“How Do You Feel?” Direct Valence Measurement Enables the Detection of Affect Shift Dynamics as Powerful Predictors of Psychological Well-Being

Carmen Goicoechea[1,2] · Pandelis Perakakis[1]

[1] *Department of Social, Work and Differential Psychology, Complutense University of Madrid, Madrid, Spain*
[2] *Department of Educational Sciences, University of La Rioja, Logroño, Spain*

Abstract Valence is central to affective experience yet remains poorly operationalized in emotion science. Most research infers valence by categorizing discrete emotions as positive or negative, which may conflate distinct aspects of emotional experience. On the other hand, direct bipolar valence measurement—asking individuals “How do you feel?” on a continuum from negative to positive—allows participants to integrate contextual complexity into their affective reports. Critically, this approach also enables the detection and quantification of transitions between positive and negative affective states, opening new possibilities for studying affect dynamics. We tested whether metrics quantifying these directional transitions predict psychological well-being more effectively than traditional intensity-based measures. Across three ecological momentary assessment datasets with 345 participants and over 30,000 assessments, the positive to negative affect shift ratio—quantifying the propensity to transition from positive to negative affect—showed stronger predictive associations with well-being outcomes than means and standard deviations derived from the same valence scale. This pattern persisted across LASSO regression, hierarchical regression, and relative importance analyses, and remained robust even with only 3 daily assessments. These findings demonstrate that direct valence measurement offers both theoretical advantages by respecting participant-integrated experience and empirical advantages by enabling more predictive dynamics metrics, while remaining practical for clinical applications.

Keywords Valence measurement · Affect dynamics · Affect shift ratio · Bistability · Ecological momentary assessment · Experience sampling method

This is the author’s submitted version of the manuscript published in the journal Affective Science as: Goicoechea, C., & Perakakis, P. (2026). “How Do You Feel?” Direct Valence Measurement Enables the Detection of Affect Shift Dynamics as Powerful Predictors of Psychological Well-Being. *Affective Science*. https://doi.org/10.1007/s42761-026-00382-9

Valence—the positive or negative quality of affective experience—is widely regarded as a fundamental dimension of emotion (Barrett, 2006; Russell, 2003; Walle & Dukes, 2023). The answer to the question “How do you feel right now?”, constitutes an indicator capable of integrating information from different psychological processes and synthesizing a person’s state at a given moment. Just as heart rate integrates signals from different organs of the body and translates them into a single parameter that reflects cardiac system functioning, the response to this question condenses available psychological information and provides an estimate of the state and evolution of the psychological system. Despite this integrative potential, valence remains “opaquely conceptualized and inconsistently applied” across both theoretical frameworks and empirical research (see introduction to this special issue). A critical source of this ambiguity lies in how valence is measured. A predominant approach across emotion science—including emotion regulation research (Gross & John, 2003), clinical assessment (Watson et al., 1988), and basic emotions research (see Ortony, 2022), assesses discrete emotions (e.g., joy, anger, sadness, fear) and subsequently categorizes them as positive or negative to derive valence.

This inference-based strategy rests on the assumption that certain emotions possess inherent, fixed valence—that anger is necessarily negative, happiness necessarily positive, and so forth. This assumption is not unreasonable if valence is defined strictly as hedonic quality—the bodily pleasantness or unpleasantness of experienced sensations. Under this definition, categorizing anger as negative and joy as positive is defensible. However, researchers often use ‘positive’ and ‘negative’ without explicitly specifying which aspect of emotional experience they reference—hedonic tone, goal congruency, moral evaluation, or behavioral tendency—each of which may vary independently (Colombetti, 2005; Walle & Dukes, 2023). When valence is left undefined, inferring it from discrete emotion categories risks conflating these distinct aspects. Moreover, even hedonic tone may be context-dependent: anger may feel satisfying when it motivates assertive problem-solving, and fear may carry positive anticipatory excitement in some contexts (Colombetti, 2005).

An alternative approach is to measure valence directly by asking individuals to report their current affective state on a bipolar continuum from negative to positive. Grounded in Russell’s (2003, 2005) concept of core affect—a basic and ever-present state in consciousness characterized by the simple experience of feeling good or bad—this method allows participants to integrate the full complexity of their emotional experience into a single coherent valence judgment. Rather than researchers inferring valence from categorical emotion labels, participants themselves integrate contextual factors, personal meanings, and potentially mixed feelings.

This approach builds on a substantial tradition of core affect research (Russell, 1980, 2003; Barrett, 2006), which has long employed direct valence measurement. Our contribution extends this tradition by demonstrating that direct bipolar measurement enables not only the assessment of momentary valence but also the quantification of affect shift dynamics—transitions between positive and negative states that prove more predictive of well-being than intensity-based measures. From a dynamical systems perspective—a framework that characterizes how systems evolve over time and identifies stable states (attractors) toward which systems return after perturbation (Scheffer et al., 2009)—the time series of valence ratings can reveal both the *structure* and *transitions* of an individual's affective landscape. Structurally, individuals may exhibit monostable dynamics—maintaining a single affective baseline—or bistable dynamics, with distinct attractors in both positive and negative regimes (Goicoechea, Dakos, et al., 2025). Bistability implies that shifts between positive and negative affect tend to be abrupt, discontinuous transitions between stable states (Scheffer et al., 2009). Interestingly, our previous work showed that approximately half of individuals in community samples exhibit such bistable patterns (Goicoechea, Dakos, et al., 2025).

Distinct from structure, the *frequency* of transitions across the valence boundary may itself be informative. The propensity to shift from positive to negative affect (P2N) may index vulnerability, while the propensity to shift from negative to positive (N2P) may reflect resilience. These directional metrics capture asymmetries that intensity-based measures cannot: a person who frequently falls into negative affect but rarely recovers differs meaningfully from one who rarely falls but recovers quickly—yet both might show similar mean levels or variability.

Previous research on affect dynamics has focused primarily on intensity-based measures. Houben et al.'s (2015) meta-analysis found that variability-based measures (instability, inertia) derived from intensity ratings were associated with well-being outcomes. Subsequently, Dejonckheere et al. (2019) demonstrated that simpler intensity-based measures—within-person mean and standard deviation of positive and negative affect—were the most robust predictors. However, whether measuring complex variability or simple aggregates, these intensity-based measures characterize typical levels and fluctuation magnitude across time; they do not capture the directional transitions between affective states that direct bipolar measurement makes possible.

We previously found that one affect shift metric in particular predicted psychological well-being more strongly than traditional intensity measures (Goicoechea, Dakos, et al., 2025). Across two independent samples from Spain ($N = 65$) and Germany ($N = 56$), the positive-to-negative affect shift ratio (P2N-ASR)—quantifying the propensity to transition from positive to negative affect—emerged as a robust predictor of multiple well-being indicators including anxiety, depression, psychological inflexibility, life satisfaction, and resilience. This metric consistently outperformed mean and standard deviation of positive and negative affect in variable selection analyses. These findings suggest that how individuals move between affective states may be more diagnostic of psychological functioning than the central tendency or variability of those states.

However, those initial studies employed relatively intensive measurement protocols, with participants reporting their affective state 5 to 6 times per day over 3 to 4 weeks. While such intensive sampling provides rich temporal resolution, it raises questions about practical feasibility, particularly for clinical populations where minimizing participant burden is essential. Can meaningful affect dynamics be captured with fewer daily assessments? Moreover, although our previous findings were consistent across two independent samples, replication with a larger sample would strengthen confidence in the superiority of shift-based over intensity-based metrics.

