To cite this article:

Dejonckheere, E.*, Mestdagh, M.*, Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (in press). Complex affect dynamic measures add limited information to the prediction of psychological well-being. *Nature Human Bevahior*. https://doi.org/10.1038/s41562-019-0555-0

This manuscript constitutes a preprint and has not been peer-reviewed.

Complex affect dynamic measures add limited information to the prediction of psychological well-being

Egon Dejonckheere*[1], Merijn Mestdagh*[1], Marlies Houben[1], Isa Rutten[1], Laura Sels[1], Peter Kuppens[1], & Francis Tuerlinckx[1]

[1] KU Leuven – Faculty of Psychology and Educational Sciences

* Egon Dejonckheere and Merijn Mestdagh contributed equally to this work.

Abstract

Over the years, increasing attention has been paid to the relation between emotion dynamics and psychological well-being. Because our emotional life is inherently time dynamic, it has been argued that, next to how positive or negative we feel on average, patterns of emotional change too, may convey important information about our mental health. This growing interest came with a surge in new affect dynamic measures, each claiming to capture a unique dynamical aspect of our emotional life that is crucial for understanding well-being. Although this accumulation of new measures may suggest scientific progress, researchers have not always evaluated (a) how different affect dynamic measures empirically interrelate, and (b) what their incremental value is in the prediction of psychological well-being. Here, we address these central questions by analyzing combined affective time series data from 15 different studies (*N* = 1,777). The results (a) map the considerable empirical interdependencies that exist between 16 commonly studied affect dynamic measures, indicating that single measures often do not convey unique information, and (b) show that dynamic measures have little incremental value over average levels of positive and negative affect (and variance in these affective states), when predicting individual differences in three prominent indicators of human well-being (life satisfaction, depressive and borderline symptoms). Although affect dynamic measures may adequately summarize individual differences in emotional time series, our findings indicate that caution is warranted in the optimism on the incremental value of these dynamics regarding their explanatory power for well-being or psychopathology.

Keywords: emotion dynamics, psychological well-being, average levels of affect, variability in affect, affective time series, Ockham's razor

Feelings change. In fact, one of the primary reasons we experience emotions is thought to lie in their time dynamic nature[1-5]. That is, affective experiences alert us about personally relevant changes in our environment that pose a threat or opportunity to our well-being, and prepare us to cope with these changes effectively[6-8]. Because this functionality of our emotional life is built on its fundamentally dynamic nature, we and others have argued that the patterns with which our emotions change over time provide unique information for our psychological well-being[9-12]. Next to how positive or negative we feel on average, investigating the temporal features of our emotional experiences is deemed crucial to explain individual differences in mental health or psychopathology[13].

A Plethora of Dynamic Measures

Over the last decade, the rising interest in the relation between emotion dynamics and psychological well-being came with a surge in new affect dynamic measures, each claiming to evaluate a unique dynamical aspect of our emotional lives (see Table 1 for a non-exhaustive overview). Despite the enthusiastic accumulation of new dynamical measures (also in our own research), the question remains to what extent all proposed measures truly add to the prediction[I] of individual differences in psychological well-being. Indeed, with the proposal of novel concepts and measures, it has often not been evaluated (a) how they empirically relate to other dynamic measures, and (b) what the incremental value of new affect dynamic measures is over other existing measures in the relation with psychological well-being (although for both claims exceptions exist[14-19]).

[I] Throughout this article, the term *prediction* is used to refer to the explanation of new unseen data points as happens when a subset of the data are set apart for validation purposes. However, it is not understood in a prospective or forecasting manner, because we only consider concurrent associations between emotion dynamics and well-being. Similarly, the term *explanatory power* should be understood in terms of *explained* variance and does not imply causality.

Empirical interdependencies among dynamic measures. Not assessing how different affect dynamic measures mutually relate leaves uncertainty about the added value of newly proposed measures[16,17,20]. If a new dynamic measure simply approaches the same data from a different angle, the danger exists that we may be merely reinventing the wheel. Put differently, disregarding any overlap with existing measures fails to evaluate the potential redundancy of new measures, which may not only lead to a disperse and scattered research field, but also creates a false sense of scientific progress in the long run.

Unique explanatory power in the prediction of psychological well-being. In a similar vein, based on Ockham's principle of parsimony, a good dynamic measure should be able to add to the prediction of well-being above and beyond other existing measures[21]. Here, we argue that affect dynamic measures should have unique explanatory power above and beyond average levels of positive (PA) and negative affect (NA). After all, research on emotion dynamics originated from the idea that psychological adjustment involves more than merely experiencing high average levels of PA and low levels of NA[1]. Mean levels of positive and negative emotionality are known to be centrally involved in human well-being and psychopathology[22], and have been historically investigated first in relation to psychological well-being[9]. Remarkably, however, the incremental value of affect dynamic measures over average levels of PA and NA has not always been assessed.

Furthermore, mathematical arguments exist to why average levels of affect are important covariates to find incremental value over. From a mathematical point of view, the mean (M) is considered the first (raw) moment to describe a distribution of scores[23], defined as the weighted sum of the values (where the weights are determined by the distribution). This implies that, in order to compute more complicated affect dynamic measures one needs to know the mean of a person's affect ratings.

The Current Research

The goal of the present paper is twofold. First, we aim to examine how 16 frequently studied (old and newer) dynamical measures (including the mean; see Table 1) empirically interrelate[II]. Second, we will investigate which affect dynamic measures display primary relations to outcomes of psychological well-being (i.e., show relations with psychological well-being independent of other measures), and which measures only show secondary associations with psychological well-being (i.e., merely show a relation by virtue of their association with measures that have a primary relation). Because average levels of positive and negative affect make the most prominent indicators of psychological well-being[24,25], we will evaluate the predictive power of all dynamical measures, with and without controlling for mean levels of positive and negative affect. In a second step, we will also add the observed variance (or standard deviation; SD) in PA and NA as controlling variables (see results for a justification).

To this end, we will analyze data from 15 different studies that have repeatedly tracked people's emotion(al changes) throughout daily life (see Table 2; N = 1,777). Across datasets, participants' levels of well-being ranged from being psychologically well adjusted, to experiencing pathological symptoms, to fulfilling the criteria of a formal clinical diagnosis. To capture the nuance that psychological well-being comprises both the presence of positive outcomes, as well as the absence of negative markers of psychological maladjustment[9], we will evaluate the unique explanatory power of emotion dynamics in both a positive indicator of psychological well-being, people's perceived life satisfaction, as well as two negative outcomes, people's levels of depressive and borderline symptoms. To assess psychopathological symptoms, we rely both on continuous measures, as well as clinical diagnoses.

[II] Other affect dynamic measures that have been proposed in the past are emotional *switching*[26], *flux, pulse*[27], and *spin*[28].Yet, due to their infrequent application, an in-depth analysis of these dynamic measures is beyond the scope of this article.

[Figure 1 around here]

[Table 1 around here]

Affect dynamic measures. Studying emotion dynamics implies measuring affective states at multiple points in time. Prominent methods to examine emotional fluctuations in daily life are experience sampling (ESM)[29] and daily diary studies[30] (see Table 2). In these protocols, participants typically carry a mobile device in their daily lives and complete momentary or daily questionnaires about their (emotional) experiences. Such methods do not only provide researchers with a particularly natural assessment of participants' emotional trajectories (i.e., highly ecologically valid), evaluating people's affective experiences close to real time also weakens the potential impact of memory biases.

Based on the same affective time series these protocols typically produce, researchers may compute different dynamic measures. Table 1 presents an overview of the most commonly studied affect dynamics in relation to psychological well-being, and the short-codes we will use to describe each measure. Figure 1 visualizes simulated data, showing how low and high values for each measure would manifest in a person's affective time series. Importantly, to compute most affect dynamic measures, it is common practice to first create a composite measure for PA and NA by averaging same-valenced emotion items at each measurement occasion. In contrast, other affect dynamic measures rely on the scores of individual emotion items. Furthermore, note that some dynamic measures are truly time-dependent (i.e., their values change when measurement occasions are permuted), while others do not carry a temporal dependency (see Table 1).

Psychological well-being. Human well-being is a broad and multifaceted construct. Here, we will examine the explanatory power of emotion dynamics in three outcomes related to psychological well-being, one positive marker (life satisfaction), and two negative ones (depressive and borderline symptoms), known to be intrinsically related to a host of emotion dynamics measures[9].

Life satisfaction is a positive indicator of psychological functioning[31],involving a global cognitive and eudaimonic evaluation of one's life[32]. Houben and colleagues concluded in their meta-analysis[9] that higher levels of life satisfaction are associated with emotional lives that are less variable and inert, and more stable. Satisfaction with life has also been linked to higher levels of emodiversity[33], and a more independent experience of positive and negative affect[34].

In contrast, the experience of depressive and borderline symptoms indicates maladaptive psychological functioning. Both symptom types have been linked to a more variable, inert and instable emotional life[9]. Yet, when controlling for overlap between these affect dynamics, depressive symptoms are independently associated with higher emotional inertia and variability, but not instability[17]. In turn, borderline patients show more affective instability than people with a diagnosis of Major Depressive Disorder (MDD)[35] and healthy controls[36]. Both borderline and depressive symptoms are also characterized by a stronger bipolar experience of positive and negative affect[15,37], and a reduced ability to differentiate between various same-valenced emotions[38,39]. Finally, depression has also been linked to a denser, more change-resistant, and less diversified emotional life[40,41].

Results

Empirical interdependencies among dynamic measures. In these analyses, we aimed to investigate the overlap that exists between all affect dynamic measures. Because the correlation matrix of all dynamic measures was highly comparable across the different datasets (ICC = .98), we z-transformed all measures per dataset and visualized these in a network[III].

[III] A correlational network without z-transforming all affect dynamic measures produced highly identical results.

[Figure 2 around here]

Correlational network of all affect dynamic measures. Figure 2 visualizes the correlations between all 16 affect dynamic measures and reveals several things. First, M positive and negative affect are moderately negatively related (r = -.39), with both G measures showing a considerable negative association with their respective means (r PA = -.43; r NA = -.64). Second, dynamic measures capturing variability in each valence (i.e., SD, SD*, and MSSD) are strongly interconnected (r's PA ≥ .79; r's NA ≥ .49), but also between valences associations among these measures appear. Third, AR measures for each valence are interrelated moderately (r = .43), and show little association with other dynamical measures ($|r|$'s ≤ .33), except for D (r's ≥ .40). Similarly, ρ is quite isolated in this network, only showing small (negative) relations with other affect dynamic measures ($|r|$'s ≤ .36).

In terms of uniqueness, both AR measures have the least variance explained by mean levels of, and SDs in, PA and NA (R^2's ≤ .08), followed by ρ (R^2 = .19). In contrast, affect dynamic measures with the most variance explained by mean levels of, and SDs in, PA and NA are both MSSD (R^2's ≥ .78) and SD* measures (R^2's ≥ .42).

[Figure 3 around here]

Principal Component Analysis (PCA) on all affect dynamic measures. Figure 3 presents the results of a rotated PCA solution with six components (Panel A). Using an orthogonal Varimax rotation, we retained six components, primarily based on arguments of intuitive interpretation. In total, this solution explained 82% of the variance in all affect dynamic measures, with each original component explaining at least 5%. Panel B visualizes an empirical example that displays the affective time series of an individual in the highest percentile of each component.

