Journal of Contextual Behavioral Science 32 (2024) 100728

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

journal homepage: www.elsevier.com/locate/jcbs

Testing the applicability of idionomic statistics in longitudinal studies: The example of 'doing what matters' ☆

Baljinder K. Sahdra [a,*], Joseph Ciarrochi [a], Korena S. Klimczak [b], Jennifer Krafft [c], Steven C. Hayes [d], Michael Levin [b]

[a] Institute for Positive Psychology and Education, Australian Catholic University, North Sydney, Australia
[b] Department of Psychology, Utah State University, USA
[c] Department of Psychology, Mississippi State University, USA
[d] Behavior Analysis Program, Department of Psychology, University of Nevada, Reno, USA

ARTICLE INFO

Keywords:
Values
Affect
Idionomic analysis
Ecological momentary assessment
Time series methods
Network analysis

ABSTRACT

This study evaluated idionomic methods for identifying within-person links between therapeutically relevant processes and outcomes, using an ecological momentary assessment dataset of valued action and hedonic well-being (participants (n) = 425; 71.76% female; age = $M(SD)$ = 22.20 (6.85); sampling design: 3–4 prompts per day; total measurements (n) = 6456). We compared the idionomic approach, integrating idiographic and nomothetic insights, with traditional multilevel modeling (MLM). Our methods included idiographic autoregressive integrative moving average models with an exogenous variable (i-ARIMAX), multivariate random-effects meta-analysis (RE-MA), deep Gaussian mixture modeling (DGMM), and multilevel vector autoregression modeling (Multilevel-VAR). The results showed that i-ARIMAX outperformed MLM in capturing within-person heterogeneity in the links between valued action and affect variables. Increases in values-based living were positively related to hedonic well-being but this effect showed a high degree of heterogeneity. A sub-group was identified, which we labeled the 'Stoics,' whose daily engagement in valued actions did not produce higher hedonic well-being (e.g., lower sadness or higher joy). Multilevel-VAR further revealed that for Stoics, stressful situations were linked to valued action, but not hedonic well-being. For Non-Stoics, valued action was less likely in stressful situations, but when valued action did occur it was associated with more joy and less sadness. The study offers initial evidence suggesting the superiority of an idionomic approach over a purely nomothetic one in capturing diverse pathways to clinically relevant outcomes. Idionomic methods may be useful or even necessary in personalizing psychological interventions, and thus may need to be considered by researchers and practitioners alike.

1. Testing the applicability of idionomic statistics in longitudinal studies: the example of 'doing what matters'

Idionomic methods—those that combine idiographic and nomothetic insights—have recently been touted to have great potential for informing individualized intervention, which is a core goal of process-based therapy (Ciarrochi et al., 2023; Hayes, Ciarrochi, Hofmann, Chin, & Sahdra, 2022; Sahdra et al., 2023). However, there is a dearth of research systematically testing how idionomic methods perform in comparison to commonly used nomothetic methods, which are arguably parsimonious. The present study aims to fill this research gap.

☆ This study involved secondary data analysis of pooled samples from previously published studies (Klimczak et al., 2023; Krafft, Klimczak, & Levin, 2021; Levin, Krafft, Pierce, & Potts, 2018). Each study was approved by the university's Institutional Review Board and informed consent was obtained from all participants. Informed consent did not include a statement regarding sharing de-identified data on open science platforms, preventing us from making the data open access. However, data is available upon reasonable request. The R code for the idionomic methods used in this paper can be found in Supplementary Materials. The study was not pre-registered. Generative artificial intelligence tools were not used in the writing or editing of this manuscript. Given his role as Editor-in-Chief, Dr. Michael Levin had no involvement in the peer-review of this article and had no access to information regarding its peer-review. We have no other conflicts of interest to disclose.

* Corresponding author. Institute for Positive Psychology and Education, Australian Catholic University, PO Box 968, North Sydney, NSW, 2059, USA.
E-mail address: baljinder.sahdra@acu.edu.au (B.K. Sahdra).

https://doi.org/10.1016/j.jcbs.2024.100728
Received 8 September 2023; Received in revised form 29 January 2024; Accepted 16 February 2024
Available online 18 February 2024
2212-1447/

1.1. The psychological homogeneity assumption

Much of psychological research focuses primarily on nomothetic links of psychological processes and their presumed outcomes. Nomothetic methods, with their emphasis on group averages, are often used in a way that suggests psychological homogeneity. Assuming psychological homogeneity implies that a process would uniformly affect an outcome for everyone, treating individual differences as negligible 'error' (Michell, 2019; Richters, 2021). The ergodic assumption—that group mean effects apply to the individual—underlies most of the nomothetic statistical tests used in psychology but is very often violated in models of human behavior (Ciarrochi et al., 2023; Molenaar, 2008; Sahdra et al., 2023). Idiographic methods can accurately characterize individual-level effects, but the challenge is to reconcile them with nomothetic effects (Lamiell, 1981), which are also important for scientific progress. There may be some aspects of psychological processes that are homogenous at the population or sub-population level (Gates & Molenaar, 2012), but determining that requires analyses at both individual and group levels.

In contrast to a purely nomothetic approach or a purely idiographic approach, an idionomic orientation toward data analysis requires that analysts first examine individual-level (idiographic) data and then build and retain group-level (nomothetic) insights only if the group-level results help clarify the individual-level insights; hence, the term, 'idionomic' (Ciarrochi et al., 2023; Hayes et al., 2022; Sahdra et al., 2023). An idionomic method does not refer to a particular statistical technique. Rather, it is a particular orientation to data, which may mean inventing new techniques or using existing statistical techniques in new ways.

As a concrete expression of that view, in the present study, we use existing analytic methods in a new way to conduct an idionomic analysis. The key feature of an idionomic orientation is that every effort is made to preserve and extend individual-level insights even while aggregating data to generate group-level insights. Nomothetic effects are treated with notable caution that increases in direct relation to the heterogeneity of the individual-level effects. Such an approach is particularly relevant for clinical decision making. Nomothetic findings, in the context of high heterogeneity, can potentially mislead clinical practitioners because the effects observed at the group level will not apply reliably at the individual level (Hayes et al., 2022; Sahdra et al., 2020, 2023).

Purely nomothetic techniques such as multilevel models (MLM)—the dominant statistical modeling approach in ecological momentary assessment (EMA) studies—are widely used to attempt to disentangle between-person and within-person effects (Wang & Maxwell, 2015). Unfortunately, MLM may not preserve the heterogeneity of the individual-level effects (Sahdra et al., 2023) and thus greater confidence may be given to such analyses when extending findings to the individual than may be warranted. In the present study, we used MLM as a benchmark to compare the value added by idionomic methods. Several statistical methods are available that can potentially be used to facilitate idionomic analysis. If these methods provide a clearer understanding of individual-level effects and variability, they may prove to be useful tools in idionomic research.

1.2. Idionomic forms of analyses

An interest in intensive longitudinal statistical methods has expanded in recent years (Gates, Chow, & Molenaar, 2023), and a recent Task Force Report specifically called for an expansion of these methods in Contextual Behavioral Science (Hayes, et al., 2021). There is a dearth of literature on idionomic methods, however, and the present study is one of the first to directly compare idionomic analyses to nomothetic analysis. For that reason, we need first to provide a rather detailed discussion and review of the state of the art of idionomic methods. We will then return to the empirical section of this paper as we apply these new methods to a real dataset to characterize the link between valued action and hedonic well-being.

1.2.1. Estimating pooled effects and heterogeneity

Powerful methods have been developed for examining within-person processes using high frequency data commonly available in neuroimaging studies (within-person $n > 100$), which use methods that combine structural equation modeling (SEM) paths, vector autoregression, and unified SEM to build idiographic and subgroup-level networks (Gates, Molenaar, Hillary, Ram, & Rovine, 2010). These methods have also been applied to psychological data (e.g., Sanford et al., 2022). However, simulation tests on network theory have shown that using existing methods a large number of within-person observations (~100) are needed to obtain reliable within-person estimates from individual networks and a large number of individuals (>500) are needed to accurately recover known paths in networks of subgroups (Gates, Henry, Steinley, & Fair, 2016). Obtaining such high-frequency within-person data on self-reported psychological processes can be challenging, especially in clinical settings. There is an urgent need to provide methodological solutions that make the most of time series data of relatively modest length (~25) to provide clinically relevant information to practitioners.

Ecological momentary assessment data of such modest length are ubiquitous in psychology and are often analyzed using MLM to account for the nested data structure, e.g., observations clustered within days clustered within persons and perhaps then conditions. MLM is a "partial pooling" compromise between the two extremes of "no pooling" of effects across clusters and "complete pooling" of the effects ignoring the nested data structure (Gelman & Hill, 2007). Therefore, it has the advantage of providing the population level estimate while accounting for clustering. However, one disadvantage of using MLM is that it can underestimate the heterogeneity of effects at the idiographic level (see Sahdra et al., 2023 for an example). Instead of focusing on why and how people respond differently, it is not uncommon for researchers to emphasize the practical and theoretical import of the mean estimate, implicitly treating nomothetic finding as the "true effect" while heterogeneity may be categorized as 'error' or 'unexplained variability' instead of potentially useful information. For their part, practitioners who consume such research and need to make decisions about individual clients are poorly served by nomothetic findings about known processes, in the absence of clear research-based guidance for how to detect processes of particular clinical relevance for one client versus another. An idionomic analytical approach is designed to detect and solve these problems.

A recent study showed how existing statistical tools can be used in an idionomic way. The study combined bivariate time series analysis with random-effects meta-analysis (RE-MA) for modest-length data (Ciarrochi et al., 2023). The analytic method involves building idiographic autoregressive integrative moving average models with an exogenous variable (i-ARIMAX; Ciarrochi et al., 2023). The AR (autoregression) component of the model involves an outcome time series being regressed on its own lagged values; the I (integrated) component involves differencing to remove trends; the MA (moving average) component involves modeling the dependency between an observed value and residual error from a moving average model applied to previous observations; and the X (exogenous) variable represents a predictor of the outcome. The 'i' in i-ARIMAX indicates that the ARIMA model with an exogenous variable is run separately for each individual (technical details of implementation of ARIMA in R can be found in Hyndman & Khandakar, 2008). For each time series, the method involves applying all model components that are appropriate, optimizing the parameters (both smoothing parameters and the initial state variable) of the model in each case, and then selecting the best fitting model according to the Akaike's Information Criterion.

The key analytic innovations by Ciarrochi et al. (2023) of combining i-ARIMAX and RE-MA will be used in the present study. For that reason, it is necessary first to describe how it is possible to analyze longitudinal data for individuals using metrics borrowed from meta-analysis to calculate estimates of the pooled effects and their heterogeneity.

