American Psychological Association
Journal of Experimental Psychology: General

ISSN: 0096-3445
http://dx.doi.org/10.1037/xge0000740

Temporal Dynamics of Real-World Emotion Are More Strongly Linked to Prediction Error Than Outcome

William J. Villano
University of Miami

A. Ross Otto
McGill University

C. E. Chiemeka Ezie, Roderick Gillis, and Aaron S. Heller
University of Miami

Primarily based on laboratory studies, theories of affect propose that emotions are driven by the valence of outcomes as well as the difference between the outcome itself and the expected outcome (i.e., the prediction error [PE]). Yet no work has assessed the drivers of emotion using real-world, personally meaningful events on timescales over which human emotion unfolds. We developed an event-triggered, ecological momentary assessment procedure measuring positive and negative affect (PA and NA, respectively) in university students as they received exam grades for which they had made predictions. We split data into exploratory and confirmatory samples, and built computational models predicting the time course of PA and NA and demonstrate that a model incorporating both exam grade and grade PE accounted for the time course of PA and NA better than a model solely using exam grades. Further, grade PEs were stronger predictors of the time course of PA and NA than the grades themselves. Similarly, the effects of PEs also persisted longer for NA than PA. These data indicate that deviations from expectations are critical determinants of the temporal dynamics of real-world emotion.

Keywords: affect, temporal dynamics, prediction error, ecological momentary assessment (EMA), computational modeling

Supplemental materials: http://dx.doi.org/10.1037/xge0000740.supp

> Happiness equals reality minus expectations.
> —Tom Magliozzi (1937–2014)

Emotion is a subjective experience comprising a coordinated affective response to personally relevant stimuli in one's environment (Frijda, Scherer, & Sander, 2009). Humans have the capacity to appraise a multitude of abstract stimuli as personally meaningful, irrespective of their relevance to our basic needs or survival (Scherer, 1984). Indeed, stimuli with implications for social standing, personal goals, or core values are often sufficient to elicit emotion (Scherer, 2005).

A substantial body of laboratory, as well as theoretical work has suggested that emotional reactions are driven, at least in part, by the degree to which an outcome deviates from one's expectation (Bhatia, Mellers, & Walasek, 2019; Carver & Scheier, 2001; Feather, 1967; Medvec, Madey, & Gilovich, 1995; Mellers, 2000; Shepperd & McNulty, 2002). At a fundamental level, the direction an outcome deviates from one's expectation has been found to drive positive or negative emotional reactions (Verinis, Brandsma, & Cofer, 1968), with better-than-expected outcomes driving positive emotion and vice versa. This principle has also been applied to theories of decision making and subjective well-being in which predicted outcomes generate a set point for future affective responding and motivate choice behaviors (Brickman & Campbell, 1971; Mellers, Schwartz, Ho, & Ritov, 1997).

More recent work using computational models to describe affect also suggest that both the outcome of an event, as well as the difference between that outcome and an expectation—commonly referred to as the prediction error (PE)—drive emotional reactions (Eldar & Niv, 2015; Eldar, Rutledge, Dolan, & Niv, 2016; Otto & Eichstaedt, 2018; Otto, Fleming, & Glimcher, 2016; Rutledge,

William J. Villano, Department of Psychology, University of Miami; A. Ross Otto, Department of Psychology, McGill University; C. E. Chiemeka Ezie, Roderick Gillis, and Aaron S. Heller, Department of Psychology, University of Miami.

We thank Andre Perwin for building the FileMaker interface. Financial support for this project included a grant from the John Templeton Foundation (to Aaron S. Heller).

Aaron S. Heller developed the study concept. Aaron S. Heller, C. E. Chiemeka Ezie, and Roderick Gillis contributed to the study design. Testing and data collection were performed by C. E. Chiemeka Ezie, William J. Villano, and Aaron S. Heller. William J. Villano, A. Ross Otto and Aaron S. Heller performed the data analysis and interpretation. William J. Villano, A. Ross Otto, and Aaron S. Heller drafted the manuscript, and Roderick Gillis and C. E. Chiemeka Ezie provided critical revisions. All authors approved the final version of the manuscript for submission.

The datasets and code used for analyses in this study are available from the authors' Github page (http://www.github.com/manateelab).

Correspondence concerning this article should be addressed to Aaron S. Heller, Department of Psychology, University of Miami, 5665 Ponce de Leon Boulevard, Coral Gables, FL 33146. E-mail: aheller@miami.edu

Skandali, Dayan, & Dolan, 2014). For example, recent laboratory-based work using financial rewards suggests that momentary happiness during decision making, measured on timescales of seconds, is driven by recent monetary outcomes, and, maximally, by PEs (Rutledge et al., 2014). This work also highlights the role of violations of expectation relative to outcomes in driving transient emotional reactions.

However, there are limitations of laboratory studies which constrain the generalizability of findings to consequential, real-life outcomes, and emotions that endure over longer timescales. First, studies to date use small financial rewards that do not have a long-lasting impact on participants' lives. Second, while emotion in the lab is typically examined on the timescale of seconds or minutes (Rutledge et al., 2014), human emotions in response to personally meaningful outcomes can last anywhere from 60 min to 8 hr (Frijda, Mesquita, Sonnemans, & Van Goozen, 1991). While naturalistic investigations have shown that unexpected real-world outcomes impact the valence and intensity of affective responses (Bhatia et al., 2019; Krupić & Corr, 2014; Otto & Eichstaedt, 2018), few, if any, studies have investigated the role of deviation from expectation versus outcome on the temporal dynamics of real-world affective responses.

Critically, few, if any, studies have compared the relative intensity and duration of positive and negative emotion in response to real-world outcomes. Moreover, in lab-based studies, the impact of outcomes and PEs on the intensity and duration of individuals' positive and negative emotion are assumed from the choice set given to them (e.g., a forced-choice task between two options, one of which has a sure outcome and the other a 50/50 outcome [or some other probability]; in these tasks, the expected value and PE are derived from this experimentally controlled choice set; Eldar et al., 2016; Rutledge et al., 2014; Tversky & Kahneman, 1992). However, real-world expectations are influenced by an individual's past experiences of the environment (Mobbs, Trimmer, Blumstein, & Dayan, 2018), and should therefore be measured rather than assumed. As such, it remains unclear whether the putative drivers of affective responses identified from the laboratory work—namely, PEs—extend to the temporal dynamics of human emotional responses to personally meaningful event outcomes in day-to-day life.

To identify the drivers of real-world emotional responses to personally meaningful outcomes, we developed an event-triggered, cell phone ecological momentary assessment (EMA; Shiffman, Stone, & Hufford, 2008) procedure. Using cell phone EMA allowed us to measure moment-to-moment positive affect (PA) and negative affect (NA) as participants went about an academic semester. A key feature of our design is that assessments of PA and NA were also specifically time-locked to occur immediately following the receipt of midterm and final exam grades. Over the course of each 17-week semester, participants received SMS messages directing them to complete a brief 10-item self-report survey that assessed momentary PA and NA every other day at pseudorandomly determined times (referred to as "baseline EMA sampling"; see the Method section). Participants completed up to four exams per semester (3 midterm exams and 1 final). Immediately following each exam (but prior to receiving the exam grade), participants reported the grade they expected to receive via SMS (their "predicted" grade). Then, after grades were made available, participants were enrolled into a "dense-sampling period" that was triggered the moment the participant logged on to the course website to check their grade (Figure 1). In these dense sampling periods, participants received up to 11 EMA prompts, which were sent every 45 min for up to 8.25 hr. The time courses of PA and NA during the event-triggered dense sampling periods were then used in a computational model that examined the extent to which PEs and grade outcomes drive affective reactions. Fundamentally, we hypothesized that positive grade PEs (i.e., performing better than one expects on an exam) would be associated with increases in PA and decreases in NA, and negative PEs would be associated with decreases in PA, and increases in NA.

