampling and Day Evaluations

A Dissertation submitted in partial satisfaction  
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Travis J. Miller

June 2020

Dissertation Committee:

Dr. Daniel J. Ozer, Chairperson

Dr. David C. Funder

Dr. Megan Robbins

Copyright by  
Travis J. Miller  
2020

The Dissertation of Travis J. Miller is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgements

I would like to thank my family for their love and support. Thank you for encouraging my curiosity, for inspiring me, for demonstrating how fulfilling it is to help others, and for teaching me that hard work pays off.

When I started graduate school at UCR, in many ways I thought I knew everything. Graduate school was a humbling experience. I feel very fortunate to have been brought to UCR by Dan Ozer. I am grateful that Dan provided me the freedom to pursue my personal academic interests. In the department, at conferences and in the academic community I have always been proud to say that you were my advisor.

I pursued a Ph.D. in psychology in large part due to Mark Otten. I joined his lab as an undergraduate at CSUN and knew that a psychology department was where I belonged after attending my first lab meeting. I have been very fortunate to have many mentors and advisors over the years. I would like to thank all those who served on my committees, gave input on various project, and contributed to my education and academic development, including David Funder, Megan Robbins, Will Dunlop, and Bob Rosenthal.

To my lab mates: Kristina Mouzakis, Jennifer Coons, Jake Gray, and later Trevor Basil and Antonio Curtis. Thank you for the social and emotional support during this process. Especially to Jake, I cannot count how many hours we spent working together in the lab, I appreciate that you were always willing to help me, even when I annoyed you. In my time at UCR the Personality Assessment Lab became the FundOzer lab. Erica Baranski, Kyle Sauerberger, Gwen Gardiner and later Daniel Lee made the lab more fun

and entertaining, and provided a reward structure for celebrating accomplishments big and small.

To the friends I made in graduate school (the list is too long to include everyone), thank you for helping me be a balanced and well-rounded person, supporting my personal as well as academic growth: Ben Miller, Alex Leung, Nicole Harake, Robert Wright, Seth Margolis, Paige Trubenstein, and Becca Tuckerman.

And lastly thank you to the people who make our whole world run, the departmental staff at UCR. While I quite literally would not have made it through the program without your hard work. Renee, Faye, Kirsten, Toya and everyone else, what you do outside of your job responsibilities (chatting, sharing dog pictures, telling stories) is what has made UCR more home than university.

## ABSTRACT OF THE DISSERTATION

Experience Sampling and Day Evaluations

by

Travis J. Miller

Doctor of Philosophy, Graduate Program in Psychology

University of California, Riverside, June 2020

Dr. Daniel J. Ozer, Chairperson

The present research examines two distinct features of the lived day of 384 undergraduates. In the first study, we investigated how well individuals remember their daily activities, the ways in which activities were misremembered, if different kinds of activities were remembered differently, and if individual differences are related to the ways in which a person remembers (or misremembers) their daily activities. In the second study we explored the ways in which the experience of affect throughout the day is related to affective evaluation of the day.

Retrospective assessments of behavior are the preferred method of data collection when the goal is to capture all behaviors. Prior research has outlined the discrepancies between experience sampling (ES) and retrospective recall of states, the current project looked at the discrepancies between these methods for reporting activities throughout the day. Participants reported what they were doing in the moment throughout the day, and in the evening retrospectively recalled their activities for each hour of that day. Nearly 75% of daily activities were recalled accurately. The remaining activities were remembered at the wrong time (13%), appear to be confabulated (10%) or were inaccurate for other

reasons (2%). Being in class or at work, which are scheduled activities often with strictly enforced start and end times, were the most accurately remembered activities. While participants misremember sleeping, shopping, recreating and housework, confabulating what they were doing when retrospectively recalling these activities. Individual difference variables were not related to how participants remembered their activities.

The second feature of the day that was explored was how the affective experience of the day relates to evaluations of the day. Various research projects have suggested that the average affect, maximum affect, affect in the final moments, or the average between the maximum and final affect are most meaningful (e.g. Hedges, Jandorf, & Stone, 1985; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Parkinson, Briner, Reynolds, & Totterdell, 1995). The current project found that when reporting on their day, each measure describing lived experience is nearly equally predictive of the affective evaluations of the day. This was the case for both positive and negative affect. These affective measures correlated appropriately with the content of day evaluations, and the various measures of affect throughout the day have similar profiles and external correlates.



## TABLE OF CONTENTS

Introduction	1
The Lived Day	1
Accuracy and errors in recall of daily activities	5
Personality and recall of daily behavior	8
Method	9
Results and Discussion	14
General Discussion	18
Affective evaluations of the day	23
Method	26
Results and Discussion	32
General Discussion	42
Conclusions	43
References	45
Appendix A	54
Appendix B	56
Appendix C	68

## List of Tables

Table 1. Sample information presented to coders	12
Table 2. Person level descriptive statistics for accuracy and reasons for inaccuracy	15
Table 3. Accurate recall and reason for inaccurate recall by activity category	16
Table 4. Randomization test results for correlations of POMP scores of accuracy and reasons for inaccuracy with individual difference variables	17
Table 5. Correlations among experience sampling affect ratings of positive affect with positive affect templates	34
Table 6. Correlations among experience sampling affect ratings of negative affect with negative affect templates	35
Table 7. Positive affect: correlations, regression and structure coefficients	36
Table 8. Negative affect: correlations, regression and structure coefficients	37
Table 9. Randomization test results for correlations of DEQ2.0 items with positive affect descriptors	38
Table 10. Correlations of DEQ2.0 items with positive affect descriptors	39
Table 11. Randomization test results for correlations of DEQ2.0 items with negative affect descriptors	40
Table 12. Correlations of DEQ2.0 items with negative affect descriptors	41

## Introduction

Human experiences with the diurnal rhythms of the natural world have led to our lives being structured around the day. The day is ingrained within us physiologically and psychologically (Aschoff, 1965; Craik, 2000). Research attempting to understand daily behavior has shied away from this inherently meaningful unit, preferring hours, events, activities, or situations (see Mehl & Conner, 2012). It has been argued that vocabulary evolves following Neo-Darwinian principles (Pinker & Bloom, 1990), that is words that are found to be useful are used more often and for longer periods of time, and as words are deemed not useful their use declines until they eventually disappear from the lexicon. Lay people have found the day to be a meaningful unit and the use of this unit is prevalent, for example, references are made to the day in our greetings — “How was your day?” or “Have a nice day”. As such it is likely a productive venture to understand daily behavior using the day as the unit of analysis. To this end, two projects on the lived day are described. In the first project, I describe an exploration into the extent to which retrospective recall of one’s daily activities are accurate (e.g. what type of events are remembered more accurately than others, what are the common errors in remembering events). In the second project, I sought to understand what makes good days (good being a positive affective evaluation).

### **The Lived Day**

Understanding individuals and their lives has been a central tenet in psychology at large, and the field of personality specifically. Allport and Odbert (1936) argued that important characteristics become part of the group or culture’s language and that the

words used to describe people must do so accurately, if they did not they would have fallen out of use or been replaced by more accurate representations. In the same way, the concepts used to describe daily behavior and experience must do so accurately, otherwise better alternatives would have replaced them in our lexicon. While there are numerous constructs individuals can use to describe their daily behavior and experience — very specific descriptions using hours, events or activities, more general descriptions of short timespans such as morning, afternoon or evening, or longer timespans week, month or year — the preferred concept for many people is the day as a unit. This can be seen across the globe and through time in daily greetings (e.g. “good day”), aphorisms (e.g. “carpe diem” or “seize the day”), colloquialisms (e.g. “the other day” or “call it a day”) and even in some threats (e.g. “make my day”). While the day is clearly a meaningful unit for describing lives, psychological research on daily behavior has preferred other descriptions (e.g. hours, events, activities and situations), with few efforts to measure the day as a unit scientifically.

Psychological researchers have been interested in daily behavior since at least the 1930s (see Verbrugge, 1980). Numerous methods have been developed which seek to further our understanding of what people do, how they act, and how they feel throughout their days. These methods include daily diaries, which ask participants to recount variables of interest (e.g. positive social interactions) in the evening (Bolger, Davis, & Rafaeli, 2003). This method is thought to be well suited for capturing infrequent but impactful events (Schwarz, 2007). Naturalistic observations of full days have been attempted, and while the data gathered through this method is rich, it has not led to the

development of theory, measures or further research questions (Barker & Wright, 1951; Craik, 2000).

Retrospectively recounting the events of the day, activity by activity (Hershey, 1999), termed the Day Reconstruction Method (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004) allows for the inclusion of a full day's events succinctly. The reflective nature of these accounts limits the domains that can be probed, as biases including recency, salience, and current state of mind call the accuracy of some kinds of information into question (Reis, Gable, & Maniaci, 2014).

To accurately capture thoughts, feelings and behaviors with more confidence researchers can collect information from participants in the moment via Ecological Momentary Assessment (EMA; Shiffman, Stone, & Hufford, 2008). EMA covers an assortment of methods which include the Experience Sampling Method (ESM; Hektner, Schmidt, & Csikszentmihalyi, 2007) in which the researcher sends brief surveys to participants at various points during the day, asking them to report on current circumstances. This method is particularly adept at probing thoughts and emotions which can vary when recalled, even hours later (Schwarz, 2007)

While each of these methods is uniquely suited to answer specific types of questions, they all share a molecular approach to understanding daily behavior by probing individual instances, activities, events, situations, or hours of the day. What is missing from the scientific community's understanding of daily behavior is how the day is understood as a unit.

To evaluate the day as a unit a representation of daily behavior must be made. The following section reviews how daily behavior is recalled, and the kinds of errors that are made when recalling daily behavior.

### **Accuracy and errors in recall of daily activities**

Over the last decade and a half, an emphasis on ecological validity and technological advancements have led to worthwhile progress in experience sampling (Hektner, Schmidt & Csikszentmihalyi, 2007; Shiffman, Stone & Hufford, 2008). As part of this trend, the extent of certain biases in other methods, namely retrospective assessments, is often emphasized. In justifications for the use of experience sampling methods, references are often made to numerous biases in participant responses based on the recall of events (Schwarz, 2007). However, the timeframe over which memories are being recalled is rarely referenced. Here, this evidence is reviewed, and recall of daily activities over the day is examined.

When seeking to understand people and their behavior, psychologists most commonly rely on self-reports (Tracy, Robins, & Sherman, 2009). It is often assumed that the subject of investigation is the most accurate and reliable source of information on their own thoughts and behaviors (past and present). Critics do not necessarily disagree with the strengths of this method, largely their concern is with the biases that can creep into assessments (e.g. Dunning, Heath, & Suls, 2005), and an overreliance on self-report as opposed to other measures (Baumeister, Vohs, & Funder, 2007). Many of the biases, which are of interest here come into play when a person recalls their thoughts, feelings or behaviors from the past (Paulhus & Vazire, 2007).

Retrospectively recalling thoughts, feelings or behaviors may be more complicated than it immediately appears and depends on the frequency of the event being recalled and the timeframe over which it is being recalled (see Schwarz, 2012 for in depth

reviews of the cognitive processes involved). The core of the issue is that what is traditionally referred to as recall, is more accurately described as reconstructing events using one or multiple heuristics, or short cuts. Additionally, participants are often asked to do more than just recall events, they are asked to aggregate or evaluate them (e.g. “how did you feel over the past year?). When asked to do this, participants do not recall each experience in sequence and then aggregate, instead they rely on heuristics to estimate an answer. These heuristics can lead to the biases present in retrospective methods (Shiffman, Stone, & Hufford, 2008). Common habitual activities are thought to be especially difficult to recall accurately (Klasnja, Harrison, LeGrand, LaMarca, Froehlich, & Hudson, 2008; Shephard 2003).

Often when assessing the accuracy of retrospective recall of thoughts feelings and behaviors the most difficult situations for recall are measured. Namely, the longer the period of recall is (e.g. asking someone to recall events over an extended amount of time), the less accurate their assessment might be. When specific experiences — such as pain, affect or cravings — are recalled over a 2-week or longer period, higher estimates are often obtained when compared to those reported via experience sampling (Carney, Tennen, Affleck, Del Boca, & Kranzle, 1998; Litt, Cooney & Morse, 2000; Shrier, Shih, & Beardslee, 2005). This tendency of events to be recalled as more intense, frequent, or long-lasting is often referred to as a saliency bias (Kahneman, Slovic, Slovic, & Tversky, 1982). While some biases may impact recall over longer periods, others have demonstrable impacts almost immediately (Redelmeier, Katz, & Kahneman, 2003).



When recall period is investigated and compared to experience sampling, high correspondence is often found between the average of momentary ratings and recall. For example, a common paradigm for examining accuracy of recall period is to have participants report on their symptoms, such as pain or fatigue, multiple times throughout the day via experience sampling and then to recall those symptoms over a period ranging from a day to multiple weeks. The average of reports made throughout the day is then compared to recall. Recalled symptoms are often higher than momentary reports, but the pattern of results are similar (e.g. a participant might consistently recall that their pain for the day was a few points higher than the average of their momentary ratings, but days of when relatively more or less pain is experienced are captured by both methods; Broderick, Schwartz, Schneider, & Stone, 2009, Broderick, Schwartz, Schneider, & Stone, 2010; Broderick, Schwartz, Vikingstad, Pribbernow, Grossman, & Stone, 2008; Ganzach & Yaor, 2019; Miron-Shatz, Stone, & Kahneman, 2009). The extent to which experience sampling reports and retrospectively recalled experiences correspond declines as the number of days recalled grows (Broderick et al., 2008, 2009, 2010)<sup>1</sup>.