In contrast to fixed-effects meta-analytical models, random-effects models do not assume that the pooled effect is the same in all studies in the meta-analysis. A random-effects meta-analysis reflects the population estimate of the effect if there is little heterogeneity, but not if there is substantial heterogeneity (Borenstein, Hedges, Higgins, & Rothstein, 2021). These same well-established meta-analytic methods can be used to see if individual effects should be aggregated into pooled effects in longitudinal psychological research. The commonly used methods of examining heterogeneity in meta-analysis include the Cochran's Q test, τ^2 and I^2. τ^2 is an estimate of the variance of the "true" effect size underlying the observed data, whereas I^2 is the percentage of total variability in an effect across participants that is due to true heterogeneity rather than chance (Harrer, Cuijpers, Furukawa, & Ebert, 2021).

The Q statistic, and whether it is statistically significant or not, is dependent on the number of effects included in the meta-analysis. In contrast, I^2 is insensitive to changes in the number of effects. There are no absolute cut-offs for interpreting I^2 but there are heuristic guidelines. The median I^2 in the Cochrane library of meta-analyses is 21% (von Hippel, 2015). If I^2 exceeds 50%, the pooled effect is commonly said to be a misleading representation of the population's "true" effect (Higgins et al., 2022). Although I^2 is not affected by variations in the number of effects, it is sensitive to the precision of those effects. τ^2 and τ (standard deviation of the "true" effect) are insensitive to the number of effects and their precision. However, they provide no straightforward clinical interpretation. For instance, if a variance of a true effect is 0.05, there is no way to know how meaningful that estimate is from the practical standpoint of generating actionable insights from the results.

These limitations can be overcome by the prediction interval (*PI*; IntHout, Ioannidis, Rovers, & Goeman, 2016). The *PI* provides a range of values within which we can expect the effects of future participants to fall based on the present evidence. For instance, if the *PI* of valued action predicting sadness falls below zero, we can expect that data from future individuals would also show a negative effect; thus, practitioners trying to reduce clients' sadness may try to improve their valued action in daily life. But if the *PI* includes both negative and positive values, there is no guarantee that data from a new person in the future would show a negative (or positive) effect; that is, the overall nomothetic effect would not easily lend itself to clinical decisions regarding new patients.

The *PI* is different in important ways from the confidence interval (*CI*) of the pooled estimate. The *CI* describes the uncertainty in the estimate of the pooled effect. It tells us how precise the pooled effect is but does not convey information about the degree of heterogeneity among the individual effects. One limitation of *PI* is that larger sample sizes are needed to assess its underlying assumption of a normal distribution of the effects. *PI* can appear spuriously wide or narrow (Higgins et al., 2022) when the number of effects are small (e.g., $n < 10$).

The importance of these methods was shown in the results of Ciarrochi et al. (2023). Across more than 50 longitudinal measurements in three samples totaling over 200 people, the great majority of i-ARIMAX longitudinal relationships from common processes of change (e.g., psychological flexibility measures) to common measure of outcomes (e.g., distress over sadness or anxiety) exceeded the amount of heterogeneity allowed when reporting central tendencies in meta-analyses. Said more simply, the values for metrics such as I^2 suggested that idiographic variability was too high for means to be meaningful. We will conduct a similar analysis in the present study. Additionally, we will compute estimates of *PIs*, which are arguably more relevant than *CIs* for any clinical decision making based on nomothetic effects.

1.2.2. Identifying meaningful subgroups

If there is substantial heterogeneity in the links of processes and outcomes under investigation, moderation tests can be conducted using any known moderators to see if known sub-groups have similar effects. If moderators are not available or do not help account for the observed heterogeneity, data-driven approaches may be used to identify potentially meaningful sub-groups of individuals. In addition to popular methods such as K-means clustering (e.g., Schubert & Rousseeuw, 2019) and latent profile analysis (e.g., Sahdra et al., 2017), recent advances in classification methods use machine learning approaches. One example is Deep Gaussian mixture modeling (DGMM), an unsupervised classification method that uses the 'deep learning' advances in machine learning in artificial neural networks (Viroli & McLachlan, 2019). Deep learning algorithms are similar to evolutionary algorithms (e.g., Sahdra, Ciarrochi, Parker, & Scrucca, 2016) in that they can 'learn' to navigate a vast number of parameters to efficiently yield solutions to complex pattern recognition problems. Deep learning computational architectures typically consist of multiple layers of latent variables, where the algorithm gradually 'learns' to characterize nonlinear interconnections between variables (Schmidhuber, 2015). Technical details of the deep learning implementation in the DGMM can be found in Viroli and McLachlan (2019). In brief, a DGMM consists of a network of layers of latent variables, which follow a mixture of Gaussian distributions at each layer. The multilayered structure amounts to nested mixtures of linear models, which deep learning algorithms characterize in flexible, nonlinear ways.

The results of a purely unsupervised classification approach, such as the DGMM, need to be interpreted with caution because machine learning methods are prone to overfitting and often require cross-validation to validate their results (James, Witten, & Hastie, 2014). Further, the differences between clusters need to be meaningful in the context of theory and prior findings, if available. An alternative would be to create meaningful sub-groups using a combination of prior knowledge, best practices in the field, and the data at hand. For instance, data from idiographic models may be used to create subgroups of individuals to separately examine the individuals who substantially deviate from the normative findings. We provide a clear example of this approach in our present study as well, in which we used the bivariate i-ARIMAX estimates to define subgroups and then extended the findings of the bivariate models using multivariate models that were conducted separately within the two idionomically defined subgroups of those who substantially deviated from the norm and the rest of the sample.

1.2.3. Modeling dynamic interrelations within subgroups

Once we have identified subgroups, we can then look at the dynamic links between processes and outcomes for those subgroups. Time series multivariate models can be run using multilevel vector autoregression modeling (multilevel-VAR), which combines MLM with vector autoregression, a multivariate extension of an autoregressive model (Bringmann et al., 2013). In a typical VAR model, variables are regressed on their own lagged versions and those of the other variables in the multivariate model. The multilevel part of multilevel-VAR allows variation in the VAR estimates across individuals. By combining VAR with MLM, it is possible to model temporal dynamics both within person and at the group level and the resulting coefficients can be used in Gaussian graphical modeling to build partial correlation networks of variables, as we will do in the present study.

Technical details of multilevel-VAR can be found in Bringmann et al. (2013). In brief, the nodes in the networks represent the variables and the edges represent the coefficients of their unique associations in the context of all other variables. A between-person network captures relations between the means of the variables across all measurement occasions. A within-person contemporaneous network plots within-person conditional links between variables within the same time point, which are estimated after conditioning on the effects of the previous time point in the time series. Exploring within-person processes in multivariate networks can be particularly helpful when there is limited prior theory about the exact structure of the complex interrelations in the networks (Borsboom et al., 2021).

1.3. Examining the utility of idionomic methods: the example of 'doing what matters'

To evaluate the utility of the methods reviewed above, we focused on the constructs of valued action and hedonic well-being in daily life. We chose these for both pragmatic and theoretical reasons. The dataset used in our demonstration was pooled from three prior EMA studies (Klimczak et al., 2023; Krafft, Klimczak, & Levin, 2021) which consistently measured valued action and hedonic well-being using the same instruments, which allowed us to maximize sample size. Theoretically, these constructs were especially appealing because increasing values-consistent behavior is a primary treatment target of many evidence-based therapies, such as acceptance and commitment therapy (ACT) or other "third wave" forms of cognitive behavior therapy (Hayes, Villatte, Levin, & Hildebrandt, 2011). In our study, we focused primarily on the valued-action item of 'doing what matters' in daily life and its relations to positive and negative affect.

Past research has shown that, on average, greater valued action is linked with reduced depression and anxiety (Tunç et al., 2023), less momentary stress (Finkelstein-Fox, Pavlacic, Buchanan, Schulenberg, & Park, 2019) and distress (Grégoire, Doucerain, Morin, & Finkelstein-Fox, 2021), greater subjective well-being (Grégoire et al., 2021), and increased well-being and quality of life (Levin et al., 2012, 2020; Villatte et al., 2016). While expecting that valued action may be positively linked to positive affect and negatively to negative affect at the group level, the psychological flexibility model central in ACT suggests a more nuanced relationship. Theoretically, and based on clinical experience, engaging in valued action often involves experiencing negative psychological reactions (Hayes, Strosahl, & Wilson, 2012). The daughter who values caring for her aging parent may have to confront her own grief and fear to do so; the teenager who values connecting with friends has to experience self-doubt and anxiety when reaching out. It is often important to engage in other processes emphasized in ACT (i.e., cognitive defusion, acceptance, present-moment awareness, and self-as-context) that allow the experiences of negative emotions as a normal part of valuing. This more nuanced relation between valued action and affect is commonly reflected in ACT training materials, such as, "values and vulnerabilities are always poured from the same vessel" (Wilson & Dufrene, 2008, p. 67). Thus, ACT theory and clinical expertise posit that valued action will be related to positive and negative affect in a way that varies based on the individual and context. Therefore, we expect to find high heterogeneity in the links of valued action with positive and negative affect.

1.4. Purpose of the present study

In the present study we examined the links between valued action and hedonic well-being as a way of examining the usefulness of the idionomic methods reviewed earlier. The following research questions were pre-specified within the team before any analyses were conducted: What are the nomothetic associations of valued action with positive and negative affect? To what extent are the nomothetic effects heterogeneous? How can we examine the links between valued action and affect idionomically, such that we obtain group-level insights based on idiographic analyses and without biasing or sacrificing individual-level insights? Based on past research examined earlier, we expected to find the following nomothetic effects: on average, valued action would be positively linked with positive affect and negatively linked with negative affect in daily life. However, we also expected the effects to be heterogeneous, such that for some individuals valued action may be unrelated to affect or even negatively linked to positive affect and/or positively linked to negative affect. People may also differ in the interconnections of valued action and affect in the context of stressful and positive life events. To examine the possible advantages of an idionomic approach, we first conducted analyses using MLM, which is a typical method used to analyze the data of EMA studies. The MLM analysis provided a nomothetic benchmark to see if our idionomic analyses added value to an understanding of the data. Thus, this study served as a preliminary test of whether an idionomic approach to data analysis might be worth pursuing.

Specifically, the following data analytical steps were designed to serve the following subgoals of the study:

(1) compare the raw within-person correlations between valued action and affect with the individual-level estimates from MLM and i-ARIMAX models to explore how well MLM and i-ARIMAX captured the distribution of the raw person-level associations;
(2) use RE-MA to create a nomothetic pooled estimate of the link between valued action and affect, and visualize and interpret the pooled effect in the context of the heterogeneity of the effect;
(3) identify subgroups of people who were similar in how values linked to affect using cluster analytical techniques, DGMM, and subgroups informed by i-ARIMAX effect sizes; and
(4) identify the ways that values dynamically linked to affect in a network of variables within the subgroups to tease apart within-person and between-person level inter-connections of the variables.