Component 1 covers measures related to average levels of positive affect. M positive affect loads highest on this component (*r* = .79), while its G measure (*r* = -.85), and M negative affect (*r* = -.24) show negative loadings.

Component 2 captures affect dynamic measures related to variability in (positive) affect, with high component loadings for SD (*r* =.94), SD* (*r* =.91), and MSSD (*r* =.87) in PA. The same dynamic measures related to NA also exhibit small positive loadings on this component (*r*'s ≤ .43).

While measures related to average levels of, and variability in, positive affect are represented by two independent components, these same measures for negative affect are reflected by a single third component. This is due to the fact that the majority of participants in each dataset was psychologically relatively well-adjusted (i.e., most people typically experienced low levels of negative affect), which leads mean and SD in NA to show stronger interrelations[20]. Related to average emotionality, M negative affect loads high on this component (*r* =.85), while its G measure (*r* = -.88), and to a lesser extent M positive affect (*r* = -.28) exhibit negative loadings on this component. Related to variability in NA, we observe high component loadings for SD (*r* =.67) and MSSD (*r* =.63) in NA.

Component 4 reflects measures that capture unique variability in negative affect, independent of mean levels of NA. SD* in NA shows the highest loading on this component (*r* = .85), followed by SD and MSSD in NA (*r*'s ≥ .44).

Component 5 includes time-related measures. The AR measures for each valence (*r*'s ≥ .82), and D measure (which is a combination of auto- and cross-regressive effects; *r* =.60) show a strong positive loading on this component. In contrast, the MSSD for PA loads negatively on this component (*r* = -.20). Emotional instability and inertia are known to be inversely related (for constant levels of variance)[16].

Finally, component 6 reflects the relation between positive and negative affect. The ρ measure show the highest loading on this component (r = -.87), followed by small yet opposite loadings for both ICC measures (r's $\geq$ .20). Strong affective bipolarity is associated with poor differentiation between various same-valenced emotions[42] [IV].

Unique explanatory power in the prediction of psychological well-being. In these analyses, we investigated the added value of all affect dynamic measures in the prediction of various outcomes related to psychological well-being.

In a first step, we aimed to evaluate the incremental value of all dynamic measures above and beyond mean levels of positive and negative affect. In a second step, we additionally controlled for the observed variance in these affective states. Since our correlational results indicated that many affect dynamic measures are directly related to SDs in PA and NA, we hypothesized that all measures would show little uniqueness once their shared variance with the SDs in PA and NA is taken into account. Although an SD carries dynamic information in itself (and may therefore be considered tantamount to other dynamic measures), one could regard this measure as the next covariate to find incremental value over. The SD is the (square root of the) second central moment to summarize a group of measurements, and therefore next in line after the mean in terms of in mathematical complexity[23]. Theoretically, the SD is also the most primitive index of all affect dynamic measures, without taking into account time[16]. It could therefore be considered a prerequisite to

[IV] To examine the unique explanatory power of these measures in the prediction of psychological well-being, we decided to use the original affect dynamic measures, not the different principal components. Our aim with PCA was to get insight in the empirical commonalities between the different measures, not to create more stable constructs (which is often the purpose of PCA). Given that these different dynamic measures are calculated from the same affective time series, we believe that combining these measures in one component makes little sense in the context of prediction. To illustrate this concern with an analogy, we believe it is illogical to average a person's observed variance and SD of positive affect to create a more stable measure for his or her affective variability, as both measures are mathematically related.

investigate, before turning to more complex affect dynamics that may play a role in psychological well-being. By additionally controlling for SDs in PA and NA, we aimed to further decrease the number of measures needed to explain individual differences in psychological well-being.

Hierarchical multiple regression. Figure 4 presents the average R^2 for each affect dynamic measure across studies, in the linear prediction of psychological well-being, with and without controlling for average levels of, and SDs in, positive and negative affect. We summarize the most important findings below.

[Figure 4 around here]

First, in each assessed outcome, most variance is explained by M positive and negative affect (R^2's $\geq .11$), except for the prediction of borderline symptoms, where the SD in negative affect is the best predictor ($R^2 = .24$). For life satisfaction and depressive symptoms, SD in negative affect is the second best predictor (R^2's $\geq .08$). Together, these findings support the claim that the mean and SD are the first most important statistics to summarize a group of scores[23], and buttress the argument that dynamics that claim to go beyond these summaries should show predictive power over them.

Second, when comparing the predictive capacity of other affect dynamic measures alone, we observe considerable differences, with relatively high R^2's for, for example, MSSD and G in NA, and D (R^2's $\geq .05$), and low R^2's for AR in PA and NA (R^2's $\leq .03$).

Yet, third, when controlling for average levels of positive and negative affect, these notable differences disappear. That is, all affect dynamic measures add little to the prediction of psychological well-being when the explanatory power of average levels of PA and NA are taken into account (depressive symptoms: R^2's $\leq .02$; borderline symptoms: R^2's $\leq .07$; life satisfaction: R^2's $\leq .03$).

Finally, when additionally controlling for SDs in PA and NA, the predictive capacity of all affect dynamic measures decreases even further (depressive symptoms: R^2's $\leq .02$; borderline symptoms: R^2's $\leq .02$; life satisfaction: R^2's $\leq .02$).

[Figure 5 around here]

Meta-analysis. Figure 5 visualizes the meta-analytic *p*-values of each affect dynamic measure in the prediction of psychological well-being. The most important findings are summarized below.

Individually, the majority of dynamic measures significantly predicts differences in depressive and borderline symptoms, and people's life satisfaction[9]. For depressive symptoms and life satisfaction, M positive and negative affect typically show the strongest significant contributions (p's $\leq 6\times10^{-12}$). For borderline symptoms, the strongest contributor is SD in negative affect ($p = 2\times10^{-21}$). In contrast, a meaningful contribution is unlikely for SD* in negative affect in the prediction of depressive symptoms ($p = .061$), G in positive affect for borderline symptoms ($p = .113$), and the AR in positive affect for life satisfaction ($p = .077$).

Next, when controlling for mean levels of positive and negative affect, considerable differences in significance appear between affective dynamic measures. For depressive symptoms, the contribution of all measures decreases (with drastic non-significance for both AR and G measures in PA and NA; p's $\geq .084$) , except for SD in NA, SD* in PA, and ρ (p's $\leq 2\times10^{-05}$). For borderline symptoms, the significance of all variability-related measures largely remains unaffected (p's $\leq 6\times10^{-05}$), together with the D and ρ measures (p's $\leq 1\times10^{-08}$). Again, both G measures show the most drastic drop in significance (p's $\geq .184$). For life satisfaction, the significance of all measures decreases remarkably, except for ρ ($p = 3\times10^{-08}$).

Finally, when additionally controlling for SDs in PA and NA, most affect dynamic measures do not surpass the $\alpha = .05$ significance threshold in all outcomes. For depressive symptoms, SD* in positive affect ($p = .021$), and ρ are the only significant predictors ($p =$

.009). For borderline symptoms and life satisfaction, the only significant predictors are the ICC in positive affect (p = .041), and ρ ($p = 9\times10^{-05}$), respectively.

Optimal combination of predictors using Lasso variable selection. To determine the optimal set of affect dynamic measures to predict participants' depressive and borderline symptoms, and their life satisfaction, we tested the predictive accuracy of three Lasso-based models. The first model comprised the set of optimal predictors out of all 16 affect dynamic measures (i.e., Lasso $_{\text{ALL}}$). The second model contained the optimal predictor selection out of 4 measures, mean levels and SDs of PA and NA (i.e., Lasso $_{\text{MEAN+SD}}$). Finally, the third model included the most favorable selection out of 2 measures, mean levels of PA and NA (i.e., Lasso $_{\text{MEAN}}$). Leave-one-out cross-validation yielded a predicted R^2 for each of the three Lasso-based models in the prediction of psychological well-being, with higher values indicating better predictive accuracy (i.e., penalized for overfitting).

First, for people's depressive symptoms and their life satisfaction, all three Lasso-based models showed an equal average predictive accuracy across all datasets (depressive symptoms: Lasso $_{\text{MEAN}}$ R^2 = .22, Lasso $_{\text{MEAN + SD}}$ R^2 = .21, Lasso $_{\text{ALL}}$ R^2 = .20; life satisfaction: Lasso $_{\text{MEAN}}$ R^2 = .18, Lasso $_{\text{MEAN + SD}}$ R^2 = .18, Lasso $_{\text{ALL}}$ R^2 = .17). These results illustrate that mean levels of PA and NA often suffice to predict individual differences in depressive symptomatology and life satisfaction, and that the added value of (other) affect dynamic measures is little.

Second, for people's borderline symptoms, Lasso $_{\text{MEAN + SD}}$ had the highest average predictive accuracy across all datasets (R^2 = .28), closely followed by Lasso $_{\text{ALL}}$ (R^2 = .27). Lasso $_{\text{MEAN}}$ had an R^2 of .21. These results indicate that affect dynamic measures show some incremental value over mean levels of PA and NA in the prediction of participants' borderline symptoms, but also that more complex (variability-related) dynamic measures (e.g., SD* and MSSD) do not outperform a simple SD in PA and NA.

Finally, we also investigated how many times each measure comprised the optimal set of predictors in the Lasso model based on all 16 affect dynamic measures. Across all psychological well-being outcomes, mean levels of PA and NA were consistently the most frequently selected measures, including the optimal set in 98% and 87% of the cases, respectively. The third most selected dynamic measure was ρ (64%). All other affect dynamic measures were only selected occasionally (≤ 38%)[V].

Discussion

Because emotions are inherently time dynamic, it has been argued that, next to how positive or negative we feel on average, investigating the dynamics of our emotions is essential to explain individual differences in psychological well-being. Crucially, however, the incremental value of affect dynamics over average levels of positive and negative affect has not always been evaluated. As a plethora of new affect dynamic measures have been introduced to the field, this raises questions about the uniqueness of these measures in understanding mental health or psychopathology.

The present paper aimed to determine the potential ramifications of both issues. Specifically, we investigated (a) how 16 commonly studied affect dynamic measures empirically relate, and (b) what their added value in the prediction of psychological well-being is, above and beyond average levels of positive and negative affect (and variability in these affective states). To reach robust conclusions, we combined affective time series data from 15 different studies, in which participants reported various levels of psychological well-being, ranging from psychologically well-adjusted, to experiencing debilitating

[V] SDs of PA and NA were only selected in 38% and 16% of the cases, respectively. This is due to overlap with other variability-related measures (e.g., SDs* and MSSDs), which evenly spreads the chance of being selected over the different measures. For the same reasons, ρ was frequently selected. This affect dynamic measure is relatively unique, sharing little overlap with other measures.

psychopathological symptoms. In the next paragraphs, we discuss the implications of our findings and provide guidelines for future research in the field of emotion dynamics.

To qualify as an adequate predictor of psychological well-being, our results suggest that an affect dynamic measure should fulfill two necessary (but individually, insufficient) conditions. First, it should summarize affective time series in a unique way, not overlapping too much with other measures. Second, it should have incremental explanatory power in the prediction of psychological well-being.