Few projects have gauged accuracy or agreement by comparing in situ assessments to retrospective assessments of the same person on the same day. More commonly, they tend to compare retrospective recall with archival data at the individual

---

<sup>1</sup> It is interesting to note here that this an example of how statistically significant differences do not mean that the differences are practically or clinically important. For example, Broderick and colleagues (2009) had participants report intensity of pain at multiple times throughout the day and recall the intensity of their pain over the day, for seven days. They found that the recall of pain intensity was significantly higher than the mean of reports of pain throughout day ( $p < .001$ ). However, across the seven days the average pain intensity reported throughout the day was 40.71 on a 100-point scale, and the average pain recalled each day was 46.97 on the same 100-point scale. I venture to say that two patients presenting to a pain clinic with pain scores of 40.71 and 46.97 would not be treated so differently.

or group level; for example, by comparing retrospective assessments to results of prior experience sampling studies (e.g. Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). Another alternative is to replicate known relationships with more objective outcomes; for example, if positive affect and health are known to be related, relating retrospective recall of positive affect to actual longevity gives the retrospective assessment of affect some credibility (e.g. Krueger, Kahneman, Schkade, Schwarz, & Stone, 2009).

### **Personality and recall of daily behavior**

Personality traits describe individual differences in tendencies of behavior and experience (John, 1989; McAdams, 1995). Traits relate time use; for example, conscientiousness is positively related to time spent on productive tasks and negatively to time spent in social or leisure activities (Hershey, 1999; Jackson, Wood, Bogg, Walton, Harms, & Roberts, 2010). There is some reason to believe that individual differences may relate to accuracy in recall of daily behavior. For example, conscientiousness is associated with features of brain structures that are linked to working memory and planning behavior (DeYoung, 2010). Definitionally, conscientiousness entails organization, productiveness and responsibility (Soto & John, 2017); as such, those higher in conscientiousness may put more effort into planning their daily activities. Douglas, Bore and Munro (2016) discuss theory and research on how all Big Five traits relate to time management behaviors (e.g. setting priorities, planning/scheduling, preference for organization), and present evidence for the relations of all but openness to experience with time management behaviors.

The current project sought to establish an estimate for accuracy of recall throughout the day by asking participants to report what they are doing throughout the day via experience sampling and recall their activities retrospectively in the evening. The accuracy of recall and the reasons for the inaccuracy were then broken down by the type of activity the participant was engaged in. The relationships of accurate recall with individual differences and features of the activities are also explored.

## **Method**

**Participants.** Participants consisted of 384 undergraduates recruited from the psychology department's subject pool at a university in Southern California. Participation was restricted to those who own a smartphone. Participants average age was 19.37 ( $SD = 1.65$ ; range: 18-31); 43% were Asian or Pacific Islander, 33% Latino/Latina, 9% White, 6% Black, 3% Middle Eastern, and 6% were mixed, other, or had missing data.

**Procedures.** Participants completed measures over two days. All surveys were administered online using Qualtrics survey software. On day 1, participants provided demographic information, completed trait measures, an affect measure, subjective well-being measures, provided their email address, cell phone number and carrier for experience sampling surveys to be text messaged and emailed to them the next day. This initial survey was completed by 384 participants at any time on day 1. On the morning of day 2, participants received a text message and email reminder about their participation in the study, and that they must complete each short survey within 20 minutes of receiving it, when it is safe, and not interrupting a class or other responsibilities. Participants received links to the online surveys via text message and email 10 times throughout the

day between 9:15am and 7:30pm. These surveys were distributed semi-randomly, such that there was a minimum of 45 minutes and a maximum of 1.5 hours between them. Each survey took less than five minutes and asked about the participant's current activity. In the evening participants received an email and text message with a link to provide a list of their daily activities (338 participants, 88%, provided this information; see appendix C for a full list of measures). 310 participants (81%) completed at least one experience sampling report and completed the evening survey and were therefore reported in this study.

### **Measures.**

***Demographics.*** Participants provided information on their age, gender, and ethnicity.

***Big Five.*** The Big Five Inventory-2 (BFI-2; Soto & John, 2017) is a 60-item measure which was used to assess the Big Five personality trait dimensions. Alpha reliabilities for the traits ranged from .80 to .88.

***Subjective well-being.*** The Riverside Life Satisfaction Scale (RLSS; Margolis, Schwitzgebel, Ozer, & Lyubomirsky, 2018) is a 6-item measure which was used to assess life satisfaction ( $\alpha = .82$ ). Additionally, the Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999), a 4-item measure, was used to assess trait happiness ( $\alpha = .86$ ). These two measures correlated  $r = .58$ .

***Affect.*** The Affect Adjectives Scale (Diener & Emmons, 1984) is a 9-item measure designed to measure the positive and negative dimensions of affect. Four items were added to the scale to ensure that the scale captures both high and low arousal

emotions (Peaceful/Serene, Dull/Bored, Excited and Tired). Participants rated how often they have experienced moods in the past year on a 7-point scale from *not at all* to *extremely*. For the 6 items of the Positive Affect Scale  $\alpha = .88$ , and  $.83$  for the 7 items of the Negative Affect Scale. The two scales correlated  $r = -.19$ .

**Activity reports: Experience sampling.** Each time a participant opened a survey throughout the day to report their behavior, they were first asked “Immediately before you started this survey, what were you doing?” with space for an open-ended response. In total, participants reported 2,284 activities ( $M = 7.37$ ;  $SD = 2.0$ ). Participants then rated the activity for “this activity is a good use of my time”, and “this activity is something I do often”, on a 5-point scale from *strongly disagree* to *strongly agree*. A pair of judges (four judges total) independently coded each activity for the activity category that best described it (Table 3 displays all activity categories excluding “forgot” and “rather not say”). The judges agreed on which category best described the activity for 76% of the activities. A fifth judge independently coded the remaining activities and agreed with one of the initial judges for all remaining activities.

**Activity recall: End of day.** Participants retrospectively recalled their activities for each hour of the day (from 7am until 1am). They provided a brief statement describing what they did that hour and picked an activity category to which their activity belonged (Table 3 displays all activity categories excluding “forgot” and “rather not say”), and identified who they were with broadly using labels like “friend” or “classmate”.

To identify whether activities were accurately recalled, judges were presented with one participant’s experience sampling activities alongside their retrospective

activities and the times that they corresponded to (Table 1). Participants retrospectively recalled activities for an hour window (e.g. 9:00 to 10:00am), however the experience sampling surveys could be completed twenty minutes after receiving them. Therefore, participants may have indicated the activity they were doing when they received the survey and not with what they were doing when they completed it. Thus, times were presented to judges for an hour and twenty-minute window (e.g. 9:00 to 10:20am). For example, the sample participant in Table 1 reported via experience sampling at 10:05am that they had been “Sleeping”. This participant may have reported what they were doing when they received the survey, which could have been as early as 9:45am. That evening, the participant retrospectively recalled that from 9:00 to 10:00am they had been asleep. Due to variability in the time that the experience sampling activity may be referring to, this information is presented to judges as having been reported from 9:00 to 10:20am.

Table 1  
*Sample information presented to coders*

<u>Retrospectively recalled activity</u>		<u>Experience sampling activity</u>		
<u>Time</u>	<u>Activity</u>	<u>Survey</u>	<u>Time</u>	<u>Activity</u>
9 to 10:20am	Sleep	1	10:05am	Sleeping
10 to 11:20am	Work	2	11:12am	Working
11 to 12:20pm	Eating Lunch	3	12:18pm	Eating
12 to 1:20pm	In class	4	1:40pm	Eating
1 to 2:20pm	Walking home	5	3:03pm	Working
2 to 3:20pm	Working	6	NA	NA
3 to 4:20pm	Working	7	5:33pm	Working
4 to 5:20pm	Working	8	6:16pm	Gym
5 to 6:20pm	Working	9	7:01pm	Gym
6 to 7:20pm	Gym	10	8:02pm	Eating
7 to 8:20pm	Gym			
8 to 9:20pm	Showering			

Each activity reported via experience sampling was coded by a pair of judges (six judges in total) on whether it was accurately recalled (full coding manual available in

Appendix A). To do so, judges identified which retrospective activity or activities corresponded to the time of the experience sampling survey. If the stated retrospective activity matched with what was stated as the experience sampling activity, the judge indicated that the experience sampling activity was accurately recalled. For example, if as in Table 1, at 11:12am a participant reported via experience sampling that just prior to starting the survey they had been “working”, the judge would look at the retrospectively activities recalled for “10am to 11:20am” and “11am to 12:20pm”. If the judge deemed that either of the retrospectively recalled activities matched with the activity reported via experience sampling, they indicated a “1” next to the activity reported via experience sampling. Otherwise, that cell was left blank.

Of the 2,284 activities reported via experience sampling by participants, judges agreed on whether or not the activity was accurately recalled 87.8% of the time (2,005 activities;  $\Phi$  for each of the six pairs of judges ranged from .65 to .81, and  $\Phi/\Phi_{\max}$  ranged from .80 to .96). The remaining 279 activities were coded by a third independent judge and was recorded as a match if the activity was coded as a match by two of the three judges.

Of the 2,284 activities reported by participants via experience sampling 1,217 (53%) were coded as having been recalled accurately. The 1,067 experience sampling activities that were coded as having been inaccurately recalled were then coded for the reason for the inaccuracy. In discussion with judges who rated accuracy, four categories were presented as potential reasons for the inaccuracy: experience sampling activity was correctly recalled but at a different time (Timing), participant retrospectively recalled an

activity that seems to be made up (Confabulation), participant used language that could be referring to the same activity (e.g. “Studying” vs “In the library”; Language), and participant stated a specific activity via experience sampling that could be a part of the hour-long retrospectively recalled activity (e.g. "waiting for class to start" rather than "in class" or "on my phone" rather than "in my room relaxing"; Specificity). A fifth category, “Other”, was provided for anything else (full coding manual available in Appendix B).

Two judges independently rated if these categories captured the reason for the inaccuracy. The two judges agreed on the reason for the inaccuracy on 968 of the 1,067 activities (91%;  $\Phi$  for each category ranged from .81 to .91, and  $\Phi/\Phi_{\max}$  ranged from .90 to .98). For the 99 activities that were not agreed on, a third independent judge was brought in, two of the three judges agreed on the reason for the inaccuracy of all 99 activities.

## **Results and Discussion**

In identifying the reason for the inaccurately recalled activities, 306 (29%) were an issue with the Timing of the activity, 227 (21%) were Confabulation, 74 (7%) were discrepancies in Language, 411 (39%) were issues with Specificity, and 49 (5%) were Other reasons (percentages provided in parentheses are of inaccurately remembered activities). Next, these were aggregated to the person level, and convert to Percent of Maximum Possible scores (POMP; Cohen, Cohen, Aiken & West, 1999) by dividing by the number of activities reported by the participant. POMP scores are calculated such that for a person who completed all 10 experience sampling reports and 5 of them were accurately recalled, the person would have a .5 ( $5 \div 10$ ) for the “Match” category, and if a



person completed 9 experience sampling reports and 3 of them were coded to be a “Confabulation”, this person would have a .33 ( $3 \div 9$ ) for the “Confabulation” category. The mean, standard deviation, and range of POMP scores for each person are displayed in Table 2.

Table 2  
*Person level descriptive statistics for accuracy and reasons for inaccuracy*

	<u>Mean (SD)</u>	<u>Range</u>
Match	.53 (.24)	0-1
Language	.04 (.09)	0-1
Specificity	.18 (.17)	0-1
Timing	.14 (.16)	0-1
Confabulation	.10 (.13)	0-.57
Other	.02 (.08)	0.67

*Note. N=310; Scores are displayed as POMP scores such that they are interpreted as the percent of experience sampling reports completed.*

By and large, people remembered most of their daily activities accurately. While Language and Specificity are not exact matches, participants were not asked to recall what they had reported in the experience sampling surveys, so these too may be considered accurate for many purposes. Among the reasons for the inaccuracies, those made due to Timing were the most common, with participants mistaking the Timing for 14% of their activities, on average. Confabulation was also prevalent with on average each participant appearing to make up 1 in 10 of their activities. These Confabulations occur in nearly every activity category with Shopping, Sleeping, Recreation and Housework being confabulated the most commonly (Table 3) meaning that people reported doing these via experience sampling and then recalled a different activity at that time retrospectively. For example, one participant reported via experience sampling that

they were “Playing online games” at 4:08pm but recalled “Doing homework” and “Eating and getting ready for class” when recalling their activities at that time retrospectively. Another participant reported via experience sampling that they were “finishing errands” at 5:22pm but recalled that they had been “Relaxing” at that time.