To ensure the robustness of findings, we split the data into two sets: an exploratory and a confirmatory set. Data from the exploratory sample was used to optimize parameters concerning the duration of affective responses in the computational model (see the Method section). We used the parameters determined from the computational model to generate linear predictors of affect based on participants' exam grade outcomes and PEs. Then, using multilevel linear models to account for the nested structure of our data (i.e., repeated measures for each exam, multiple exams for each participant), data from the confirmatory sample was subsequently used to test: (a) whether outcomes alone were sufficient to predict the time course of emotion following these personally meaningful events (i.e., the delivery of exam grades), or whether a model containing both outcomes and PEs better fit the time course of emotion; and (b) whether PEs or the outcomes more strongly impacted the time course of PA and NA.

Method

Participants

Participants were undergraduate students enrolled in an introduction to psychology course. Over five semesters, we recruited 318 individuals (167 female; 3 did not specify gender; Table 1 for sample demographics) to participate in a semester-long EMA study to assess positive and negative affect (PA; NA). These data were split into an exploratory sample (Semesters 1 and 2; 108 participants) and a confirmatory sample (Semesters 3, 4, and 5; 210 participants). Participants who did not respond to any EMA surveys of PA and NA during any of the postexam, dense sampling periods were excluded from analyses. These exclusion criteria yielded a final exploratory sample of 93 participants (61 female, 2 did not specify gender), and a final confirmatory sample of 156 participants (106 female, 1 did not specify gender). Data from the exploratory sample were used to test competing models and to tune model parameters; the confirmatory sample was used for statistical testing.

Procedure

At the start of each semester, participants completed an initial laboratory session where they provided informed consent per study protocol approved by the Institutional Review Board at the University of Miami. At this visit, participants also provided contact information (i.e., cell phone number, e-mail address) for purposes of EMA survey distribution. Participants were incentivized to participate in the study to complete mandatory research credits in

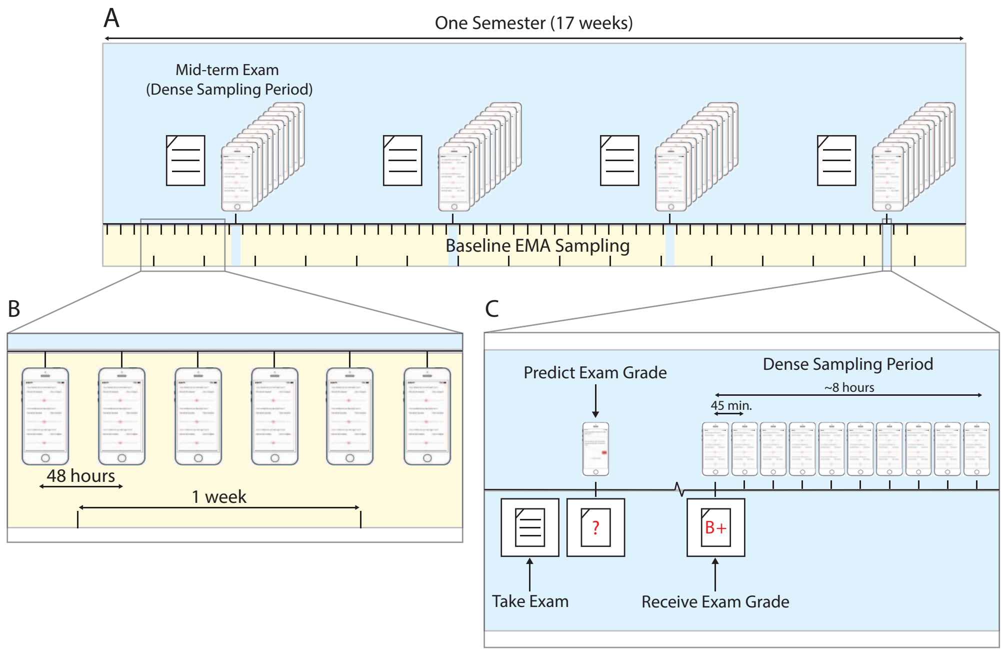

Figure 1. Semester-long experience sampling design. (A) The 17-week- long semester, during which participants completed brief ecological momentary assessment (EMA) self-reports to assess baseline positive affect (PA) and negative affect (NA). A set of more frequent EMA samples (a "dense-sampling period") was yoked to the precise moment when a participant viewed their exam grade (i.e., outcome) for the first time. Dense-sampling periods, which lasted up to 8.25 hr, afforded more fine-grained assessment of the time course of PA and NA in the aftermath of exam outcomes. (B) Baseline EMA self-report surveys, which consisted of 10 items that assessed a broad range of PA and NA. These surveys were distributed to participants once every two days. Momentary PA and NA scores were derived from the mean of PA and NA items from each EMA self-report survey. (C) The sequence of events following an exam. In the 1–2 hr following an exam, participants were prompted to report the exam grade they expected to receive. See the online article for the color version of this figure.

the course. The number of research credits a student received was yoked to their EMA survey completion rate.

EMA self-report surveys. As they went about their daily lives, participants received SMS text messages at pseudorandomly determined times (baseline EMA sampling) every other day, and then more frequently (every 45 min for up to 8.25 hr) immediately after viewing each of their midterm and final exam grades for the first time (termed the dense-sampling period). Each SMS contained a URL to a brief self-report survey that contained items derived from the Positive Affect/Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). These self-reports asked participants to rate the current intensity (0–100) of their feeling of a broad range of emotions on separate visual analog scales (VAS; i.e., "slider bars"). Items were selected from the larger PANAS set for PA and NA composite scores. We selected items that sampled across the full dimensions of affective valence and arousal, but strategically elected to use only a subset of items to minimize participant burden, given the high number of assessments used in this study. Momentary PA scores were comprised of the mean score from questions assessing the following emotions: "happy," "excited," "attentive," and "relaxed." Momentary NA scores were comprised of the mean score from questions assessing the emotions: "upset," "irritable," and "anxious" (Table 2 displays item-level descriptive statistics for composite affect scales). Though anxious was not included in the original PANAS measures, we elected to add this item to EMA self-reports as we had found through qualitative interviews with participants during piloting that anxious better captured students' experience of NA in this exam context than "nervous" or "jittery."

Determinants of affect.

Grade outcomes. Participants provided consent for the course instructor to share their exam grades directly with the researchers. Following each exam, per the course instructor's (author R.G.) method of sharing grades with students, grades were presented to students on a 12-point scale. On this 12 point scale, each point corresponds to a discrete letter grade, including "+" and "−" letter grades. Thus, on this 12-point scale, a 12 represents an A+, an 11 represents an A, and a 0 represents an F. As a result, on this scale, a grade of 0 represented an F, irrespective of whether a student received a 59% or a 0% on the exam. Given that this was the method of sharing grades with students, we adhered to the identical method by which students were informed of their exam scores.