Overall, our PCA results reveal that all 16 affect dynamic measures can be adequately summarized by six independent components. This reduction indicates that considerable commonalities between measures exist, which illustrates that many dynamic measures may not convey unique information. Interestingly, four of the underlying components directly relate to average levels of, and/or variability in both affective states, which suggests that all dynamic measures are more or less function of these primary summaries[23]. For instance, both positive and negative emodiversity show considerable empirical overlap with their respective mean level of that affective state[38], which may challenge the added value of this measure in explaining differences in people's emotional experiences[21]. Similarly, differences in emotional instability empirically largely coincide with differences in emotional variability, which shows that the temporal order of affect ratings plays a minor role in determining this dynamic.

Other emotion dynamics, however, summarize affective time series relatively independent of other measures, a claim we see supported by two unique components in our PCA. Emotional inertia, for example, shows little empirical relations with other dynamics, except for emotion-network density. Yet, this overlap is to be expected, as emotion-network density directly includes information about a person's auto-regressive slopes[37]. Similarly, the PA-NA correlation informs us relatively little about other dynamics. This measure only exhibits small associations with emotion differentiation. That is, stronger affective bipolarity (weakly) correlates with less fine-grained distinctions between same-valenced emotions[39].

Finally, the overlap between different emotion dynamics is substantially stronger in negative affect, compared to positive affect. The high interconnectedness between all NA dynamics is the result of people typically experiencing low levels of negative affect. In the context of bounded affect ratings, low levels of NA heavily restrict the amount of variability one can observe in that affective state, which leads many emotion dynamics to be more confounded with the mean[20,25]. This finding may also explain why effect sizes of dynamical measures in the prediction of well-being are generally stronger for NA dynamics than for PA dynamics[9]: NA measures share more information with their mean. This brings us to the second criterion.

Affect dynamic measures should have adequate and incremental explanatory power in the prediction of psychological well-being. With reference to adequacy, we indeed replicate the finding in our meta-analysis that most affect dynamics, individually, show a significant relation with various indicators of psychological well-being[9]. However, in terms of effect sizes, most measures explain remarkably little variance in well-being, compared to average levels of positive and negative affect. This not only illustrates that mean levels of affect are indeed the most prominent indicators of psychological well-being[24,25], it also puts the added value of more complex emotion dynamic measures in perspective.

Not surprisingly then, when controlling for average levels of affect, all emotion dynamic measures show little incremental value in the prediction of psychological well-being. This finding suggests that many emotion dynamics merely predict differences in well-being by virtue of their association with mean levels of positive and negative affect. Emodiversity and emotion differentiation in negative affect, for example, explain differences in depressive symptoms reasonably well[35,38]. Yet, once their shared overlap with average levels of negative affect is taken into account, their added value is no longer significant. Similarly, emotional instability in negative affect is strongly linked to the experience of borderline symptoms[32]. However, controlling for differences in average emotionality drastically impacts the explanatory power of this dynamic[20].

In a second step, we additionally controlled for SDs in positive and negative affect. Adding these covariates reduced the predictive capacity of all other affect dynamic measures even further, rendering almost all effect sizes non-significant. One may argue, however, that an SD conveys dynamic information in itself, and is therefore on par with other affect dynamic measures, questioning its nomination as the next most relevant covariate to find incremental value over. Nevertheless, several arguments exist to select variance in positive and negative affect as the second most important covariates in the prediction of well-being. Mathematically, an SD is a one-to-one function of the second central moment to summarize a group of scores next to the mean[23]. Empirically, this claim is supported by the fact that an SD in (negative) affect typically shows the highest explanatory power in well-being after mean levels of PA and NA (except in borderline symptoms, where it is the best predictor). Finally, given it is also the most parsimonious measure of all affect dynamic measures, we believe these arguments justify its selection as the next covariate to control for in the prediction of well-being.

One could argue that merely explaining variance in individual differences in psychological well-being informs us little about the possible (dynamic) processes underlying psychological (mal)adjustment. That is, although affective dynamics may have little predictive capacity, they do provide insight in the potential mechanisms involved in mental health and psychopathology; we call this a theoretical or process explanation.

Although statistical and process explanation are related, they are also distinct phenomena[40]. Merely establishing an association between affect dynamic measures and well-being does not necessarily mean that these affect dynamics contribute to well-being or are causally related to it. We believe that, in the current context of affect dynamics, researchers should adopt a similar parsimonious approach when it comes to theoretical explanation. That is, if two emotion dynamic measures are equally good at explaining individual differences in well-being, emotion researchers should go with the simpler one. We illustrate our case with two examples.

Emotional inertia, for example, has been investigated extensively in people with depressive symptoms[9,12,17,41]. This dynamic is believed to explain how a depressed individual's emotional system has slowed down and is resistant to change. Although this theory seems plausible and is intuitively appealing, once average levels of positive and negative affect are taken into account (which are, nota bene, equivalent to MDD's two main criteria, anhedonia and depressed mood[22]), emotional inertia no longer meaningfully describes a depressed person's affective life.

In a similar vein, emotional instability is considered a core feature of borderline personality disorder[42]. Emotional instability refers to rapid and extreme fluctuations in affect, with BPD patients, for example, feeling very positive one time, yet deeply depressed the next. To effectively describe these moment-to-moment affect oscillations in BPD, researchers justly favor the MSSD over the SD, as this measure acknowledges the temporal dependency between consecutive affect ratings[16,32,43]. Although this theoretical reasoning nominates the MSSD indeed as the preferred measure, our empirical results show that an SD (in negative affect) describes differences in BPD symptomatology equally well, which may suggest that the temporal order in affect ratings in BPD is empirically not as crucial as previously hypothesized.

What value should these dynamic measures be attributed then? When it comes to describing an individual's emotional trajectory, we believe affect dynamics do have empirical value. After all, mere average levels of positive and negative affect often overlook the meaningful ups and downs of a person's emotional life that come and go with the ebb and flow of daily life[3].

For example, there may exist considerable individual differences in the auto-regressive slope of an affective state[47]. This indicates that a temporal dimension is certainly important in modeling emotional processes and understanding individual differences therein. Nevertheless, when relating this dynamic to individual differences in psychological well-being, this measure has little incremental explanatory value.

Finally, the question remains whether current measures of psychological well-being (e.g., typical self-reports or clinician-rated assessments) reflect a valid representation of their underlying construct. Recently, it has been argued that psychiatric disorders may be complex dynamical systems too, with various symptoms mutually interacting with each other[48]. In contrast, traditional symptom questionnaires approach psychopathology rather *static*, conceiving psychiatric disorders as discrete entities, made up by the unweighted sum of their individual symptoms[49]. As many assessments typically cover a broad time span (e.g., request a retrospective average of a symptom during the last two weeks), this may explain why average levels of affect predominantly explain individual differences in psychopathology and well-being.

In summary, although some affect dynamic measures convey unique information about people's emotional lives, many dynamics interrelate considerably. Furthermore, once the explanatory power of mean levels of positive and negative affect (and variability in these affective states) is taken into account, all affect dynamic measures show little incremental value in the prediction of various psychological well-being outcomes. In sum, our findings put the added value of these measures in perspective, and illustrate that caution is warranted for optimism in developing novel and complex measures of emotion dynamics for the prediction / explanation of well-being and emotion disorder.

Methods

To answer our research questions, we combined data from 15 different ESM and daily diary studies. Table 2 presents a detailed summary for each dataset, including information about the participants, protocol, predictors and outcomes assessed in each study. All studies complied with local ethical regulations and were approved by an institutional ethics committee (UNSW Human Research Ethics Advisory Board[77], Medical Ethics Committee UZ Leuven[78,81], KU Leuven[17,26,79,80,83,84], Hospital Network Antwerp[18], Stanford University[19], and Max Planck Institute[82]). All participants provided informed consent.

[Table 2 around here]

Data preprocessing. To prepare the data for analysis, we undertook several steps. Preprocessing of the data had to guarantee a valid score for all participants on each affect dynamic measure (which allowed a fair comparison between all measures). We started with an initial pool of 1,869 participants. First, we excluded participants who showed poor compliance with the ESM or daily diary protocol (i.e., < 50% measurement occasions completed), as a low response rate raised doubts about the validity of their affect scores (n = 38). Second, we eliminated individuals who still completed 20 measurement occasions or less, to ensure a reliable estimation of all affect dynamic measures (n = 39). Next, we removed participants who displayed no variability in their PA or NA composite scores, to enable a valid calculation of participants' intra-class correlations (ICC) and relative variabilities (SD*) in PA and NA, as well as their PA-NA correlation (ρ; n = 12). Finally, to compute valid Gini-coefficient (G) for each individual, we excluded people who never rated the intensity of any same-valenced emotion above 10% of the response scale (n = 3). In the end, this left us with a final sample of 1,777 individuals[VI].

Affect dynamic measures. Several specifics regarding the calculation of each affect dynamic measure are noteworthy. First, for all auto-regressive (AR), mean squared successive differences (MSSD) and density (D) measures, we removed overnight lags, to ensure a relatively equal time window between all measurement occasions[51]. Second, we derived participants' AR and D scores from multilevel models (i.e., slopes drawn from an estimated normal distribution), and did not calculate these per person[47,51]. In contrast, ρ scores were computed separately for each individual[15], as the assignment of one valence as predictor and the other one as outcome in a multilevel regressive context is arbitrary[52]. For the computation of people's ICCs, we relied on consistency measures, rather than measures based on absolute agreement, yet both type of ICCs are known to be interrelated strongly[53]. Finally, to calculate G scores for each valence, we concluded an emotion to be present when the intensity was rated above 10% of the measurement scale[VII].

Psychological well-being. Across studies, we investigated the explanatory role of affect dynamics in three indicators of psychological well-being: people's life satisfaction, and depressive and borderline symptoms.

Satisfaction with life. In all studies, people's perceived life satisfaction was assessed with the Satisfaction With Life scale (SWL; see Table 2)[54]. This 5-item questionnaire captures a global judgment of one's life satisfaction (e.g., *The conditions of my life are excellent*.).

[VI] Two other data checks were performed, but did not result in the exclusion of any participants. First, we verified whether all participants' ICC scores were positive, as the meaning of negative values in an emotion differentiation context is debatable[50]. In total, 79 negative cases were observed, which we set to .00. Second, we examined the number of people with a positive ρ score. In total, 185 individuals displayed a positive correlation between positive and negative affect. For both checks, analyses with and without these distinctive cases yielded identical conclusions.

[VII] Conclusions remained identical when different cut-offs were used (i.e., 5 or 20% of the measurement scale).

Item scales range from 1 (*strongly disagree*) to 7 (*strongly agree*). Internal consistency was high in all studies (mean α = .85).

Depressive symptoms. Different continuous scales were used to assess participants' depressive symptom levels. In some studies, a categorical outcome was used in the form of a clinical MDD diagnosis (see Table 2).

CES-D. The Center for Epidemiologic Studies Depression scale (CES-D)[55] is a 20-item questionnaire to evaluate the frequency of a range of depression complaints participants experienced during the last week (e.g., *I felt depressed.*), ranging from 0 (*rarely or none of the time*) to 3 (*most or all of the time*). Internal consistency was high in all studies (mean α = .89).

PHQ-9. The Patient Health Questionnaire (PHQ-9)[56] assesses the frequency of nine prominent depression symptoms experienced over the last two weeks (e.g., *feeling down, depressed, or hopeless*). Item scales range from 0 (*not at all*) to 3 (*nearly every day*). Internal consistency was acceptable (α = .64).