Table 3  
*Accurate recall and reason for inaccurate recall by activity category*

	Act.	Accurate				Timing	Confab.	Other
		Match	Lang.	Spec.				
Class	270	.79	.02	.06	.09	.04	.00	
Studying	421	.62	.07	.05	.14	.11	.01	
Sleep	101	.54	.03	.09	.10	.20	.04	
Friends	59	.39	.03	.32	.12	.14	.00	
Relaxation	255	.37	.03	.25	.15	.15	.05	
Eating	302	.52	.01	.20	.20	.05	.02	
Locomotion	260	.43	.02	.32	.15	.06	.02	
Working	69	.86	.01	.00	.09	.04	.00	
Grooming	98	.37	.02	.34	.16	.09	.02	
Partner	4	.50	.00	.00	.00	.00	.50	
Exercise	39	.69	.00	.00	.26	.05	.00	
Shopping	35	.34	.00	.29	.14	.20	.03	
Recreation	159	.47	.03	.20	.09	.18	.03	
Family	15	.60	.00	.33	.00	.07	.00	
Housework	76	.39	.08	.18	.16	.17	.01	
Other	121	.45	.03	.37	.05	.07	.02	
Total <sup>2</sup>	2284	.53	.03	.18	.13	.10	.02	

*Note. N=310 participants; Act. refers to the number of activities in each category; Total refers to the proportion of each reason for miss of the total number of activities (2284); Lang. = Language; Spec. = Specificity; Confab. = Confabulation*

Table 3 displays the number of activities reported via experience sampling and how they were recalled by activity category. The activity categories that matched the most often were working and in class. These activities are likely scheduled in advance,

<sup>2</sup> The numbers presented in the “Total” row differ from those presented in Table 2 as Table 2 values are aggregated to the person level from which descriptive statistics were derived, while Table 3 presents the values of all activities. This discrepancy is only seen at two decimal places for Confabulation.

their timing may be strictly enforced, and students may have the times of these on hand in calendars when completing the retrospective assessment. Errors in recall due to specificity were most common when participants had reported that they were grooming, engaging in activities with their friends, on their way somewhere (locomotion), or when the activity was not coded as being best described by none of the content categories (other; e.g. “procrastinating”). Inaccuracies due to timing were most frequent when participants reported that they were exercising and eating.

**Correlates of accurate recall.** To identify whether recall accuracy and reasons for inaccuracy were related to individual differences, randomization tests of the correlations between these constructs as described in Sherman and Funder (2009) and Sherman and Serfass (2015) were used. This procedure evaluated whether the distribution of correlations differed from what would be expected by chance. Table 4 displays the results of the randomization tests of the correlations between the POMP scores for accuracy and reasons for inaccuracy with each Big Five trait, Happiness, Life Satisfaction and affect (positive and negative). How participants remembered their daily activities was not related to these individual difference variables.

Table 4  
*Randomization test results for correlations of POMP scores of accuracy and reasons for inaccuracy with individual difference variables*

	<u>E</u>	<u>A</u>	<u>C</u>	<u>N</u>	<u>O</u>	<u>Happiness</u>	<u>RLSS</u>	<u>Affect</u>
Average absolute r	.021	.066	.067	.042	.046	.048	.069	.050
r expected by chance	.045	.046	.045	.045	.046	.046	.045	.046
Standard error	.015	.016	.015	.015	.016	.015	.015	.011
P-value	.958	.110	.091	.550	.466	.415	.075	.340

*Note.* N= 310.

Next, we sought to identify if recall accuracy was related to features of the situation, if the activity was something that the participant did often, or if the participant felt that the activity was a good use of their time, as these may make the activities more salient to the participant when recalling them retrospectively. Logistic multilevel models were used with activities nested within persons, predicting whether the activity was accurately recalled (accuracy here includes the Match, Language and Specificity categories). The between-person intraclass correlation — estimated using the unconditional model, with no predictors — yielded an ICC = .09, suggesting that some of the variability in recall accuracy is between-person, thus multilevel models are appropriate. Statistical significance was determined by calculating confidence intervals around slopes (Pinheiro & Bates, 2000), which were then converted to odds ratios. Neither participant ratings of how often they engage in the activity ( $b = .026$ ,  $SE = .053$ ,  $OR = 1.03$ , 95% CI [.916, 1.132]), nor whether the activity was a good use of their time ( $b = .081$ ,  $SE = .051$ ,  $OR = 1.08$ , 95% CI [.979, 1.203]) were related to accurate recall.

### **General Discussion**

The present study is the first to directly examine the accuracy of retrospective assessments of daily activities in a large sample. In research on other experiences (e.g. pain), evidence suggests that many common biases may have undue influence on recall. In the present study, an estimate of the extent to which recall is accurate in a day was identified, as well as the prevalence of several errors in recall. These errors in recall occur across activity categories, but some activities are recalled more accurately than others. Lastly, how accurate a person remembers their daily activities is not related to the

individual difference variables measured here, nor to the degree to which the participant deemed the activity a good use of their time or how often they engage in the activity.

By comparing retrospectively recalled daily activities to in situ reports of activities an estimate of how accurately participants remember their daily activities was identified. While only 53% of daily activities were recalled exactly, differences in the responses due to language and specificity can be considered accurate recall for many purposes. Including these as accurate brings the percentage of accurately recalled activities up to 74%. For some purposes, one may be interested in participants remembering their daily activities in the correct order, and not necessarily at the exactly right time. In this case, some of the activities coded as being incorrectly recalled due to timing may be considered accurate as well. These estimates are considered upper bound due to the design of the study requiring participants to report and then recall events, the initial reporting may cue participants to recall their events with more accuracy. Additionally, participants may have used aides, such as a calendar, when recalling their daily activities. However, even an event that has been recalled frequently is prone to decays in memory over longer periods, this may be true in the short term as well (Talarico & Rubin, 2003).

It appears that different kinds of activities are remembered with different degrees of accuracy. Time spent in class and working<sup>3</sup> were both accurately recalled 87% of the time, when matches as well as errors due to language and specificity are considered accurate. These activity categories may be more likely than other activities to be

---

<sup>3</sup> While time with family was remembered accurately 93%, it was only reported by participants 15 times.

scheduled far in advance, occur at regular times for weeks or months at a time. Time spent exercising, sleeping, relaxing, doing housework and shopping<sup>4</sup> were recalled accurately less than 70% of the time.

The degree to which the amount of accurate recall detailed here is impressive or good may be debated and may depend on context. For example, in one study, 56% of law enforcement personnel believed that an alibi generated 24-hours after a crime was very or extremely unlikely to contain mistakes (Dysart & Strange, 2012), these 56% may be surprised by the findings presented here that participants only recalled 74% of their activities accurately. However, it is much more common to collect alibis a week rather than a day after a crime (Burke, Turtle, & Olson, 2017).

These findings viewed in another context, might be expected or even high. Forgetting is often viewed as a feature of memory and not as a failure of memory. From this perspective, forgetting is an adaptive mechanism (see Fawcett & Hulbert, 2020 for a review) with benefits that include allowing us to maintain a positive and stable self-image, to curate our memories for efficient recall, and to identify similarities between situations for problem solving.

The frequency with which activities were confabulated may warrant further attention. These confabulations occurred when recalling nearly every activity category. On average, participants confabulated one out of every ten activities recalled. However, these too may be considered an upper bound estimate, as the event being recalled may

---

<sup>4</sup> As with time with family, time spent with a partner was reported very rarely and is not interpreted.

have taken place between experience sampling reports, thus appearing to be confabulated when it was not<sup>5</sup>.

The present study was not able to find correlates of recall accuracy. How a participant recalled their activities was not related to their Big Five traits, happiness, or life satisfaction. Nor was accuracy related to how often the participant described having done that activity or if they thought the activity was a good use of their time.

**Limitations.** The findings presented here are novel and as such a metric to interpret them on is not readily available. The results presented here should be interpreted as estimates from a single sample of undergraduates who were each assessed on a single day. Additionally, the retrospective assessment used here required participants to name an activity that happened for each hour of the day. This may not allow for a perfect recall of lived experience as not all activities take exactly one hour, starting on the hour, and some activities may best be described by multiple activity categories, which the method did not allow (e.g. exercising with a friend).

**Future research.** First and foremost, attempts need to be made to replicate these novel results in a variety of populations and with a variety of methods. Attempts to do so can answer questions including how recall accuracy changes with age, if some methods allow for more accurate recall than others, how recall of a day degrades over time (recalling the day today or the day two weeks ago), and if certain lifestyles facilitate

---

<sup>5</sup> For example, a participant may have reported in successive experience sampling surveys that at 10:05am they were in class, and at 11:30am they were eating. When retrospectively recalling what they had done from 10 to 11:00am they may have stated that they were exercising. To a judge coding these responses, this may appear to be confabulated, however the participant may have exercised from 10:10-11:20 and was not sent an experience sampling survey during that activity.

accurate recall (e.g. students vs working adults). Future projects may seek to utilize alternative methods for capturing behavior in situ (e.g. via the Electronically Activated Recorder or wearable cameras; Brown, Blake & Sherman, 2017; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). Additionally, research on memory and the self suggest that individuals may distort their memories to cohere with current goals and self-perceptions (Conway, 2005; Greenwald, 1980). Perhaps some of the discrepancies that were seen between reported and recalled activities can be explained by motivational factors (e.g. participants wanting to exercise more may over report how long they exercised for when assessed retrospectively). There has been a growing interest among memory researchers in understanding memory outside of controlled setting (Crozier, Moeck, Weinsheimer, McDonald & Baldassari, 2016; Gruneberg & Herrmann, 2003; Wright, 1997). For the most part, this research has centered on single events that are often of outsized importance (e.g. eye witness testimony; Loftus & Palmer, 1996). The results reported here suggest that memory of mundane daily events may be an interesting avenue for research on memory outside of the research laboratory.

The current study sought a better understanding of how daily activities are remembered. Daily activities have the advantage of being relatively verifiable, we are comfortable assuming that the activities reported via experience sampling are accurate, and those activities can either be recalled accurately or not. The following study investigated a different process that unfolds over the day and the retrospective assessment of it, however this phenomenon operates very differently — the affective experience throughout the day and the affective evaluation of the day as a unit.



### **Affective evaluations of the day**

“I cannot make my days longer so I strive to make them better” -Paul Theroux

Affect has an omnipresent influence on behavior and experience. It influences how we go about our daily lives (e.g. decision making, how we perceive others and how others perceive us; Forgas, 2003; Loewenstein & Lerner, 2003), simultaneously the experience of affect is impacted by several environmental factors (e.g. anticipating pain, social interaction; Lamb, 1976; Watson & Clark, 1997).

Positive and negative affect are independent from one another, with distinct patterns of relations with daily behavior (Diener & Emmons, 1984). Stressful, aversive daily experiences bring on heightened negative affect. Anxiety is increased when individuals anticipate pain (Hodges & Spielberger, 1966; Lamb, 1976), or when being evaluated by others (Johnson, 1968). These negative experiences influence negative affect, while positive affect is largely unaltered (Watson & Pennebaker, 1989).

Conversely, social interactions largely increase positive affect; time spent in social interactions is related to higher positive affect (but not less negative affect; Watson, 1988). Although the nature of social activity certainly matters, as social interactions can be in the form of arguments, which are typically unpleasant (Berry & Hansen, 1996). Exercise has also been shown to improve positive affect but not necessarily decrease negative affect (Peluso & Andrade, 2005).

Affect not only flows from daily behavior but influences it as well. Positive affect may buoy intrinsic motivation, leading workers to spend more time engaged in work related tasks (Isen & Reeve, 2005). Individuals are more willing to make purchases when

positive affect is high (Brown, Homer, & Inman 1998), while experiencing more negative affect increases impulsive consumption (Tice, Bratslavsky, & Baumeister, 2001). In addition to affect influencing and being influenced by daily experiences, there seems to be a rhythm to the affect experienced throughout the day, such that positive affect increases steadily throughout the course of the day (Clark, Watson, & Leeka, 1989), while negative affect does not show a consistent pattern (Thayer, Takahashi, & Pauli, 1988).

While much can be gleaned by looking at the affective experience in the lab, trends in affect and affect as it is recalled, the experience of affect is a complex and transitory phenomenon that varies over the day (Schwarz, 2007). Measuring affect as it is recalled largely mirrors the measurement of affect as it experienced (Dockray, Grant, Stone, Kahneman, Wardle & Steptoe, 2010), however there are many reasons why we would not expect these measures to be the same, including the numerous biases that influence recall of fleeting phenomena like affect (Reis, Gable & Maniaci, 2014).