The present study addresses these questions through three complementary contributions. First, we analysed a new dataset ($N = 216$) using a reduced measurement protocol of only 3 assessments per day. This tests whether the predictive advantage of affect shift metrics persists with minimal measurement burden, a critical consideration for clinical translation. Second, we aggregate analyses across all three datasets (total $N = 345$ participants, over 30,000 momentary assessments) to replicate and extend our previous findings, providing further evidence regarding the relative importance of shift versus intensity metrics in predicting psychological well-being. Third, by situating this empirical work within ongoing debates about valence operationalization (Walle & Dukes, 2023), we articulate how direct bipolar measurement addresses fundamental conceptual challenges in affective science while simultaneously enabling more powerful prediction of meaningful psychological outcomes.

Method

Design

We analyzed data from three ecological momentary assessment (EMA) studies to compare the predictive validity of affect shift metrics versus traditional intensity-based measures for psychological well-being outcomes. The Spanish 1 study and the German study (Studies 1 and 2, respectively) were included in our previous work (Goicoechea, Dakos, et al., 2025), whereas the Spanish 2 study (Study 3) was analyzed for the first time for the present purposes (Banos et al., 2024). The design is correlational, with affect dynamic predictors derived from EMA time series and well-being outcomes assessed via validated questionnaires administered at baseline. The three studies were approved by the ethics committees of their respective institutions: the two Spanish studies by the Human Research Ethics Committee of the University of Granada in Spain (Reference: 2214/CEIH/2021), and the German study by the Department of Psychology of the Philipps-University of Marburg in Germany (Reference: 2022-22v).

Participants

Participants of both Spanish studies were recruited by a survey company (QUANTICA Marketing Research) to represent the Spanish population based on gender, age, location, and annual income. Inclusion criteria were being over 18 years of age and an Android user, as the custom app used for data collection was only available on this operating system. The initial sample size of 103 participants in the Spanish 1 study, determined by funding constraints, was reduced to 73 after data preprocessing (see relevant section). In the Spanish 2 study, data were obtained

Table 1. Summary of the studies

Study	Participants				EMA Protocol		Well-being Indicators				
	N	Female %	Age	Bistable %	Days	Occ.	Life Sat.	Resil.	Flour.	Depression	Anxiety
1. Spanish 1	73	48	44.1 ± 17 (18–69)	61.5	29	6	SWLS.1	BRS	FS	PHQ-9	GAD-7
2. German	56	79	23.9 ± 2.7 (19–32)	46.4	21	5	SWLS	BRS	–	DASS-21	DASS-21
3. Spanish 2	216	60	46.7 ± 14.7 (20–74)	48.4	29	3	SWLS.1	BRS	FS	PHQ-9	GAD-7

Note. Psychological inflexibility was assessed with the AAQ-II in all three studies. For the German study the depression and anxiety subscales of the DASS-21 were used. *SWLS* Satisfaction with Life Scale, *BRS* Brief Resilience Scale, *FS* Flourishing Scale, *PHQ-9* Patient Health Questionnaire–9, *GAD-7* Generalized Anxiety Disorder Scale–7, *AAQ-II* Acceptance and Action Questionnaire–II.

from 384 participants, but the final analytic sample consisted of 216 participants (see data preprocessing section).

The German study involved 56 participants from Germany, recruited through the online portal used by the Department of Psychology at the Philipps-University of Marburg, as well as flyers and posters displayed on the university campus. The sample size was determined by the time frame for recruitment based on the availability and resources of the research team. Participants were required to be over 18 years of age and either native German speakers or possess native-level German language proficiency, since the questionnaires were presented in German.

All participants provided informed consent after receiving complete information about the studies. Table 1 presents a detailed summary of each data set, including participant information, EMA protocol, and available well-being indicators.

Procedure

In all three studies, participants completed EMA protocols via smartphone applications. At each prompt, participants responded to questions including a bipolar valence rating (see Affect Measures below). The Spanish 1 study administered 6 daily prompts over 29 days (174 possible assessments) between November 16 and December 14, 2021. The German study administered 5 daily prompts over 21 days (105 possible assessments) between September 14 and December 8, 2022. The Spanish 2 study used a reduced protocol of 3 daily prompts over 29 days (87 possible assessments) administered between July 10 and August 8, 2023. Table 1 summarizes the protocol details and available measures for each dataset.

Affect Measures

In the three EMA studies included in our analysis, valence, a primary dimension of affect referring to the pervasive subjective experience underlying any emotionally charged episode and consisting simply of feeling good or bad (Russell, 1980), was assessed via the prompt "How do you feel right now?" using a single bipolar visual analogue scale ranging from -50 (very bad) to $+50$ (very good). Although debate continues regarding bipolar versus bivariate structures of affect—with some models proposing that positive and negative affect can operate independently (Cacioppo & Berntson, 1994) and others advocating hierarchical structures incorporating both bipolar and independent components (Watson & Tellegen, 1999)—there is substantial support for single-item bipolar scales as efficient measures for studying intraindividual dynamic patterns of momentary valence (Cloos et al., 2023; Russell, 2017; Russell & Carroll, 1999).

The within-person mean and standard deviation of PA and NA were derived from the valence variable. PA values were obtained by averaging all positive valence observations (valence > 0), while NA values were derived from averaging all negative valence observations (valence < 0). Standard deviations were computed separately for PA and NA using the same partitioning of the valence time series. These parameters were chosen for comparison with the novel affect shift measures based on the results of the meta-analysis by Dejonckheere et al. (2019), which identified them as having superior predictive power compared to more elaborate, time-dependent measures of affect.

To quantify the frequency of valence transitions, we computed two Affect Shift Ratio (ASR) metrics: positive-to-negative (P2N) and negative-to-positive (N2P). The P2N represents the ratio of transitions from positive to negative affect relative to the total number of measurement occasions in the positive affect regime. This metric quantifies the propensity for positive affect to shift to negative affect at the next measurement occasion, with values ranging from 0 (no transitions) to 1 (certain transition). Similarly, N2P represents the ratio of transitions from negative to positive affect relative to the total number of measurement occasions in the negative affect regime, reflecting the probability that negative affect will transition to positive affect at the next measurement instance. Both metrics provide intuitive indices of affect dynamics that require only the determination of affect valence (positive vs. negative) at each measurement occasion, without requiring intensity ratings. For detailed mathematical formulations of these metrics, see Goicoechea, Dakos, et al. (2025).

In the three studies, participants were prompted to answer multiple questions using similar visual scales. However, for the purposes of this study, only the valence scale was analyzed.

Psychological Well-Being Indicators

We investigated the predictive value of affect shift metrics on six indicators of psychological well-being, five of which were common in the three studies (see Table 1). These indicators encompass both positive and negative aspects of well-being, including measures of life satisfaction, resilience, and flourishing (positive), as well as symptoms of depression, anxiety, and psychological inflexibility (negative).

Life Satisfaction

Two different instruments were used to assess people's perceived life satisfaction.

Satisfaction With Life Scale. The German study used the

Satisfaction with Life Scale (SWLS; Diener et al., 1985), which includes five statements (e.g., "In most ways my life is close to my ideal") that participants rate based on their level of agreement, in a range from 1 (strongly disagree) to 7 (strongly agree). Cronbach's α was 0.82.

Single-Item. The Spanish studies used a single-item life satisfaction scale ("In general, how satisfied are you with your life?") ranging from 1 (very dissatisfied) to 4 (very satisfied), which has been found to have comparable performance to the SWLS scale (Cheung & Lucas, 2014).