Predictions and PEs. Shortly after each exam but prior to receiving their grade (e.g., 30 min after finishing the exam), participants reported the grade they expected to receive (0–100) via a VAS on an SMS-distributed online survey. With this experimental design, we could calculate a grade PE as the difference between a participant's expected grade and their actual grade for each exam. We chose the 0–100 scale as this was more intuitive to students in the class. Because grades were presented to students on

Table 1
Sample Demographics for Exploratory and Confirmatory Samples

Demographic	Exploratory sample	Confirmatory sample
n	93	156
Age		
M	19.31	19.14
SD	1.02	1.04
Minimum	17.67	17.63
Maximum	23.08	24.83
Gender (%)		
Female	67.03	68.83
Male	32.97	31.17
Race (%)		
White or Caucasian	68.13	72.19
Black or African-American	12.09	9.93
Asian or Asian-American	9.89	5.3
Multiracial	5.49	7.28
Other	4.4	5.3
Ethnicity (%)		
Non-Hispanic	71.43	71.9
Hispanic/Latino	28.57	28.1

a 12-point scale, we scaled predictions to the same 12-point scale, and computed an outcome PE as the difference between a participant's expected grade and their actual grade for each exam. To minimize any effects of social desirability on predictions, we assured all participants that their grade predictions would not be shared with anyone outside of the research team, especially the course professor. During the informed consent process, participants were told that all survey response data would be anonymized and viewable only by members of the research team. Moreover, consent forms contained the following sentence in boldface type: "Your name will not be associated with any of the data when it is analyzed" to reiterate that all responses would anonymized.

Dense-sampling period. On days when grades were posted to the course website (Blackboard Learn), the start of a dense-sampling period was yoked to the precise moment participants viewed their grade for the first time. To do this, when viewing their grade, participants submitted their cell phone number in a pop-up online dialogue box. This cell phone number was automatically compared to the database of cell phone numbers of participants for the semester. If the cell phone number matched that of a current study participant, they were auto-enrolled into the dense-sampling period. Thus, immediately after seeing an exam grade, 10 to 11 EMA prompts were sent every 45 min for up to 8.25 hr to examine the emotional dynamics in response to these personally relevant events. There was a total of 200 dense-sampling periods for the 93 participants in the exploratory sample (mean number of exams = 2.15, SD = 0.69, min = 1, max = 3), and 453 dense-sampling periods for the 156 participants in the confirmatory sample (mean number of exams = 2.90, SD = 0.99, min = 1, max = 4).

Calculation of baseline PA and NA and baseline-corrected PA and NA. Using the PA and NA scores from the nondense, baseline EMA period, we computed mean baseline PA and NA (baseline affect) for each participant. PA and NA measures from dense sampling periods were excluded in the calculation of baseline affect. This ensured that baseline affect was a stable individual difference metric, not impacted by possible transient state effects of grade anticipation and exam feedback. Momentary PA and NA scores from the dense-sampling period were centered to that participant's mean baseline affect. We did this by subtracting baseline affect from each momentary PA and NA score from the dense-sampling period: $\text{Affect}_{\text{BC}_i}(t) = \text{Affect}_i(t) - \text{Affect}_{\text{BL}_i}$ where t refers to time point and BL is the mean of PA or NA from the baseline period for participant i. These baseline-corrected scores represent momentary affect as difference relative that participant's baseline affect.

Preprocessing of postexam feedback PA and NA time series. Using up to 8.25 hr of dense EMA sampling collected in the aftermath of exams, we constructed affect time series for each exam, for each participant. PA and NA measures from each dense-sampling period were sorted into discrete, 30-min time bins. The start of the dense-sampling period corresponded to the moment a participant viewed his or her grade. When multiple survey completions occurred within the same 30-min bin, only the earliest response in the bin was kept. Although messages were sent out every 45 min, it was possible for participants to submit responses to survey prompts well after they were received. Thus, to capture late responses while preserving temporal resolution, we used a bin size that was smaller than the sampling rate (30 min). The number of bins for each dense sampling period was limited to 16, which

Table 2
Descriptive Statistics for Ecological Momentary Assessment Survey Items Comprising Positive Affect (PA) and Negative Affect (NA) Composite Scores

	PA items				NA items		
Descriptive statistic	Happy	Excited	Attentive	Relaxed	Upset	Irritable	Anxious/nervous
Baseline samples							
M	60.72	48.99	54.31	55.05	31.45	34.74	40.26
SD	22.91	25.44	23.96	25.07	24.68	25.61	27.57
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Dense samples							
M	58.23	49.62	52.44	55.04	37.02	40.04	42.16
SD	23.01	24.73	23.08	24	25.65	26.32	26.55
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Note. Statistics are displayed for surveys collected during baseline and dense sampling.

limited the sampling window for analyses to 8 hr, and served to screen out excessively delayed dense sampling responses. With this, we constructed affective time series data from responses with an inherently uneven sampling rate.

Computational model selection and validation. To examine the relative impact of exam grade outcome and grade PE on the time course of PA and NA, we modified a computational model previously used by Rutledge et al. (2014) which modeled changes in momentary emotion during a decision-making task. Given our set of exam-related predictors, our computational model did not contain a certain reward term used in Rutledge et al. (2014):

$$\mathrm{PA}(t) = w_0 + w_1\gamma_{\mathrm{PA}}^{t-j}\mathrm{outcome}_j + w_2\gamma_{\mathrm{PA}}^{t-j}\mathrm{PE}_j \tag{1}$$

$$\mathrm{NA}(t) = w_0 + w_1\gamma_{\mathrm{NA}}^{t-j}\mathrm{outcome}_j + w_2\gamma_{\mathrm{NA}}^{t-j}\mathrm{PE}_j \tag{2}$$

As prior work suggests that the influence of events on mood state decay over time (Eldar et al., 2016; Rutledge et al., 2014), we employed forgetting factors (γ_{PA}, γ_{NA}) to decay the influence of model parameters as time following exam feedback elapsed. Using data from the exploratory sample, we separately optimized forgetting factor parameters for models predicting PA (γ_{PA}) and NA (γ_{NA}). To estimate the best-fitting γ values for PA and NA computational models, we identified the γ value ($0.01 \leq \gamma \leq 0.99$) that minimized the root-mean-square error for each participant in separate fixed-effects linear regressions: one predicting baseline-corrected PA, and another predicting baseline-corrected NA. In each fixed-effects model, a participant's baseline-corrected dense sampling measures were regressed onto their decayed exam-specific grade outcomes and PEs:

$$\begin{gathered}\mathrm{PA}_{\mathrm{BC}} \sim \mathrm{outcome} + \mathrm{PE};\ \mathrm{PA}_{\mathrm{BC}} \in \{\mathrm{PA}_{\mathrm{BC}j}, \ldots, \mathrm{PA}_{\mathrm{BC}n}\},\\ \mathrm{outcome} \in \{\gamma_{\mathrm{PA}}^{t-j}\mathrm{outcome}_j, \ldots, \gamma_{\mathrm{PA}}^{t-n}\mathrm{outcome}_j\},\\ \mathrm{PE} \in \{\gamma_{\mathrm{PA}}^{t-j}\mathrm{PE}_j, \ldots, \gamma_{\mathrm{PA}}^{t-n}\mathrm{PE}_j\},\\ \gamma_{PA} \in \{0.01, \ldots, 0.99\}\end{gathered} \tag{3}$$

$$\begin{gathered}\mathrm{NA}_{\mathrm{BC}} \sim \mathrm{outcome} + \mathrm{PE};\ \mathrm{NA}_{\mathrm{BC}} \in \{\mathrm{NA}_{\mathrm{BC}_j}, \ldots, \mathrm{NA}_{\mathrm{BC}_n}\},\\ \mathrm{outcome} \in \{\gamma_{\mathrm{NA}}^{t-j}\mathrm{outcome}_j, \ldots, \gamma_{\mathrm{NA}}^{t-n}\mathrm{outcome}_j\},\\ \mathrm{PE} \in \{\gamma_{\mathrm{NA}}^{t-j}\mathrm{PE}_j, \ldots, \gamma_{\mathrm{NA}}^{t-n}\mathrm{PE}_j\},\\ \gamma_{\mathrm{NA}} \in \{0.01, \ldots, 0.99\}\end{gathered} \tag{4}$$

In the full exploratory sample, the medians of the best-fitting γ_{PA}, and γ_{NA} values were 0.94 and 0.96, respectively. To avoid bias issues stemming from simultaneously fitting decay rate parameters and estimating effects as a function of these decay rates, we used the same γ estimates in our PA- and NA-predicting models in the confirmatory sample.