Q-IDS. The Quick Inventory of Depressive Symptomatology (Q-IDS)[57] evaluates the severity and frequency of 14 depressive symptoms experienced in the last week (e.g., *feeling sad*). Item scales range from 0 (e.g., *I do not feel sad.*) to 3 (e.g., *I feel sad nearly all of the time.*). In both studies, internal consistency was high (mean α = .88).

BDI-II. The Beck Depression Inventory-II (BDI-II)[58] assesses the intensity of 20 depression symptoms experienced in the past two weeks (e.g., *sadness*). Item scales range from 0 (*absent*) to 3 (*very severe*). Internal consistency was high (α = .88).

SCID-I MDD. The Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I)[59] was used in three studies to determine the presence of a formal MDD diagnosis. In this semi-structured interview, a trained clinician assessed whether participants fulfilled five or more DSM-IV criteria for depression for at least two weeks.

Borderline symptoms. Two continuous surveys were used to assess participants' borderline symptom levels. Two studies had a categorical outcome in the form of a clinical BPD diagnosis (Borderline Personality Disorder; see Table 2).

ADP-IV BPD. The Assessment of DSM-IV Borderline Personality Disorder subscale (ADP-IV BPD)[60] consists of 10 trait items that evaluate the DSM-IV criteria for BPD. Participants indicate how much each trait applies on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Internal consistency was high in all studies (mean $\alpha = .82$).

PAI-BOR. The Personality Assessment Inventory Borderline scale (PAI-BOR)[61,62] is a 24-item survey that assesses four important domains of impairment related to BPD (i.e., *affective instability, identity problems, negative relationships, self-harm*). Participants rate each trait on a scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). Internal consistency was high ($\alpha = .94$).

SCID-II BPD. In two studies, the Structured Clinical Interview for DSM-IV Axis II personality disorders (SCID-II)[63] was used to determine the presence of a clinical BPD diagnosis. In this semi-structured interview, a qualified clinician assessed whether participants displayed five or more DSM-IV borderline personality traits.

Statistical Analyses. All tests performed in this article are two-sided. For each study, scatter-plots that visualize the data distribution and relation between all affect dynamic measures and psychological well-being outcomes can be found online in Supplemental Materials (http://osf.io/zm6uw). Additionally, we provide the reader with 5 randomly drawn affective time series for PA and NA, giving further insight in the raw data structure of each dataset. Finally, we present all specific correlation coefficients and effect sizes, test-statistics, degrees of freedom, *p*-values, confidence intervals and standard errors of our meta-analysis.

Code and data availability. All analyses reported in this paper were conducted in MATLAB (R2017a), except our meta-analysis and visualisation of the correlational network, which were performed in R (3.4.0). A reproducible MATLAB and R-code, as well as two complementary datasets[17,79], are available online in Supplemental Materials (http://osf.io/zm6uw). The other datasets used in this article are available upon reasonable request from the original sources referenced in Table 2, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Empirical interdependencies among dynamic measures. In a first step to examine the empirical relations between all affect dynamic measures, we calculated *Pearson correlations*. Next, to get further insight in the communalities between measures, we performed *Principal Component Analysis*.

Unique explanatory power in the prediction of psychological well-being. In a first step to evaluate the predictive value of all affect dynamic measures, we conducted *hierarchical multiple regression*[VIII]. In these analyses, we predefined average levels of positive and negative affect as the primary covariates to find incremental value over. Thus, for each dataset, we investigated the explanatory power (a) of each measure alone, (b) when controlling for mean levels of PA and NA, and (c) when controlling for mean levels of, and SDs in PA and NA, in the prediction of people's depressive and borderline symptoms, and their satisfaction with life (see results for justification).

In the hierarchical multiple regression approach, we evaluated the average R^2 for each measure across datasets to get an indication of the effect size. We also performed a mixed-model *meta-analysis* with random slopes and intercepts, to test for significance.

In a series of secondary analyses, we did not define an a priori set of covariates (e.g., mean levels of, or SDs in, PA and NA), but used prediction-based variable selection to

[VIII] In this section, we only present the results of our *linear* regression analyses (i.e., with continuous psychological well-being outcomes). Results for our *logistic* regression analyses (i.e., with a categorical outcome in the form of a clinical MDD or BPD diagnosis) can be found in Appendix A. Linear and logistic results support the same conclusions.

determine the optimal combination of affect dynamic measures in the prediction of psychological well-being[64]. In this way, we aimed (a) to establish the best set of affect dynamic predictors to explain individual differences in people's life satisfaction, depressive and borderline symptomatology, and (b) to find converging evidence that mean levels of (and variance in) PA and NA are indeed the primary covariates to find incremental value over. For each dataset, we performed a *Lasso-analysis* (Least Absolute Shrinkage and Selection Operator)[65], in which we selected the most favorable set of affect dynamic predictors of well-being, penalized for the number of predictors to prevent overfitting. In total, we compared three different Lassos per dataset. The first Lasso could make an optimal selection out of all 16 dynamic measures. In the second and third Lasso, we forced certain coefficients to be set to zero, and pursued an optimal combination out of mean levels and SDs of positive and negative affect (i.e., 4 measures), or out of mean levels only (i.e., 2 measures), respectively. Next, we evaluated the predictive accuracy of each of these Lasso-selections using leave-one-out cross-validation (LOO)[66]. In this cross-validation procedure, we partitioned the sample of each dataset in a training and validation subset (with $n = 1$; i.e., leave-one-out). In the training subset, each of the three optimal Lasso-selected predictors sets were used to predict psychological well-being. In the validation subset, we determined the predictive accuracy of each of the three Lasso-based models by evaluating the squared error between the actual and fitted scores. We repeated this procedure until each participant once served as validation dataset, and averaged the squared error across all cross-validations. This yielded a mean squared error (MSE) for each of the three Lasso-based models per dataset. We divided each MSE by the total variance observed in the outcome, and then subtracted this value from 1.

This produced a predicted R^2 for each of the Lasso-based models, approximately ranging from 0 to 1.00. The model with the highest predicted R^2 showed the best predictive accuracy[IX].

[IX] We performed similar analyses using *random forests*[67], a machine learning technique that is equally capable of dealing with many predictor variables, and inherently corrects for overfitting. This black-box supervised learning technique additionally models non-linear relations. Yet, these models did not outperform the less complex linear models with a Lasso-penalty.

Acknowledgements

This research was supported by the Research Fund of KU Leuven (GOA/15/003; OT/11/031). Merijn Mestdagh and Marlies Houben are supported by the Fund of Scientific Research Flanders. We sincerely thank the following researchers that provided data for this project: Annette Brose, Brock Bastian, Ian Gotlib, John Jonides, Elise Kalokerinos, Peter Koval, Ulman Lindenberger, Martin Lövdén, Madeline Pe, Florian Schmiedek, Renee Thompson, Tim Trull, and Kathleen Van der Gucht. This article uses data from the COGITO study, supported by a grant from the Innovation Fund of the President of the Max Planck Society to Ulman Lindenberger. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government - department EWI.

Author Contributions

E.D. and M.M. contributed equally to this project. M.M. performed the analyses and E.D. drafted the manuscript. Both authors conceptualized the study project and interpreted the results under P.K and F.T.'s supervision. I.R. independently re-analysed parts of the data with different statistical software to achieve converging results. M.H. and L.S. critically revised the manuscript. All authors approved the final version of the article.

Conflict of Interest

The authors declare no conflict of interest with respect to the authorship or the publication of this article.

Materials and Correspondence

Correspondence concerning this article and material requests should be addressed to Egon Dejonckheere, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Leuven, 3000, Belgium. E-mail: egon.dejonckheere@kuleuven.be.

Tables

Table 1. Overview of the most commonly studied affect dynamic measures.

Measure (abbreviation)	Affect dynamic feature	Substantive description	Mathematical description	Composite score?	Time reliant?	Example references
1. Mean PA and NA (M)	Average levels of emotion	Captures one's average level of positive or negative affect.	The sum of all affect intensity ratings divided by the total number of completed measurement occasions.	Yes	No	22,24,68
2. Variance or standard deviation PA and NA (SD)	Emotional variability	Captures the average emotional deviation from one's mean levels of positive or negative affect.	The sum of all squared differences between a particular affect intensity rating and the mean level of that affective state, divided by the total number of completed measurement occasions. The square root of the variance produces an SD for that affective state.	Yes	No	25, 28
3. Relative variance or standard deviation PA and NA (SD*)	Relative emotional variability	Captures the average emotional deviation from one's mean levels of positive or negative affect, taking into account the maximum possible variability given the mean of that affective state.	The SD of an affective state divided by the maximum possible SD of that affective state, given a certain mean level of that affective state.	Yes	No	20,69,70
4. MSSD PA and NA (MSSD)	Emotional instability	Captures the average change in emotional intensity between two successive measurement occasions for positive or negative affect.	The mean of all squared differences between two successive intensity ratings of an affective state. The square root of this measure produces the MSSD for that affective state.	Yes	Yes	16, 35,71
5. Autoregression PA and NA (AR)	Emotional inertia	Captures the degree to which positive or negative affect carries over from one moment to the next, is self-predictive, and resistant to change.	The person-specific (within-person centered) autoregressive slope in a multilevel AR(1) model, in which the intensity rating of an affective state at time *t*-1 predicts the intensity rating of that state at time *t*.	Yes	Yes	10,12,44
6. Emotion-network density (D)	Emotional interdependency across time	Captures the degree to which various positive and negative emotions predict each other over time, reflecting how one's entire emotional system is resistant to change.	The mean of all absolute person-specific (within-person centered) auto- and cross-regressive slopes in a series of multilevel VAR(1) models. In each model, the intensity of one emotion rating at time *t* is predicted once by the intensity ratings of all other emotions at time *t*-1, including the intensity rating of that emotion itself.	No, discrete emotions	Yes	40,51
7. ICC PA and NA (ICC)	Emotional granularity or differentiation	Captures one's ability to differentiate between various positive or negative discrete emotions.	The intra-class correlation between various same-valenced emotion intensity ratings. This measure reflects the degree to which different emotion intensity ratings converge. A low ICC reflects high emotion differentiation.	No, discrete same-valenced emotions	No	53,72,73
8. PA-NA correlation (ρ)	Affective bipolarity, valence focus, or emotional dialecticism	Captures the degree to which one experiences positive and negative affect rather independent, or as bipolar opposites.	The within-person correlation between positive and negative affect.	Yes	No	15, 74, 75
9. Gini-coefficient PA and NA (G)	Emodiversity	Captures the variety of one's emotional repertoire for positive and negative emotions.	The weighted sum of the frequencies of various same-valenced emotions (with the weight being the order of the emotion frequencies), divided by the product of the total frequency of all same-valenced emotions and the total number of emotion categories. A low Gini-coefficient reflects a high emodiversity.	No, discrete same-valenced emotions	No	21,41,76

Table 2. Overview of the datasets used in the current study.