In sum, we aimed to characterize the known nomothetic links between valued action and affect in new ways to minimize the loss of individual-level information, which is a necessary step for personalizing interventions.

2. Method

2.1. Participants

We pooled three samples of college students, each collected from a separate longitudinal EMA study: combined $n = 425$; Sample 1 $n = 70$ (Levin et al., 2018); Sample 2 $n = 187$ (Krafft et al., 2021); Sample 3 $n = 168$ (Klimczak et al., 2023). All participants were recruited from the same predominantly White university located in the Western region of the United States, through an online research participation platform in exchange for research credit. Eligibility criteria were the same across samples, in which participants had to be at least 18 years old, a current student at the university we recruited from, and own a smartphone they could use to respond to EMA surveys. Participants identified primarily as being women (71.76%), not Hispanic or Latino (95.53%), White (93.65%), and were typically young adults ($M = 22.20$, $SD = 6.85$, range = 18–58). See Table 1 for the characteristics of the three samples.

Table 1
Demographic characteristics by sample.

	Total Sample	Sample 1	Sample 2	Sample 3
EMA design	–	3 prompts per day for 7 days	4 prompts per day for 7 days	3 prompts per day for 7 days
Participants (*n*)	425	70	187	168
EMA surveys (*n*)	6456	1337	2868	2251
Age (*M* (*SD*))	22.20 (6.85)	21.84 (7.16)	20.76 (3.51)	23.96 (8.86)
Women (%)	71.76	64.29	71.12	75.60
POC (%)	6.35	10.00	4.28	7.14

Note. POC: Person of color, as indicated by self-reported race other than White/Caucasian.

We conducted a one-way ANOVA followed by the Tukey HSD for age and a chi-square tests for gender, ethnicity, and race to determine whether demographic characteristics varied between the three samples. Age was found to be significantly greater for sample 3 ($M = 23.96$, $SD = 8.86$) as compared to sample 2 ($M = 20.76$, $SD = 3.51$; $p < 00.001$). No other statistically significant differences among demographics were found (all $p > 00.05$).

2.2. Measures

Table 2 contains all EMA items used in this study. Three valued-action items (Klimczak et al., 2023; Krafft et al., 2021; Levin et al., 2018) and four affect items (Watson & Clark, 1994) were available in all three samples (combined $n = 425$). Valued-action items all began with the prefix "Since the last prompt," and assessed whether participants were 'doing what matters', whether they were being 'the kind of person you want to be', and 'how content were you.' Affect items all followed the format "Right now, how ___ do you feel?" and assessed one facet of positive affect (joyful) and three facets of negative affect (sad, angry, ashamed). The two contextual items assessed the degree to which stressful and positives events and situations were experienced since the last prompt (Klimczak et al., 2023; Levin et al., 2018) were available in Samples 1 and 3 (combined $n = 238$).

All EMA items were rated on a 5-point Likert scale from 1 (*not at all*) to 5 (*very much so*). We used intraclass correlation coefficient-1 or *ICC*(1) to estimate the amount of variance that was between persons rather than within persons and *ICC*(2) from a one-way analysis-of-variance model to estimate the reliability of each of the EMA items (Bliese, 2000). These are reported in Table 2. The *ICC*(1) of the items ranged from 0.34 to 0.44, suggesting relatively low between-person variance, that is, relatively high clustering of responses within persons. Further, all items were highly reliable: *ICC*(2) ranging from 0.92 to 0.95.

2.3. Procedure

Each of the three studies used in this analysis were approved by the university's Institutional Review Board and informed consent was obtained from all participants. Sample 1's ($n = 70$) study investigated the immediate effects of engaging in experiential avoidance in the moment and took place between September 2015 and April 2016 (Levin et al., 2018). Sample 2's ($n = 187$) study examined how cognitive restructuring and cognitive defusion are naturally used with no formal training and took place between September 2017 and April 2018 (Krafft et al., 2021). Sample 3's ($n = 168$) study explored psychological inflexibility subprocesses predicting affect and valued action at the state and trait levels, with data being collected between February 2020 and April 2021 (Klimczak et al., 2023). For this study, we focused on the valued action and affect variables that were administered in all three samples. We also examined the context of stressful and positive events in daily life, which were measured in two of the three samples (Klimczak et al., 2023; Levin et al., 2018).

Data collection procedures for the three samples were largely similar, with some differences. Sample 1 included an initial in-person appointment for completing informed consent and baseline assessment, so that participants could be oriented to the EMA procedures by a research assistant. All procedures for Sample 2 and 3 were completed online. Participants from all three samples received prompts to complete EMA surveys for the next seven days, with Sample 1 using the MetricWire mobile platform, Sample 2 using the LifeData app, and Sample 3 receiving text messages with a link to the EMA Qualtrics survey.

A semi-random schedule was used to prompt participants to complete EMA surveys for each study. For Sample 1, participants were randomly prompted three times between the hours of 9 a.m. and 9 p.m., with a minimum of 1 h passing between each of the prompts and a reminder prompt being sent after 10 min of non-response. Participants could respond to the prompt at any time before the next prompt was delivered. For Sample 2, participants were prompted at a random time between the hours of 10 a.m.-1 PM, 1–4 PM, 4–7 PM, and 7–10 p.m. for a total of four prompts each day with a minimum of 30 min between prompts, with participants being given a window of 15 min to respond. At baseline and post-assessment, participants were asked "Overall, how would you describe your participation in this survey today?" with clarification that their response would not affect their compensation or participation and could respond on a scale of 1 "I answered every question carefully and honestly" to 5 "I randomly responded and/or did not respond honestly to any questions." All participants met the required cut-off for establishing non-random responses by answering 3 or lower. For Sample 3, participants were prompted at a random time between the hours of 9 a.m.-1 PM, 1–5 PM, and 5–9 PM for a total of three prompts each day with a minimum of 2 h between prompts, with participants being given a window of 2 h to respond. To limit the inclusion of data that may have resulted from careless responding within this third sample, responses must have been completed within 72 s or more in order to be included in analyses, which was determined by allocating 2 s per item.

The survey items that were administered varied by study. Relevant for the purposes of our study, the EMA items that were administered to all three samples included three items assessing the degree to which the participant engaged with valued action since the last prompt, one positive affect item, and three negative affect items (see Table 2). On the second day of receiving prompts, participants were given a reminder to keep responding to the prompts and were asked if they had any questions, with this being delivered through a phone call for Sample 1 and an email for Samples 2 and 3. Further details regarding the procedures for the three samples can be found in Klimczak et al., (2023); Krafft et al., (2021); Levin et al., (2012).

Table 2
Ecological momentary assessment items and their intraclass correlation coefficients 1 and 2.

	Construct	EMA Item	*ICC*(1) [95% *CI*]	*ICC*(2) [95% *CI*]
1.	Valued action - matters	Since the last prompt, were you able to do what matters to you?	0.35 [0.32, 0.39]	0.93 [0.92, 0.94]
2.	Valued action - content	Since the last prompt, how content were you with the amount and types of things you did?	0.35 [0.32, 0.39]	0.93 [0.92, 0.94]
3.	Valued action - person	Since the last prompt, were your actions in line with the kind of person you want to be?	0.44 [0.40, 0.48]	0.95 [0.94, 0.96]
4.	Positive affect - joyfulness	Right now, how joyful do you feel?	0.42 [0.30, 0.46]	0.95 [0.94, 0.95]
5.	Negative affect - sadness	Right now, how sad do you feel?	0.40 [0.37, 0.44]	0.94 [0.93, 0.95]
6.	Negative affect - anger	Right now, how angry do you feel?	0.38 [0.35, 0.42]	0.94 [0.93, 0.95]
7.	Negative affect - ashamed	Right now, how ashamed do you feel?	0.43 [0.39, 0.47]	0.95 [0.94, 0.96]
8.	Stressful context[a]	Since the last prompt, how much did you experience stressful events and situations?	0.34 [0.30, 0.39]	0.92 [0.90, 0.93]
9.	Positive context[a]	Since the last prompt, how much did you experience positive events and situations?	0.35 [0.31, 0.40]	0.92 [0.91, 0.94]

Note. *ICC*(1): the amount of variance that was between persons rather than within persons; *ICC*(2): reliability of the EMA item from a one-way analysis-of-variance model.

[a] Stressful and positive context items were administered in Samples 1 and 3 (combined $n = 238$).

2.4. Data analyses

Analyses were conducted in R version 4.3.1 (R Core Team, 2023). Participants were not asked for their consent to make their data publicly available, thus the data cannot be made available on an open access repository but can be made available to individual researchers upon reasonable request. Nevertheless, in the interest of scientific transparency (Cumming, 2013; Johnson, 2021), we report all data-analytical steps we used in this study in the order in which we conducted them, moving some of the details to Supplementary Materials as appropriate for brevity of the main text. Further, sample R code for the idionomic methods used in this paper can be found in Section 3 of Supplementary Materials, which may benefit other researchers wishing to use our methods.

Preliminary analyses showed that the three samples were comparable in terms of the mean, standard deviation, and skewness estimates of the EMA items (see Table S1 for the descriptive statistics of the within-person means study variables in the three samples); and also in terms of the mean within-person correlations of valued action items and affect, and the spread around the mean (see Fig. S1 for density plots). It was therefore reasonable to pool the data of the three samples. Missing data patterns of the EMA items in each of the samples indicated an overall missingness of 17% in Sample 1, 37% in Sample 2, and 30% in Sample 3 (see missingness maps in Figs. S2–S4 in the Supplementary Materials). We handled missing data using longitudinal imputation (Genolini, Écochard, & Jacqmin-Gadda, 2013).

2.4.1. Multilevel modeling: a nomothetic benchmark

Considering substantial within-person clustering indicated by relatively low *ICC*(1) values reported in Table 2, we first conducted the commonly used nomothetic method of MLM, with which most EMA researchers would be familiar. We ran models with each of the valued-action items predicting each of the affect items separately. We also ran models for each affect item where the three valued-action items were entered together. These models tested if the valued-action items uniquely predicted the outcomes, which would suggest that pooling the scores of the three items into a single score may not be appropriate for an idionomic approach. The results of MLM provided a nomothetic comparison to our subsequent idionomic analyses to help clarify the value added by our novel analytical approach.