Resilience

Resilience, understood as the ability to recover from stress, was assessed in the three studies by the Brief Resilience Scale (Smith et al., 2008). This questionnaire consists of six statements (e.g., "I tend to bounce back quickly after hard times") that participants rate based on their level of agreement, ranging from 1 (strongly disagree) to 5 (strongly agree). Cronbach's α was 0.72 in the Spanish 1 study, 0.85 in the German study, and 0.84 in the Spanish 2 study.

Flourishing

Flourishing, the eudaimonic aspect of well-being understood as the self-perceived success in important areas such as relationships, self-esteem, purpose, and optimism, was assessed in the Spanish studies by the Flourishing Scale (Diener et al., 2010). This self-report instrument includes eight items (e.g., "I lead a purposeful and meaningful life" or "My social relationships are supportive and rewarding"), ranging from 1 (strongly disagree) to 7 (strongly agree). Cronbach's α was 0.91 in the Spanish 1 study and 0.85 in the Spanish 2 study.

Depression Symptoms

Two different continuous scales were used to assess participants' depression symptom levels. In the Spanish studies, 27.69% of participants in the first study and 21.76% in the second showed significant symptoms of depression (Patient Health Questionnaire ≥ 10). For the German study, 10.71% of participants exceeded the cut-off of the Depression subscale of the Depression, Anxiety and Stress Scale (score ≥ 7).

Patient Health Questionnaire–9. The Spanish studies used the Patient Health Questionnaire (Kroenke et al., 2001), which assesses the frequency of nine prominent depression symptoms experienced over the last 2 weeks (e.g., "Feeling down, depressed, or hopeless"). Item scales range from 0 (not at all) to 3 (nearly every day). Cronbach's α was 0.87 in both the Spanish 1 and the Spanish 2 studies.

Depression, Anxiety and Stress Scales–21 (Depression). The German study used the depression subscale of the Depression, Anxiety and Stress Scales (DASS-21; Lovibond & Lovibond, 1995), which consists of seven items (e.g., "I felt downhearted and blue") and is scored on a range from 0 (not at all) to 3 (very much or most of the time). Cronbach's α was 0.81.

Anxiety Symptoms

Two distinct instruments were used to assess participants' anxiety symptom levels. In the Spanish studies, significant symptoms of anxiety (Generalized Anxiety Disorder Scale ≥ 10) were indicated by 21.53% in the first study and 15.28% in the second, while in the German study 17.85% of participants met the criterion (Anxiety subscale of the Depression, Anxiety and Stress Scale ≥ 5).

Generalized Anxiety Disorder Scale–7. The Spanish studies used the Generalized Anxiety Disorder Scale–7 (Spitzer et al., 2006), which is a seven-item screening tool for general anxiety symptoms in diverse settings and populations. Each item is a statement about the presence of bothersome anxiety-related symptoms during the last 2 weeks (e.g., "Not being able to stop or control worrying"). Item scales range from 0 (not at all) to 3 (nearly every day). Cronbach's α was 0.91 in the Spanish 1 study and 0.89 in the Spanish 2 study.

Depression, Anxiety and Stress Scales–21 (Anxiety). The German study used the anxiety subscale of the DASS-21 (Lovibond & Lovibond, 1995), which consists of seven items (e.g., "I felt scared without any good reason") and is scored on a range from 0 (not at all) to 3 (very much or most of the time). Cronbach's α was 0.64.

Psychological Inflexibility

Psychological inflexibility refers to the level of experiential avoidance and was assessed in the three studies by Acceptance and Action Questionnaire–II (Bond et al., 2011). This questionnaire has seven statements (e.g., "I worry about not being able to control my worries and feelings") that are rated from 1 (never true) to 7 (always true). Cronbach's α was 0.95 in the Spanish 1 study, 0.89 in the German study, and 0.93 in the Spanish 2 study.

Data Preprocessing

To ensure the quality of the affect data, we performed a series of preprocessing steps on all three data sets. First, we excluded participants who showed poor compliance with the EMA protocol. Specifically, we removed participants with less than 75% of measurement occasions completed. Compliance was defined as the proportion of non-missing responses for the valence variable. This resulted in the exclusion of 0 participants from the initial 56 in the German study, 30 participants from the initial 103 in the Spanish 1 study, and 168 participants from the initial 384 in the Spanish 2 study. The remaining 56 participants of the German study had a mean compliance rate of 94% ($SD = 5\%$), the 73 participants of the Spanish 1 study had a mean compliance rate of 90% ($SD = 6\%$), and the 216 participants of the Spanish 2 study had a mean compliance rate of 90% ($SD = 9\%$). The 75% threshold was selected to ensure adequate temporal resolution for computing reliable affect shift metrics, since participants with sparse observations may yield unstable estimates of transition probabilities. To verify that our findings were not dependent on this exclusion criterion, we conducted sensitivity analyses including all participants regardless of compliance (see Supplementary Materials).

Data Analyses

Prevalence of Bistability in Affect Dynamics

To address the prevalence of bistability in the three data sets, we reconstructed the probabilistic stability landscape of each

Table 2. Descriptive statistics for affect dynamic predictors

Predictor	*N*	*M*	*SD*	*Mdn*	Min	Max	Skew.
mPA	344	26.28	10.76	25.21	1.48	50.00	0.20
mNA	310	16.02	8.69	15.00	1.00	50.00	1.14
sdPA	343	8.70	3.36	8.52	0.00	21.48	0.17
sdNA	286	8.09	3.99	7.68	0.00	21.26	0.46
P2N	305	0.13	0.14	0.08	0.01	0.80	2.14
N2P	304	0.66	0.26	0.67	0.02	1.00	−0.24

mPA/mNA mean positive/negative affect, *sdPA/sdNA* standard deviation of positive/negative affect, *P2N* positive-to-negative affect shift ratio, *N2P* negative-to-positive affect shift ratio.

Table 3. Between-person correlation matrix

	mPA	mNA	sdPA	sdNA	P2N	N2P
mPA	—					
mNA	0.42	—				
sdPA	0.12	0.13	—			
sdNA	0.16	0.53	0.32	—		
P2N	−0.45	0.11	0.01	0.27	—	
N2P	0.40	−0.07	0.01	−0.27	−0.57	—

Correlations based on pairwise complete observations ($n = 286$–344). *mPA/mNA* mean positive/negative affect, *sdPA/sdNA* standard deviation of positive/negative affect, *P2N* positive-to-negative affect shift ratio, *N2P* negative-to-positive affect shift ratio.

participant using potential analysis (Livina et al., 2010) that is based on fitting a polynomial function to the probability density of individual valence time series. We used the *liv_potential()* function of the *earlywarnings* package (Dakos, 2013) in R (version 4.4.2; R Core Team, 2024) to carry out this analysis. Next, we identified attractor basins as local minima of the fitted potential using a sliding window approach. For each data point, we selected a window spanning three adjacent samples centered at that point and identified the minimum value within that window. A data point was considered a local minimum if its value was smaller than all other values within the window. Participants were classified as "bistable" if at least one local minimum was found in each of the two affective regimes (PA and NA). Non-bistable participants fell under five distinct categories: "positive-monostable", if they presented only one minimum in the PA regime; "negative-monostable", with one minimum in the NA regime; "positive-multistable", if they exhibited more than one minima only in the PA regime; "negative-multistable", if they displayed more than one minima only in the NA regime; and "undefined", when no identifiable local minima could be detected in the potential analysis.