To bypass these biases, offering a better window into the minute-to-minute fluctuations in affect, experience sampling studies have become more popular (e.g. Bolger, Davis & Rafaeli, 2003; Csikszentmihalyi, & Larson, 2014). Evidence has been mixed on how experience sampling of affect relates to the affective evaluation of the day as a whole. Discrepancies between the measurement of affect throughout the day and the affective evaluation of the day is not solely the result of biases in retrospective reports (Neubauer, Scott, Sliwinski & Smyth, 2019). These measurements can be considered different conceptualizations of affect, with different uses. While momentary ratings of

experiences, such as vacations or medical procedures, might be more valid measurements of how they were experienced, retrospective evaluations are better predictors of future behaviors, such as desires to go on similar vacations or returning for a follow-up procedure (Redelmeier, Katz, & Kahneman, 2003; Wirtz, Kruger, Scollon, & Diener, 2003). Additionally, differences between the average affect over experience sampling reports and overall evaluations of that time period are expected (e.g. Ganzach & Yaor, 2019; Miron-Shatz, Stone, & Kahneman, 2009). At the same time, affect measured via experience sampling and retrospectively evaluated should correspond (Röcke, Hoppmann, & Klumb, 2011). For example, if positive affect is measured throughout the day and evaluated retrospectively, the average of affect measured throughout the day will likely be different than the affective evaluation of the day, however days where more positive affect is experienced should be detectable through both methods (for a review of the processes involved and the research on it, see Neubauer, Scott, Sliwinski, & Smyth, 2019).

What it is about the affective experience throughout the day that impacts how the day is evaluated as a whole is still not well understood. Some argue that peak affect — rather than mean or most recently reported — relates the most strongly to global affective ratings of the day (e.g. Hedges, Jandorf, & Stone, 1985). While others provide evidence that the average affect reported via experience sampling most strongly relates to daily affect ratings (e.g. Ganzach & Yaor, 2019; Neubauer, Scott, Sliwinski, & Smyth, 2019), and still others suggest that the peak and the most recent affect reported are important as well (Fredrickson, 2000; Parkinson, Briner, Reynolds, & Totterdell, 1995). Research on

recall of affect and pain taken out of the context of the day provides evidence that recall of affect over specific episodes (e.g. a medical procedure) is predicted by the average of peak and end affect (Do, Rupert & Wolford, 2008; Fredrickson, 2000; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Redelmeier & Kahneman, 1996).

One explanation for the lack of convergence in findings across studies is that the emphasis in the studies is not on generalizability. In studies that directly assess affect throughout the day and affect over the day, the same scale is often used with minor changes made to the instructions (e.g. the same items are rated for the extent to which the participants feels them in the moment versus how much they felt them that day). This design may allow for biases to increase congruence between ratings. People tend to want to appear consistent and this tendency is thought to be most impactful when people are asked to describe their accounts of their feelings and behaviors (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff & Organ, 1986; Schmitt, 1994).

The current project sought to assess how the affective experience throughout the day manifests in day evaluations using different measures to assess affect in the moment and at the end of the day. Toward this end, affect throughout the day was quantified in several ways, including mean, peak, and most recent. These were then compared to evaluations of the day overall.

## **Method**

### **Participants**

Participants consisted of 384 undergraduates (the same sample as Study 1) recruited from the psychology department's subject pool at a public university in

Southern California. Participation was restricted to those who own a smartphone.

Participants average age was 19.37 ( $SD = 1.66$ ; range: 18-31); 39% were Asian or Pacific Islander, 33% Latino/Latina, 9% White, 7% Middle Eastern or Indian, and 8% were mixed, other, or had missing data.

### **Procedures**

Participants completed measures over two days. On day 1, participants provided demographic information, a broad assessment including personality traits (see appendix C for a full list of measures) and provided their email address, cell phone number and carrier for experience sampling surveys to be text messaged and emailed to them the next day. This initial survey was completed by 384 participants at any time on day 1. On the morning of day 2, participants received a text message and email reminder about their participation in the study, and that they must complete each short survey within 20 minutes of receiving it, when it is safe, and not interrupting a class or other responsibilities. Participants received links to the surveys via text message and email 10 times throughout the day between 9:15am and 7:30pm. These surveys were distributed semi-randomly, such that there was a minimum of 45 minutes between observations and a maximum of 1.5 hours. Each survey took no more than five minutes to complete and asked about participants current affect. 326 participants (85%) completed at least two of these experience sampling surveys with 298 (78%) completing 5 or more. In the evening participants received an email and text message reminder to provide their affect at the end of the day (338 participants, 88%, provided this information) and complete a day evaluation (302 participants, 79%, completed this assessment). The day evaluation was

completed using qsortware.net (Pruned, 2013) and all other questions were administered using Qualtrics survey software. 276 participants (72%) completed at least two experience sampling surveys and the day evaluation, data from these 276 participants are reported here.

## Measures

**Demographics.** Participants provided information on their age, gender, and ethnicity.

**Experience sampling of affect.** The Affect Adjectives Scale (Diener & Emmons, 1984) was used to assess positive and negative affect in the moment. Four items were added to the scale to ensure that the scale captures both high and low arousal emotions (Peaceful/Serene, Dull/Bored, Excited and Tired). Participants indicated the extent to which each of 13 items described their current emotions on a 7-point scale from *not at all* to *extremely*. For each of the 11 times surveys were sent to participants<sup>6</sup>, alpha reliability for Positive Affect ranged from .87 to .92, and for Negative Affect ranged from .79 to .84. For each survey sent, Positive and Negative Affect correlated between -.46 and -.32.

Descriptive statistics derived from the experience sampling reports of affect include average, median, minimum, maximum, standard deviation, range, last report (end), the average between the maximum and last report (peak-end), intercept, and slope for positive and negative affect for each participant.

---

<sup>6</sup> Surveys were matched across participants by survey sent such that if a participant completed the first and third experience sampling surveys sent to them but not the second, they would have scores recorded for the first and third, and the second would contain missing data.

**Day Evaluation.** The Day Evaluation Q-sort (DEQ2.0) was developed to work toward a scientific understanding of the day as a unit. Days are described by recounting their most salient, most personally important, and most unusual features. When seeking to describe a day, one or more variables can be rated for how descriptive they are of the day, or one or more days can be identified, and the most descriptive variables of those days can be determined, Q-sorts describe their subject in this way (Ozer, 1993; Block, 1961).

DEQ items were created by identifying themes in open ended responses in two studies, then refined after an initial investigation using the item pool. The first open ended responses were elicited from 225 undergraduates recruited for a study on daily behavior. These students were asked to describe their day in a few words. A second set of participants responded to two prompts paralleling the life story narrative and the high point prompts (McAdams, 2001) both adapted for the day. Themes were identified from the responses to these questions, and 39 items were created and administered 192 adult participants via Amazon's Mechanical Turk (MTurk). Results of this study suggested that item content needed to be revised and a new version of the Day Evaluation Q-sort (DEQ2.0) was adapted such that bipolar items stating that high and low placement meant opposites, and not characteristic or uncharacteristic [e.g. "Had a significant failure (Note- low placement indicates having a significant success)"], were split into two items (e.g. "Had a significant failure" and "Had an important success"). Additionally, the item "Felt discriminated against (Note- low placement indicates feeling as if you benefitted from

favoritism)” was made less extreme due to low endorsement (“Felt that I was treated unfairly”).

The DEQ2.0 (Miller & Ozer, in prep) consists of 42 items which cover a broad range of constructs present in the day including affect (e.g. “A calm day” and “A good day”), accounts of events (e.g. “Had many satisfying social interactions” and “Filled with a variety of activities”), and items noting accomplishments (e.g. “Difficult to get anything accomplished” and “A productive day”). Participants first sorted the items into three categories: Uncharacteristic, Neutral and Characteristic. Items presented in these categories, were then sorted by participants into the 7-category distribution from *extremely uncharacteristic* to *extremely characteristic* with a forced quasi-normal distribution (4-6-7-8-7-6-4).

Previous research on the DEQ (Miller & Ozer, in prep) has demonstrated that it can be used to describe daily experience, that day evaluations relate meaningfully to personality traits and to time spent in various daily activities. Additionally, finding that two people spending their time similarly may evaluate the day differently. Day evaluations relate to how people tend to experience affect (participants indicated how often they experienced moods over the past year). For example, Positive Affect was significantly correlated with items such as “An interesting day” (.16), and Negative Affect is related to items such as “A tiring day” (.17), while both related (in opposite directions) to items “A good day” (.22,-.21 respectively) and “A day filled with worries” (-.21,.21 respectively).



*Affect templates.* Three graduate students with expertise in affective science independently completed DEQ2.0 templates. The first template asked them to “describe a day in which much positive affect is experienced,” and the second used the same prompt but about negative affect. For positive affect, the judges correlated at .85-.91 ( $M = .87$ ),  $\alpha = .95$ , for negative affect judges correlated at .88-.89 ( $M = .88$ ),  $\alpha = .96$ , and an average across judges for each template was created. These two templates were correlated at  $r = -.95$ . Each participant’s Q-sort was then correlated with these templates to get an estimate of the degree to which each participants’ days were characterized by positive and negative affect. For positive affect, on average participants correlated with the judge template  $r = .30$  ( $SD = .36$ ; *Range*: -.80 to .81), and for negative affect on average participants correlated with the judge template  $r = -.29$  ( $SD = .36$ ; *Range*: -.82 to .79). These two measures were correlated extremely highly,  $r = -.99$ .

Due to the strong correlation between judge templates, two other metrics were used to evaluate the degree to which participants’ days were characterized by positive and negative affect instead. First, participant mean item placement on the items that judges sorted into the “extremely characteristic” and “fairly characteristic” categories (the ten highest placed items) were calculated. These ten positive affect items are noted in Table 10 with a superscript “+10” and the ten negative affect items are noted in Table 12 with a superscript “-10”. For positive affect, on average participants  $Mean = 4.4$  ( $SD = .81$ ; *Range*: 1.8 to 5.9;  $\alpha = .69$ ), for negative affect on average participants  $Mean = 3.25$  ( $SD = .89$ ; *Range*: 1.8 to 6.0;  $\alpha = .75$ ), and these two composites correlate at  $r = -.83$ . Lastly, participant mean item placement on the items judges placed in the “extremely

characteristic” (the four highest placed) items were calculated. For positive affect, these four items “A fun day”, “Had many satisfying social interactions”, “A rewarding day”, and “A happy day” (also noted with a superscript “+4” in Table 10) had  $Mean = 4.6$  ( $SD = 1.1$ ;  $Range: 1.75$  to  $6.75$ ;  $\alpha = .65$ ). For negative affect, the four items “A bad day”, “A stressful day”, “Had negative interactions with others”, and “An upsetting day” (also noted with a superscript “-4” in Table 12) had  $Mean = 3.1$  ( $SD = 1.1$ ;  $Range: 1.5$  to  $6.5$ ;  $\alpha = .64$ ). These two composites correlate at  $r = -.62$ . This correlation between negative and positive affect using the four highest placed items, while still high, is the lowest among the templates and will be used to evaluate experience sampling affect descriptors. The correlations among these metrics are displayed in Tables 5 and 6.

## **Results and Discussion**

### **Evaluating affect measures**

Table 5 displays the correlations among the descriptive statistics derived from the experience sampling affect ratings and positive affect composites derived from the DEQ2.0. Examination of this correlation matrix and principal components analysis (excluding DEQ2.0 composites) of it suggests that the experience of positive affect throughout the day is described similarly by the mean, median, minimum, maximum, peak-end, intercept and last report (end). While the standard deviation, range, and slope were related to the other positive affect descriptors and affect templates to a much lesser degree.

Table 6 displays the correlations among the descriptive statistics derived from experience sampling affect ratings and negative affect templates. By examining this

correlation matrix and using principal components analysis (excluding DEQ2.0 composites) of it the experience of negative affect throughout the day is described similarly by the mean, median, minimum, maximum, standard deviation, range, peak-end, intercept and end. While the slope was related to the other negative affect descriptors and affect templates to a much lesser degree.

Table 5

*Correlations among experience sampling affect ratings of positive affect with positive affect templates*

	<u>Top four mean</u>	<u>Mean</u>	<u>Median</u>	<u>Min</u>	<u>Max</u>	<u>SD</u>	<u>Range</u>	<u>End</u>	<u>Peak-end</u>	<u>Intercept</u>	<u>Slope</u>
<u>Top 10 mean</u>	.88	.53	.50	.48	.43	.02	.00	.44	.47	.46	.09
<u>Top 4 mean</u>	-	.52	.49	.47	.43	.03	.01	.46	.49	.42	.16
<u>Mean</u>		-	.98	.84	.85	.13	.10	.78	.89	.88	.13
<u>Median</u>			-	.79	.80	.15	.11	.77	.86	.86	.11
<u>Min</u>				-	.55	-.35	-.38	.67	.67	.74	.12
<u>Max</u>					-	.55	.57	.66	.90	.70	.17
<u>SD</u>						-	.96	.10	.35	.07	.08
<u>Range</u>							-	.07	.33	.05	.07
<u>End</u>								-	.92	.56	.40
<u>Peak-end</u>									-	.69	.32
<u>Intercept</u>										-	-.35

*Note.*  $N=276$ ; Slope and intercept are affect rating regressed on time (in hours) intercept is centered on the time of each participant's first response.