Statistical Analyses

To confirm the distribution of the data, we first computed skewness and kurtosis to confirm that all computational model predictors were approximately normally distributed (e1071 package for R; Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2019). We considered all variables with skewness or kurtosis values above 2 or below −2 to be nonnormally distributed (George & Mallery, 2010). Next, to account for the multilevel nature of these data, we analyzed the impact of grades (our operationalization of "outcome") and grade PEs in a mixed-effects linear regression (lme4 package for the R programming language; Bates et al., 2018). Using the best-fitting γ values from the exploratory sample, continuous PE and outcome predictors were modeled as both fixed and random effects in linear mixed-effects models with random effects for exam (first, second, third, or final exam), nested within participant, and nested within semester. To determine the final random effects structure used in analyses, we tested the goodness-of-fit of several competing models to choose the optimal random effects structure for our data (See online supplemental materials and Table S1):

$$\mathrm{PA}_{\mathrm{BC}} \sim \mathrm{outcome}\ + \mathrm{PE} + (1 + \mathrm{outcome} + \mathrm{PE} \mid \mathrm{semester/id/exam}) \tag{5}$$

$$\mathrm{NA}_{\mathrm{BC}} \sim \mathrm{outcome} + \mathrm{PE} + (1 + \mathrm{outcome} + \mathrm{PE} \mid \mathrm{semester/id/exam}) \tag{6}$$

For both the exploratory and confirmatory samples, we estimated regression weights for the exam grade outcome and grade PE terms in both the PA and NA computational models using these linear mixed-effects models. Baseline-corrected PA and NA were used as dependent variables. We computed 95% confidence intervals (CIs) for grade outcome and grade PE weights from the fixed and random effects in both PA and NA models, and computed linear contrasts to test for differences between weights using the esticon function (doBy package for R; Højsgaard & Halekoh, 2018). The *p* values for the PE and outcome terms were estimated via Satterthwaite's degrees of freedom method (lmerTest package for R; Kuznetsova, Brockhoff, & Christensen, 2018) in the confirmatory sample. Model goodness of fit was assessed using the Akaike information criterion (AIC) estimates (Akaike, 1998; stats package for R; R Core Team, 2017), which penalizes a model's goodness-of-fit for model complexity (operationalized as the number of parameters). We performed a Wilcoxon signed-rank test, and computed bootstrapped 95% CIs (1,000 samples; boot package for R; Canty & Ripley, 2019) to test for significant differences between the decay rates for PA and NA fit to our exploratory sample.

Results

Descriptive Statistics

EMA measures of momentary affect. Cronbach's alpha for the composite PA scale was 0.73 for responses collected during baseline sampling, and 0.74 for responses collected during dense sampling. For the NA composite scale, Cronbach's alpha was 0.80 for baseline responses, and 0.83 for dense sampling responses (see Table 2 for descriptive statistics for composite affect scale survey items, and Table 3 for covariance matrix of composite affect survey items). On average, participants in both samples reported higher levels of PA than NA at baseline (exploratory: mean $\mathrm{PA}_{\mathrm{baseline}} = 55.93$, mean $\mathrm{NA}_{\mathrm{baseline}} = 34.03$; confirmatory: mean $\mathrm{PA}_{\mathrm{baseline}} = 54.26$, mean $\mathrm{NA}_{\mathrm{baseline}} = 37.32$; Table 4 displays full descriptive statistics). There was no significant difference between baseline PA in the exploratory and confirmatory samples, $t(329.10) = 1.83$, 95% CI $[-0.12, 3.46]$, $p = .07$. There was,

Table 3
Covariance Matrix for Positive Affect (PA) and Negative Affect (NA) Ecological Momentary Assessment (EMA) Survey Items

	PA items				NA items		
EMA item	Happy	Excited	Attentive	Relaxed	Upset	Irritable	Anxious/nervous
Baseline samples							
Happy	1.00	0.60	0.35	0.56	−0.57	−0.50	−0.40
Excited	0.60	1.00	0.31	0.38	−0.28	−0.28	−0.19
Attentive	0.35	0.31	1.00	0.22	−0.21	−0.22	−0.10
Relaxed	0.56	0.38	0.22	1.00	−0.44	−0.44	−0.51
Upset	−0.57	−0.28	−0.21	−0.44	1.00	0.67	0.53
Irritable	−0.50	−0.28	−0.22	−0.44	0.67	1.00	0.50
Anxious	−0.40	−0.19	−0.10	−0.51	0.53	0.50	1.00
Dense samples							
Happy	1.00	0.61	0.33	0.58	−0.53	−0.47	−0.39
Excited	0.61	1.00	0.33	0.41	−0.27	−0.26	−0.17
Attentive	0.33	0.33	1.00	0.24	−0.16	−0.16	−0.09
Relaxed	0.58	0.41	0.24	1.00	−0.45	−0.44	−0.48
Upset	−0.53	−0.27	−0.16	−0.45	1.00	0.69	0.60
Irritable	−0.47	−0.26	−0.16	−0.44	0.69	1.00	0.54
Anxious	−0.39	−0.17	−0.09	−0.48	0.60	0.54	1.00

however, a significant difference in baseline NA between both samples, $t(351.41) = -2.62$, 95% CI [−5.75, −0.822], $p = .01$. For both samples, baseline PA and baseline NA were both normally distributed (see the Method section; see Figure S1 in the online supplemental materials).

Table 4
Descriptive Statistics for Positive Affect (PA) and Negative Affect (NA) Baseline Scores, Exam Grade Outcomes, and Exam Grade Prediction Errors (PEs) for Exploratory and Confirmatory Samples

Descriptive statistic	Exploratory sample	Confirmatory sample
PA baseline score		
M	55.93	54.26
SD	11.24	9.47
Minimum	36.11	28.59
Maximum	92.91	93.07
Skewness	0.57	0.374
Kurtosis	0.41	0.98
NA baseline score		
M	34.03	37.32
SD	15.15	13.83
Minimum	2.51	2.73
Maximum	67.57	69.21
Skewness	−0.09	−0.28
Kurtosis	−0.71	−0.40
Exam grade outcomes		
M	8.32	8.73
SD	2.98	2.88
Minimum	0.00	0.00
Maximum	12.00	12.00
Skewness	−0.97	−1.05
Kurtosis	0.15	0.71
Exam grade PEs		
M	0.17	0.11
SD	2.40	2.33
Minimum	−6.75	−9.00
Maximum	7.00	8.00
Skewness	−0.07	−0.78
Kurtosis	0.24	1.76