An idionomic analysis first focuses on individual-level data and then moves to nomothetic methods only if they help clarify individual-level insights. We first examined the heterogeneity in the raw within-person associations of valued action and affect items. We then examined the same associations at the group level using MLM and compared the MLM-implied individual slopes to the raw associations. The goal was to see how well the distribution of MLM-implied slopes captured the distribution of the observed idiographic within-person associations.

2.4.2. Idiographic modeling using the i-ARIMAX algorithm

Raw within-person correlations ignore the temporal nature of the data, so we used a technique more suitable for time series data. We used i-ARIMAX with valued action as the exogenous variable, separately for each valued-action and affect pair of items, separately for each individual. The within-person data of each EMA variable in our study were standardized separately for each participant before running the i-ARIMAX algorithm. For any valued action and affect pair, individuals for whom the within-person variance for any item was less than 0.01 were omitted from the i-ARIMAX computation of that pair of items to avoid non-convergence of models. Such individuals with little variance in their time series data were a minority. The models of valued action items predicting joyfulness successfully converged for 417 individuals (98.12% of the sample), sadness for 386 individuals (90.82%), anger for 351 (82.59%), and ashamed for 327 (76.94%). Section 3 in Supplementary Materials contains an annotated R code we created to automate the i-ARIMAX algorithm. It consists of the following steps: (1) person-level standardizing of the time series data, (2) filtering out time series with almost zero variance, (3) computing separate ARIMAX models for each person, and (4) extracting an effect size and standard error from the ARIMAX model for each person. The distribution of the i-ARIMAX estimates was compared with the distributions of raw within-person associations and MLM-implied random slopes to see the extent to which the i-ARIMAX estimates matched the raw associations.

2.4.3. Random-effects meta-analysis (RE-MA): estimating the nomothetic effect and its heterogeneity

As mentioned earlier, an idionomic approach begins with individual-level data, which are aggregated to yield nomothetic effects if it helps to improve individual-level insights. For that, it is crucial to closely examine the heterogeneity of the nomothetic effects. The i-ARIMAX step described above helps us achieve the initial step of creating individual estimates. To derive nomothetic insights from the individual-level estimates, the i-ARIMAX estimates for each valued-action and affect pair were meta-analyzed in random-intercept meta-analyses to calculate estimates of pooled effects and their heterogeneity. We ran RE-MA models using the number of EMA sessions completed as a weighting variable to give more weight to estimates from individuals who completed more EMA measurements. This allowed us to adjust for any potential bias in estimates from individuals who provided fewer measurements, but sensitivity tests of models without weights were also conducted to see if the results were consistent. For each RE-MA, we conducted the Cochran's Q test of heterogeneity. We examined the heterogeneity of the effects using different visualizations and by calculating τ^2, I^2 and 95% prediction intervals.

Where evidence of substantial heterogeneity exists in a meta-analysis, moderation tests are recommended to account for the variability in the effects (Harrer et al., 2021). We ran moderation tests using the source of data (sample), gender, age, race, and ethnicity. To minimize the total number of tests, we used an omnibus test for the i-ARIMAX models of the three valued action items linking to an affect item in a multivariate RE-MA where effects were nested within participants (each person had three effects for three valued action items predicting the affect item). Multivariate RE-MA also allowed partitioning of the total heterogeneity (I^2) in between-person and within-person components, the within-person part characterizing the heterogeneity in the pooled effect due to the differences in the way that the three valued action items related to the relevant affect item within a person. Evidence of within-person heterogeneity in multivariate RE-MA models would suggest that the three valued action items function differently for different people in terms of their links with affect. In that case, aggregating the three items in a single score per person—a suitable strategy for purely nomothetic goals—would not be appropriate in an idionomic approach such as ours.

2.4.4. Identifying subgroups: cluster analysis, DGMM and i-ARIMAX informed subgroups

To aggregate the idiographic effects into subgroups, we used deep Gaussian mixture modeling (DGMM), an unsupervised classification method that uses the 'deep learning' advances in machine learning in artificial neural networks (Viroli & McLachlan, 2019). In sensitivity tests, we also used the more familiar methods of K-means clustering algorithms, using variations of Euclidean or Manhattan distance measures, means or medians as centroids, and multiple imputations for handling missing data (Basagaña, Barrera-Gómez, Benet, Antó, & Garcia-Aymerich, 2013). We also used the PAM (Partitioning Around Medoids) algorithm, which is considered a more robust version of K-means (Schubert & Rousseeuw, 2019). The sensitivity tests allowed us to explore any method-related bias in classification: if different classification algorithms yield comparable solutions, that can increase confidence in the results. From an idionomic standpoint, our goal was to see if a purely data-driven unsupervised approach to identifying clusters could help us aggregate the idiographic effects from the i-ARIMAX models in meaningful subgroups without sacrificing important insights gained from the heterogeneity of the idiographic effects.

In addition to an unsupervised approach to identifying clusters, we used a novel i-ARIMAX informed classification approach to the aggregation of idiographic effects. We identified individuals for whom valued action was negatively linked with joy and positively linked with sadness. Using a heuristic from Funder and Ozer (2019), in which an effect size of 0.30 is considered as a 'large' effect with potentially potent short- and long-term effects, we pulled data from individuals who had effect sizes equal to or smaller than −0.30 in the sadness i-ARIMAX models and the effect sizes equal to or larger than 0.30 in the joyfulness models with 'doing what matters' as a predictor. These individuals clearly deviated from the norm both in terms of the direction of the effect (±) and its size. In the next analytical step, our goal was to compare the interrelations of the valued-action items with the affect items for these individuals with those of the rest of the sample, especially in the context of daily experiences of stressful and positive events.

2.4.5. Multilevel vector autoregression modeling (Multilevel-VAR): creating within-person and between-person networks of variables

A network analytical approach was particularly suitable for the final step of our study because it allowed us to explore the interrelations between multiple variables in the absence of a prior theory about how the variables should be related (Borsboom et al., 2021). The nomothetic findings of past research (reviewed above) suggested that for many people, valued action would be positively linked with joyfulness and negatively with sadness. Therefore, we expected to find this pattern of interrelations between variables at both between- and within-person levels for many individuals in our sample. For those who deviated from the norm, we did not have any strong a-priori hypotheses about the structures of their within- and between-person interrelations between the variables, but considered it reasonable to build exploratory networks. Specifically, the networks from multilevel-VAR models provided a glimpse of how 'doing what matters' in daily life, the key focus of our study, operated in the constellation of other valued action variables, positive and negative affect, and contextual variables in the two groups of people (i.e., those who deviated from the norm, and the rest of the sample).

3. Results

3.1. Raw between-person and within-person associations

Table S2 in Supplementary Materials reports the mean, standard deviation, and correlations between mean (averaged across time) scores of all the EMA variables. Averaged across all measurements across the week, valued action items were linked positively with joyfulness and positive events and negatively with sadness, anger, shame, and stressful events. These correlations of means of EMA items ignore the within-person variability in the inter-relations of the variables across time. Table S3 reports the descriptive statistics of the raw within-person correlations of the EMA items. All correlations had a wide range. For instance, the within-person correlations of the valued-action item of 'doing what matters' with joyfulness had a mean of 0.39 ($SD = 0.32$) but a wide range from −0.94 to 0.99, and similarly with sadness: $M = -0.23$ ($SD = 0.34$), range = −0.98—0.97 (see gray highlighted rows in Table S3 of associations of 'doing what matters' with all other variables).

3.2. Nomothetic results from multilevel modeling

Preliminary unconditional (no predictor) random intercepts models were run for each outcome that showed that there was statistically significant between-person variation in the outcome to justify MLM. We conducted 2-level models with observations nested within persons and 3-level models with observations nested within days within persons (see Table S4 on Supplementary Materials). Log-likelihood model comparison tests showed that the 3-level models fitted the data better than the 2-level models ($p < 0.001$), thus 3-level nesting was used in all subsequent conditional models with valued action items as predictors of affect. In all conditional models, person-mean centering was used for valued-action items. Therefore, the intercept represented the expected value of affect when valued action was at each participant's own mean. For each affect outcome, we ran three separate models, each with one of the three items of valued action predicting the affect item. We ran a fourth model with all three valued action items entered together, competing against each other. When the valued-action items were added as predictors one at a time, a model with random slopes and random intercepts fit the data better than a model with random intercepts alone ($p < 0.001$). For the fourth model with all three valued-action items entered together, adding random slopes to random intercepts made it difficult for the models to converge due to the increased model complexity, so random-intercepts only models were used.

The fourth model allowed us to test whether the three valued-action items uniquely predicted affect. They did so significantly, thus justifying our decision to analyze the three valued action items separately in all subsequent analyses. Tables S5–S8 in Supplementary Materials report the details of the results of all conditional models of joyful, sad, angry, and ashamed outcomes, respectively. As expected, valued action items positively predicted joyfulness and negatively predicted sadness, anger, and ashamed outcomes. The three valued action items also uniquely predicted the outcomes, suggesting that they were not redundant.

For the model with the valued action item of 'doing what matters' predicting joyfulness, the fixed effect estimate was: 0.40 [95% *CI*: 0.37, 0.43]. For the model with 'doing what matters' predicting sadness, the fixed effect estimate was: −0.21 [95% *CI*: −0.24, −0.18]. These fixed-effect nomothetic estimates closely match the mean of the raw within-person correlations we reported earlier. The 95% *CI*s of the MLM estimates were relatively narrow, suggesting that the estimates were fairly precise. However, that result did not tell us anything about the heterogeneity of the nomothetic effect, nor how well the heterogeneity resembled the heterogeneity observed in the raw within-person correlations.

The intercepts and slopes estimates for each person from the multilevel models were extracted to compute each person's MLM-implied estimates of the links between valued action and affect. As shown in the top panel of Fig. S5, in the model with 'doing what matters' predicting joyfulness, all the individual slopes were positive, that is, individuals with negative raw within-person correlations were also 'assigned' positive slopes in the multilevel model. Similarly, the bottom panel of Fig. S5 shows that all the individual slopes from the model of sadness predicted by 'doing what matters' were negative such that people with positive raw within-person associations were "assimilated" to the nomothetic effect. We also compared the distribution of the individual-level slope estimates from the MLM with the distribution of the raw within-person correlations. As shown in Fig. 1, the density, and hence also the expected cumulative distribution function (ECDF) of the MLM-implied individual slopes and those of the raw within-person correlations were dramatically different, with MLM shrinking the distribution closer to the nomothetic effect.