Table 6

*Correlations among experience sampling affect ratings of negative affect with negative affect templates*

	<u>Top four</u>	<u>Mean</u>	<u>Median</u>	<u>Min</u>	<u>Max</u>	<u>SD</u>	<u>Range</u>	<u>End</u>	<u>Peak-end</u>	<u>Intercept</u>	<u>Slope</u>
	<u>mean</u>										
<u>Top 10 mean</u>	.88	.60	.56	.53	.58	.34	.38	.51	.60	.52	.08
<u>Top 4 mean</u>	-	.51	.47	.44	.51	.33	.36	.46	.53	.44	.07
<u>Mean</u>		-	.98	.90	.87	.50	.50	.84	.94	.88	.09
<u>Median</u>			-	.89	.78	.38	.39	.85	.89	.85	.12
<u>Min</u>				-	.67	.16	.15	.77	.78	.75	.18
<u>Max</u>					-	.82	.74	.67	.93	.82	-.04
<u>SD</u>						-	.97	.32	.65	.55	-.17
<u>Range</u>							-	.32	.65	.55	-.17
<u>End</u>								-	.90	.62	.36
<u>Peak-end</u>									-	.80	.16
<u>Intercept</u>										-	-.36

*Note.*  $N=276$ ; Slope and intercept are affect rating regressed on time (in hours) intercept is centered on the time of each participant's first response.

Examination of these correlation matrices did not aid in identifying what it is about the experience of affect throughout the day that impacts affective evaluations of the day. To help identify what aspect of the affective experience throughout the day is most impactful on the affective evaluation of the day, structure coefficients were used (Meredith, 1965; Ziglari, 2017). Structure coefficients are the bivariate correlation between the predictor ( $X$ ) and the predicted value of the criterion ( $\hat{Y}$ ; Cooley & Lohnes, 1971; Thompson & Borrello, 1985). These coefficients can further our understanding of the importance of each predictor in a model as they indicate how well each predictor can explain the criterion. For positive and negative affect, the four item composites were regressed on all experience sampling affect variables, respectively (Tables 7 and 8). For both models, the range and peak-end were not defined because of singularities and were thus excluded.

Table 7  
*Positive affect: correlations, regression and structure coefficients*

	$r$	$b$ (SE)	$\beta$	Structure Coefficients
<u>(Intercept)</u>	-	2.72 (.24)***	-	-
<u>Mean</u>	.52*-.**	.56 (.62)	.55	.95
<u>Median</u>	.49***	-.42 (.30)	-.44	.89
<u>Min</u>	.47***	.20 (.23)	.20	.88
<u>Max</u>	.43***	-.18 (.22)	-.20	.80
<u>SD</u>	.03	.46 (.53)	.26	.06
<u>End</u>	.46***	.06 (.08)	.08	.85
<u>Intercept</u>	.42***	.32 (.36)	.33	.77
<u>Slope</u>	.16**	2.23 (1.96)	.22	.30

Note.  $N=276$ ; \*  $p \leq .05$ , \*\*  $p \leq .01$ , \*\*\* $p \leq .001$

Table 8

*Negative affect: correlations, regression and structure coefficients*

	<u>r</u>	<u>b (SE)</u>	<u>β</u>	<u>Structure Coefficients</u>
<u>(Intercept)</u>	-	1.39 (.18)***	-	-
<u>Mean</u>	.51***	.14 (.94)	.10	.92
<u>Median</u>	.47***	-.19 (.51)	-.15	.87
<u>Min</u>	.44***	-.26 (.45)	-.15	.80
<u>Max</u>	.51***	.77 (.28)**	.80	.94
<u>SD</u>	.33***	-1.27 (.78)	-.35	.61
<u>End</u>	.46***	.21 (.13)	.18	.84
<u>Intercept</u>	.44***	.01 (.41)	.01	.80
<u>Slope</u>	.07	.23 (2.10)	.02	.14

Note.  $N=276$ ; \*  $p \leq .05$ , \*\*  $p \leq .01$ , \*\*\*  $p \leq .001$

Examination of the structure coefficients do not point to a small set of variables. As many variables seem equally appropriate, subsequent analyses use mean, max, end, and peak-end, as prior studies have focused on these (e.g. Hedges, Jandorf, & Stone, 1985; Parkinson, Briner, Reynolds, & Totterdell, 1995; Kahneman et al., 1993).

### **Relations between DEQ2.0 affect and affective experience throughout the day**

To understand how the affective experience of the day impacts the content of day evaluations, the descriptions of the mean, max, end, and peak-end for both positive and negative affect were related to descriptions of the day using the DEQ2.0. Correlations of DEQ2.0 items with affect variables were evaluated using randomization tests described in Sherman and Funder (2009) and Sherman and Serfass (2015) were used. This procedure evaluated whether the distribution of DEQ2.0 item correlations with each experience sampling affect descriptor differed from what would be expected by chance.

For positive affect the randomization tests between DEQ2.0 items and mean, max, end and peak-end positive affect are shown in Table 9. All four show relationships between the constructs. Table 10 displays the correlations between individual DEQ2.0

items and each positive affect descriptor derived from experience sampling. The pattern of relationships is similar across all measures, profile correlations of these correlations (e.g. taking the vector of item correlations with mean positive affect and correlating that with the vector of item correlations with maximum positive affect) range from .97 to .993 ( $M = .99$ ). The four strongest positive correlations for all four experience sampling of positive affect descriptors being “A happy day”, “A fun day”, “A good day” and “A carefree day”. While the strongest negative correlations include “A day filled with worries”, “A bad day” and “A stressful day”.

Table 9  
*Randomization test results for correlations of DEQ2.0 items with positive affect descriptors*

	<u>Mean</u>	<u>Max</u>	<u>End</u>	<u>Peak-end</u>
Average absolute r	.223	.174	.179	.193
r expected by chance	.048	.048	.048	.048
Standard error	.009	.009	.009	.009
P-value	<.0001	<.0001	<.0001	<.0001

*Note.*  $N=276$ ;



Table 10  
*Correlations of DEQ2.0 items with positive affect descriptors*

<u>Item</u>	<u>Mean</u>	<u>Max</u>	<u>End</u>	<u>Peak-end</u>
A fun day <sup>+10, +4</sup>	.45***	.34***	.44***	.43***
A busy day	-.02	-.05	-.06	-.06
A good day <sup>+10</sup>	.41***	.33***	.31***	.35***
A bad day	-.35***	-.27***	-.33***	-.33***
A productive day	.11	.10	.00	.05
Difficult to get anything accomplished	-.18**	-.14*	-.14*	-.15*
Full of small annoyances	-.23***	-.19**	-.19**	-.21***
Did not feel well	-.35***	-.23***	-.24***	-.26***
Felt healthy and strong <sup>+10</sup>	.31***	.22***	.25***	.26***
A boring day	-.27***	-.21***	-.28***	-.28***
Progress was made toward accomplishing an important goal	.22***	.21***	.19***	.22***
An unusual day- not at all routine	.03	.06	.00	.03
A stressful day	-.36***	-.26***	-.31***	-.32***
Time passed quickly <sup>+10</sup>	.07	.03	.00	.02
As expected; there were no surprises	-.09	-.05	-.07	-.07
A tiring day	-.23***	-.18**	-.23***	-.22***
Had many satisfying social interactions <sup>+10, +4</sup>	.21***	.20***	.20***	.22***
A lonely day	-.19***	-.14*	-.19***	-.18***
Spent much of the day interacting with a romantic partner <sup>+10</sup>	.13*	.14**	.11	.14*
Had negative interactions with others	-.24***	-.21***	-.22***	-.23***
An eventful day	.13*	.09	.15*	.13*
An emotional day (filled with emotional highs and lows)	-.16**	-.09	-.07	-.09
An upsetting day	-.29***	-.24***	-.22***	-.25***
A disappointing day	-.30***	-.25***	-.21***	-.26***
An interesting day	.19**	.10	.17**	.15*
Finished a task I had been working on	.07	.07	.05	.06
A rewarding day <sup>+10, +4</sup>	.28***	.22***	.22***	.24***
Able to do what I wanted	.31***	.26***	.23***	.27***
Filled with a variety of activities <sup>+10</sup>	.09	.02	.07	.05
A happy day <sup>+10, +4</sup>	.50***	.45***	.43***	.48***
An active day	.15*	.07	.08	.08
A calm day	.22***	.16**	.17**	.18**
A carefree day <sup>+10</sup>	.28***	.21***	.26***	.26***
Had a significant failure	-.17**	-.14*	-.05	-.10
Will remember this day better than most	.17**	.09	.26***	.19**
A frustrating day	-.37***	-.32***	-.27***	-.32***
Felt I was treated unfairly	-.10	-.04	-.10	-.08
The weather had a negative impact on me	.09	.05	-.01	.02
Had an important success	.27***	.25***	.22***	.26***
Failed to complete a task I had planned to complete	-.17**	-.16**	-.11	-.14*
A day filled with worries	-.45***	-.37***	-.34***	-.39***
A setback was experienced when working toward an important goal	-.10	-.07	-.06	-.07

Note.  $N = 276$ ; \*  $p \leq .05$ , \*\*  $p \leq .01$ , \*\*\* $p \leq .001$

For negative affect the randomization tests between DEQ2.0 items and mean, maximum, last reported (end) and peak-end negative affect are shown in Table 11. All five show relationships between the constructs. Table 12 displays the correlations between individual DEQ2.0 items and negative affect descriptors. As was the case with positive affect, the pattern of relationships across the four experience sampling of negative affect descriptors is similar, profile correlations range from .97 to .997 ( $M = .99$ ). The four strongest positive correlations for all four descriptors are “A frustrating day”, “An upsetting day”, “A stressful day”, and “A day filled with worries”, while the strongest negative relationships were with “A good day”, “A happy day”, and “Able to do what I wanted”.

Table 11  
*Randomization test results for correlations of DEQ2.0 items with negative affect descriptors*

	<u>Mean</u>	<u>Max</u>	<u>End</u>	<u>Peak-end</u>
Average absolute r	.236	.227	.204	.237
r expected by chance	.048	.048	.048	.048
Standard error	.009	.009	.009	.009
P-value	<.0001	<.0001	<.0001	<.0001

*Note.*  $N=276$ ;

Table 12  
*Correlations of DEQ2.0 items with negative affect descriptors*

<u>Item</u>	<u>Mean</u>	<u>Max</u>	<u>End</u>	<u>Peak-end</u>
A fun day	-.32***	-.32***	-.32***	-.35***
A busy day	.03	-.02	.09	.04
A good day	-.50***	-.45***	-.41***	-.48***
A bad day <sup>-10, -4</sup>	.27***	.29***	.28***	.31***
A productive day	-.24***	-.25***	-.19**	-.25***
Difficult to get anything accomplished	.27***	.25***	.22***	.26***
Full of small annoyances	.28***	.26***	.21***	.26***
Did not feel well	.21***	.23***	.14*	.21***
Felt healthy and strong	-.36***	-.33***	-.34***	-.36***
A boring day	.04	.04	.06	.06
Progress was made toward accomplishing an important goal	-.22***	-.23***	-.19**	-.23***
An unusual day- not at all routine	-.02	.00	.01	.01
A stressful day <sup>-10, -4</sup>	.42***	.40***	.34***	.41***
Time passed quickly	-.27***	-.25***	-.25***	-.27***
As expected; there were no surprises	-.15*	-.15*	-.18**	-.17**
A tiring day	.15*	.11	.13*	.14*
Had many satisfying social interactions	-.22***	-.15*	-.24***	-.21***
A lonely day <sup>-10</sup>	.17**	.13*	.17**	.16**
Spent much of the day interacting with a romantic partner	.04	.02	.09	.06
Had negative interactions with others <sup>-10, -4</sup>	.29***	.30***	.26***	.31***
An eventful day	-.08	-.05	-.11	-.08
An emotional day (filled with emotional highs and lows)	.16**	.20***	.14*	.19**
An upsetting day <sup>-10, -4</sup>	.43***	.44***	.39***	.46***
A disappointing day <sup>-10</sup>	.36***	.38***	.24***	.34***
An interesting day	-.13*	-.15*	-.08	-.13*
Finished a task I had been working on	-.18**	-.18**	-.10	-.16**
A rewarding day	-.32***	-.35***	-.27***	-.34***
Able to do what I wanted	-.41***	-.38***	-.35***	-.40***
Filled with a variety of activities	-.08	-.11	-.03	-.08
A happy day	-.50***	-.43***	-.46***	-.49***
An active day	-.17**	-.23***	-.07	-.17**
A calm day	-.33***	-.30***	-.26***	-.31***
A carefree day	-.32***	-.29***	-.23***	-.29***
Had a significant failure <sup>-10</sup>	.32***	.29***	.29***	.32***
Will remember this day better than most	-.07	-.08	-.13*	-.12
A frustrating day <sup>-10</sup>	.53***	.50***	.43***	.51***
Felt I was treated unfairly	.20**	.23***	.20***	.24***
The weather had a negative impact on me	.07	.03	.09	.06
Had an important success	-.14*	-.14*	-.15*	-.16**
Failed to complete a task I had planned to complete	.09	.09	.01	.05
A day filled with worries <sup>-10</sup>	.36***	.35***	.32***	.37***
A setback was experienced when working toward an important goal <sup>-10</sup>	.18**	.17**	.11	.16**

Note. N = 276; \* p ≤ .05, \*\* p ≤ .01, \*\*\*p ≤ .001

## General Discussion

Research on relations between experience sampling and retrospective evaluations of affect has yet to identify what it is about the affective experience of the day that is most impactful for the evaluation of the day, with some advocating for the mean, maximum, most recent, or an average of the maximum and most recent (e.g. Hedges, Jandorf, & Stone, 1985; Kahneman et al., 1993; Parkinson, Briner, Reynolds, & Totterdell, 1995). The current project found that for both positive and negative affect, all four descriptions of affect throughout the day related to how individuals described their day. Each related appropriately with positive and negative affect. However, meaningful differences in how each related with the description of positive and negative affect were not found.