Exam grades and PEs. Average exam grade outcomes for both samples corresponded to approximately a B/B+ letter grade (exploratory: $M = 8.32$; confirmatory: $M = 8.73$ [scores on 12-point grade scale]). Participants in the exploratory sample achieved scores on exams that were very close, but marginally higher than their expectations ($M = 0.17$ [out of 12], $SD = 2.40$, min = −6.75, max = 7.00). The same was true of participants in the confirmatory sample ($M = 0.11$ [out of 12], $SD = 2.33$, min = −9.00, max = 8.00). Distributions of exam grade outcomes in both exploratory and confirmatory samples were approximately normal, though slightly negatively skewed (see Figure S2 in the online supplemental materials). Exam grade PEs were normally distributed in the exploratory sample, and approximately normally distributed in the confirmatory sample, though negatively skewed and leptokurtotic within an acceptable range (skewness = −0.778, kurtosis = 1.76). In dense sampling periods, the mean number of responses was 9.73 ($SD = 7.36$, min = 1, max = 57; Table 5), and mean latency to respond to a prompt was 7.18 min ($SD = 9.50$, min = 0.02, max = 45.00). For baseline sampling, the mean number of responses was 28.34 ($SD = 14.89$, min = 1, max = 79),

Table 5
Descriptive Statistics for Ecological Momentary Assessment Compliance During Baseline Sampling and Postexam Feedback Dense Sampling

Descriptive statistic	Baseline sampling	Dense sampling
Response latency (min)		
M	54.81	7.18
SD	193.49	9.50
Minimum	0.60	0.02
Maximum	2,857.27	45.00
Number of survey responses		
M	28.34	9.73
SD	14.89	7.36
Minimum	1	1
Maximum	79	57

and mean latency to respond to a prompt was 54.81 min (SD = 193.49, min = 0.60, max = 2,857.27). As can be seen from the summary time course figure (Figure 2), affect tended to regress toward baseline levels over the 8-hr dense sampling period.

We then constructed linear mixed-effects models to test whether exam grade predictions, grade outcomes, or grade PEs changed across the semester. This is relevant to examine because if participants' grade predictions do not change appreciably over the course of the semester, we can conclude that it is reasonable to model exam grade outcomes and PEs (regardless of their order in the semester) as independent trials of a similar event. Linear mixed-effects models revealed no significant change in exam grade outcomes in the exploratory sample ($B = 0.132$, $SE = 0.160$, $p = .412$), or in the confirmatory sample ($B = -0.031$, $SE = 0.080$, $p = .701$). The same was true of exam grade predictions in the exploratory sample ($B = -0.162$, $SE = 0.143$, $p = .261$), and in the confirmatory sample ($B = -0.082$, $SE = 0.071$, $p = .261$). In the exploratory sample, a slight positive slope in exam grade PEs across the semester trended toward significance ($B = 0.311$, $SE = 0.163$, $p = .060$), but there was no such significant change in PEs in the confirmatory sample ($B = 0.109$, $SE = 0.088$, $p = .232$). Thus, in general, there were limited changes in grades or PEs over the course of the semester. Although there was a slight numeric difference in decay rates between PA and NA (γ_{PA} = 0.94; γ_{NA} = 0.96), a Wilcoxon signed-ranks test revealed no significant difference between the PA and NA decay rates fit to the exploratory sample ($V = 847.5$, $p = .6893$). This was further confirmed by examining the 95% CIs for the median PA (Mdn = 0.94, 95% CI [0.89, 0.99]) and NA (Mdn = 0.96, 95% CI [0.81, 0.97]) forgetting factors, which overlapped.

Computational Models of Affect

Model comparison. After tuning model parameters on the exploratory dataset, in which the best-fitting forgetting factors were 0.94 and 0.96 for γ_{PA}, and γ_{NA}, respectively, we utilized data from the confirmatory dataset to, (a) test whether exam grade outcomes alone or grade outcomes in addition to grade PEs yielded a better fit to the PA and NA postexam dense sampling data, and (b) test whether exam grade PEs or outcomes more strongly influenced the time course of PA and NA. To test our first aim, we fit mixed-effects linear regressions to test whether exam grade PE significantly improved model fit over a model that only included exam grade (outcome). Our first model (Equation 7) predicted the time course of dense sampling PA and NA (triggered upon viewing the exam grade), solely on the basis of exam grade outcome:

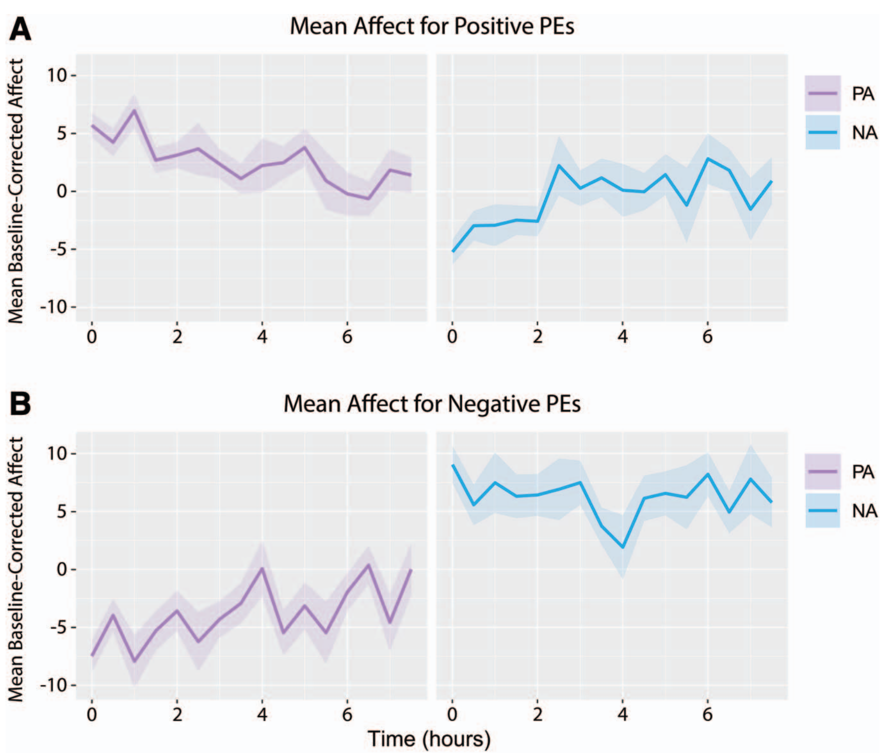

Figure 2. Mean baseline-corrected positive affect (PA) and negative affect (NA) observations from confirmatory participants following exams where participants reported positive and negative exam grade prediction errors (PEs). (A) The average time courses of PA observations (left) and NA observations (right) following exams in which participants outperformed their predictions (i.e., positive prediction errors). (B) Plots the average time courses of PA observations (left) and NA observations (right) following exams in which participants' grades fell short of their predictions (i.e., negative PEs). Error ribbons for each line represent the standard error of the mean. See the online article for the color version of this figure.

$$\text{Affect} \sim \text{outcome} + (1 + \text{outcome} \mid \text{semester/id/exam});$$
$$\text{Affect} \in \{\text{Affect}_1 \ldots \text{Affect}_n\},$$
$$\text{Outcome} \in \{\gamma^{t-i}\text{Outcome}_j, \ldots, \gamma^{t-n}\text{Outcome}_j\} \tag{7}$$

Equation 8 predicted the time course of PA and NA from exam grade outcomes as well as grade PEs:

$$\text{Affect} \sim \text{outcome} + \text{PE} + (1 + \text{outcome} + \text{PE} \mid \text{semester/id/exam})$$
$$\text{Affect} \in \{\text{Affect}_1 \ldots \text{Affect}_n\},$$
$$\text{Outcome} \in \{\gamma^{t-i}\text{Outcome}_j, \ldots, \gamma^{t-n}\text{Outcome}_j\},$$
$$\text{PE} \in \{\gamma^{t-i}\text{PE}_j, \ldots, \gamma^{t-n}\text{PE}_j\} \tag{8}$$

As PA and NA-predicting models were separately estimated, Affect in Equations 7 and 8 refer to two separate models, one predicting the time course of PA and one predicting the time course of NA.