3.3. Idiographic results of i-ARIMAX models

Fig. 1 also shows the distribution of the estimates obtained from the i-ARIMAX models in which separate models were run on the time series of each of the valued action and affect pairs for each individual. For the links of 'doing what matters' with joyfulness and sadness, the densities of i-ARIMAX estimates were almost identical to the respective densities

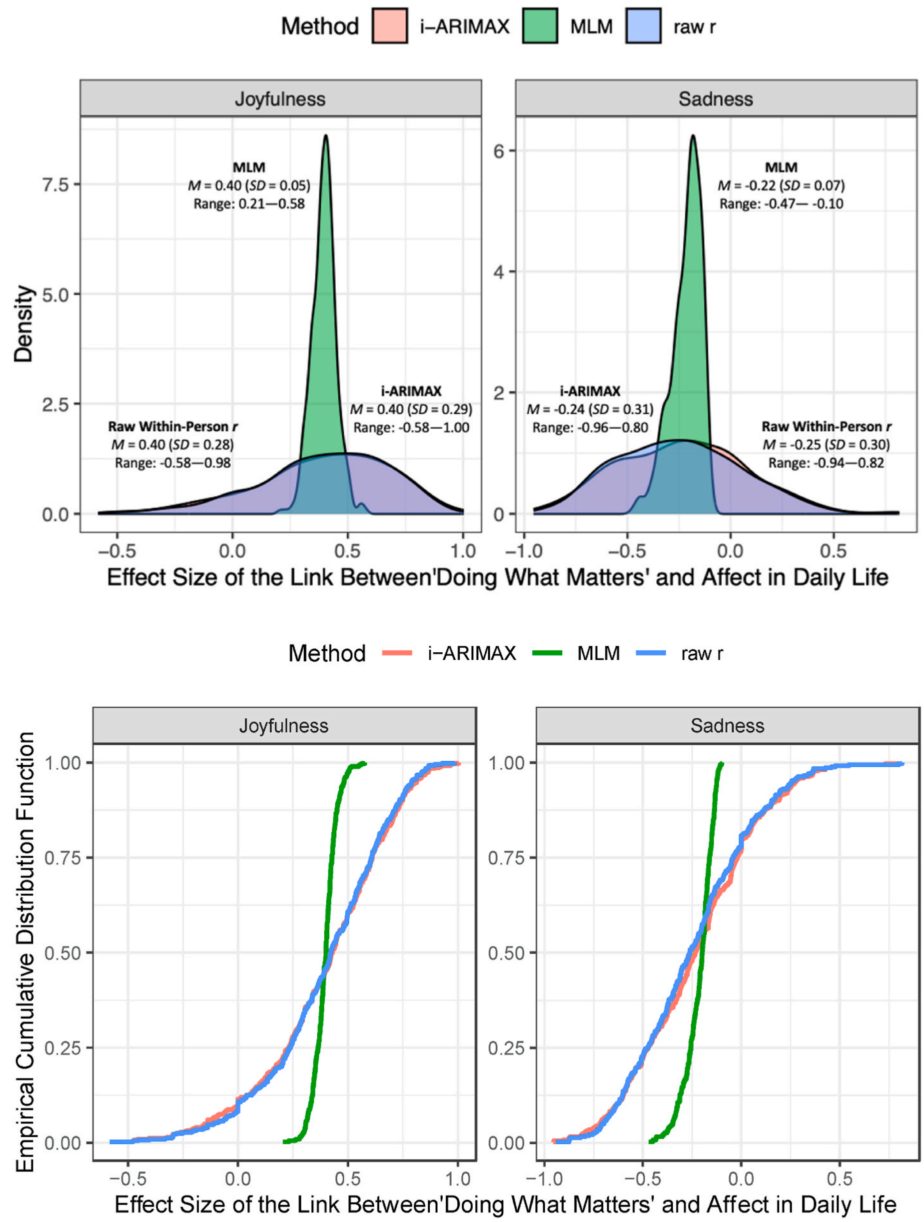

Fig. 1. The distributions of within-person links of ‘doing what matters’ with joyfulness and sadness calculated using three methods: raw within-person correlations, individual slopes estimates from multilevel modelling, and i-ARIMAX estimates.

of raw within-person correlations.[1] The bottom panel of Fig. 1 shows that the ECDFs of the three methods intersect at the overall mean effect, which is not surprising. The three methods produced identical nomothetic conclusions: the link of ‘doing what matters’ with joyfulness in daily life was positive, with the effect size of 0.40, which is considered a very large effect size that can be powerful short- and long-term; and the link with sadness was negative: −0.22, a medium effect size that may be important short-term Funder and Ozer, 2019). However, the i-ARIMAX estimates better preserved the heterogeneity of the raw observed within-person associations, so they were probably closer to reality than the MLM-based individual slope estimates.[2]

i-ARIMAX reduced “noise” due to autocorrelations and trends in the time series of the Y variable, that is, it related X at time t to the variance in Y at time t that was not explainable by previous values of Y or any trends in the time series of Y. The within-person correlations and MLM analysis were similar in that they did not control for the noise in the time series of Y. i-ARIMAX and within-person correlations were similar in that they were both idionomic. As can be seen in Fig. 1, both the idionomic methods yielded much more variable estimates than the “smoothed” MLM based results. The benefit of using i-ARIMAX estimates was that they took into account the temporal dynamics and time series trends, which raw within-person correlations did not. Therefore, the i-

[1] The *M* and *SD* of raw correlations reported in Fig. 1 differed very slightly from those reported in Table S3 because only the individuals for whom i-ARIMAX converged were used in the comparisons of the three methods in Fig. 1.

[2] A similar pattern of densities of the estimates obtained using the three methods was observed for all links between all valued action items and all affect items but only the results for the link of ‘doing what matter’ with joyfulness and sadness are reported here for brevity.

ARIMAX estimates were used in all subsequent analyses.

Critically, MLM results greatly underestimated the categorical implications of the within-person relationships. Both the i-ARIMAX and raw within-person associations showed many individuals for whom the direction (sign) of the process-outcome association was the opposite to that of the nomothetic effect, while the MLM results did not. If this were representative, a practitioner presented with only the MLM results would be surprised to find many clients on their caseload for whom the within-person relationship was opposite to what was expected nomothetically. The idionomic analyses would contain no such surprises for practitioners. We will discuss the importance of this difference in a later section.

3.4. Visualizations of heterogeneity of the idiographic effects

We conducted separate RE-MAs of the i-ARIMAX estimates for each of the valued-action item predicting each of the affect items, which allowed us to create radial and GOSH plots.[3] Radial plots, also known as Galbraith plots (Galbraith, 1994), show the association of the inverse of variability (the square root of the sampling variance (V_i) of the effect sizes plus the heterogeneity estimate (τ^2) from the RE-MA) on the *x*-axis to the standardized effect sizes on the *y*-axis. Fig. 2 shows the radial plots from the RE-MAs of the i-ARIMAX estimates of 'doing what matters' predicting joyfulness (top panel) and sadness (bottom panel). An arrow has been drawn in each plot to show the pooled effect and its 95% *CI* in the context of an arc on the right on which the observed i-ARIMAX effect sizes fall. The slope of the line in the middle of the gray region represents the pooled effect estimate, which is 0.38 [95% *CI*: 0.34, 0.41] for joyfulness and −0.23 [95% *CI*: −0.27, −0.19] for sadness. The gray shaded region in each plot represents the 95% limits (also known as Galbraith limits) of 2 units above and below the line in the middle. The points outside the limits represent individuals for whom the 95% *CI* of their respective i-ARIMAX estimate did not contain the pooled estimate from the RE-MA. Such individuals clearly deviated from the norm. These can be difficult to visualize in a forest plot in a meta-analysis with more than 400 effect sizes, but those who prefer forest plots can find them in Figs. S6 and S7 in Supplementary Materials. In the radial plot, several points can be easily spotted outside the gray region for both the joyfulness and sadness RE-MA models, making it hard to ignore the heterogeneity of the observed effects. The pooled effect sizes, despite their impressively narrow *CI*s, were dwarfed by the big picture, which showed many individuals who clearly deviated from the norm. Many of the points in the radial plot (and in the forest plot) represent multiple individuals with similar estimates, so merely counting the points within or outside the Galbraith limits would not be accurate in a plot with a large number of overlapping points. It is important to note that a radial plot (or a forest plot) on its own cannot help us determine the meaningfulness of the effects that deviate from the norm. Our subsequent analyses of identifying subgroups and the patterns of associations in the within-person and between-person networks better serve the goal of determining the meaningfulness of the deviations.

In meta-analysis, GOSH plots can help identify clusters of studies (or individuals in our present study) with potentially different effect sizes and amounts of heterogeneity (Olkin, Dahabreh, & Trikalinos, 2012). For instance, there may be subsets of individuals with little heterogeneity. To empirically search for such potential sub-groups, the same meta-analysis model was fit to all possible subsets of individuals. With k being the total number of effects (individuals) in a meta-analysis, a model for all $2^{(k-1)}$ possible combinations of effects was fitted, with a maximum of 1 million randomly selected models to make it computationally tractable (Viechtbauer, 2010). As shown in Fig. S8, the GOSH plot of the i-ARIMAX estimates of 'doing what matters' predicting joyfulness showed a roughly symmetric, homogeneous distribution. That is, there was no evidence of sub-groups of individuals who showed little heterogeneity in their effects.[4] See additional details in the note of Fig. S8 in Supplementary Materials.

3.5. Overall group-level aggregation: multivariate random-effects meta-analyses

To characterize the heterogeneity of the effects numerically, we ran multivariate RE-MA models in which the three effects of the three valued-action items linking to an affect item were nested within persons. Log-likelihood ratio tests were run for each multivariate RE-MA of each affect to compare the fit of a 3-level ('Full') model that allowed the effects at all levels to vary with the fit of a 2-level ('Reduced') model where the participant level was forced to be fixed. A 3-level model fit the data better than a 2-level model ($p < 0.001$). Therefore, a 3-level version was used for all multivariate RE-MA tests. We ran four models for the four affect outcomes.

Multivariate RE-MA allowed us to summarize the meta-analyses results concisely, partition the total heterogeneity into within-person and between-person components, and minimize the number of total tests for moderation analyses. Table 3 summarizes the results of the four multivariate RE-MA models. The results are reported from the models that used the number of measurements individuals completed as weights, thus giving more weight to the i-ARIMAX estimates from individuals who provided more measurements in the time series. Sensitivity tests of models without weights yielded comparable results.