While the present study was not designed to identify absolute agreement between measures of affective experience and daily evaluations of affect, in line with previous research, these descriptions of affect corresponded (e.g. Dockray, Grant, Stone, Kahneman, Wardle & Steptoe, 2010). Positive affective evaluations of the day correlated with affective experience measures as high as .50 (“A happy day”), and negative affective evaluations of the day correlate with affective measures as high as .53 (“A frustrating day”).

The various measures of affect experienced throughout the day had similar profiles of correlates with day evaluations. Meaning that how the day is evaluated relates similarly to the average affect over the day, the most extreme (positive or negative) affective experience over the day, the most recent affect, and the average of most extreme

and most recent affect. While this finding requires replication, there are reasons why this might be the case. Among them, the DEQ assesses and describes the general experience of the day and does not capture trends or patterns of experience, as such, it is better suited for capturing variation between days than it is for describing variation of phenomena within days.

**Limitations and future research.** Replication of the results presented here are necessary, both through direct and conceptual replications, using similar and varied samples. Results presented here are from a single undergraduate sample on a single day. Perhaps the novelty of reporting on their thoughts, feelings and behaviors impacted the associations between participants experience of affect and the affective evaluation of the day. Additionally, the novelty and frequency of reporting throughout the day may be viewed as intrusive and lead to less valid in situ assessment (Aldwin, 2010). As such, future projects should seek to collect similar data over multiple days. Additionally, future projects should seek to identify if affective experience and affective evaluations differentially predict future behavior. For example, as has been demonstrated with other phenomena, the affective evaluation of the day, might predict future affective experience better than the aggregate of momentary ratings.

### **Conclusions**

The studies presented here describe two distinct features of the lived day. The first examined the accuracy of the recall of daily activities by comparing retrospectively recalled activities to activities reported in the moment. This is a novel description of memory as it is used in daily life. This study provided upper bound estimates that nearly

75% of daily activities were recalled accurately, 13% were recalled at the wrong time, and 10% were confabulated. The second study explored how aspects of affective experience relate to the affective evaluation of the day. This study found that mean, peak, most recent and peak-end affect throughout the day relate appropriately to day evaluations, however there is little discrepancy between how each of these variables related to day evaluations.

Taken together, these studies present evidence that experience sampling and retrospective assessment might capture different conceptualizations of similar phenomena. As prior literature has suggested, clear discrepancies exist when measuring the same construct in multiple ways (e.g. Redelmeier, Katz, & Kahneman, 2003; Wirtz, Kruger, Scollon, & Diener, 2003). For example, while recall of daily activities might not correspond precisely to what a person does throughout the day, perhaps the recall of daily activities predicts some aspects of future behavior better than what the person in fact did. Similarly, assessment of affect over a longer period might predict future affect better than in the moment reports of affect do.

These findings may provide context for recall of other phenomena. For example, recall of pain or other symptoms suggests that when reported in the moment and recalled there is correspondence in the relative levels of symptoms, but there are differences in the mean levels (e.g. Broderick et al., 2008, 2009; Miron-Shatz et al., 2009). While this may be due to salience, sense making, or other biases that are often referenced (Reis, Gable, & Maniaci, 2014), it may also be due to some of the reasons for inaccuracy identified in the current project. For example, recalled symptoms may be higher than symptoms reported

in the moment due to an activity being remembered at the wrong time, or on the wrong day, or being confabulated. Conversely, biases may be leading to the inaccuracies.

## References

- Aldwin, C. M. (2010). *Stress, coping, and development: An integrative perspective* (2nd ed.). New York, NY: Guilford Press
- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, 47(1), i.
- Aschoff, J. (1965). Circadian rhythms in man. *Science*, 148(3676), 1427-1432.
- Barker, R. G., & Wright, H. F. (1951). *One boy's day; a specimen record of behavior*. Oxford, UK: Harper Press.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on Psychological Science*, 2(4), 396-403.
- Berry, D. S., & Hansen, J. S. (1996). Positive affect, negative affect, and social interaction. *Journal of Personality and Social Psychology*, 71(4), 796.
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research* (Vol. 457). Springfield, IL: Thomas.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1), 579-616.
- Broderick, J. E., Schneider, S., Schwartz, J. E., & Stone, A. A. (2010). Interference with activities due to pain and fatigue: accuracy of ratings across different reporting periods. *Quality of Life Research*, 19(8), 1163-1170.
- Broderick, J. E., Schwartz, J. E., Schneider, S., & Stone, A. A. (2009). Can End-of-day reports replace momentary assessment of pain and fatigue?. *The journal of pain: official journal of the American Pain Society*, 10(3), 274-281. doi:10.1016/j.jpain.2008.09.003
- Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain*, 139(1), 146-157. doi:10.1016/j.pain.2008.03.024
- Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science*, 8(5), 592-600.



- Brown, S. P., Homer, P. M., & Inman, J. J. (1998). A meta-analysis of relationships between ad-evoked feelings and advertising responses. *Journal of Marketing Research*, 35(1), 114-126.
- Burke, T. M., Turtle, J. W., & Olson, E. A. (2017). Alibis in criminal investigations and trials. In *the Handbook of Eyewitness Psychology: Volume I* (pp. 157-174). Psychology Press.
- Carney, M. A., Tennen, H., Affleck, G., Del Boca, F. K., & Kranzler, H. R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time" electronic interviews". *Journal of studies on alcohol*, 59(4), 447-454.
- Clark, L. A., Watson, D., & Leeka, J. (1989). Diurnal variation in the positive affects. *Motivation and Emotion*, 13(3), 205-234.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate behavioral research*, 34(3), 315-346.
- Conway, M. A. (2005). Memory and the self. *Journal of memory and language*, 53(4), 594-628.
- Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: John Wiley & Sons.
- Craik, K. H. (2000). The lived day of an individual: A person-environment perspective. *Person-environment psychology: New directions and perspectives*, 2, 233-266.
- Crozier, W. E., Moeck, E. K., Weinsheimer, C. C., McDonald, D. L. L., & Baldassari, M. J. (2016). A history of the Society for Applied Research in Memory and Cognition. *Journal of Applied Research in Memory and Cognition*, 5(2), 103-109.
- Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35-54). Springer, Dordrecht.
- DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological science*, 21(6), 820-828.
- Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of personality and social psychology*, 47(5), 1105.

- Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*(1), 96-98.
- Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J., & Steptoe, A. (2010). A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Social Indicators Research*, *99*(2), 269-283.
- Douglas, H. E., Bore, M., & Munro, D. (2016). Coping with university education: The relationships of time management behaviour and work engagement with the five factor model aspects. *Learning and individual Differences*, *45*, 268-274.
- Dunning, D., Heath, C., & Suls, J. M. (2005). Flawed self-assessment: Implications for education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69-106.
- Dysart, J. E., & Strange, D. (2012). Beliefs about alibis and alibi investigations: A survey of law enforcement. *Psychology, Crime & Law*, *18*(1), 11-25.
- Fawcett, J. M., & Hulbert, J. C. (2020). The many faces of forgetting: Toward a constructive view of forgetting in everyday life. *Journal of Applied Research in Memory and Cognition*.
- Forgas, J. P. (2003). Affective influences on attitudes and judgments. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science. Handbook of affective sciences* (pp. 596-618). New York, NY, US: Oxford University Press
- Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*(4), 577-606.
- Ganzach, Y., & Yaor, E. (2019). The retrospective evaluation of positive and negative affect. *Personality and Social Psychology Bulletin*, *45*(1), 93-104.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*, 603-618.
- Hedges, S. M., Jandorf, L., & Stone, A. A. (1985). Meaning of daily mood assessments. *Journal of Personality and Social Psychology*, *48*(2), 428.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage.

- Hershey, J.W. (1999). *Personality traits and personal goals in everyday behavior* (Unpublished doctoral dissertation). University of California, Riverside, Riverside, CA
- Hodges, W. F., & Spielberger, C. D. (1966). The effects of threat of shock on heart rate for subjects who differ in manifest anxiety and fear of shock. *Psychophysiology*, 2(4), 287-294.
- Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and emotion*, 29(4), 295-323.
- Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality*, 44(4), 501-511.
- John, O. P. (1989). Towards a taxonomy of personality descriptors. In *Personality psychology* (pp. 261-271). Springer, New York, NY.
- Johnson, D. T. (1968). Effects of interview stress on measure of state and trait anxiety. *Journal of Abnormal Psychology*, 73(3p1), 245.
- 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. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), 1776-1780.
- Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Klasnja, P., Harrison, B. L., LeGrand, L., LaMarca, A., Froehlich, J., & Hudson, S. E. (2008). Using wearable sensors and real time inference to understand human recall of routine activities. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 154-163).
- Krueger, A. B., Kahneman, D., Schkade, D., Schwarz, N., & Stone, A. A. (2009). National time accounting: The currency of life. In *Measuring the subjective well-being of nations: National accounts of time use and well-being* (pp. 9-86). University of Chicago Press.

- Lamb, D. H. (1976). On the distinction between psychological and physical stressors. *Psychological Reports, 38*(3), 797-798.
- Litt, M. D., Cooney, N. L., & Morse, P. (2000). Reactivity to alcohol-related stimuli in the laboratory and in the field: predictors of craving in treated alcoholics. *Addiction, 95*(6), 889-900.
- Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. *Handbook of affective science, 619*(642), 3.
- Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research, 46*(2), 137-155.
- Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2018). A New Measure of Life Satisfaction: The Riverside Life Satisfaction Scale. *Journal of personality assessment, 1-1*.
- McAdams, D. P. (1995). What do we know when we know a person? *Journal of personality, 63*(3), 365-396.
- Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. New York, NY, US: Guilford Press.
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers, 33*(4), 517-523.
- Meredith, W. (1964). Canonical correlations with fallible data. *Psychometrika, 29*(1), 55-65.
- Miller, T. J., & Ozer, D. J. (in prep). The Day Evaluation Q-sort.
- Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion, 9*, 885-891.
- Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2019). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of personality and social psychology*.
- Ozer, D. J. (1993). The Q-sort method and the study of personality development. *Studying lives through time: Personality and development, 147-168*.

- Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, 21(4), 331-339.
- Paulhus, D. L. & Vazire, S. (2007). The self-report method. In R.W. Robins, R.C. Fraley, & R.F. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York: The Guilford Press.
- Peluso, M. A. M., & Andrade, L. H. S. G. D. (2005). Physical activity and mental health: the association between exercise and mood. *Clinics*, 60(1), 61-7.
- Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3-56.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences*, 13(4), 707-727.
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12, 69–82.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879.
- Pruneddu, A. (2013) Implicit person theories and Q-sort: Personality change in emerging adults. PhD thesis, University of York.
- 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.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, 104(1-2), 187-194.
- Reis, H. T., Gable, S. L., & Maniaci, M. R. (2014). Methods for studying everyday experience in its natural context. *Handbook of research methods in social and personality psychology*, 2, 373-403.
- Röcke, C., Hoppmann, C. A., & Klumb, P. L. (2011). Correspondence between retrospective and momentary ratings of positive and negative affect in old age: Findings from a one-year measurement burst design. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences*, 66, 411–415.

- Schmitt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior*, 15, 393–398
- Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. *The science of real-time data capture: Self-reports in health research*, 11, 26.
- Schwarz, N. (2012). Why researchers should think “real-time”: A cognitive rationale. *Handbook of research methods for studying daily life*, 22-42.
- Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British journal of sports medicine*, 37(3), 197-206.
- Sherman, R. A., & Funder, D. C. (2009). Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, 43(6), 1053-1063.
- Sherman, R. A., & Serfass, D. G. (2015). The comprehensive approach to analyzing multivariate constructs. *Journal of Research in Personality*, 54, 40-5.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4, 1-32.
- Shrier, L. A., Shih, M. C., & Beardslee, W. R. (2005). Affect and sexual behavior in adolescents: A review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics*, 115(5), e573-e581.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113, 117-143.
- Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological science*, 14(5), 455-461.
- Thayer, R. E., Takahashi, P. J., & Pauli, J. A. (1988). Multidimensional arousal states, diurnal rhythms, cognitive and social processes, and extraversion. *Personality and Individual Differences*, 9(1), 15-24.
- Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and psychological measurement*, 45(2), 203-209.