Descriptive Statistics and Correlations

Table 2 presents descriptive statistics for the affect dynamic predictors. Mean positive affect ($M = 26.28$) was higher than mean negative affect ($M = 16.02$), reflecting the general positivity of affective experience in these community samples. The shift metrics showed a striking asymmetry: participants transitioned from positive to negative affect on only 13% of opportunities (P2N: $M = 0.13$, $Mdn = 0.08$), while they recovered from negative to positive affect on 66% of opportunities (N2P: $M = 0.66$, $Mdn = 0.67$). This pattern indicates that negative affective states were relatively transient, with participants typically returning to positive states. The P2N distribution was positively skewed (skewness $= 2.14$), with most participants showing low values and a tail of individuals with elevated propensity to fall into negative affect.

Table 3 presents between-person correlations among predictors. P2N and N2P were negatively correlated ($r = -.57$), indicating that individuals who transition into negative affect more readily also tend to recover less easily. P2N was negatively associated with mean positive affect ($r = -.45$), suggesting that higher typical positive affect is associated with fewer falls into negativity. Standard deviation of negative affect showed opposite associations with the two shift metrics: positively correlated with P2N ($r = .27$) and negatively with N2P ($r = -.27$), indicating that greater variability in negative affect is associated with more frequent falls into negativity and less frequent recovery. These patterns support the theoretical interpretation of P2N as an index of affective vulnerability and N2P as an index of recovery capacity.

Predictive Performance of Affect Dynamics

To evaluate the comparative predictive value of different affect dynamic measures, we conducted LASSO regression analyses (Least Absolute Shrinkage and Selection Operator) for each psychological well-being indicator available in the three datasets. This analytical framework replicates Dejonckheere et al. (2019), who tested whether complex variability measures add predictive value beyond mean affect. Here, we adapt this framework to test whether intensity measures add value beyond shift metrics. LASSO regression applies L1 regularization, which penalizes the absolute values of regression coefficients and can shrink some coefficients to zero, thereby reducing model complexity and mitigating overfitting.

We compared two predictor sets to determine whether conventional affect intensity measures add predictive value beyond affect shift metrics. The first model ("bench") included only the affect shift rate metrics P2N and N2P. The second model ("all") included these shift metrics together with the within-person mean and standard deviation of PA and NA (mPA, mNA, sdPA, sdNA, P2N, N2P). This comparison allowed us to assess whether more elaborate affect measures improve prediction of psychological well-being outcomes beyond what can be achieved with shift metrics alone.

For each model and outcome, we estimated out-of-sample prediction accuracy using leave-one-out cross-validation (LOOCV). In each fold, we held out one participant and fit a LASSO model to the remaining participants, using cross-validation on the training data to select the optimal regularization parameter (lambda) that minimized cross-validated error. We then used this fitted model to predict the held-out participant's outcome value. After obtaining predictions for all participants, we computed the cross-validated R^2 as $1 - SS_{res}/SS_{tot}$, where SS_{res} represents the sum of squared residuals between observed and predicted values, and SS_{tot} represents the total sum of squares. All predictors were standardized prior to analysis.

Additionally, we examined which individual predictors were most frequently selected by LASSO when all six predictors were available. For each LOOCV fold using the "all" model, we identified which predictors had nonzero coefficients at the optimal lambda value. Selection frequency for each predictor was calculated as the proportion of folds in which it was retained in the model. This analysis provided insight into which affect dynamic measures LASSO consistently identified as most informative for predicting psychological well-being.

To aggregate results across datasets and outcomes, we computed weighted averages of R^2 values and selection frequencies, with weights corresponding to the number of observations in each dataset-outcome combination. We report weighted means and standard errors for overall well-being (all outcomes combined) as well as for specific outcome types. All LASSO analyses were conducted using the *glmnet* package (Friedman et al., 2010) in R.

To complement the LASSO analyses and evaluate the individual and incremental contributions of each affect dynamic measure, we conducted a series of linear regression analyses. All affect predictors and outcome variables were standardized prior to analysis to facilitate comparison of effect sizes across predictors and outcomes.

For each psychological well-being indicator, we fitted three types of regression models. First, we estimated solo models in which each of the six affect predictors (mPA, mNA, sdPA, sdNA, P2N, N2P) was used individually to predict the outcome. These models provided a baseline measure of each predictor's individual explanatory power. Second, we fitted hierarchical models to assess the incremental explanatory power of each predictor beyond P2N alone. Specifically, for each of the remaining five predictors, we fitted a model including both P2N and the focal predictor, and computed the incremental R^2 (ΔR^2) as the difference between this model's R^2 and the R^2 of the P2N-only model. Third, we repeated this hierarchical approach using both shift metrics (P2N and N2P) as the base model, fitting models that included P2N, N2P, and each of the four intensity measures (mPA, mNA, sdPA, sdNA) separately, and computing the incremental R^2 for each intensity measure beyond the shift metrics. To provide inferential support for R^2 comparisons, we computed 95% bootstrap confidence intervals (1000 resamples) for all solo R^2 values and incremental ΔR^2 values.

This hierarchical modeling approach allowed us to determine whether conventional affect intensity measures (means and standard deviations) provide additional explanatory value beyond what can be achieved with the simpler shift rate metrics alone. Given that shift metrics require only binary classification of momentary affect (positive vs. negative) rather than intensity ratings, demonstrating comparable predictive performance would support the use of these more practical and intuitive measures.

We aggregated results across datasets and outcomes using weighted averages, where weights corresponded to the number of observations in each dataset-outcome combination. We computed weighted means and standard errors for R^2 values for overall well-being (all outcomes combined) and for each individual outcome separately. All R^2 values reported from linear regression analyses represent in-sample explained variance.

Finally, to determine the relative contribution of each affect dynamic measure when all predictors are considered simultaneously, we conducted relative importance analyses using the LMG method (Sen et al., 1981). Unlike the hierarchical regression models that assess incremental contributions in a specified order, relative importance analysis partitions the total explained variance (R^2) among all predictors by averaging their contributions across all possible orderings. This approach accounts for shared variance among correlated predictors and provides a fair assessment of each predictor's importance when all compete in the same model.

For each psychological well-being indicator, we fitted a linear model including all six affect predictors (mPA, mNA, sdPA, sdNA, P2N, N2P) and computed the relative importance metric for each predictor using the *relaimpo* package (Groemping, 2006) in R. The LMG values represent the proportion of total R^2 attributable to each predictor, with all LMG values summing to 1.0 within each model. Higher LMG values indicate that a predictor accounts for a larger share of the explained variance in psychological well-being outcomes.

This analysis complements the previous LASSO and linear regression analyses by addressing which predictors account for more of the explained variance when intensity metrics and shift metrics are included together in the same model. We aggregated results across datasets and outcomes using weighted averages, where weights corresponded to the number of observations in each dataset-outcome combination. We computed weighted means and standard errors for LMG values for overall well-being (all outcomes combined) and for each individual outcome separately.

As a supplementary analysis addressing whether our partitioned intensity measures (means and standard deviations computed separately for positive and negative valence observations) might disadvantage intensity-based prediction, we also conducted regression analyses using overall valence mean and standard deviation computed directly on the full -50 to $+50$ scale. These results are reported in the Supplementary Materials.

Results

Prevalence of Bistability in Affect Dynamics

In the Spanish 1 study, we classified 61.5% of participants as "bistable", 18.5% as "positive-monostable", 18.5% as "positive-multistable", and 1.5% as "negative-multistable". The German study showed a slightly lower prevalence of bistability, with 46.4% of participants classified as "bistable", 39.3% as "positive-monostable", 5.4% as "negative-monostable", 7.1% as "positive-multistable", and 1.8% as "undefined". The Spanish 2 study showed intermediate prevalence, with 48.4% of participants

Table 4. Cross-validated R^2 for LASSO models predicting overall psychological well-being

Model	Weighted R^2	*SE*
Bench (P2N, N2P)	0.065	0.024
All (6 predictors)	0.078	0.026

Weighted R^2 values are averaged across all datasets and outcomes, weighted by sample size ($N = 1{,}631$ total observations). *SE* standard error of weighted R^2.