Not surprisingly, and consistent with the fixed effects of MLMs, the pooled effect for joyfulness was positive whereas the pooled effects for the negative affect outcomes were negative. However, there was substantial heterogeneity in all models. The within-person I^2 in the four models ranged from 35.13% to 45.80%, suggesting that the three valued-action items related to the affect items differently within persons. For instance, a given individual may have a strong association between one of the valued-action items linking to joyfulness but a moderate, weak, or null association for the other valued-action items linking to joyfulness. The between-person I^2 ranged from 47.47% to 56.64%, suggesting that the effect of valued action on affect varied substantially between people, taking into account all three items of valued action. The total heterogeneity ranged from 88.71% to 93.27%. Such high levels of heterogeneity suggested that the pooled effects, despite being very precise given the narrow *CIs*, needed to be interpreted with extreme caution. Were this study a meta-analysis of effects from RCTs, pooled effect might not even have been allowed to be reported. Further, the 95% prediction intervals for all four models included both negative and positive values, suggesting that in any new sampling of data from new persons, the effect would be in the range of the prediction interval 95% of the cases, that is, *it could be negative or positive*. This is another face of the clinically relevant categorical finding mentioned earlier. Further, when we ran tests of moderation, there was no evidence of moderation by gender, age, sample, race or ethnicity ($p > 0.05$; see all *F*-tests of moderators in Table S9).

Taken together, the results in Table 3 showed that the nomothetic effects of valued activity benefiting affect, while true at the group level, did not apply at the individual level. Such a clear violation of the ergodic assumption (Molenaar, 2008) can easily misinform clinical practice and research. Despite that, violations of ergodic assumptions appear to be both common and invisible in mainstream psychological research (Hayes et al., 2022; Molenaar, 2004).

[3] For brevity, only the plots from the RE-MAs of the i-ARIMAX effects of 'doing what matters' predicting joyfulness and sadness are reported but all plots of all valued action and affect item pairs showed heterogeneity of the pooled effects, consistent with the numerical estimates of heterogeneity obtained from multivariate RE-MA models reported in Table 3.

[4] GOSH plots of other RE-MAs also showed symmetrical distributions with no evidence of sub-groups with little heterogeneity.

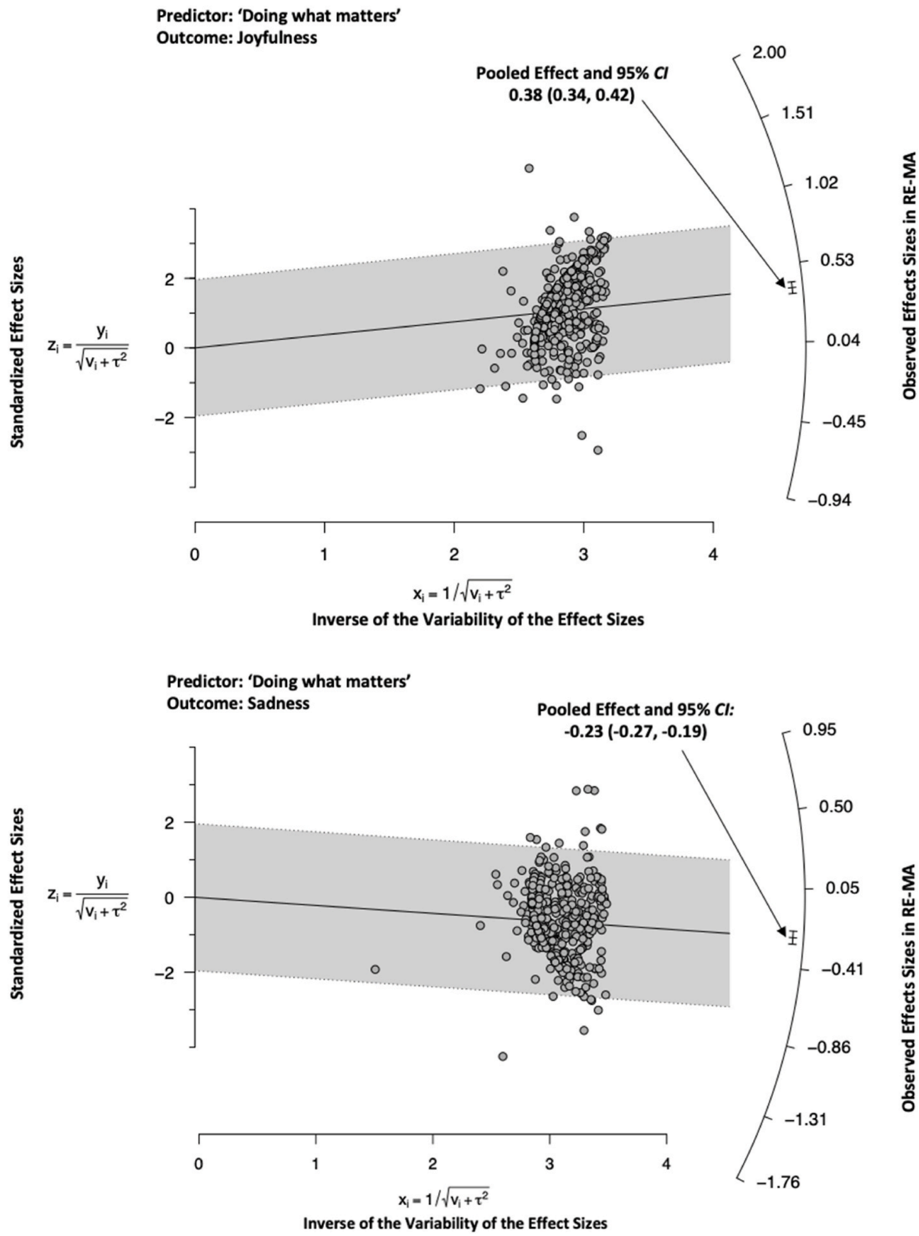

Fig. 2. Radial plot of the random-effects meta-analyses of the i-ARIMAX estimates of 'doing what matters' predicting joyfulness (top) and sadness (bottom).

Table 3
Summary of the findings of the four multivariate RE-MA models of the four affect outcomes with valued action items nested within persons.

Outcome	Pooled Effect	*SE*	95% *CI*	95% *PI*	Heterogeneity (I^2)		
					Total	Within	Between
Joyful	0.35	0.02	[0.32, 0.39]	[-0.30, 1.00]	88.71%	36.11%	52.59%
Sad	−0.22	0.02	[-0.26, −0.19]	[-0.87, 0.42]	91.91%	35.26%	56.64%
Angry	−0.20	0.02	[-0.23, −0.17]	[-0.85, 0.45]	93.27%	45.80%	47.47%
Ashamed	−0.21	0.02	[-0.24, −0.17]	[-0.83, 0.41]	90.88%	35.13%	55.74%

Note. SE = standard error; *CI* = confidence interval; *PI* = prediction interval.

3.6. Subgroup level aggregation: DGMM and cluster analyses

The i-ARIMAX estimates for each person for each valued-action and affect item pair were merged in a single dataset to run the DGMM. Not all participants had i-ARIMAX estimates for all 12 pairs. In the data of the 12 variables of i-ARIMAX estimates, 13% of the data were missing, which we handled using multiple imputation (Honaker, King, & Blackwell, 2011). The DGMM models were run in 50 imputations. A majority vote across the imputations was used to determine classification of individuals in the clusters. The DGMM yielded a 2-cluster solution with 314 participants (74.58% of the sample) in one cluster and 107 (25.42%) in the other cluster. Sensitivity tests using the cluster analysis methods of K-means cluster analysis and the PAM algorithm also indicated a 2-factor solution (see Fig. S9).

Chi-square tests indicated that DGMM-derived cluster membership did not differ by gender, race, or ethnicity ($p > 0.05$; details reported in Fig. S10). When we examined the means of the i-ARIMAX estimates (and 95% *CIs*), the difference between the two clusters was unremarkable. As shown in Fig. S10, the 95% *CIs* of all i-ARIMAX variables, except for the estimates of sadness predicted by 'doing what matters', overlapped. This suggests that the two groups were largely similar. The mean levels of 'doing what matters' predicting sadness for both clusters were negative, with one cluster's estimate slightly smaller than the other cluster. When we used cluster membership as a moderator in the multivariate RE-MA, it did not help reduce the observed heterogeneity. In short, cluster analysis did not serve our idionomic goals of aggregating up from the idiographic effects to yield meaningful nomothetic insights without loss of idiographic information. For instance, individuals who had positive i-ARIMAX links of valued action with sadness and those who had negative associations of valued action with joyfulness were "assimilated" in the clusters, just like they were in the nomothetic effects in MLMs.

3.7. *Subgroup-level aggregation: i-ARIMAX estimates binned by different effect sizes*

To better capture the information about individuals who deviated from the norm, we looked at the heterogeneity of i-ARIMAX estimates in a slightly aggregated form by plotting the percentage of people with i-ARIMAX estimates in different effect size ranges. We binned the data using the cut-offs based on the effect size interpretation guidelines from Funder and Ozer (2019): 0.05: very small effect; 0.10: small; 0.20: medium; 0.30: large; and 0.40 or greater: very large. Fig. 3 contains the effect size plots for the i-ARIMAX estimates for 'doing what matters' linked with joyfulness and sadness and Fig. S11 contains the plots for all pairs of valued-action and affect items, including anger and shame.

The left panel of Fig. 3 shows the results for the outcome of joyfulness. While more than half the individuals (~59%) had large to very large sized positive associations of 'doing what matters' with joyfulness, there was a substantial proportion of participants (~12%) who had null effects and a minority (~8%) who had negative effects. The proportion of people with large to very large negative effects was 2.7% for the 'doing what matters' item (and 4.27% for 'the kind of person you want to be' item, and 5.15% for 'how content were you' item, as shown in Fig. S11).

As shown in the right panel of Fig. 3, for nearly 20% of the sample, 'doing what matters' was unrelated with sadness. The percentage of people with a large to very large negative association of valued action with sadness were 41.25% for the 'matters' item (and 36.81% for the 'content' item of valued action and 43.43% for the 'person' item, as shown in Fig. S11). On the other extreme, the percentage of people with a large to very large positive association of valued action with sadness was 3.99% for the 'matters' item (and 6.01% for the 'content' item and 6.7% for the 'person' item, as shown in Fig. S11).

As a final step of subgroup-level aggregation, we selected the individuals who showed a large to very large negative effect (i-ARIMAX estimate $=< -0.30$) of 'doing what matters' predicting joyfulness and a large to very large positive effect (i-ARIMAX estimate ≥ 0.30) for sadness, as identified in the previous step. We tentatively labeled these individuals as 'Stoics' because their engagement in valued action in daily life was not diminished by either sadness or reduced joyfulness, which is consistent with the Stoic philosophy of choosing to live a 'good' life despite the varying circumstances of life (Aurelius, 2nd Century AD; 2002). These individuals are particularly interesting from a clinical standpoint because applying the nomothetic findings to increase valued action in them to promote their wellbeing may be counterproductive, at least in the short run. Consistent with ACT theory and practice, the very meaning of engaging in valued action may be different for these individuals, relative to the rest of the sample.