- Tice, D. M., Bratslavsky, E., & Baumeister, R. F. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it!. In *Self-Regulation and Self-Control* (pp. 275-306). Routledge.
- Tracy, J. L., Robins, R. W., & Sherman, J. W. (2009). The practice of psychological science: Searching for Cronbach's two streams in social–personality psychology. *Journal of Personality and Social Psychology*, *96*(6), 1206.
- Verbrugge, L. M. (1980). Health diaries. *Medical care*, *18*(1), 73-95.
- Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: their relation to health complaints, perceived stress, and daily activities *Journal of personality and social psychology*, *54*(6), 102.
- Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In *Handbook of personality psychology* (pp. 767-793).
- Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: exploring the central role of negative affectivity. *Psychological review*, *96*(2), 234.
- Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*(5), 520-524.
- Ziglar, L. (2017). Interpreting multiple regression results:  $\beta$  weights and structure coefficients. *General Linear Model Journal*, *43*(2), 13-22.

## Appendix A

### Accuracy of recalled activities coding manual

In this project, participants were text messaged surveys throughout the day and asked what they were doing just prior to starting the survey. In the evening, they also recalled what they had been doing throughout the day. We are interested in how accurately participants were able to recall their activities.

Please start by orienting yourself to the excel sheet. Column A contains the Participant ID, which is for organizational purposes only, do not edit or delete it. Column B is the time of the retrospective activity, Column C is the activity the participant stated at the time indicated in Column B. Column D is the text message number, which is for organizational purposes only, do not edit or delete it. Column E is time that each text message was responded to. Column F is the activity the participant stated at the time indicated in E. Column G, labeled “Match”, is where you will be entering codes. The activities are organized by participant such that each participant takes up 12 rows, and their retrospective and text messages activities are laid out together. On the next page is a hypothetical participant and the coding instructions which will walk you through the coding using the hypothetical participant.

	A	B	C	D	E	F	G
1	ID	Retrospective Activity		Text #	Text Message Activity		Match
2	1234	9am to 10:20am	sleep	1	10:05 AM	Sleeping	
3	1234	10am to 11:20am	Working	2	11:12 AM	Working	
4	1234	11am to 12:20pm	Eating lunch	3	12:18 PM	Eating	
5	1234	12pm to 1:20pm	in class	4	1:40 PM	eating	
6	1234	1pm to 2:20pm	walking home	5	3:03 PM	Working	
7	1234	2pm to 3:20pm	Working	6	NA	NA	
8	1234	3pm to 4:20pm	working	7	5:33 PM	Working	
9	1234	4pm to 5:20pm	working	8	6:16 PM	Gym	
10	1234	5pm to 6:20pm	working	9	7:01 PM	Gym	
11	1234	6pm to 7:20pm	Gym	10	8:02 PM	Doing this survey	
12	1234	7pm to 8:20pm	Gym				
13	1234	8pm to 9:20pm	showering				

#### Coding:

1. For each text message (not including NA’s), identify the retrospective activity (or activities) for that participant that matches with the time stated for the text message.

EXAMPLE: text message #1 above took place at 10:05am, both the first and second retrospective activity include this time.

NOTE: You are matching the TIME of the text message with the time range of the retrospective report. Each column is in order of reporting and the rows are not meant to match times.

2. If the stated retrospective activity (or activities) matches with the text message activity, put a “1” next to that text message in Column G (“Match”). If there are two retrospective activities that match with the time of the text message, it does not matter if either or both matches, if one does, it should be coded as a 1



If the activity stated in the text message does not match the retrospective activity exactly but is similar, use your best judgement for if it should be a match or not.

EXAMPLE: Text message #1 happened at 10:05, the retrospective activity from 9am to 10:20am matches, so text message #1 should be coded with a “1”.

3. If the stated (or activities) do not match with the text message activity, or the text message activity was left blank (indicated by “NA”), leave Column G (“Match”) next to that text message blank.

## Appendix B

### Accuracy of recalled activities coding manual

In this project, participants were text messaged surveys throughout the day and asked what they were doing just prior to starting the survey. In the evening, they also recalled what they had been doing throughout the day. In the data you will be coding, it has already been determined that the two descriptions did not match. We are now interested in why some activity reports do not match.

Please start by orienting yourself to the excel sheet. Column A and B contain the row number, and Participant ID which are for organizational purposes only, do not edit or delete them. Column C is the time of the retrospective activity, Column D is the activity the participant stated at the time indicated in Column C. Column E is the text message number, which is for organizational purposes only, do not edit or delete it. Column F is time that each text message was responded to. Column G is the activity the participant stated at the time indicated in F. Column H, labeled “Match”, indicates whether that text message was accurately recalled (1) or not (0). The activities are organized by participant such that each participant takes up 12 rows, and their retrospective and text messages activities are laid out together. On the next page is a hypothetical participant and the coding instructions which will walk you through the coding using the hypothetical participant.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Row	ID	Retrospective Activity		Text #	Text Message Activity		Match	Timing	Confabulation	Language	Specificity	Other
2	2000	1234	9am to 10:20am	Sleep	1	10:05 AM Studying		0					
3	2001	1234	10am to 11:20am	Working	2	11:12 AM Driving to work		0					
4	2002	1234	11am to 12:20pm	Eating lunch	3								
5	2003	1234	12pm to 1:20pm	In class	4	1:40 PM eating		0					
6	2004	1234	1pm to 2:20pm	walking home	5	3:03 PM Studying		0					
7	2005	1234	2pm to 3:20pm	Working	6								
8	2006	1234	3pm to 4:20pm	working	7								
9	2007	1234	4pm to 5:20pm	working	8	6:16 PM Nap		0					
10	2008	1234	5pm to 6:20pm	working	9								
11	2009	1234	6pm to 7:20pm	Gym	10	8:02 PM Doing this survey		0					
12	2010	1234	7pm to 8:20pm	Gym									
13	2011	1234	8pm to 9:20pm	showering									

#### Coding:

Only the text messaged activity descriptions that did not match with the retrospective report for that time are included.

We are now interested in WHY these activities did not match. We have identified a few reasons why they might not be matches, your task is to identify which of these reasons best explains the reason why the text message was not accurately recalled.

Each of the following reasons has its own column for you to indicate in the excel document:

Timing: Retrospective report of activity seems to be correctly recalled, but the time does not match the time of the text message.

Confabulation: Participant retrospectively recalls an activity that seems to be made up.

Language: Participant uses vague language that could be referring to the same activity (e.g. “Studying” vs “In the library”)

Specificity: Participant states a specific activity in the text message that could be a part of the hour-long retrospective activity (e.g. "waiting for class to start" rather than "in class" or "on my phone" rather than "in my room relaxing")

Other: Any other reason

4. When you identify the reason for the miss match, put a 1 in the column that corresponds with the reason you identified.

## Appendix C

### Instrument Table of Contents

---

#### Consent to Act as a Human Research Subject

##### Day 1

Demographic Information

Big Five Inventory 2 (BFI-2)

Affect Adjectives Scale

Subjective Happiness Scale

Riverside Life Satisfaction Scale

Twenty Item Values Inventory

Plastic Behavior Questionnaire

##### Day 2

Texted surveys:

Activity questionnaire

Affect Adjectives Scale

End of day emailed questions:

Daily Activities Measure

Day Evaluation Q-Sort 2.0

##### Debrief

UNIVERSITY OF CALIFORNIA, RIVERSIDE  
CONSENT TO ACT AS A HUMAN RESEARCH SUBJECT

Experience sampling and day evaluations

Investigators: Travis J. Miller, Graduate Student, UCR                      Email:  
Tmill009@ucr.edu  
Daniel J. Ozer, Professor, UCR

**Purpose of study:** The purpose of this study is to investigate how our behavior and experiences throughout the day influences how we evaluate the day as a whole.

**Procedures:** If you consent, you will participate over two days. Part 1 will start immediately following this consent procedure and will last approximately twenty minutes. Then, tomorrow morning you will receive an email and text message about the study and throughout the day you will receive 10 links to the survey via text message and email between 8:30am and 6:30pm. Each survey will take at most five minutes, you must complete the survey within 20 minutes of receiving it, when it is safe, and not interrupting a class or other responsibilities to receive credit. At least 6 of these surveys must be completed to receive credit. Finally, in the evening tomorrow you will receive an email with a final survey which will take approximately twenty minutes. Once you complete the final session you will receive two research participation credits.

Participation involves filling out surveys consisting of questions about your personality, demographic characteristics, providing information about what you are doing throughout your day, writing a passage about your day, providing a list of your daily activities, and completing a measure to describe your day.

You are free to withdraw from the study at any time. However, you must complete the entire study to receive compensation. A total of 350 participants will be recruited for this study.

**Compensation and reimbursement:** After completion you will receive 1.5 research participation credits.

**Risks:** Participation in this study will create no foreseeable risk, harm or discomfort to you.

**Benefits:** By participating in this study you will have the opportunity to learn about psychological research methods, as well as have the chance to engage in self-reflection about yourself and your days.

**Withdrawal or termination from the study:** If at any point you decide that you do not wish to participate in this study, you are free to withdraw by either closing the browser

window or leaving the webpage. If you withdraw you will not receive research participation credit.

**Confidentiality:** Any identifiable information collected during this study will be discarded once data collection is complete. If you consent, your de-identified data may be shared with other researchers.

In some instances, a representative of Office of Research Integrity may review research-related records for quality assurance in order to ensure that relevant laws and guidelines are followed. All information accessed by ORI will be held to the same level of confidentiality that has been stated by the research team.

**If you have questions:** If, during the course of this study, you have any health concerns or if you feel sad or generally unhappy, you are encouraged to call your doctor or mental health practitioner. To find a mental health practitioner near you, please visit <https://www.find-a-therapist.com> (International).

If you have questions about your rights or complaints as a research subject, please contact the IRB Chairperson at (951) 827 - 4802 during business hours, or to contact them by email at [irb@ucr.edu](mailto:irb@ucr.edu). If you have any questions specific to this study please email the primary investigator, Travis J. Miller, at [tmill009@ucr.edu](mailto:tmill009@ucr.edu).

Voluntary participation statement: I understand that participation in this study is voluntary, and I must be at least 18 years of age. I may refuse to answer any question or discontinue my involvement at any time.

By clicking the “Next” button below, you have indicated your consent to participate in this study and will begin filling out the surveys

## Demographics

Please provide the following information about yourself:

1. Age: \_\_\_\_\_
2. Gender: \_\_\_\_\_
3. Ethnicity: \_\_\_\_\_
4. Parent/Guardian 1's level of education:  NA/unknown  Did not finish high school  High school diploma  some college  trade school  Associate's degree  Bachelor's degree  Master's degree  doctorate
  - a. In what country/state did this parent/guardian complete this education:  
\_\_\_\_\_
5. Parent/Guardian 1's level of education:  NA/unknown  Did not finish high school  High school diploma  some college  trade school  Associate's degree  Bachelor's degree  Master's degree  doctorate
  - a. In what country/state did this parent/guardian complete this education:  
\_\_\_\_\_
6. Grandparent 1's level of education:  NA/unknown  Did not finish high school  High school diploma  some college  trade school  Associate's degree  Bachelor's degree  Master's degree  doctorate
  - a. In what country/state did this grandparent complete this education:  
\_\_\_\_\_

7. Grandparent 2's level of education:  NA/unknown  Did not finish high school  
 High school diploma  some college  trade school  Associate's degree  
 Bachelor's degree  Master's degree  doctorate

a. In what country/state did this grandparent complete this education:

\_\_\_\_\_

8. Grandparent 3's level of education:  NA/unknown  Did not finish high school  
 High school diploma  some college  trade school  Associate's degree  
 Bachelor's degree  Master's degree  doctorate

a. In what country/state did this grandparent complete this education:

\_\_\_\_\_

9. Grandparent 4's level of education:  NA/unknown  Did not finish high school  
 High school diploma  some college  trade school  Associate's degree  
 Bachelor's degree  Master's degree  doctorate

a. In what country/state did this grandparent complete this education:

\_\_\_\_\_

10. Growing up, did your family receive a daily newspaper or read the news daily?

YES/NO

11. Growing up, about how many books were in your home, available for you to read?

None, 1 or 2, Around 10, Around 20, Around 50, Around 100, Around 200,

Around 500, 1000 or more



12. Growing up, was there a musical instrument (for example, a piano, drum, guitar)  
for you to use at home? YES/NO

The following information is needed to distribute surveys to you tomorrow:

Email Address: \_\_\_\_\_

Cell phone number: \_\_\_\_\_

Cell phone carrier: \_\_\_\_\_

## Big Five Inventory 2 (BFI-2)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

Scale: 1. Disagree Strongly, 2. Disagree a little, 3. Neutral: no opinion, 4. Agree a little, 5. Agree strongly

### *I am someone who...*

1. Is outgoing; sociable.
3. Tends to be disorganized.
5. Has few artistic interests.
7. Is respectful, treats others with respect.
9. Stays optimistic after experiencing a setback.
11. Rarely feels excited or eager.
13. Is dependable, steady.
15. Is inventive, finds clever ways to do things.
17. Feels little sympathy for others.
19. Can be tense.
21. Is dominant, acts as a leader.
23. Has difficulty getting started on tasks.
25. Avoids intellectual, philosophical discussions.
27. Has a forgiving nature.
29. Is emotionally stable, not easily upset.
31. Is sometimes shy, introverted.
33. Keeps things neat and tidy.
35. Values art and beauty.
37. Is sometimes rude to others.
39. Often feels sad.
41. Is full of energy.
43. Is reliable, can always be counted on.
45. Has difficulty imagining things.
47. Can be cold and uncaring.
49. Rarely feels anxious or afraid.
51. Prefers to have others take charge.
53. Is persistent, works until the task is finished.
55. Has little interest in abstract ideas.
57. Assumes the best about people.
59. Is temperamental, gets emotional easily.
2. Is compassionate, has a soft heart.
4. Is relaxed, handles stress well.
6. Has an assertive personality.
8. Tends to be lazy.
10. Is curious about many different things.
12. Tends to find fault with others.
14. Is moody, has up and down mood swings.
16. Tends to be quiet.
18. Is systematic, likes to keep things in order.
20. Is fascinated by art, music, or literature.
22. Starts arguments with others.
24. Feels secure, comfortable with self.
26. Is less active than other people.
28. Can be somewhat careless.
30. Has little creativity.
32. Is helpful and unselfish with others.
34. Worries a lot.
36. Finds it hard to influence others.
38. Is efficient, gets things done.
40. Is complex, a deep thinker.
42. Is suspicious of others' intentions.
44. Keeps their emotions under control.
46. Is talkative.
48. Leaves a mess, doesn't clean up.
50. Thinks poetry and plays are boring.
52. Is polite, courteous to others.
54. Tends to feel depressed, blue.
56. Shows a lot of enthusiasm.
58. Sometimes behaves irresponsibly.
60. Is original, comes up with new ideas.