Table 5. LASSO selection frequencies for affect dynamic predictors in overall psychological well-being

Predictor	Selection Frequency	*SE*
P2N	0.958	0.044
mPA	0.514	0.107
sdNA	0.510	0.106
N2P	0.326	0.103
mNA	0.263	0.098
sdPA	0.209	0.091

Selection frequency is the proportion of leave-one-out cross-validation folds in which each predictor received a nonzero coefficient. *SE* standard error.

classified as "bistable", 32.6% as "positive-monostable", 5.6% as "negative-monostable", 11.6% as "positive-multistable", 0.5% as "negative-multistable", and 1.4% as "undefined". Across all three datasets, over half of participants (52.1%) exhibited bistable affect dynamics, suggesting that bistability is a prevalent feature of affective experience in these samples (Table 1).

Predictive Performance of Affect Dynamics

LASSO Regression

To evaluate whether conventional affect intensity measures improve prediction beyond affect shift metrics, we compared LASSO models using two predictor sets: the "bench" model including only P2N and N2P, and the "all" model including these shift metrics plus the four intensity measures (mPA, mNA, sdPA, sdNA). Table 4 presents the cross-validated R^2 values for overall psychological well-being across the three datasets.

The addition of intensity measures to the shift metrics resulted in minimal improvement in predictive accuracy. The "all" model explained 7.8% of variance in psychological well-being outcomes compared to 6.5% for the "bench" model. Given the standard errors of these estimates ($SE = 0.026$ and 0.024, respectively), this difference of 1.3% points does not represent a meaningful improvement in out-of-sample prediction.

To examine which predictors LASSO consistently selected when all six were available, we calculated selection frequencies across all cross-validation folds. Table 5 shows the proportion of folds in which each predictor received a nonzero coefficient for overall well-being.

P2N was by far the most consistently selected predictor, retained in 96% of cross-validation folds. Among intensity measures, mean positive affect and standard deviation of negative affect showed moderate selection frequencies (51% each), while other intensity measures were selected less frequently. These results suggest that when regularization forces the model to prioritize predictors, the positive-to-negative affect shift rate emerges as the most reliable indicator of psychological well-being.

Selection frequencies for individual well-being outcomes (Fig. 1A) revealed similar patterns across different well-being indicators, with P2N consistently showing the highest selection frequency in most outcomes.

Linear Regression Analysis

To examine the individual and incremental contributions of each affect dynamic measure, we conducted linear regression analyses for all psychological well-being outcomes. Table 6 presents the R^2 values for solo models (each predictor alone) and the incremental R^2 (ΔR^2) when adding each predictor to baseline models for overall well-being.

Among solo models, P2N emerged as the strongest individual predictor, explaining 11.9% of variance in psychological well-being outcomes. Mean positive affect (mPA) was the second strongest solo predictor (7.1%), followed by N2P (5.4%). The remaining intensity measures explained considerably less variance individually (2.3–3.7%).

When examining incremental contributions beyond P2N alone, intensity measures added modest but meaningful explanatory power. Mean positive affect provided the largest incremental contribution ($\Delta R^2 = 2.6\%$), followed by mean and standard deviation of negative affect (both $\Delta R^2 = 2.1\%$). Adding N2P to P2N yielded a smaller increment ($\Delta R^2 = 1.0\%$).

When both shift metrics (P2N and N2P) served as the baseline, intensity measures showed small incremental contributions. Mean positive affect provided the largest increment ($\Delta R^2 = 2.8\%$), with the other intensity measures adding approximately 2% each. These results indicate that while intensity measures do account for some additional variance, P2N remains the dominant predictor, and the practical gain from adding intensity measures is modest relative to the increased measurement burden they require.

Patterns across individual outcomes (Fig. 1B) showed consistency in the relative ordering of predictors, with P2N typically showing the highest solo R^2 and intensity measures providing variable incremental contributions depending on the specific well-being indicator.

Supplementary analyses using overall valence mean and standard deviation (rather than partitioned positive and negative affect measures) yielded consistent results: P2N remained the strongest predictor ($R^2 = 0.117$), outperforming mean valence ($R^2 = 0.095$) despite the latter capturing both intensity and time-balance information (see Supplementary Materials, Table S1).

Sensitivity analyses including all participants regardless of compliance ($N = 543$) yielded consistent results: P2N remained the strongest predictor ($R^2 = 0.095$), and the pattern of findings was unchanged despite somewhat attenuated effect sizes (see Supplementary Materials, Table S2).

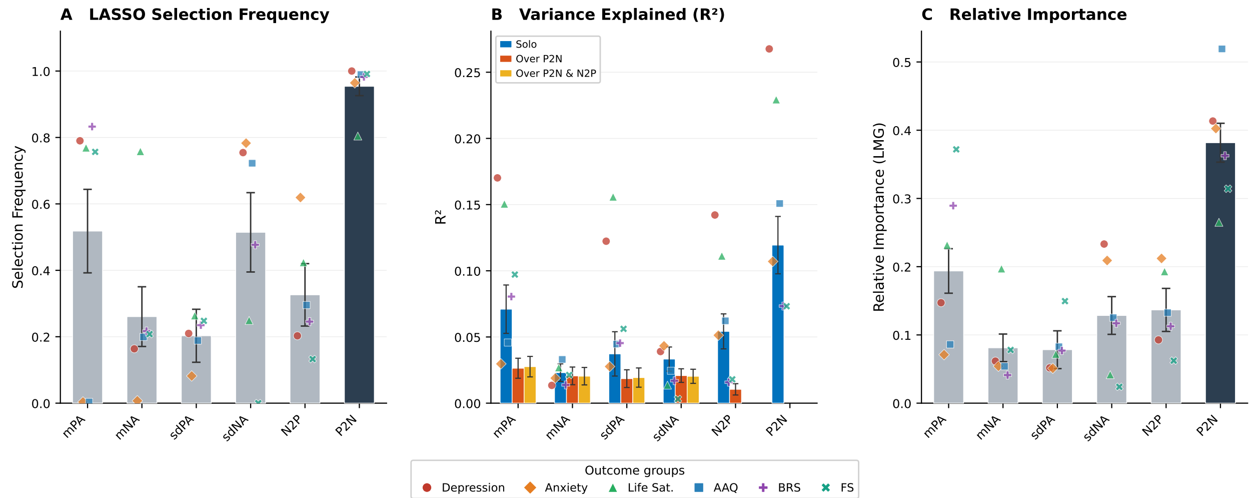

Figure 1. Positive-to-negative affect shift ratio (P2N) emerges as the strongest predictor of psychological well-being across three converging analytical methods. *Note.* Convergence across three analytical methods identifying P2N (positive-to-negative affect shift ratio) as the most robust predictor of psychological well-being outcomes. (A) LASSO selection frequencies representing the proportion of leave-one-out cross-validation folds in which each predictor received a nonzero coefficient when all six predictors were available. Bars indicate weighted averages across all outcomes; error bars indicate standard errors. (B) Variance explained (R^2) for three model types: Solo (blue) shows the variance explained by each predictor alone; Over P2N (orange) shows the incremental R^2 when adding each predictor to a model already containing P2N; Over P2N & N2P (yellow) shows the incremental R^2 when adding each predictor to a model already containing both P2N and N2P. Bars represent weighted averages across all outcomes; error bars indicate standard errors. (C) Relative importance values (LMG metric) representing the proportion of total R^2 attributable to each predictor when all six predictors are included simultaneously in a linear regression, averaged across all possible orderings of predictors. Bars indicate weighted averages across all outcomes; error bars indicate standard errors. In all panels, dots represent outcome-group averages (weighted by sample size): Depression (DASS-D, PHQ-9), Anxiety (DASS-A, GAD-7), Life Satisfaction (SWLS, SWLS1), and three individually displayed outcomes (AAQ = Acceptance and Action Questionnaire–II; BRS = Brief Resilience Scale; FS = Flourishing Scale). All values are weighted averages across datasets. Predictor abbreviations: mPA/mNA = mean positive/negative affect; sdPA/sdNA = standard deviation of positive/negative affect; P2N = positive-to-negative affect shift ratio; N2P = negative-to-positive affect shift ratio.