3.8. *Within-person vs. between-person networks of variables: Multilevel-VAR*

We conducted separate multilevel-VAR models for the Stoics and the rest of the sample, labeled 'Non-Stoics,' to compare their interrelations between the three items of valued action and the affect items of joyfulness and sadness. To avoid model convergence issues, anger and ashamed items were not added in these models. Sensitivity tests showed that replacing sadness with either anger or shame in the multilevel-VAR models yielded networks that were virtually identical to the ones reported here.

Specifically, we examined the within-person contemporaneous networks and between-person networks in the two groups. Although temporal networks based on user-defined lags can also be built using multilevel-VAR, we were wary of interpreting any temporal networks

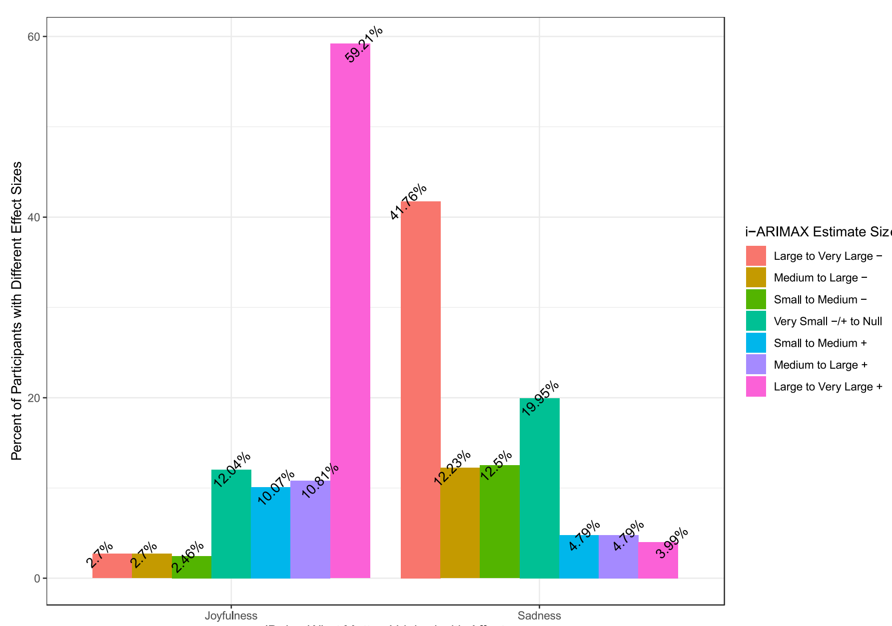

Fig. 3. Distribution of i-ARIMAX effects in different categories of effect sizes.

with these data because we had no way of determining the right temporal lag for examining the lagged associations between valued action and affect. For those interested, lag-1 temporal networks can be found in Figs. S12 and S13.

Fig. 4 shows the networks of within-person contemporaneous associations between valued-action items, joy, and sadness based on two multilevel-VAR models, one for the Stoics ($n = 17$) and another for the Non-Stoics ($n = 405$). The red colors represent negative associations and green represent positive associations. The thickness of the lines represents the strength of the associations. The affect items were dissociated from the valued action items among the Stoics (Fig. 4 left panel) but interlinked among the Non-Stoics (right panel). Although we had selected the group of Stoics based on their positive i-ARIMAX estimates for sadness and negative estimates for joyfulness, in the network of all valued action and affect variables taken together, there was no association between joy or sadness with any of the valued action items. Instead, the associations within the valued action items and within the affect items dominated the network.[5]

For the Non-Stoics, the within-person network in Fig. 4 (right panel) showed that the valued action items were positively linked with joy and negatively linked with sadness. Their between-person network shown in Fig. 5 (right panel) was largely similar to their within-person network in Fig. 4 (right panel). That is, the effects observed at the group level were generally true at the within-person level too for the Non-Stoics.

The between-person network of the Stoics shown in Fig. 5 (left panel) was largely similar to the between-person network of Non-Stoics (right panel). If anything, the strength of the associations in the between-person networks was stronger among the Stoics relative to the Non-Stoics. There was one potential difference worth noting in the between-person networks. Stoics' sense of self ('the kind of person they wanted to be') was largely linked to 'doing what matters', whereas Non-Stoics' sense of self was more driven by feeling content. Type III multivariate analysis of variance (MANOVA) tests showed that the Stoics and the Non-Stoics did not differ in terms of the mean (averaged across measurement occasions) levels of any of the valued-action items, sadness, or joyfulness ($F\,(1, 5) = 1.37$, $p = 0.23$, *Wilk's lambda* $= 0.98$). However, unlike the Non-Stoics, the interrelations of valued action variables with the affect variables observed at the between-person level (Fig. 5 left panel) were not there at the within-person level (Fig. 4 left panel) for the Stoics.

In the two of the three samples in our study, contextual variables were available, which allowed us to extend the within-person networks by adding the variables of stressful and positive events in daily life of the Stoics ($n = 9$) and the Non-Stoics ($n = 228$). We focus only on within-person networks here. Due to the reduced sample size in the Stoics group, the between-person network did not converge because the number of parameters to be estimated was greater than the number of observations. The between-person network of the Non-Stoics is reported in Fig. S14. As shown in Fig. 6, adding the context variables did not alter the pattern of associations of valued-action items with the affect items in the within-person networks of larger samples in Fig. 4: valued action items were unrelated to joy and sadness in the network of the Stoics but were related positively with joy and negatively with sadness in the network of the Non-Stoics. Importantly, Fig. 6 also showed that positive events were associated with increased joy and reduced sadness among the Stoics and Non-Stoics alike. Further, in both groups, positive events were positively linked with 'doing what matters.' However, stressful events were negatively related with 'doing what matters' among the Non-Stoics, but positively related among the Stoics. Type III MANOVA tests showed that the Stoics and the Non-Stoics did not differ in the mean (averaged across occasions) levels of stressful and positive events they experienced ($F\,(1, 2) = 0.93$, $p = 0.40$, *Wilk's lambda* $= 0.99$). It was not the amount of stress but how they related to stress in daily life that distinguished the Stoics from the Non-Stoics.

It may be worth reiterating that the link between stressful events and 'doing what matters,' like any other link in the network, represents the *unique* association between the two variables taking into account the levels of all other variables as well as the levels of all variables at the previous occasion in the time series. That is, taking all else into account in the network, stressful events seemed to have a dampening effect on 'doing what matters' among the Non-Stoics, whereas the Stoics were especially likely to do what mattered to them when they were experiencing stressful events in daily life. Put differently, for Stoics, stressful situations were linked to valued action, but not hedonic well-being. For Non-Stoics, valued action was less likely in stressful situations, but when valued action did occur it was associated with more joy and less sadness.

4. Discussion

Etymologically speaking, the word "method" derives from the Greek *meta* (in pursuit of) and *hodos* (a way or path; Online Etymological Dictionary, 2023). A path has a purpose, and in applied psychology, the purpose has been to develop organized knowledge of the relationships among events that allows particular people to live more vital and meaningful lives. For over 150 years, the tools of normative statistics and carefully controlled measurement have been central to the pursuit of a scientific pathway toward that end – that is, they are at the very hub of scientific methods in applied psychology.

In the present study, we examined what might happen if a less traveled pathway was pursued. Our conclusions can be simply stated as follows: purely nomothetic approaches appear to be systematically distorting the kind of knowledge practitioners need to serve their clients, and an idionomic statistical approach provides a new and useful way forward. This puts empirical "meat on the bones" of the recent conclusions of the Association for Contextual Behavioral Science Task Force that concluded that a more pragmatic methodological approach may be needed in applied psychology so as to focus on "functionally important processes of change, meaningful intervention goals, and user-friendly methodological and statistical approaches that meet its underlying assumptions" (Hayes et al., 2021, p. 181).

In our study, we reviewed statistical methods that can be applied to EMA studies with at least 20 within-person observations for examining inter-connections between variables at the between-person and within-person levels. We demonstrated an innovative use of existing statistical methods to address practically important questions about how valued action relates to positive and negative affect in an EMA design while simultaneously examining the importance of heterogeneity of within-person associations. Ultimately, we aimed to obtain individual-level insights into how valued action and affect were interrelated, as well as group-level nomothetic estimates that were complementary to individual-level estimates and did not treat them as errors.

We began by comparing distributions of raw within-person processes to outcome correlations with the random slopes from normative multilevel models and the idiographic estimates from i-ARIMAX models. All three methods yielded comparable group-level nomothetic estimates. However, MLM failed to capture the heterogeneity observed in the raw within-person associations while i-ARIMAX models succeeded. Shrinkage of individual estimates towards the mean is a necessary feature of MLM (Gelman & Hill, 2007; Sahdra et al., 2023), which was never designed for idionomic analyses despite its ability to examine how individual factors interact with group level influences. The problem is that practitioners need to know, *for this person they are treating,* whether particular processes of change are helpful or hurtful in moving toward the agreed upon goal. Despite its research popularity, MLM is not adequate for that task. In contrast, the i-ARIMAX algorithm was

[5] The i-ARIMAX estimates in our study are akin to zero-order correlations whereas the associations in the network are akin to partial correlations, except both the i-ARIMAX and multilevel-VAR estimates take into account the temporal nature of the data by conditioning the contemporaneous effects on the effects of previous time in the time series.

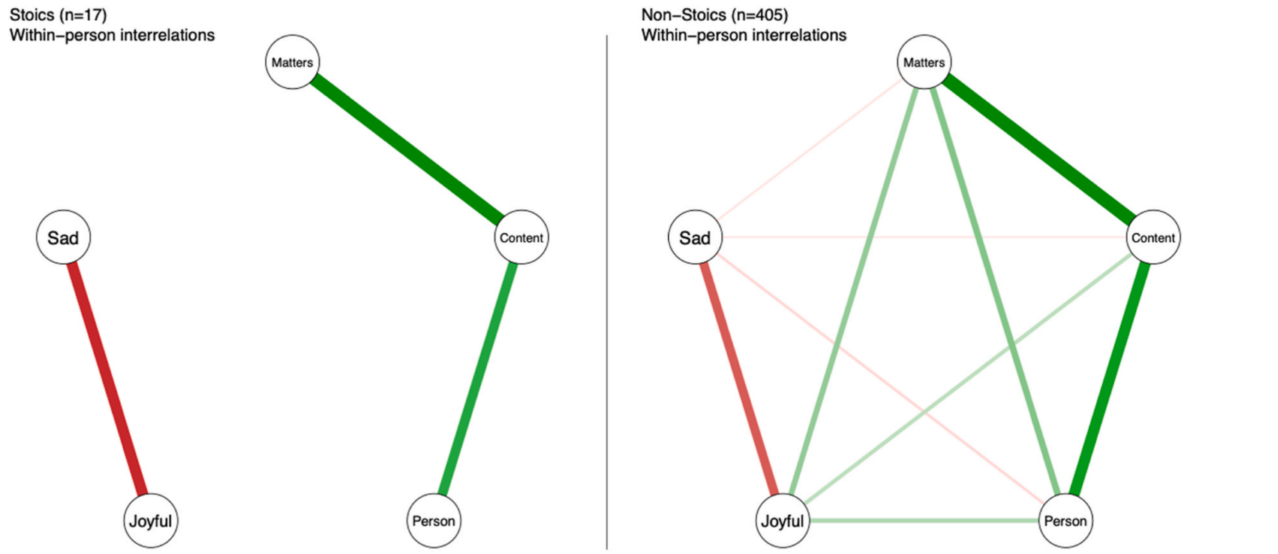

Fig. 4. Within-person contemporaneous networks from multilevel-VAR models of valued action, joy and sadness.
Note. These networks are based on the pooled data from three samples. Matters = 'Doing what matters' item of valued action; Content = 'How content were you' item; and Person = 'The kind of person you want to be' item..