#### Affect Adjectives Scale

Using the seven-point scale below, please describe your emotional experience in the past week, think of how you felt throughout the week in general.

1. Not at all, 2. Very Slightly, 3. Somewhat, 4. Moderately, 5. Much, 6. Very Much, 7. Extremely

1. Happy
2. Worried
3. Pleased
4. Angry/Hostile
5. Frustrated
6. Depressed/Blue
7. Relaxed/Calm
8. Unhappy
9. Enjoyment/Fun
10. Peaceful/Serene
11. Dull/Bored
12. Joyful
13. Tired

### Subjective Happiness Scale

For each of the following statements and/or questions, please circle the point on the scale that you feel is most appropriate in describing you.

1. In general, I consider myself:

not a very happy person    1       2       3       4       5       6       7    a very happy person

2. Compared with most of my peers, I consider myself:

less happy    1       2       3       4       5       6       7    more happy

3. Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you?

not at all    1       2       3       4       5       6       7    a great deal

4. Some people are generally not very happy. Although they are not depressed, they never seem as happy as they might be. To what extent does this characterization describe you?

not at all    1       2       3       4       5       6       7    a great deal

### Riverside Life Satisfaction Scale

Please rate your agreement with each of the statements below. Use the 7-point scale provided.

- 1: Strongly disagree
- 2: Moderately disagree
- 3: Slightly disagree
- 4: Neither agree nor disagree
- 5: Slightly agree
- 6: Moderately agree
- 7: Strongly agree

- 1. I like how my life is going.
- 2. If I could live my life over, I would change many things.
- 3. I am content with my life.
- 4. Those around me seem to be living better lives than my own.
- 5. I am satisfied with where I am in life right now.
- 6. I want to change the path my life is on.

## Twenty Item Values Inventory

Here we briefly describe some people. Please read each description and think about how much each person is or is not like you. Using a 6-point scale from “not like me at all” to “very much like me,” choose how similar the person is to you.

6	5	4	3	2	1
very much like me	like me	somewhat like me	a little like me	not like me	not like me at all

### HOW MUCH LIKE YOU IS THIS PERSON?

- \_\_\_\_\_ 1. S/he believes s/he should always show respect to his/her parents and to older people. It is important to him/her to be obedient
- \_\_\_\_\_ 2. Religious belief is important to him/her. S/he tries hard to do what his religion requires.
- \_\_\_\_\_ 3. It's very important to him/her to help the people around him/her. S/he wants to care for their well-being.
- \_\_\_\_\_ 4. S/he thinks it is important that every person in the world be treated equally. S/he believes everyone should have equal opportunities in life.
- \_\_\_\_\_ 5. S/he thinks it's important to be interested in things. S/he likes to be curious and to try to understand all sorts of things.
- \_\_\_\_\_ 6. S/he likes to take risks. S/he is always looking for adventures.
- \_\_\_\_\_ 7. S/he seeks every chance he can to have fun. It is important to him/her to do things that give him/her pleasure.
- \_\_\_\_\_ 8. Getting ahead in life is important to him/her. S/he strives to do better than others.
- \_\_\_\_\_ 9. S/he always wants to be the one who makes the decisions. S/he likes to be the leader.
- \_\_\_\_\_ 10. It is important to him/her that things be organized and clean. S/he really does not like things to be a mess.
- \_\_\_\_\_ 11. It is important to him/her to always behave properly. S/he wants to avoid doing anything people would say is wrong.
- \_\_\_\_\_ 12. S/he thinks it is best to do things in traditional ways. It is important to him/her to keep up the customs s/he has learned.
- \_\_\_\_\_ 13. It is important to him/her to respond to the needs of others. S/he tries to support those s/he knows.
- \_\_\_\_\_ 14. S/he believes all the worlds' people should live in harmony. Promoting peace among all groups in the world is important to him/her.
- \_\_\_\_\_ 15. Thinking up new ideas and being creative is important to him/her. S/he likes to do things in his/her own original way.
- \_\_\_\_\_ 16. S/he thinks it is important to do lots of different things in life. S/he always looks for new things to try.
- \_\_\_\_\_ 17. S/he really wants to enjoy life. Having a good time is very important to him/her.
- \_\_\_\_\_ 18. Being very successful is important to him/her. S/he likes to impress other people.
- \_\_\_\_\_ 19. It is important to him/her to be in charge and tell others what to do. S/he wants people to do what s/he says.
- \_\_\_\_\_ 20. Having a stable government is important to him/her. S/he is concerned that the social order be protected.

The Nature Relatedness Scale (Nisbet et al. 2009)

Instructions: For each of the following, please rate the extent to which you agree with each statement, using the scale from 1 to 5 as shown below. Please respond as you really feel, rather than how you think “most people” feel.

1	2	3	4	5
Disagree strongly	Disagree a little	Neither agree or disagree	Agree a little	Agree strongly

<p>1. I enjoy being outdoors, even in — unpleasant weather.</p> <p>2. Some species are just meant to — die out or become extinct.</p> <p>3. Humans have the right to use — natural resources any way we want.</p> <p>4. My ideal vacation spot would be a — remote, wilderness areas.</p> <p>5. I always think about how my — actions affect the environment.</p> <p>6. I enjoy digging in the earth and — getting dirt on my hands.</p> <p>7. My connection to nature and the — environment is part of my spirituality.</p> <p>8. I am very aware of environmental —</p>	<p>12. I am not separate from nature, — but a part of nature</p> <p>13. The thought of being deep in the — woods, away from civilization, is frightening</p> <p>14. My feelings about nature do not — affect how I live my life</p> <p>15. Animals, birds, and plants — should have fewer rights than humans.</p> <p>16. Even in the middle of the city, I — notice nature around me.</p> <p>17. My relationship to nature is an — important part of who I am.</p> <p>18. Conservation is unnecessary — because nature is strong enough to recover from any human impact.</p> <p>19. The state of non-human species —</p>
---	---

<p>issues.</p> <p>9. I take notice of wildlife wherever I____ am.</p> <p>10. I don't often go out in nature. _____</p> <p>11. Nothing I do will change problems____ In other places on the planet.</p>	<p>is an indicator of the future for humans.</p> <p>20. I think a lot about the suffering _____ of animals.</p> <p>21. I feel very connected to all _____ living things and the earth.</p>
--	--



### Activity Questionnaire

Immediately before you started this survey, what were you doing?

---

Please rate how much you agree with the following statements:

1. Strongly disagree   2. Disagree   3. Neither agree or disagree   4. Agree   5.  
Strongly agree

This activity is a good use of my time

This activity is something I do not do often

I am not feeling well right now

### Affect Adjectives Scale

Using the seven-point scale below, please describe your emotional experience right NOW.

1. Not at all, 2. Very Slightly, 3. Somewhat, 4. Moderately, 5. Much, 6. Very Much, 7. Extremely

1. Happy
2. Worried
3. Pleased
4. Angry/Hostile
5. Frustrated
6. Depressed/Blue
7. Relaxed/Calm
8. Unhappy
9. Enjoyment/Fun
10. Peaceful/Serene
11. Dull/Bored
12. Joyful
13. Tired

### Daily Activities Measure

Instructions: We are interested in what you did earlier today. For each hour of the day we ask you to report:

What were you doing?

Write a brief sentence describing your activities during each hour. If you cannot remember or would rather not say what you were doing, write “forgot” or “rather not say.”

Activity Category

Please pick the category that best summarizes the activity you engaged in during that hour. If you were engaged in more than one activity, please select the primary one.

List of Activities: 1. In class, 2. Studying, 3. Sleeping, 4. Activity with friend, 5. Relaxation, 6. Eating, 7. Locomotion (on the way to somewhere), 8. Work, 9. Grooming, 10. Activity with romantic partner, 11. Exercise, 12. Shopping/errands, 13. Recreation, 14. Activity with family, 15. Housework/chores, 16. Forgot, 17. Rather not say, 18. Other

Who were you with

Do not use names here, just labels like “friend” or “classmate.”

## Day Evaluation Q-sort 2.0

### Initial instructions:

We are interested in your description of your day today. There are many ways to describe a day, and we have listed 39 different statements that might be used to describe a day.

To begin, please sort the items below into three piles by clicking on the wording and dragging it into the desired pile. Place items that seem characteristic of your day on the right, those that seem uncharacteristic on the left and all others in the middle (There is no limit to how many items can be in each pile). You can move the items after you have placed them if you change your mind, and order within a category does not matter.

### Main sort instructions:

Please sort the items into more specific categories, by clicking on the wording and dragging it into the desired category. The items you already rated as “characteristic” should fall to the right side of the scale, such that the most characteristic items are in the 7<sup>th</sup> category (“extremely characteristic”). Only a limited number of items can be placed in each category, with the number allowed in each category in parentheses in the category title. You can move the items after you have placed them if you change your mind, and order within a category does not matter. Below each category is a note on the current status of that category (i.e. how many more items are needed, when the category is full, and when there are too many items in the category).

*Seven categories will be displayed below this instruction: “extremely uncharacteristic”, “fairly uncharacteristic”, “somewhat uncharacteristic”, “neutral”, “somewhat characteristic”, “fairly characteristic” and “extremely characteristic”.*

1. A fun day
2. A busy day
3. A good day
4. A bad day
5. A productive day
6. Difficult to get anything accomplished
7. Full of small annoyances
8. Did not feel well
9. Felt healthy and strong
10. A boring day
11. Progress was made toward accomplishing an important goal
12. An unusual day- not at all routine
13. A stressful day
14. Time passed quickly
15. As expected; there were no surprises

16. A tiring day
17. Had many satisfying social interactions
18. A lonely day
19. Spent much of the day interacting with a romantic partner
20. Had negative interactions with others
21. An eventful day
22. An emotional day (filled with emotional highs and lows)
23. An upsetting day
24. A disappointing day
25. An interesting day
26. Finished a task I had been working
27. A rewarding day
28. Able to do what I wanted
29. Filled with a variety of activities
30. A happy day
31. An active day
32. A calm day
33. A carefree day
34. Had a significant failure
35. Will remember this day better than most
36. A frustrating day
37. Felt I was treated unfairly
38. The weather had a negative impact on me
39. Had an important success
40. Failed to complete a task I had planned to complete
41. A day filled with worries
42. A setback was experienced when working toward an important goal

UNIVERSITY OF CALIFORNIA, RIVERSIDE  
DEBRIEFING STATEMENT

Experience sampling and day evaluations

Investigators: Travis J. Miller  
Daniel J. Ozer

Email: [tmill009@ucr.edu](mailto:tmill009@ucr.edu)

Thank you for your participation in this study.

This study is attempting to examine how traits and affect throughout the day is associated with evaluations of the day as a whole. The reports of your affect and activities along with the measures of personality will help us better understand how university students spend their time and how they evaluate their days.

All the information you have provided will remain anonymous, and any identifying information will be deleted once you receive research participation credit. You will also receive one (2) research participation credits for completing the study.

If you have questions about your rights or complaints as a research subject, please contact the IRB Chairperson at (951) 827 - 4802 during business hours, or contact them by email at [irb@ucr.edu](mailto:irb@ucr.edu). If you have any questions specific to this study please email the primary investigator, Travis J. Miller, at [tmill009@ucr.edu](mailto:tmill009@ucr.edu).

If you have any other questions or would like to receive information on the general and collective (not your individual responses) findings of this study, please email the primary investigator, Travis J. Miller, at [tmill009@ucr.edu](mailto:tmill009@ucr.edu).

Sincerely,  
Travis J. Miller, M.A.