Relative Importance Analysis

To examine how variance is partitioned among predictors when all compete simultaneously in the same model, we conducted relative importance analyses using the LMG method. This approach complements the previous analyses by showing each predictor's proportional contribution to the total explained variance. Table 7 presents the LMG values for overall psychological well-being.

P2N accounted for 38.2% of the total explained variance in psychological well-being, substantially more than any other predictor. This finding converges with results from LASSO (where P2N was selected in 96% of cross-validation folds) and linear regression (where P2N showed the highest solo R^2 of 11.9%). Among intensity measures, mean positive affect made the largest contribution (19.4%), while other intensity measures each accounted for less than 13% of explained variance. The two shift metrics combined (P2N and N2P) contributed 51.9% of the total explained variance, demonstrating that affect transitions between positive and negative states capture a substantial portion of the predictable variance in psychological well-being outcomes.

Patterns across individual outcomes (Fig. 1C) showed that P2N consistently accounted for the largest or second-largest share of explained variance across most well-being indicators, though the relative contributions of intensity measures varied depending on the specific outcome.

Discussion

Our findings support the value of direct bipolar valence measurement for advancing understanding of affect dynamics and psychological well-being. Across three independent datasets totalling 345 participants and over 30,000 momentary assessments, we demonstrated that affect shift metrics—particularly P2N-ASR, quantifying the propensity to transition from positive to negative affective states—predict psychological well-being more strongly than traditional intensity-based measures derived from the same bipolar valence scale. Critically, this predictive advantage persisted even with a reduced measurement protocol of only 3 daily assessments, addressing a key barrier to clinical translation. These results replicate and substantially extend our previous findings (Goicoechea, Dakos, et al., 2025), providing robust evidence that how individuals move between affective states may be more diagnostic of psychological functioning than the typical intensity or variability of those states.

The prevalence of bistable affect dynamics—observed in approximately 52.1% of participants across all three datasets—provides important complementary information about the structure of valence dynamics in these samples. Bistability indicates

Table 6. R^2 values for solo and hierarchical linear regression models predicting overall psychological well-being

Predictor	Solo R^2	95% CI	Over P2N (ΔR^2)	95% CI	Over P2N & N2P (ΔR^2)	95% CI
P2N	0.119	[0.028, 0.277]	—	—	—	—
mPA	0.071	[0.017, 0.190]	0.026	[0.004, 0.107]	0.028	[0.004, 0.106]
N2P	0.054	[0.009, 0.159]	0.010	[< 0.001, 0.067]	—	—
sdPA	0.037	[0.007, 0.130]	0.019	[0.001, 0.094]	0.019	[0.001, 0.094]
sdNA	0.033	[0.006, 0.117]	0.021	[0.001, 0.094]	0.020	[0.002, 0.094]
mNA	0.023	[0.003, 0.100]	0.021	[0.004, 0.093]	0.020	[0.004, 0.092]

Solo R^2 represents variance explained by each predictor alone. Over P2N shows incremental R^2 (ΔR^2) when adding each predictor to a model containing P2N. Over P2N & N2P shows incremental R^2 when adding each intensity measure to a model containing both shift metrics. All values are weighted averages across datasets and outcomes ($N = 1{,}631$ total observations). 95% CIs derived from 1,000 bootstrap resamples. *mPA/mNA* mean positive/negative affect, *sdPA/sdNA* standard deviation of positive/negative affect, *P2N* positive-to-negative affect shift ratio, *N2P* negative-to-positive affect shift ratio.

Table 7. Relative importance (LMG) of affect dynamic predictors for overall psychological well-being

Predictor	LMG	*SE*
P2N	0.382	0.028
mPA	0.194	0.033
N2P	0.137	0.032
sdNA	0.129	0.028
mNA	0.081	0.020
sdPA	0.078	0.028

LMG values represent the proportion of total R^2 attributable to each predictor, averaged across all possible orderings of predictors. Values sum to 1.0. All values are weighted averages across datasets and outcomes ($N = 1{,}631$ total observations). *SE* standard error.

that roughly half of individuals have distinct positive and negative attractors, implying that their transitions between affective states tend to be abrupt regime shifts rather than gradual fluctuations. This structural characterization is theoretically meaningful in its own right, connecting human affect to broader dynamical systems phenomena. However, bistability classification did not significantly predict well-being outcomes in our previous work (Goicoechea, Dakos, et al., 2025), whereas shift metrics—particularly P2N—did. This dissociation is informative: it suggests that the frequency of crossing the valence boundary matters for well-being regardless of whether such crossings reflect true regime shifts between attractors or gradual drift across a monostable landscape. This indicates that the neutral point on a valence scale represents a psychologically meaningful threshold—the boundary between feeling good and feeling bad—and crossing this boundary may be more consequential for well-being than the underlying attractor structure. In this sense, dichotomizing valence into positive and negative extracts a meaningful signal rather than merely discarding intensity information.

Why Shift Metrics Outperform Intensity Measures

The superior predictive validity of affect shift metrics, particularly P2N-ASR, reflects a fundamental difference in what these metrics capture about affective experience. Intensity-based measures—whether simple aggregates like mean and standard deviation or more complex variability indices—characterize the typical level and magnitude of fluctuation in affective states. They tell us where someone typically resides affectively and how much they vary around that central tendency. Shift metrics, in contrast, capture regulatory dynamics: the propensity to fall into negative affect (P2N) versus the propensity to recover from it (N2P). These metrics reveal asymmetries in affective transitions that intensity measures obscure.

Our analyses revealed a convergent pattern across multiple methodological approaches. In LASSO regression, models including only shift metrics (P2N, N2P) performed comparably to models that also included all four intensity measures (mean and standard deviation of positive and negative affect), suggesting that the additional measurement burden of intensity ratings provides minimal incremental predictive value. Relative importance analyses further revealed that when all predictors compete simultaneously, shift metrics account for a larger proportion of explained variance in well-being outcomes. These converging results indicate that the dynamics of valence—how people move between feeling good and bad—may be more psychologically meaningful than the levels of valence.

From a theoretical perspective, this pattern makes sense when we consider what psychological well-being entails. Well-being is not merely about experiencing more positive affect or less negative affect on average; it involves the capacity to regulate affective states effectively—to maintain positive states when appropriate and to recover from negative states when they occur. The propensity to shift from positive to negative affect (P2N-ASR) may index vulnerability or weak regulatory capacity, reflecting difficulty sustaining positive affective states in the face of challenges. Conversely, the propensity to shift from negative to positive affect (N2P-ASR) may reflect resilience and effective emotion regulation, capturing the ability to "bounce back" from negative experiences. These directional propensities provide insight into regulatory processes that aggregate intensity measures cannot access.