																		Outcomes				
	Sample characteristics			Protocol characteristics					Predictor characteristics									Depressive symptoms		Borderline symptoms		Life satisfaction
Dataset	*N* participants	% Female	*M* age	Type	Occasions / day	*N* days	Max. occasions	*M* compliance	Items PA	Items NA	ICC PA	ICC NA	α PA $_{item}$	α PA $_{person}$	α NA $_{item}$	α NA $_{person}$	Scale	Classification	Regression	Classification	Regression	Regression
1. Dejonckheere et al. (2017)[77]	94	51%	34.12	Daily diary	1	30	30	95.04%	happy, relaxed	sad, anxious, depressed, angry, stressed	39.86%	47.81%	.56	.95	.64	.96	1 – 7		CES-D, PHQ-9			SWL
2. Dejonckheere et al. (*in preparation*)[78]	118	66%	35.69	ESM	10	7	70	87.11%	happy, relaxed, euphoric	anxious, depressed, angry, stressed	49.85%	70.87%	.31	.98	.59	.99	0 – 100	SCID-I MDD	Q-IDS	SCID-II BPD	ADP-IV BPD, PAI-BOR	SWL
3. Houben et al. (2016)[26]	50	84%	29.96	ESM	10	8	80	79.86%	relaxed, cheerful	anxious, depressed, angry, stressed	58.39%	70.12%	.61	.99	.53	.99	0 – 100				ADP-IV BPD	
4. Kalokerinos et al. (*in preparation*)[79]	100	77%	24.12	ESM	7	14	98	88.80%	happy, relaxed	sad, angry, stressed	25.18%	29.59%	.70	.97	.42	.97	0 – 100		CES-D, BDI-II			SWL
5. Koval et al. (2013)[17]	95	62%	19.06	ESM	10	7	70	90.34%	happy, relaxed	sad, anxious, depressed, angry, stressed	34.19%	48.73%	.65	.97	.63	.98	1 – 100		CES-D			SWL
6. Pe et al. (2016)[80] – Wave 1	200	55%	18.32	ESM	10	7	70	87.27%	happy, relaxed, cheerful	sad, anxious, depressed, angry, stressed	24.99%	38.93%	.68	.98	.63	.98	0 – 100		CES-D		ADP-IV BPD	SWL
7. Pe et al. (2016)[80] – Wave 2	190	56%	18.64	ESM	10	7	70	87.87%	happy, relaxed, cheerful	sad, anxious, depressed, angry, stressed	32.93%	43.00%	.57	.97	.65	.98	0 – 100		CES-D		ADP-IV BPD	SWL
8. Pe et al. (2016)[80] – Wave 3	176	55%	19.28	ESM	10	7	70	88.57%	happy, relaxed, cheerful	sad, anxious, depressed, angry, stressed	32.28%	40.49%	.55	.97	.63	.98	0 – 100		CES-D		ADP-IV BPD	SWL
9. Provenzano et al. (*in preparation*)[81]	34	59%	23.85	ESM	10	7	70	87.35%	happy, relaxed, excited, content	sad, anxious, angry, stressed	29.38%	36.34%	.82	.96	.39	.97	0 – 100		CES-D, Q-IDS			
10. Schmiedek et al. (2010)[82]	175	50%	48.01	Daily diary	1	±101	±101	not applicable	happy, excited, alert, attentive, interested	distressed, upset, scared, afraid, nervous	79.45%	63.87%	.46	1.00	.53	.99	0 – 7		CES-D			SWL
11. Sels et al. (2017)[83]	100	50%	27.75	ESM	10	8	80	80.14%	happy, relaxed, cheerful, content	sad, anxious, depressed, angry, stressed	39.61%	39.95%	.81	.98	.65	.98	0 – 100		CES-D			SWL
12. Sels et al. (2018)[84]	185	50%	26.30	ESM	6-14*	8	64	92.94%	happy, relaxed	sad, anxious, angry	29.21%	31.13%	.76	.96	.62	.96	0 – 100		CES-D			SWL
13. Thompson et al. (2012)[19]	83			ESM	±8	±7	±56	not applicable	happy, excited, alert, active	sad, anxious, angry, guilty, frustrated	45.34%	59.19%	.64	.97	.67	.98	1 – 4	SCID-I MDD				
14. Trull et al. (2008)[16]	131			ESM	±5	±28	±140	not applicable	excited, alert, attentive, interested, enthusiastic	distressed, upset, scared, afraid, nervous	39.81%	48.45%	.75	.99	.75	.99	1 – 5	SCID-I MDD		SCID-II BPD		
15. Van der Gucht et al. (2017)[18]	46	76%	41.96	ESM	10	4	40	87.39%	happy, relaxed, cheerful, content, calm	sad, anxious, depressed, angry, stressed	46.12%	40.66%	.79	.97	.54	.96	0 – 100					SWL

Note. ± some participants diverged from the original protocol, leaving a variable number of measurement occasions per individual; * participants were sampled 14 times during the weekend, yet the time interval between two consecutive occasions remained identical; ICCs reflect the amount of variance explained by between-person differences; Internal consistencies for PA and NA were calculated following[85].

Figures

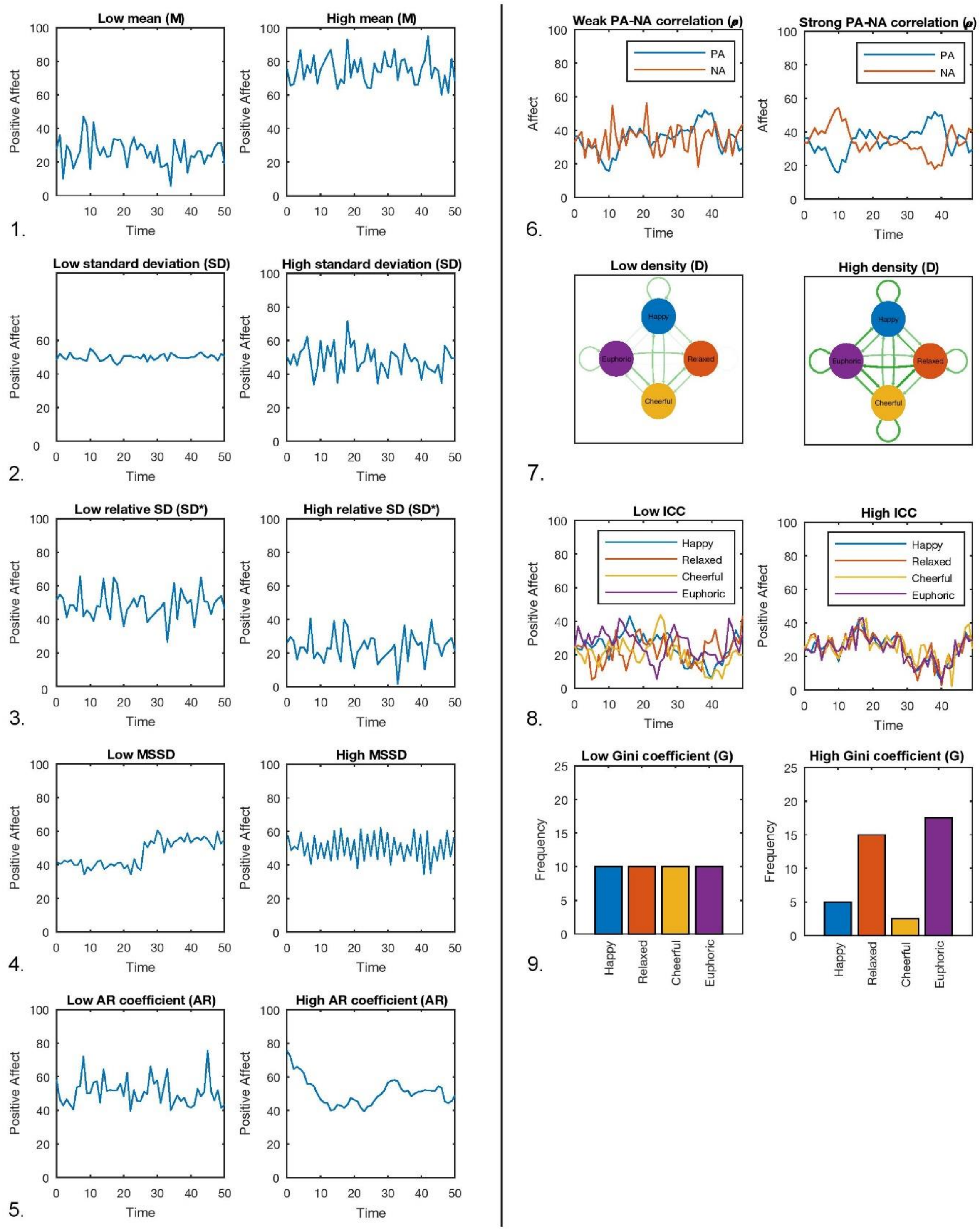

Figure 1. Simulated data for each affect dynamic measure, showing how low and high values for each measure would manifest in a person's affective time series.

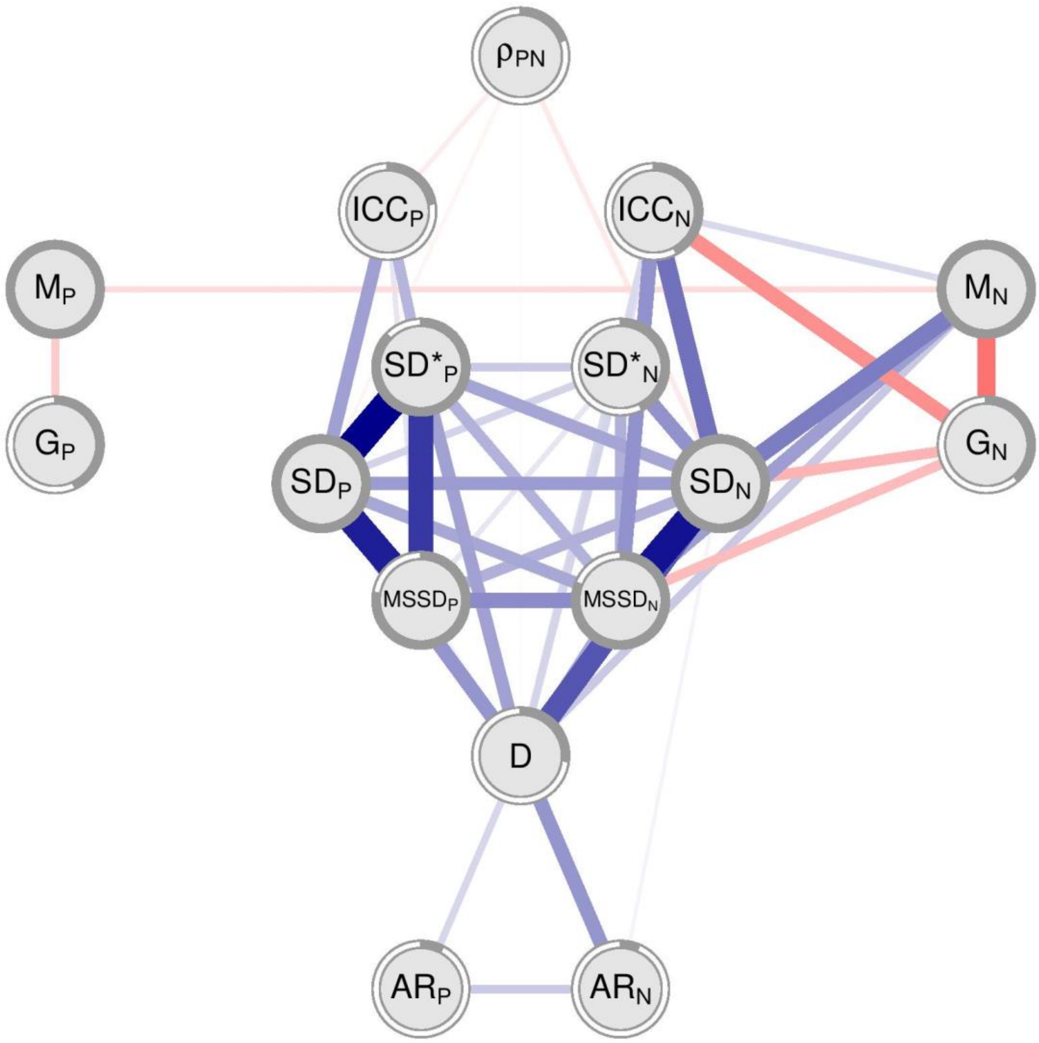

Figure 2. Empirical network of all affect dynamic measures based on their pairwise correlations. Blue lines represent positive relations, red lines indicate negative ones. Edge thickness and transparency correspond with the degree of association. For clarity, correlations below |.30| are not shown. The grey color in the ring around each node depicts how much variance PA and NA mean levels and SDs explain in each affect dynamic measure. See Table 1 for affect dynamic short-codes. Subscripts refer to positive (P) or negative affect (N).