We then tested whether both PE and grade outcome yielded better model fits than either one alone. For models predicting PA, including the PE term in addition to the grade outcome (Equation 8) yielded a significantly better fit than the model with the grade outcome predictor alone (Equation 7; $\chi(10) = 155.97$, $p < .001$; $\Delta\text{AIC} = -135.96$). The same result emerged for NA ($\chi(10) = 141.33$, $p < .001$; $\Delta\text{AIC} = -121.33$). We then tested additional models containing only a PE term (i.e., no outcome predictor), to determine whether including the grade outcome, in addition to PE, provided a better fit to the data than a model with PE alone. In the confirmatory sample, models with both PE and outcome predictors yielded better fits than models with no outcome predictor (PA model: $\Delta\text{AIC} = -79.7$, NA model: $\Delta\text{AIC} = -127.36$). Similar results emerged for the exploratory sample (Table 6). These results indicate that, in addition to the grade outcomes themselves, the extent to which grade outcomes deviate from the expected grade was associated with PA and NA.

We also assessed whether use of a forgetting factor at all improved model fit as compared with a model that assumed no decay at all in affect. To do this, we constructed additional models in which predictors were not decayed to test whether forgetting factors to account for the temporal course of affect following exam grade feedback. The inclusion of a forgetting factor significantly improved model fit over models with no forgetting in the exploratory sample (PA model: $\chi(2) = 132.89$, $p < .001$; NA model: $\chi(2) = 128.37$, $p < .001$), and in the confirmatory sample (PA model: $\chi(2) = 188.47$, $p < .001$; NA model: $\chi(2) = 198.00$, $p < .001$).

Exam-related determinants of affect. To test whether exam grade PEs or outcomes more strongly influenced the time course of affect, we then examined the parameter estimates from the linear mixed-effects models that included both grade outcomes and grade PE as predictors. First, we assessed whether grade outcome and grade PE significantly predicted PA and NA. Within the confirmatory sample, the parameter estimate for grade PE, but not the grade outcome, was a significant predictor of PA in the hours following exam feedback (Figure 3; Table 7; outcome: $B = 0.21$, $SE = 0.21$, $p = .38$; PE: $B = 3.10$, $SE = 0.43$, $p < .001$). In the model predicting NA, the weights for grade PE and grade outcome both were significant and in the predicted direction (outcome: $B = -0.69$, $SE = 0.27$, $p = .037$; PE: $B = -2.69$, $SE = 0.48$, $p < .001$). The size of the parameter estimates for these same models using the exploratory sample were very similar to those of the confirmatory sample. In all cases, the parameter estimates for the grade PE predictors were numerically larger than the parameter estimates for grade outcome predictors. Thus, to test whether these differences were statistically significant, we performed linear contrasts between grade PE and grade outcome parameter estimates for both PA and NA-predicting models. For both PA and NA models, PE − outcome contrasts were significant (PA model: $B = 2.89$, $SE = 0.577$, $p < .001$; NA model: $B = -2.01$, $SE = 0.673$, $p = .003$), suggesting that for both PA and NA, grade PE is more robustly associated with emotional responses than the exam grade outcome. Indeed, we observed that the sign of an exam grade PE—that is, whether a participant underperformed (negative PE) or overperformed (positive PE) with respect to their expectations for an exam—produced discernible differences in group level time courses for PA and NA following exam feedback (Figure 3).

Table 6
Model Comparison via Akaike Information Criterion (AIC) of Mixed-Effects Models With Parameters for Exam Grade Outcome, Exam Grade Prediction Error (PE), or Both Outcome and PE

Model dependent variable	Model predictors	AIC	ΔAIC (model – final model)
Exploratory sample			
PA	Outcome	9,428.14	75.90
NA	Outcome	9,957.26	60.24
PA	PE	9,429.06	76.82
NA	PE	9,989.34	92.31
PA	Outcome + PE[a]	9,352.24	0.00
NA	Outcome + PE[a]	9,897.03	0.00
Confirmatory sample			
PA	Outcome	22,416.48	135.96
NA	Outcome	22,917.64	121.33
PA	PE	22,360.22	79.70
NA	PE	22,923.67	127.36
PA	Outcome + PE[a]	22,280.52	0.00
NA	Outcome + PE[a]	22,796.31	0.00

Note. PA = positive affect; NA = negative affect.
[a] Denotes final model.

Table 7
Results From Exploratory and Confirmatory Linear Mixed-Effects Models Predicting Positive Affect (PA) and Negative Affect (NA) Dense-Sampling Observations From Exam Grade Outcome and Exam Grade Prediction Error (PE)

Parameter	Estimate	*SE*	*p* value
Exploratory PA model			
Outcome	−0.2926	0.2743	
PE	3.2766	0.6274	
Exploratory NA model			
Outcome	−0.6595	0.4362	
PE	−2.1798	0.8336	
Confirmatory PA model			
Outcome	0.2092	0.2060	.380
PE	3.0964	0.4299	<.001
Confirmatory NA model			
Outcome	−0.6854	0.2741	.037
PE	−2.6913	0.4824	<.001

Note. *SE* = standard error of fixed effect parameter estimates from linear mixed-effects models.

Analyses addressing possible confounds. We tested a series of additional models to confirm that these effects are unbiased by factors such as trait-level PA and NA, EMA compliance following exam grade feedback, and time of day.

Impact of trait-level affect. To examine whether participants' baseline affect (i.e., mean of affect measures collected outside of postexam feedback dense sampling periods) could account for relations between PEs, grades, and momentary affect, we ran additional models that included baseline PA or NA as additional indicators in models predicting the temporal course of postexam PA and NA, respectively. Thus, within the confirmatory sample, we included baseline PA as a predictor in the PA model. Adding baseline PA did not improve model fit (AIC with baseline PA = 22,282, AIC without baseline PA = 22,281). The same result emerged for the NA model (AIC with baseline NA = 22,797, AIC without baseline NA = 22,796). Moreover, the significance of PE and outcome terms did not change with the inclusion of baseline affect in the PA-predicting model (outcome: $B = 0.21$, $SE = 0.21$, $p = .39$; PE: $B = 3.07$, $SE = 0.43$, $p < .001$), or in the NA predicting model (outcome: $B = -0.68$, $SE = 0.27$, $p = .039$; PE: $B = -2.67$, $SE = 0.48$, $p < .001$). This suggests that the particular associations between grade PE and PA and NA are robust to mean levels of PA and NA, and that deviations from expectation are predictive of affective responses regardless of one's trait levels of positive and negative affect.

Impact of EMA compliance. To test whether exam grade PEs were associated with compliance during dense sampling periods, we tested a linear mixed-effects model in which a participant's number of responses in a dense sampling period were predicted by their grade PE. We specified a random intercept and a random slope for PE and used participants as a grouping level to account for multiple dense sampling periods for each participant. There was no significant linear relationship between PE and the number of dense sampling responses in the exploratory sample (PE: $B = -0.1280$, $SE = 0.1303$, $p = .333$) nor in the confirmatory sample (PE: $B = -0.07948$, $SE = 0.1287$, $p = .539$).