Relationship to Emotional Inertia

Our affect shift metrics bear a conceptual relationship to emotional inertia—the tendency for emotional states to persist over time, typically operationalized as the autocorrelation of affect ratings across successive measurements (Kuppens et al., 2010). Both approaches capture temporal dynamics rather than static levels. However, important differences exist. Inertia is computed on continuous intensity ratings and reflects the degree to which any affective state—regardless of valence—carries over to the next moment. High inertia indicates that affect is "sticky," whether positive or negative. In contrast, P2N and N2P are directional and regime-specific: they quantify the probability of crossing the valence boundary in a particular direction. A person could show high inertia (strong autocorrelation) while rarely transitioning between valence regimes, or low inertia (rapid fluctuations) with frequent regime crossings. Moreover, our metrics distinguish vulnerability (P2N: falling into negative affect) from recovery capacity (N2P: returning to positive affect)—an asymmetry that undifferentiated inertia measures cannot capture. In this sense, shift ratios can be viewed as directional, dichotomized indices of regime transitions that complement rather than duplicate inertia-based approaches.

Implications for Valence Measurement and Conceptualization

These findings have important implications for how valence should be operationalized in affective science. As Colombetti (2005) and Walle and Dukes (2023) have articulated, valence remains poorly defined despite its centrality to emotion theory, with researchers often conflating different aspects of emotional experience under the valence label. Direct bipolar valence measurement addresses this problem by allowing participants themselves to integrate the complexity of their affective experience into a coherent valence judgment. Rather than researchers imposing assumptions about what should be positive or negative, participants report how they actually feel at each moment, naturally incorporating contextual factors, personal meanings, and potentially mixed feelings into their response. In essence, participants are better at integrating the complexity of their affective experience than researchers are at inferring it from categorical emotion labels. This approach aligns with a growing recognition that valence is not a property of emotions themselves but rather a dynamic evaluation process—a subjective aspect of emotional experience whose positive or negative character depends on individual concerns and contexts (Colombetti, 2005).

Beyond its theoretical soundness and strong empirical validation (Cloos et al., 2023), direct bipolar measurement offers substantial practical advantages. It is cognitively less demanding than rating multiple discrete emotions, avoiding the challenge of discriminating between similar intensity levels on emotion scales without absolute anchors. It reduces participant burden in intensive longitudinal designs—a critical consideration demonstrated by our Spanish 2 dataset, where meaningful affect dynamics were captured with only 3 daily assessments. It may also reduce some challenges arising from translation and cross-cultural interpretation of discrete emotion labels, though it is acknowledged that even simple questions like "How do you feel?" can carry different connotations across languages and cultures (Wierzbicka, 1999). Nevertheless, a single bipolar item requires translating one concept rather than multiple emotion terms with culture-specific meanings. Finally, it may align with how people naturally communicate their affective state in everyday interactions (Barrett, 2004), potentially making it accessible across diverse populations including different ages, educational backgrounds, and clinical conditions (Lucas & Donnellan, 2012; Goicoechea, Dakos et al., 2025). In this sense, "How do you feel?" may be both the simplest and most powerful question we can ask in affective science.

Limitations and Future Directions

Several limitations warrant consideration. First, our approach relies on self-report, raising considerations about response biases common to all subjective measures. Participants may exhibit social desirability effects, recall limitations, or individual differences in scale interpretation. However, we note that these concerns apply equally—or perhaps more so—to multi-item discrete emotion assessments, which require participants to distinguish between similar emotional states without absolute anchors. The simplicity of a single bipolar valence question may actually reduce some response biases by minimizing cognitive load and interpretation demands. Nevertheless, future research could strengthen these findings by incorporating physiological or behavioral correlates of affective transitions, providing convergent validity for the self-reported shift metrics.

Second, our positive and negative affect variables were derived from the bipolar valence continuum rather than assessed as independent dimensions. While some theories propose that positive and negative affect can be experienced simultaneously as separate dimensions (Cacioppo & Berntson, 1994), we view this as a feature rather than a limitation of our approach. As we have argued, the bipolar operationalization avoids imposing researcher assumptions about which emotions should be categorized as positive or negative, instead allowing participants to integrate their experience into a unified valence judgment. Importantly, however, this also means that the intensity-based measures used for comparison—means and standard deviations of positive and negative affect—were derived from the same single-item bipolar scale as the shift metrics. Our findings therefore demonstrate the predictive advantage of shift metrics within a direct bipolar measurement framework; whether this advantage generalizes to comparisons with intensity measures derived from independent multi-item instruments, such as the PANAS, remains an open empirical question. Future research could compare shift metrics derived from bipolar valence measurement with those derived from independent positive and negative affect ratings to examine whether the predictive advantages we observed extend across different operationalizations.

Relatedly, our single-item bipolar measure constrains valence to a unidimensional structure, which may oversimplify affective experience. The circumplex model of affect (Russell, 1980) posits that core affect varies along two dimensions: valence and arousal. When these axes are rotated 45°, they yield alternative valenced dimensions (Watson & Tellegen, 1985) that cannot be captured separately by a single bipolar item. Participants reporting that they feel 'good' may be experiencing calm contentment or energized excitement—states with different arousal profiles and potentially different implications for well-being. Our approach assumes that participants integrate these dimensions into a single summary judgment, which may be appropriate for some research questions but not others. Future research could examine whether shift metrics computed separately for high-arousal and low-arousal affective states yield differential predictions, or whether arousal moderates the relationship between valence shifts and well-being outcomes.

Third, all three datasets were collected in Western populations (Spain and Germany), limiting the generalizability of our findings in ways that extend beyond sampling. At the interpretive level, our finding that lower P2N predicts better well-being implicitly assumes that remaining in positive affective states is generally adaptive. This reflects a limitation shared with most affect-well-being research, which has been developed primarily within Western cultural frameworks that value positive affect, a preference not necessarily shared by all cultures (Tsai, 2017). However, shift metrics make weaker assumptions than intensity-based measures, which presume not only that positive states are desirable but that more intense positive affect is better. Nevertheless, we need to acknowledge that neither approach captures contextual appropriateness as a person who remains positive in circumstances warranting negative affect would appear well-regulated by our metrics. Future empirical validation across diverse cultural contexts remains therefore essential, as cultural differences in emotion regulation norms, display rules, and the meaning attributed to affective states could influence both the prevalence of bistability and the relationship between shift metrics and well-being outcomes.

Fourth, our analyses are correlational and cannot determine whether affect shifts cause changes in well-being or vice versa. It is plausible that diminished well-being increases vulnerability to negative affect (higher P2N) or impairs recovery capacity (lower N2P). Alternatively, individual differences in shift propensities may represent relatively stable traits that confer risk or resilience over time. Disentangling these possibilities requires longitudinal designs with dense temporal sampling to establish temporal precedence, or experimental and intervention studies that test whether enhancing recovery propensity (e.g., through emotion regulation training) causally improves well-being outcomes. Additionally, the convergence of results across independent datasets and methods is reassuring, but the analyses should be understood as partly exploratory; future confirmatory work with preregistered hypotheses would strengthen the conclusions drawn here.

Finally, while our shift metrics quantify the propensity to transition between affective states, they do not identify the psychological processes or contextual factors that trigger these transitions. Understanding what precipitates shifts from positive to negative affect, or what facilitates recovery to positive states, represents a critical next step for translating these findings into clinical interventions. We are currently addressing this gap through ecological momentary assessment of biopsychosocial processes (as measured by the Process-Based Assessment Tool; Ciarrochi et al., 2022) in relation to affect shift propensities. Using network analysis approaches similar to those described in our cross-cultural validation work (Goicoechea, Wallman-Jones, et al., 2025), this ongoing effort aims to identify both idiographic (person-specific) and nomothetic (group-level) patterns linking specific psychological processes—such as attention, cognitive flexibility, and emotion regulation—to individual differences in P2N and N2P. Such research could ultimately inform precision interventions tailored to the specific processes that drive vulnerability or recovery in individual patients.