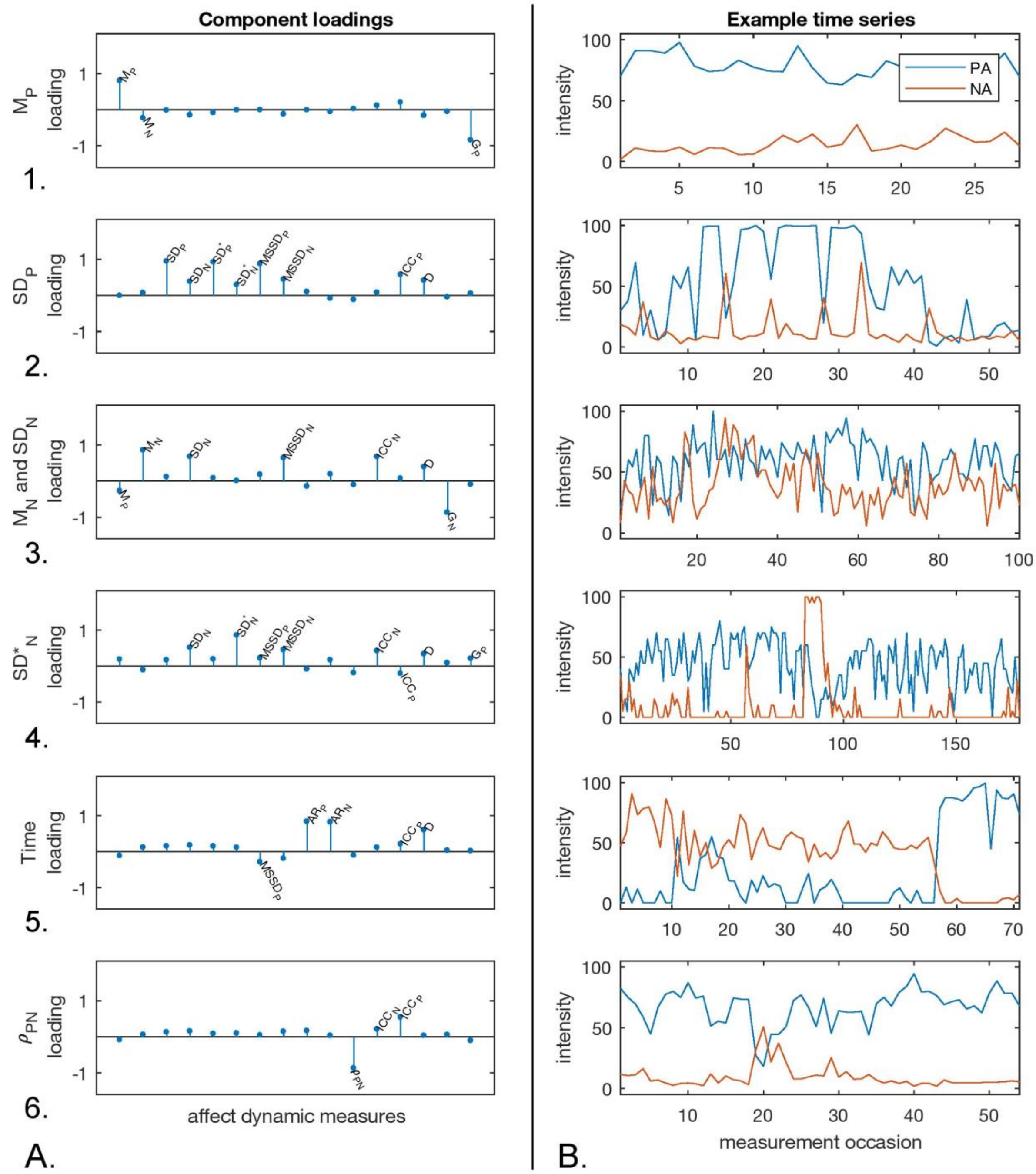

Figure 3. The six components retrieved from a PCA on all affect dynamic measure (Panel A). For clarity, component loadings below |.20| are not shown. Each component is accompanied by an empirical example that displays the affective time series of an individual in the highest percentile of that component (Panel B). See Table 1 for affect dynamic short-codes.

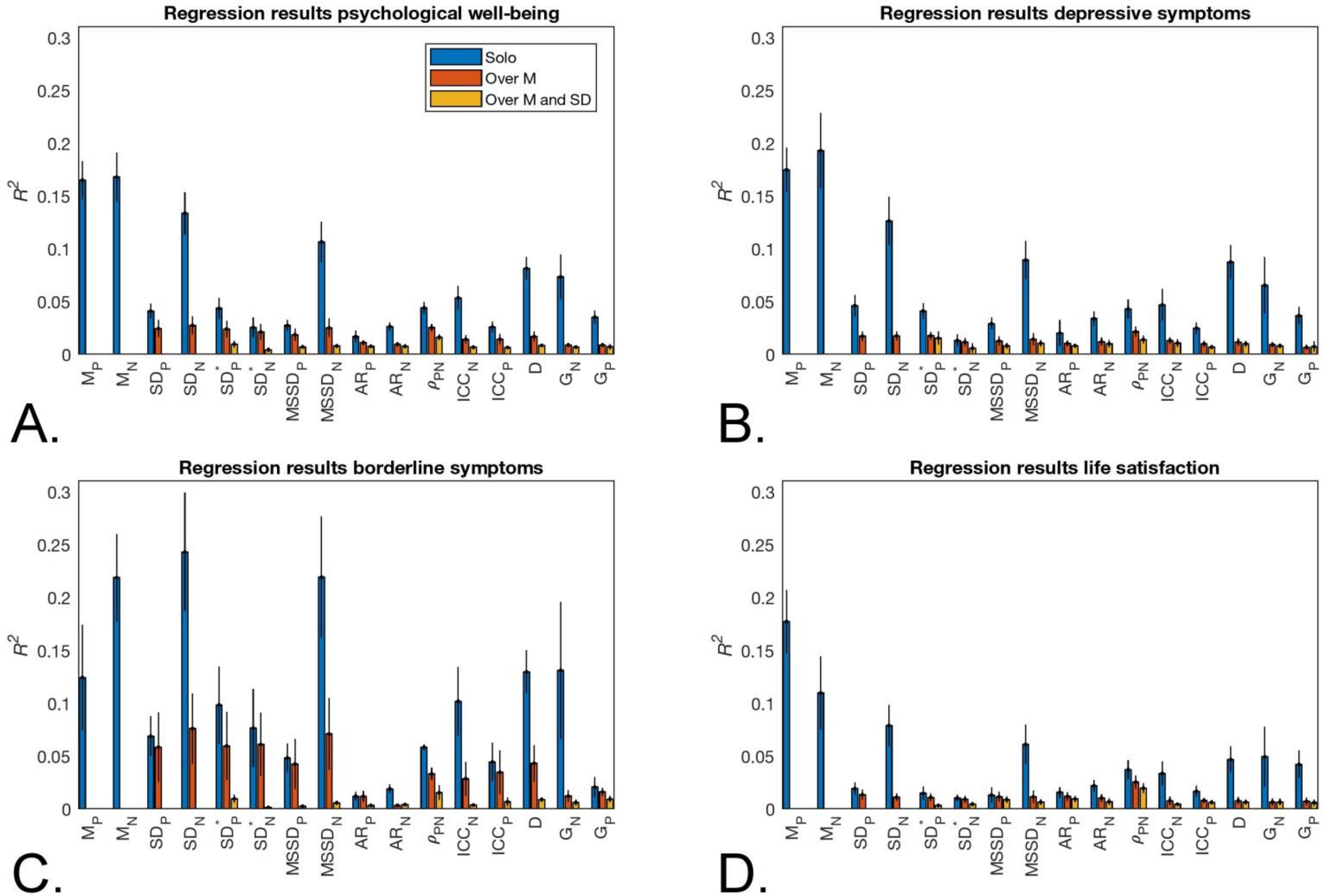

Figure 4. The bars in each plot visualize the average R^2 for all affect dynamic measures across studies in the prediction of a combined measure of psychological well-being (Panel A), depressive symptoms (Panel B), borderline symptoms (Panel C), or life satisfaction (Panel D). Blue, red, and yellow bars reflect the average R^2 for each measure alone, when controlling for mean levels of PA and NA, and when controlling for mean levels of, and SDs in PA and NA, respectively. The error bars reflect the standard error around the average R^2. See Table 1 for affect dynamic short-codes.