Impact of time of day. To test whether the effects of outcome and PE on the time course of affect could be explained by the time of day when participants completed dense sampling periods (Figure S3 in online supplemental materials), we constructed additional models with a fixed effect covariate representing the hour-of-day for each dense sampling response, and tested these models with data from our confirmatory sample. With the inclusion of an hour-of-day covariate, the significance of outcome and PE parameter estimates was unchanged in both the PA model (outcome: $B = 0.23$, $SE = 0.21$, $p = .35$; PE: $B = 3.08$, $SE = 0.43$, $p < .001$) and the NA model (outcome: $B = -0.69$, $SE = 0.27$, $p = .04$; PE: $B = -2.68$, $SE = 0.48$, $p < .001$). This suggests that exam grade outcomes and predominantly, PEs, were robustly associated with the time course of PA and NA, irrespective of the time of day when a participant checked their grade and completed a dense sampling period.

Discussion

Overall, these results demonstrate how unexpected, personally relevant outcomes are associated with measurable changes in the time course of individuals' affective responses. Moreover, the deviation from expected outcomes appears to be a more robust driver of self-reported emotion than the outcome itself. Furthermore, the decay rate estimates suggest that these PE effects persist on the order of hours—indeed, per the exponential rate of decay for a forgetting factor of 0.94, the effect of PE on PA diminishes to one half of its initial magnitude within 6–6.5 hr of a personally meaningful event. At the same time, the larger decay rates for NA (0.96) suggest that the effect of PEs on NA tend to persist over time, as a PE of equal magnitude reaches half of its initial magnitude within 8–8.5 hr, notably longer than PA. This dif-

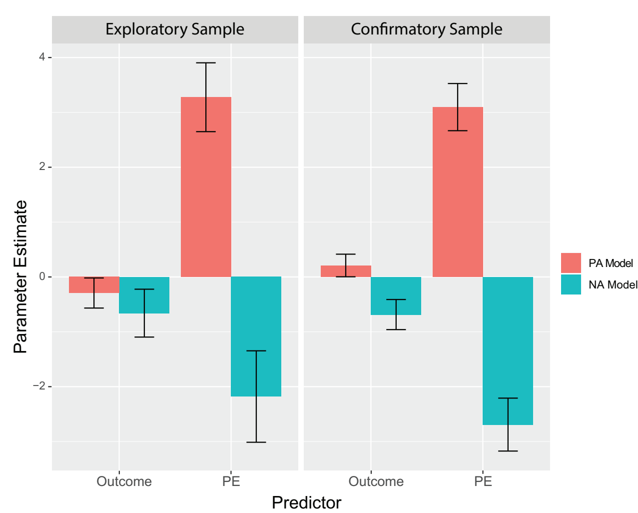

Figure 3. Exam grade prediction error (PE) and grade outcome effect estimates derived from computational models. Across positive affect (PA) and negative affect (NA) models, and in both the exploratory and confirmatory sample, exam grade PE exhibited a greater effect on momentary affect than exam grade outcome. The computational models that best explained PA and NA ratings had positive weights for outcome and PE terms in the models predicting PA, and negative weights in the models predicting NA. Confidence intervals for each bar depict the standard error of fixed effect parameter estimates from linear mixed-effects models. See the online article for the color version of this figure.

ference suggests that while PEs may exert a larger-magnitude effect on PA compared with NA, the impact of personally meaningful events and PEs on PA may be more fleeting than on NA, similar to other published work (Bonanno et al., 2002; Lucas, 2007). These findings demonstrate that using naturalistic paradigms to measure affect precisely when personally meaningful events occur is important, as both in-the-moment designs using EMA (Larson, Csikszentmihalyi, & Graef, 1980; Sbarra, 2006; van Eck, Nicolson, & Berkhof, 1998) and designs using retrospective (Frijda et al., 1991) recall indicate that real-world human emotional responses tend to last on the scale of hours (or longer) and not seconds.

In line with other emerging experimental research, these data corroborate recent laboratory research identifying the PE as an important driver of subjective well-being on the timescale of seconds and point to avenues for future research. As a result, these findings suggest that the drivers of affect on the timescale of seconds to minutes and minutes to hours might be similar. However, the rate at which emotion returns to baseline (the forgetting factor) is likely to differ across paradigms and settings. For example, the experimental paradigm used by Rutledge et al. (2014), which employed relatively small financial incentives and assessed emotion frequently (on the timescale of tens of seconds), yielded a smaller (i.e., quicker) forgetting factor of 0.61. While forgetting factors fit to data with different sampling frequencies cannot be directly compared on the same scale, the combination of a wider sampling interval and larger best-fitting forgetting factors (0.94 and 0.96) relative to Rutledge et al. (2014) suggests that decay rates of emotion may differ depending on the salience of an outcome.

These data are in line with a long-standing research literature that deviations from expectation cause changes in emotion (Verinis et al., 1968). While doing much better or much worse than expected impacts emotion, it has been suggested that there are additional mechanisms for emotion generation that may be a fruitful avenue for future work. One, in particular, may be via counterfactual processes (Roese, 1997), or the idea that a bad outcome feels less disappointing when the mental representation of the alternative was worse (Johnson, 1986). Counterfactual thinking is thought to reflect internal cognitive simulations of "what might have been" (e.g., "I would have done better on the exam had I studied more," or "I would have done better on the exam had I not gone to the party last night"), and it has been suggested that these simulations can amplify the affective impact of events (Kahneman & Tversky, 1981). A classic example of this effect is that in the Olympics, viewers perceive bronze medalists to be happier than silver medalists (Medvec et al., 1995), as the silver medalist's counterfactual alternative is gold, and that of the bronze medalist is not receiving a medal at all. Counterfactual thinking following these "near misses" typically drive negative affective processes such as regret (Bell, 1982; Loomes & Sugden, 1982) and disappointment (Bell, 1985; Loomes & Sugden, 1986). Thus, counterfactual thinking appears to be most strongly engaged when individuals "nearly miss" their goals, which does not necessarily covary with a PE. A counterfactual thinking predictor could thus be operationalized in our framework as the "proximity to a desired grade." For example, a student, whose desired grade in the class was an 80, predicted after taking the final that he would receive a 60, but his actual score on the final exam was a 79. In this example, such a counterfactual predictor would be highly negative (very close to desired grade, but a "near miss"), and importantly would be orthogonal to both grade PE (highly positive: 20) and grade outcome (79). We hypothesize that incorporating such a counterfactual thinking predictor could further enrich computational models of emotional responses and potentially account for unexplained variance. Finally, it is worth noting that our analyses assume that affective responses to grade outcomes are linear—that is, a one unit change of grade outcome engenders a fixed change in affect regardless of grade outcome—it may be the case that a nonlinear response function may better account for the impact of grade outcomes on affect. Future investigations should compare different scaling functions for real-world outcomes to better understand the nature of their impact on the time course of affect.

These results highlight the importance of deviations from expectation in driving human emotional responses on the scale of hours—affective science must continue to incorporate measures of expectation to appropriately model emotion. In addition, these results present an experimental paradigm to combine experience sampling of affect in the real world in response to personally meaningful outcomes with precise, experimentally determined timing to identify the drivers of affect. While our usage of a university student sample facilitated a novel, temporally controlled, and ecologically valid experimental design, it has yet to be demonstrated whether these effects generalize to a more diverse (i.e., less White, Educated, Industrialized, Rich, and Democratic [WEIRD]; Henrich, Heine, & Norenzayan, 2010) sample, or to a different set of real-world outcomes. Follow-up work in other populations is needed to assess this.

Recently, researchers across a variety of disciplines have commented on the uncertain ecological validity of traditional laboratory paradigms (Mobbs et al., 2018). These data present a novel approach that yields concordant findings to emerging laboratory based computational models of human emotion. We believe that future work, using these types of sampling approaches with ambulatory physiological recording (Eldar, Roth, Dayan, & Dolan, 2018), can further elucidate the processes that drive our emotional lives and help pinpoint the specific locus of aberrant affective responses in clinical populations (Eldar et al., 2016; Rutledge et al., 2017).