Conclusion

Valence is central to affective experience, yet its operationalization has remained problematic in emotion science. We argue that measuring valence directly through a simple bipolar question—"How do you feel?"—offers both theoretical and empirical advantages over approaches that infer valence from discrete emotion categories. This direct measurement approach respects the context-dependent and participant-integrated nature of valence, avoiding researcher assumptions about which emotions should be considered positive or negative. Moreover, it enables the quantification of affect shift dynamics that may be more diagnostic of psychological well-being than the typical levels or variability of affective experience. The propensity to fall into negative affect and the propensity to recover from it capture regulatory processes central to psychological functioning—processes that aggregate measures of affective intensity cannot access. Our findings across 216 participants in the second Spanish study demonstrate that these advantages hold even with minimal measurement burden, making this approach feasible for clinical populations and applied contexts where participant burden is a critical concern. The dynamics of valence—how people move between feeling good and bad—may be more diagnostic of psychological well-being than the typical levels or variability of affective experience. As affective science continues to refine its understanding of valence, we hope this work contributes not only to conceptual clarity but also to methodological progress. Sometimes the most powerful insights come from the simplest questions, and "How do you feel?" may offer a clearer window into affective experience than the complex emotion rating batteries that have dominated the field.

Additional Information

Funding. This research was supported by grant PID2022-138249NB-I00 from the Spanish Ministry of Science and Innovation.

Competing Interests. The authors declare no competing interests.

Data Availability. The data and software code can be accessed on the Open Science Framework repository (https://doi.org/10.17605/OSF.IO/QFKHJ).

Ethics Approval. The two Spanish studies were approved by the Human Research Ethics Committee of the University of Granada in Spain (Reference: 2214/CEIH/2021), and the German study by the Department of Psychology of the Philipps-University of Marburg in Germany (Reference: 2022-22v).

Consent to Participate. Informed consent was obtained from all individual participants included in the study.

Author Contributions. P.P. and C.G.: conceptualization,

methodology, formal analysis, writing.

References

Banos, O., Damas, M., Goicoechea, C., Perakakis, P., Pomares, H., Rodriguez, C., Sanabria, D., & Villalonga, C. (2024). MobileWell400+: A large-scale multivariate longitudinal mobile dataset for investigating individual and collective well-being. In *Zenodo*. https://doi.org/10.5281/zenodo.11060596

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology, 87*(2), 266–281. https://doi.org/10.1037/0022-3514.87.2.266

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality, 40*(1), 35–55. https://doi.org/10.1016/j.jrp.2005.08.006

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.

Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research, 23*(10), 2809–2818. https://doi.org/10.1007/s11136-014-0726-4

Ciarrochi, J., Sahdra, B., Hofmann, S. G., & Hayes, S. C. (2022). Developing an item pool to assess processes of change in psychological interventions: The Process-Based Assessment Tool (PBAT). *Journal of Contextual Behavioral Science, 23*, 200–213. https://doi.org/10.1016/j.jcbs.2022.02.001

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment, 35*(3), 189–204. https://doi.org/10.1037/pas0001200

Colombetti, G. (2005). Appraising valence. *Journal of Consciousness Studies, 12*(8–10), 103–126.

Dakos, V. (2013). *earlywarnings: Early warning signals for critical transitions in time series*. R package version 1.1.29. https://cran.r-project.org/package=earlywarnings

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour, 3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Goicoechea, C., Dakos, V., Sanabria, D., Heshmati, S., Westhoff, M., Banos, O., Pomares, H., Hofmann, S. G., & Perakakis, P. (2025). Bistability and affect shift dynamics in the prediction of psychological well-being. *Emotion, 25*(4), 982–996. https://doi.org/10.1037/emo0001454

Goicoechea, C., Wallman-Jones, A., Ciarrochi, J., Hayes, S. C., Hofmann, S. G., & Perakakis, P. (2025). *PBAT network analysis: Spanish validation and cross-cultural perspectives*. https://doi.org/10.31234/osf.io/sxhwt_v1

Groemping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software, 17*(1), 1–27. https://doi.org/10.18637/jss.v017.i01

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141*(4), 901–930. https://doi.org/10.1037/a0038822

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science, 21*(7), 984–991. https://doi.org/10.1177/0956797610372634

Livina, V. N., Kwasniok, F., & Lenton, T. M. (2010). Potential analysis reveals changing number of climate states during the last 60 kyr. *Climate of the Past, 6*(1), 77–82.

Lovibond, P. F., & Lovibond, S. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research, 105*(3), 323–331. https://doi.org/10.1007/s11205-011-9783-z

Ortony, A. (2022). Are all basic emotions emotions? A problem for the (basic) emotions construct. *Perspectives on Psychological Science, 17*(1), 41–61. https://doi.org/10.1177/1745691620985415

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2005). Emotion in human consciousness is built. *Journal of Consciousness Studies, 12*(8–10), 26–42.

Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review, 9*(2), 111–117. https://doi.org/10.1177/1754073916639658

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*(1), 3–30. https://doi.org/10.1037/0033-2909.125.1.3

Scheffer, M., Bascompte, J., Brock, W., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature, 461*(7260), 53–59. https://doi.org/10.1038/nature08227

Sen, P. K., Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1981). Introduction to bivariate and multivariate analysis. *Journal of the American Statistical Association, 76*(375), 752.

Smith, B., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The Brief Resilience Scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Tsai, J. L. (2017). Ideal affect in daily life: Implications for affective experience, health, and social behavior. *Current Opinion in Psychology, 17*, 118–128. https://doi.org/10.1016/j.copsyc.2017.07.004

Walle, E. A., & Dukes, D. (2023). We (still!) need to talk about valence: Contemporary issues and recommendations for affective science. *Affective Science, 4*(3), 463–469. https://doi.org/10.1007/s42761-023-00217-x

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*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin, 125*(5), 601–610. https://doi.org/10.1037/0033-2909.125.5.601

Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press. https://doi.org/10.1017/CBO9780511521256

Supplementary Material

Supplementary Tables

Table S1. *Solo R^2 values for linear regression models using overall valence measures*

Predictor	Solo R^2	95% CI	*SE*
P2N	.117	[.028, .266]	.022
M_Valence	.095	[.024, .224]	.018
N2P	.054	[.010, .156]	.013
SD_Valence	.026	[.002, .107]	.009

Note. R^2 values represent variance explained by each predictor alone, averaged across all dataset-outcome combinations weighted by sample size. 95% CIs derived from 1,000 bootstrap resamples. *M_Valence* within-person mean of valence ratings on the full -50 to $+50$ scale; *SD_Valence* within-person standard deviation of valence ratings.

Table S2. *Sensitivity analysis: solo R^2 values including all participants regardless of compliance ($N = 543$ participants)*

Predictor	Solo R^2	95% CI
P2N	.095	[.022, .213]
M_PosA	.059	[.014, .149]
N2P	.058	[.013, .146]
SD_PosA	.030	[.004, .101]
SD_NegA	.023	[.004, .084]
M_NegA	.018	[.001, .079]

Note. Analysis includes all participants from all three datasets regardless of EMA compliance rate ($N = 543$ participants). The pattern of results—with P2N as the strongest predictor—remains consistent with the compliant sample analysis.