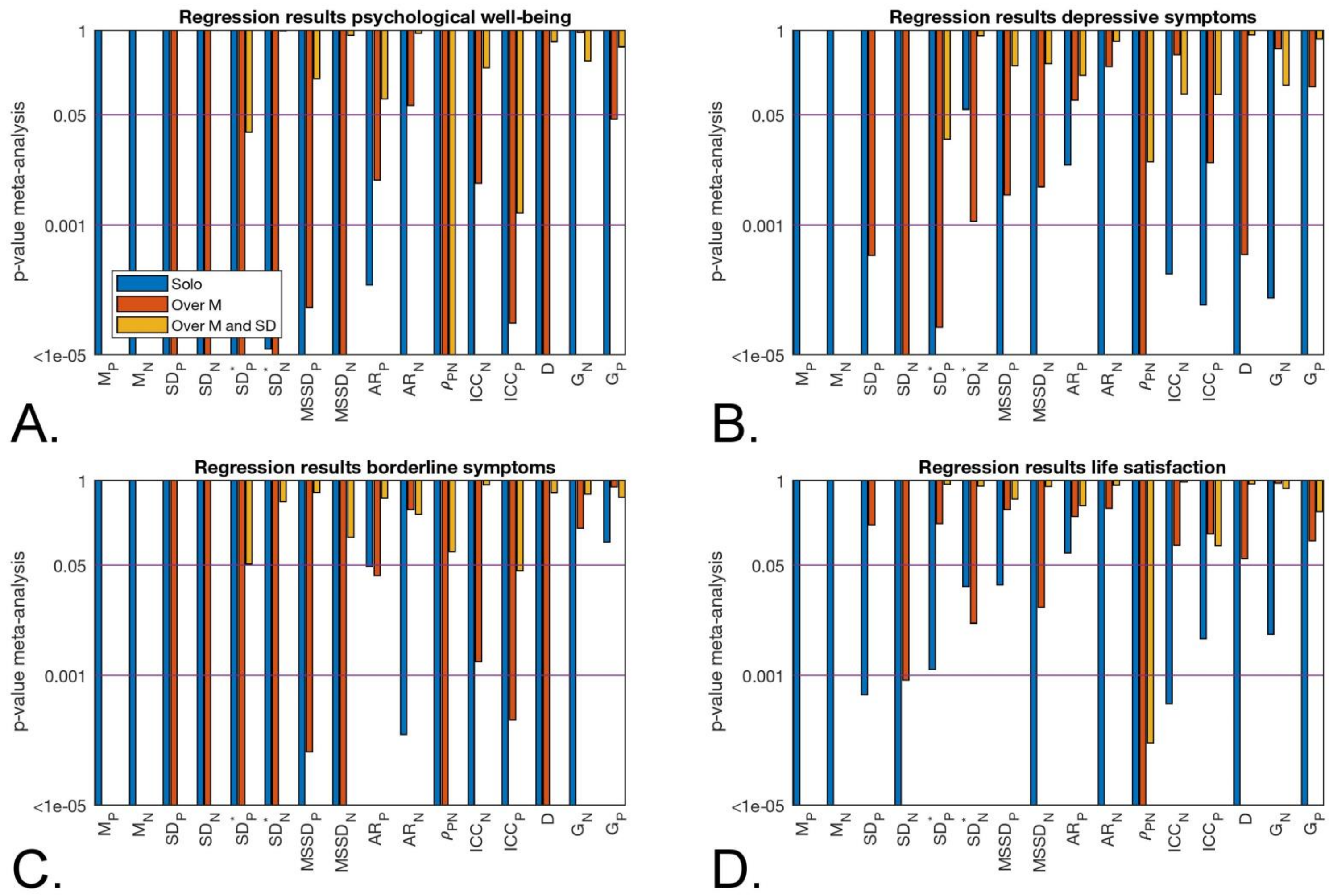

Figure 5. The bars in each plot indicate the meta-analytic *p*-value for each affect dynamic measure in the prediction of a combined measure of psychological well-being (Panel A), depressive symptoms (Panel B), borderline symptoms (Panel C), or life satisfaction (Panel D). Blue, red, and yellow bars reflect the level of significance for each measure alone, when controlling for mean levels of PA and NA, and when controlling for mean levels of, and SDs in PA and NA, respectively. All tests were two-sided. See Table 1 for affect dynamic short-codes.

References

1. Davidson, R. J., Jackson, D. C., & Kalin, N. H. Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. *Psychological Bulletin*, *126,* 890-909 (2000). doi:10.1037/0033-2909.126.6.890

2. Frijda, N. H. *The laws of emotion* (NJ: Erlbaum, Hillsdale, 2007).

3. Kuppens, P. It's about time: A special section on affect dynamics. Emotion Review, 7, 297–300 (2015). doi:10.1177/1754073915590947

4. Larsen, R. J. Towards a science of mood regulation. *Psychological Inquiry, 11*, 129–141 (2000). doi:10.1207/S15327965PLI1103_01

5. Scherer, K. R. The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*, 1307–1351 (2009). doi:10.1080/02699930902928969

6. Kuppens, P., Oravecz, Z., & Tuerlinckx, F Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*, 1042–1060 (2010). doi:0.1037/a0020962

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

8. Scherer, K. R. Emotion. In M. Hewstone & W. Stroebe (Eds.), *Introduction to social psychology: A European perspective* (3rd ed., pp. 151–191) (Blackwell, Oxford, England, 2000).

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

10. Koval, P., Sütterlin, S., Kuppens, P. Emotional inertia is associated with lower well-being when controlling for differences in emotional context. *Frontiers in Psychology, 6,* 1–11 (2015). doi:10.3389/fpsyg.2015.01997

11. Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J. G., & Allen, N. B. Emotional inertia prospectively predicts the onset of depressive disorder in adolescence. *Emotion*, *12*, 283–289 (2012). doi:10.1037/a0025046

12. van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., … Scheffer, M. Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111, 87–92 (2014). doi:10.1073/pnas.1312114110

13. Trull, T. J., Lane, S. P., Koval, P., Ebner-Priemer, U. W. Affective Dynamics in Psychopathology. *Emotion Review, 7,* 355–361 (2015). doi:10.1177/1754073915590617

14. Grühn, D., Lumley, M. A., Diehl, M., & Labouvie-vief, G. Time-based indicators of emotional complexity: Interrelations and correlates. *Emotion*, *13*, 226–237 (2013). doi:10.1037/a0030363.Time-Based

15. Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., Brose, A., Bastian, B., & Kuppens, P. The bipolarity of affect and depressive symptoms. Journal of Personality and Social Psychology, 114, 323–341 (2018). doi:10.1037/pspp000018

16. Jahng, S., Wood, P. K., & Trull, T. J. Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*, 354–375 (2008). doi:10.1037/a0014173.

17. Koval, P., Pe, M. L., Meers, K., & Kuppens, P. Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*, 1132–1141 (2013). doi:10.1037/a0033579

18. Van der Gucht, K., Dejonckheere, E., Erbas, Y., Takano, K., Vandemoortele, M., Maex, E., Raes, F., & Kuppens, P. An experience sampling study examining the potential impact of a mindfulness-based intervention on emotion differentiation. *Emotion* (2018). Advance online publication. doi:10.1037/emo0000406

19. Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. The everyday emotional experience of adults with major depressive disorder: examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology, 121*, 819-829 (2012). doi:10.1037/a0027978

20. Mestdagh, M., Pe, M. L., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods* (*in press*).

21. Brown, N. J. L., & Coyne, J. C. Emodiversity: Robust predictor of outcomes or statistical artifact? *Journal of Experimental Psychology: General, 146,* 1372–1377 (2017). doi:10.1037/xge0000330

22. Clark, L. A., & Watson, D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316–336 (1991). doi:10.1037/0021-843X.100.3.316

23. von Mises, R. *Mathematical theory and probability and statistics*. (Academic Press, New York, NY, 1964).

24. Diener, E., Sandvik, E., & Pavot, W. Happiness is the frequency, not the intensity, of positive versus negative affect. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being: An interdisciplinary perspective* (pp. 119-139) (Pergamon, New York, NY, 1991).

25. Eid, M., & Diener, E. Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76*, 662–676 (1999). doi:10.1037/0022-3514.76.4.662

26. Houben, M., Vansteelandt, K., Claes, L., Sienaert, P., Berens, A., Sleuwaegen, E., & Kuppens, P. Emotional switching in borderline personality disorder: A daily life study. *Personality Disorders, 7*, 50–60 (2016). doi:10.1037/per0000126

27. Moskowitz, D. S., & Zuroff, D. C. Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology*, *86*, 880–893 (2004). doi:10.1037/0022-3514.86.6.880

28. Kuppens, P., & Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion, 2*, 262–274 (2007). doi:10.1037/1528-3542.7.2.262

29. Csikszentmihalyi, M., & Larson, R. Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disorders*, *175*, 526–536 (1987). doi:10.1097/00005053-198709

30. Bolger, N., Davis, A., & Rafaeli, E. Diary methods: capturing life as it is lived. *Annual Review of Psychology*, *54,* 579–616 (2003). doi:10.1146/annurev.psych.54.101601.145030

31. Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*, 276–302 (1999). doi:10.1037/0033-2909.125.2.276

32. Waterman, A. S. Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology, 64*, 678–691(1993). doi:10.1037/0022-3514.64.4.678

33. Vuillier, L., Brooks, A. W., Gruber, J., Sun, R., Norton, M. I., Samson, M. J., … Keltner, D. Amount and diversity of digital emotional expression predicts happiness. *Harvard Business School*, *18,* 2–42 (2018).

34. Davis, M. C., Zautra, A. J., & Smith, B. Chronic pain, stress, and the dynamics of affective differentiation. *Journal of Personality*, *72*, 1133–1159 (2004). doi:10.1111/j.1467-6494.2004.00293.x

35. Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology, 117,* 647–661 (2008). doi:10.1037/a0012532

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

37. Coifman, K. G., Berenson, K. R., Rafaeli, E., & Downey, G. From negative to positive and back again: Polarized affective and relational experience in borderline personality disorder. *Journal of Abnormal Psychology, 121*, 668–679 (2012). doi:10.1037/a0028502000-00004

38. Demiralp, E., Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Barrett, L. F., … Jonides, J. Feeling blue or turquoise? Emotional differentiation in Major Depressive Disorder. *Psychological Science, 23*, 1410–1416 (2012). doi:10.1177/0956797612444903

39. Wolff, S., Stiglmayr, C., Bretz, H. J., Lammers, C. H., & Auckenthaler, A. Emotion identification and tension in female patients with borderline personality disorder. *The British Journal of Psychology, 46*, 347–360 (2007). doi:10.1348/014466507X17373

40. Pe, M. L., Kircanski, K., Thompson, R.J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., … Gotlib, I. H. Emotion-network density in major depressive disorder. *Clinical Psychological Science, 3*, 292–300 (2015).

41. Quoidbach, J., Gruber, J., Mikolajczak, M., Kogan, A., Kotsou, I., & Norton, M. I. Emodiversity and the emotional ecosystem. *Journal of Experimental Psychology: General, 143*, 2057–2066 (2014). doi:10.1037/a0038025

42. Erbas, Y., Ceulemans, E., Koval, P., & Kuppens, P. The role of valence focus and appraisal overlap in emotion differentiation. *Emotion, 15,* 373–382 (2015). doi:10.1037/emo0000039

43. Shmueli, G. To explain or to predict? *Statistical Science, 25*, 289–310 (2010). doi:10.1214/10-STS330

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

45. Linehan, M. *Cognitive behavioral treatment of borderline personality disorder* (Guilford Press, New York, NY, 1993).

46. Ebner-Priemer, U. W., Houben, M., Santangelo, P., Kleindienst, N., Tuerlinckx, F., Oravecz, Z., . . . Kuppens, P. Unraveling affective dysregulation in borderline personality disorder: A theoretical model and empirical evidence. *Journal of Abnormal Psychology, 124*, 186–198 (2015). doi:10.1037/abn0000021

47. Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods* (2018). Advance online publication. doi:10.1037/met0000178

48. Fried, E. I. Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology, 6,* 1–11. (2015). doi:10.3389/fpsyg.2015.00309

49. Fried, E. I., & Nesse, R. M. Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *Current Controversies in Psychiatry, 13,* 1–11. (2015). doi:10.1186/s12916-015-0325-4

50. Erbas, Y., Ceulemans, E., Kalokerinos, E., Houben, M., Koval, P., Pe, M. L., & Kuppens, P. Why I don't always know what I'm feeling: The role of stress in within-person fluctuations in emotion differentiation. *Journal of Personality and Social Psychology* (*in press*).

51. Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. A network approach to psychopathology: New insights into clinical longitudinal data. *Plos One*, *8*, e60188, 1–13 (2013). doi:10.1371/journal.pone.0060188

52. Ong, A. D., Zautra, A. J., & Finan, P. H. Inter- and intra-individual variation in emotional complexity: Methodological considerations and theoretical implications. *Current Opinion in Behavioral Sciences*, *15*, 22–26 (2017). doi:10.1016/j.cobeha.2017.05.018

53. Erbas, Y., Ceulemans, E., Pe, M. L., Koval, P., & Kuppens, P. Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition and Emotion, 28,* 1196–1213 (2014). doi:10.1080/02699931.2013.875890

54. Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75 (1985). doi:10.1207/s15327752jpa4901_13

55. Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 384–401 (1977). doi:10.1177/014662167700100306

56. Kroenke, K., Spitzer, R. L., &Williams, J. B. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine,16*, 606–613 (2001).

57. Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C., ... & Fava, M. An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. Biological Psychiatry, 59(6), 493-501 (2006). DOI: 10.1016/j.biopsych.2005.08.022

58. Beck, A. T., Steer, R. A., & Brown, G. K. *Manual for the Beck Depression Inventory–II.* (Psychological Corporation, San Antonio, TX, 1996).

59. First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B.W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)* (Biometrics Research, New York, NY, 2002).

60. Schotte, C. K. W., de Doncker, D., Vankerckhoven, C., Vertommen, H., & Cosyns, P. Self-report assessment of the DSM–IV personality disorders. Measurement of trait and distress characteristics: The ADPIV. *Psychological Medicine, 28*, 1179–1188 (1998). doi:10.1017/S0033291798007041

61. Distel, M. A., de Moor, H. M., & Boomsma, D. I. Dutch translation of the Personality Assessment Inventory Borderline features scale (PAI-BOR): norms, factor structure and reliability, *Psychology and Health, 37,* 38–46 (2009).

62. Morey, L.C. *The Personality Assessment Inventory: Professional manual*. (Psychological Assessment Resources , Odessa, FL, 1991).

63. First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II).* (American Psychiatric Press, Washington, D.C., 1997).

64. Meinshausen, N., & Bühlmann, P. Stability selection. *Royal Statistical Society*, *72*, 417–473 (2010). doi:10.1111/j.1467-9868.2010.00740.x

65. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, *58*, 267–288 (1996).

66. Geisser, S. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*, 320–328 (1975). doi: 10.2307/2285815

67. Ho, T. K. Random decision forest. *Proceedings of the 3rd international conference on document analysis and recognition*, Montreal, QC: 278–282 (1995).

68. 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. doi:10.1037/0022-3514.54.6.1063

69. Baird, B. M., Le, K., & Lucas, R. E. On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90*, 512–527 (2006). doi:10.1037/0022-3514.90.3.512

70. Kalmijn, W., & Veenhoven, R. Measuring inequality of happiness in nations: In search for proper statistics. *Journal of Happiness Studies, 6*, 357–396 (2005). doi:10.1007/s10902-005-8855-7

71. Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. Analytic strategies for understanding affective (in)stability and other dynamic processes in

psychopathology. *Journal of Abnormal Psychology, 118,* 195–202 (2009). doi:10.1037/a0014868.

72. Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion, 15,* 713–724 (2001). doi:10.1080/02699930143000239

73. Kashdan, T. B., Barrett, L. F., & McKnight, P. E. Unpacking emotion differentiation transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science, 24,* 10–16 (2015). doi:10.1177/0963721414550708

74. Feldman, L. A. Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology, 69,* 153–166 (1995). doi:10.1037/0022-3514.69.1.153

75. Lindquist, K. A., & Barrett, L. F. Emotional complexity. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *The handbook of Emotions* (pp. 1–31). (The Guilford press, New York, NY, 2008).

76. Benson, L., Ram, N., Almeida, D., Zautra, A., & Ong, A. D. Fusing biodiversity metrics into investigations of daily life: Illustrations and recommendations with emodiversity. *Journal of Gerontology: Psychological* Sciences, *73*, 74–86 (2018). doi:10.1093/geronb/gbx025

77. Dejonckheere, E., Bastian, B., Fried, E. I., Murphy, S., & Kuppens, P. Perceiving social pressure not to feel negative predicts depressive symptoms in daily life. *Depression and Anxiety, 34*, 836–844 (2017). doi:10.1002/da.22653

78. Dejonckheere, E., Houben, M., Bastian, B., & Kuppens, P. The role of social expectancies not to feel negative in Major Depressive Disorder (*in preparation*).

79. Kalokerinos, E. K., Tamir, M., & Kuppens, P. Motivated emotion regulation in daily life (*in preparation*).

80. Pe, M. L., Brose, A., Gotlib, I. H., & Kuppens, P. Affective updating ability and stressful events interact to prospectively predict increases in depressive symptoms over time. *Emotion*, *16*, 73–82 (2016). doi:10.1037/emo0000097

81. Provenzano, J., Dejonckheere, D., Verduyn, P., & Kuppens, P. The Role of trait and state intrinsic functional connectivity in emotional flexibility (*in preperation*).

82. Schmiedek, F., Lövdén, M., Lindenberger, U. Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience, 2,* 1–27 (2010). doi:10.3389/fnagi.2010.00027

83. Sels, L., Ceulemans, E., & Kuppens, P. Partner-expected affect: How you feel now is predicted by how your partner thought you felt before. *Emotion, 17,* 1066–1077 (2017). doi:10.1037/emo0000304

84. Sels, L., Cabrieto, J., Butler, E., Reis, H., Ceulemans, E., & Kuppens, P *The occurrence and correlates of emotional Interdependence in romantic relationships.* Manuscript submitted for publication (2018).

85. Nezlek, J. B. A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality* (2016). doi:10.1016/j.jrp.2016.06.020

86. Robert, G. P. J., & Kangping, S. The total operating characteristic to measure diagnostic ability for multiple thresholds. *International Journal of Geographical Information Science*, 28, 570–583 (2014). doi:10.1080/13658816.2013.862623

87. Mason, S. J., & Graham, N. E. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society, 128,* 2145–2166 (2006). doi:10.1256/003590002320603584

Appendix A

We also aimed to investigate the unique explanatory power of all affect dynamic measures in the prediction of categorical well-being outcomes, which were operationalized as the presence or absence of either a clinical MDD or BPD diagnosis. To this end, we performed similar analyses as described in the main text, but now in a logistic regression context.

[Appendix Figure 1 around here]

***Hierarchical multiple regression*.** To evaluate the explanatory power of all affect dynamic measures in the prediction of psychopathological diagnoses, we performed hierarchical multiple logistic regression. Since outcomes were categorical, we estimated Receiver Operating Characteristic (ROC)[86] curves for each logistic model with (a) each affect dynamic measure alone, (b) when controlling for mean levels of PA and NA, and (c) when controlling for mean levels of, and SDs in, PA and NA. For all affect dynamic measures, we then evaluated the Area Under the Curve (AUC)[87], which typically ranges between .00 and 1.00, and where values higher than .50 indicate a correct classification better than chance.

Panel A of Figure 1A visualizes the average AUC for each affect dynamic measure across the different datasets. Individually, most measures are able to correctly differentiate between the presence and absence of a clinical MDD or BPD diagnosis, with the highest predictive accuracies for M and SD negative affect (AUC's $\geq$ .81), and lowest accuracies for both AR measures (AUC's $\leq$ .53). However, when controlling for average levels of positive affect, all affect dynamic measures add little to an accurate prediction of both psychopathological diagnoses (Δ AUC's $\leq$ .06). When additionally taking into account the explanatory power of both SDs, the added predictive accuracy of all measures even becomes practically non-existent or worse (Δ AUC's $\leq$ .01).

Meta-analysis. Panel B of Figure 1A presents the meta-analytic *p*-values for each affect dynamic measure in the prediction of a clinical diagnosis. We summarize the most important findings below.

Individually, considerable differences between measures appear in terms of significantly predicting the presence of an MDD or BPD diagnosis. Both M positive and negative affect show a meaningful contribution (p's $\leq .032$). For the SDs, however, we only observe a significant contribution for the SD in negative ($p = .014$), but not positive affect ($p = .819$). A similar distinction in valence is apparent for other affect dynamic measures, with a significant contribution for MSSD ($p = .021$), ICC ($p = 4\times10^{-04}$), and G measures ($p = .001$) in NA, but not in PA (p's $\geq .243$).

Next, when taking into account the explanatory power of mean level of PA and NA, these differences decrease considerably. Only G in NA ($p = .022$) shows a significant contribution above mean levels of affect in the prediction of clinical diagnoses.

Finally, when additionally controlling for SDs in PA and NA, no affect dynamic measure surpasses the $\alpha = .05$ significance level (p's $\geq .143$).

Optimal combination of predictors using Lasso variable selection. In a last step, we determined the best set of affect dynamic measures to predict clinical diagnoses, without an a priori selection of covariates. We used the same procedure as described in the main text of this article, comparing the predictive accuracy of three Lasso-based optimal predictor sets.

All three Lasso-based models showed an equal average predictive accuracy across all datasets (Lasso $_{\text{MEAN}}$ LOO correct = 79%, Lasso $_{\text{MEAN + SD}}$ LOO correct = 83%, Lasso $_{\text{ALL}}$ LOO correct = 84%). This illustrates again that a combination of mean levels of PA and NA (and variance in these affective states) is sufficient to predict the absence or presence of a clinical MDD or BPD diagnosis.

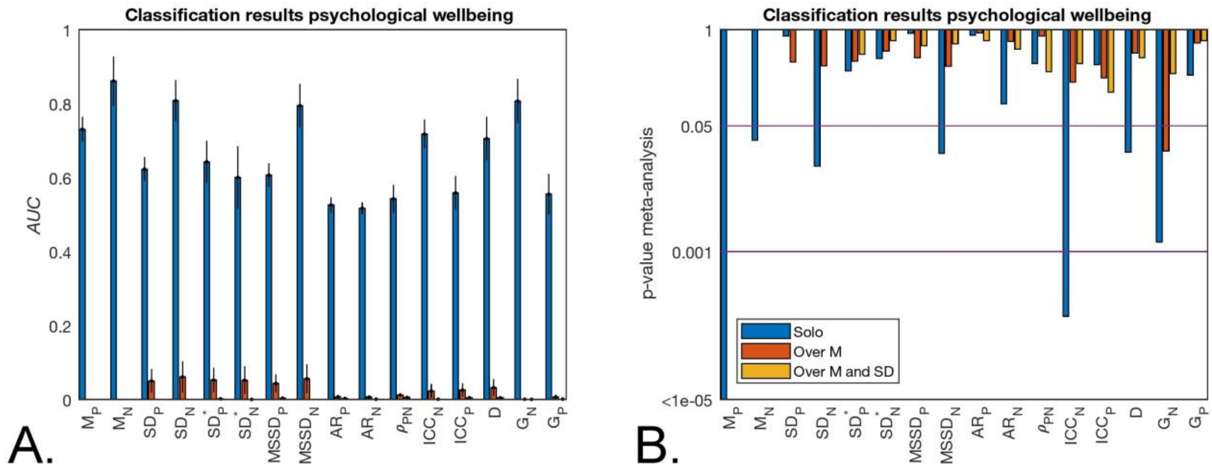

Figure 1 – Appendix A. The bars in each plot visualize the AUC (Panel A), and the meta-analytic *p*-value (Panel B) for each affect dynamic measure in the logistic prediction of an MDD or BPD diagnosis. Blue, red, and yellow bars reflect models with each measure alone, when controlling for mean levels of PA and NA, and when controlling for mean levels of, and SDs in PA and NA, respectively. The error bars reflect the standard error around the average R^2. All tests were two-sided. See Table 1 in the main text for affect dynamic short-codes.