Context of Research

The genesis for this work emerged from a motivation to utilize personally meaningful stimuli as the "events" to measure affective reactions and to begin to build computational models of naturalistic emotion. We are hopeful that utilizing this type of design can determine the ecological validity of current research examining relatively brief affective reactions induced in the laboratory. Future directions also include collecting additional predictors of the temporal course of emotion, including, for example, the confidence one has about their prediction, among others. The goals are to build better and more complete models of the determinants of the temporal course of emotion and ultimately link them to risk for psychopathology.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4612-1694-0_15

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Green, P. (2018). lme4: Linear mixed-effects models using "Eigen" and S4 (Version R package version 1.1–18-1). Retrieved from https://CRAN.R-project.org/package=lme4

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research, 30,* 961–981. http://dx.doi.org/10.1287/opre.30.5.961

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research, 33,* 1–27. http://dx.doi.org/10.1287/opre.33.1.1

Bhatia, S., Mellers, B., & Walasek, L. (2019). Affective responses to uncertain real-world outcomes: Sentiment change on Twitter. *PLoS ONE, 14,* e0212489. http://dx.doi.org/10.1371/journal.pone.0212489

Bonanno, G. A., Wortman, C. B., Lehman, D. R., Tweed, R. G., Haring, M., Sonnega, J., . . . Nesse, R. M. (2002). Resilience to loss and chronic grief: A prospective study from preloss to 18-months postloss. *Journal of Personality and Social Psychology, 83,* 1150–1164. http://dx.doi.org/10.1037/0022-3514.83.5.1150

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 287–305). New York, NY: Academic Press.

Canty, A., & Ripley, B. D. (2019). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-24. Retrieved from https://CRAN.R-project.org/package=boot

Carver, C. S., & Scheier, M. F. (2001). *On the self-regulation of behavior*. New York, NY: Cambridge University Press.

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. *Nature Communications, 6,* 6149. http://dx.doi.org/10.1038/ncomms7149

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of reward learning signals predicts mood fluctuations. *Current Biology, 28,* 1433–1439.e7. http://dx.doi.org/10.1016/j.cub.2018.03.038

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). Mood as representation of momentum. *Trends in Cognitive Sciences, 20,* 15–24. http://dx.doi.org/10.1016/j.tics.2015.07.010

Feather, N. T. (1967). Valence of outcome and expectation of success in relation to task difficulty and perceived locus of control. *Journal of Personality and Social Psychology, 7,* 372–386. http://dx.doi.org/10.1037/h0025184

Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions. In K. T. Strongman (Ed.), *International review of studies on emotion* (Vol. 1, pp. 187–225). New York, NY: Wiley.

Frijda, N. H., Scherer, K. R., & Sander, D. (2009). Emotion definitions (psychological perspectives). In D. Sander & K. R. Scherer (Eds.), *The Oxford companion to emotion and the affective sciences* (pp. 142–144). Oxford, UK: Oxford University Press. Retrieved from http://dare.uva.nl/personal/pure/en/publications/emotion-definitions-psychological-perspectives(7108a16a-cc0c-4e6a-96e9-eda594d4bb87).html

George, D., & Mallery, P. (2010). *SPSS for Windows step by step. A simple study guide and reference*. Boston, MA: Pearson Education.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83. http://dx.doi.org/10.1017/S0140525X0999152X

Højsgaard, S., & Halekoh, U. (2018). doBy: Groupwise statistics, LSmeans, linear contrasts, utilities (Version R package version 4.6–2). Retrieved from https://CRAN.R-project.org/package=doBy

Johnson, J. T. (1986). The knowledge of what might have been. *Personality and Social Psychology Bulletin, 12,* 51–62. http://dx.doi.org/10.1177/0146167286121006

Kahneman, D., & Tversky, A. (1981). *The simulation heuristic* (Report No. TR-5). Retrieved from the Stanford University Department of Psychology website http://www.dtic.mil/docs/citations/ADA099504

Krupić, D., & Corr, P. J. (2014). Individual differences in emotion elicitation in university examinations: A quasi-experimental study. *Personality and Individual Differences, 71,* 176–180. http://dx.doi.org/10.1016/j.paid.2014.08.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest: Tests in linear mixed effects models (Version R package version 3.0–1). Retrieved from https://CRAN.R-project.org/package=lmerTest

Larson, R., Csikszentmihalyi, M., & Graef, R. (1980). Mood variability and the psychosocial adjustment of adolescents. *Journal of Youth and Adolescence, 9,* 469–490. http://dx.doi.org/10.1007/BF02089885

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal, 92,* 805–824. http://dx.doi.org/10.2307/2232669

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies, 53,* 271–282. http://dx.doi.org/10.2307/2297651

Lucas, R. E. (2007). Adaptation and the set-point model of subjective well-being. *Current Directions in Psychological Science, 16,* 75–79. http://dx.doi.org/10.1111/j.1467-8721.2007.00479.x

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology, 69,* 603–610. http://dx.doi.org/10.1037/0022-3514.69.4.603

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin, 126,* 910–924. http://dx.doi.org/10.1037/0033-2909.126.6.910

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8,* 423–429. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00455.x

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2019). E1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071; Version R package version 1.7–1). Retrieved from https://CRAN.R-project.org/package=e1071

Mobbs, D., Trimmer, P. C., Blumstein, D. T., & Dayan, P. (2018). Foraging for foundations in decision neuroscience: Insights from ethology. *Nature Reviews Neuroscience, 19,* 419–427. http://dx.doi.org/10.1038/s41583-018-0010-7

Otto, A. R., & Eichstaedt, J. C. (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PLoS ONE, 13,* e0206923. http://dx.doi.org/10.1371/journal.pone.0206923

Otto, A. R., Fleming, S. M., & Glimcher, P. W. (2016). Unexpected but incidental positive outcomes predict real-world gambling. *Psychological Science, 27,* 299–311. http://dx.doi.org/10.1177/0956797615618366

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121,* 133–148. http://dx.doi.org/10.1037/0033-2909.121.1.133

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., . . . Dolan, R. J. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. *Journal of the American Medical Association Psychiatry, 74,* 790–797. http://dx.doi.org/10.1001/jamapsychiatry.2017.1713

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 12252–12257. http://dx.doi.org/10.1073/pnas.1407535111

Sbarra, D. A. (2006). Predicting the onset of emotional recovery following nonmarital relationship dissolution: Survival analyses of sadness and anger. *Personality and Social Psychology Bulletin, 32,* 298–312. http://dx.doi.org/10.1177/0146167205280913

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information/Information Sur les Sciences Sociales, 44,* 695–729. http://dx.doi.org/10.1177/0539018405058216

Shepperd, J. A., & McNulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science, 13,* 85–88. http://dx.doi.org/10.1111/1467-9280.00416

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1–32. http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–323. http://dx.doi.org/10.1007/BF00122574

van Eck, M., Nicolson, N. A., & Berkhof, J. (1998). Effects of stressful daily events on mood states: Relationship to global perceived stress. *Journal of Personality and Social Psychology, 75,* 1572–1585. http://dx.doi.org/10.1037/0022-3514.75.6.1572

Verinis, J. S., Brandsma, J. M., & Cofer, C. N. (1968). Discrepancy from expectation in relation to affect and motivation: Tests of McClelland's hypothesis. *Journal of Personality and Social Psychology, 9,* 47–58. http://dx.doi.org/10.1037/h0025672

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Received May 15, 2019
Revision received December 16, 2019
Accepted December 23, 2019 ■