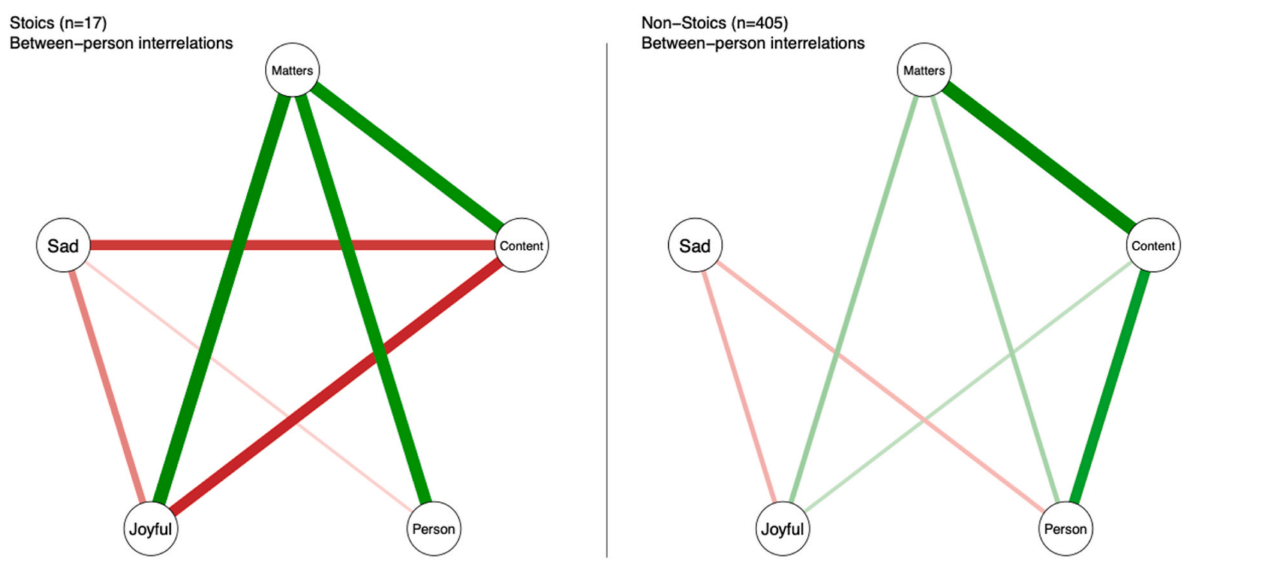

Fig. 5. Between-person networks from multilevel-VAR models of valued action, joy and sadness.
Note. These networks are based on the pooled data from three samples. Matters = 'Doing what matters' item of valued action; Content = 'How content were you' item; and Person = 'The kind of person you want to be' item..

specially designed for idionomic analysis (Ciarrochi et al., 2023) and seems to perform well in our study too.

We next attempted to characterize the mean and heterogeneity of the idiographic effects using random-effects meta-analyses (RE-MAs) models. We computed estimates of the pooled effect and its precision (95% *CI*), variance (τ), a statistical test of heterogeneity of the effects (Q), heterogeneity (I^2), and a range of values for an expected future data (95% *PI*). This meta-analytic approach asked a simple but critical question: do the individuals represent a population for which a pooled effect may be interesting? The I^2 values were extremely high (>80%). At this level of heterogeneity, it is often recommended that individual-level estimates should not be pooled at all (Borenstein, et al., 2021). Rather, a narrative review may be more appropriate (Harrer et al., 2021; Higgins et al., 2022). Applying the same logic to our study, in which idiographic results take on the same role as RCTs in standard meta-analysis, the pooled effects from our multivariate RE-MA of the links of the three valued-action items with each of the various affect items cannot be interpreted as reliable indicators of the "true" population level effects because in all models heterogeneity was too high. It is worrisome to note that these kinds of pooled effects are precisely what would normally be published in similar research studies. But researchers and readers alike would largely be unaware of the risk of self-deception that such pooled effects carry due to violations of ergodicity.

An idionomic analytic strategy provides a new pathway to identify such risks. The radial plots (Fig. 2) showed that the nomothetic pooled effects were dwarfed by the staggering heterogeneity in the data (see also Figs. S6 and S7 for the forest plots similarly showing high heterogeneity). Further, although the 95% *CIs* of the pooled effects were impressively narrow, the 95% *PIs* included both negative and positive values in all models. That is, the overall pooled effects in our study were impressively precise, so we can confidently predict the likely pooled effect of a new sample of a similar size. However, this "precision"

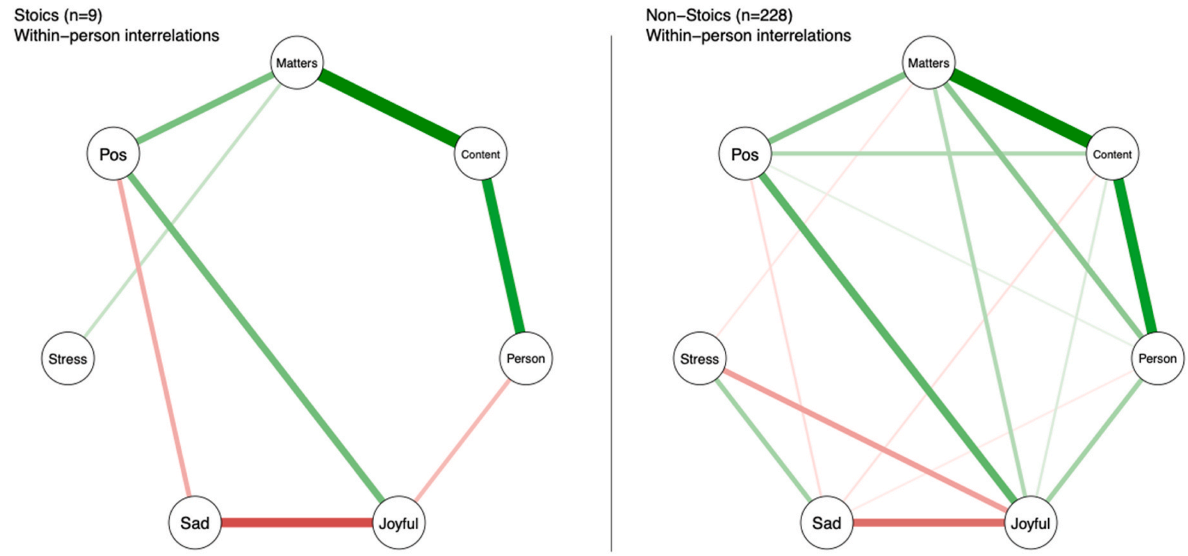

Fig. 6. Within-person networks from multilevel-VAR models of valued action, joy, sadness, stressful events and positive events.
Note. These networks are based on the pooled data from two samples in which the context variables were available. Matters = 'Doing what matters' item of valued action; Content = 'How content were you' item; Person = 'The kind of person you want to be' item; Stress = The context of stressful events in daily life; and Pos = The context of positive events in daily life..

disguised an alarmingly high uncertainty about expected future value of the effects for a new *individual* in a fresh sample. Thus, group averages may be best at predicting future group average behavior, but not future individual behavior. An idionomic strategy, conversely, allowed these associations to be examined in a way that did not distort idiographic information.

A purely idiographic research program could never be fully mounted since every individual would be a "psychology of one." It would also not be scientifically or practically useful, because concepts need nomothetic scope to be practically generalized to others. Thus, in an idionomic approach it is critical to determine how to aggregate idiographic knowledge to the nomothetic level without undermining idiographic fit and even possibly improving it. In our study, we attempted several aggregation methods to try to characterize meaningful subgroups of individuals. The i-ARIMAX-informed classification method seemed to work better than the unsupervised machine learning approach. In the two idionomically defined groups of Stoics and Non-Stoics, network analysis yielded meaningful nomothetic conclusions that preserved the insights gained from the idiographic models. For a minority of individuals, the Stoics, idiographic estimates substantially deviated from the norm. The Stoics engaged in valued action in daily life in the face of sadness and reduced joyfulness, which is consistent with the Stoic philosophy of living a 'good' life even when experiencing the vicissitudes of life (Aurelius, 2nd Century AD; 2002).

It is one thing to identify nomothetic groups. It is another to show that these groupings are meaningful. We found that within-person networks of the interrelations of the study's variables among the Stoics were meaningfully different from those of the Non-Stoics. The pattern of interrelations between valued action, affect, and stressful and positive events among the Non-Stoics was consistent with past nomothetic findings about valued action being linked with less distress and greater well-being (Levin et al., 2012, 2020; Villatte et al., 2016). These patterns held at both the between-group level and the within-group level. The same was not true for the Stoics. The Stoics differed from Non-Stoics in their within-person interrelations of variables—a difference that was even more striking when within-person networks were extended by adding contextual variables of stressful and positive events.

Positive contextual events were positively linked with joy and 'doing what matters' in the within-person networks of both groups, but stressful events were negatively linked to 'doing what matters' among the Non-Stoics but positively linked among the Stoics. The mean levels of stressful and positive events averaged across time were comparable in the two groups—the key difference in the two groups was how they related to stressful events in terms of engaging in activities that mattered to them. In stressful times, the Stoics tended to engage in more valued action while the Non-Stoics engaged in less. These results are particularly notable given that the context variables of stressful and positive events were not used to define the two groups before constructing their respective networks. Stated more technically, the subgroup-level aggregation approach combining the insights from i-ARIMAX idiographic models with multilevel-VAR modeling allowed us to characterize the nomothetic interconnections between valued action and affect, both at the between-person and within-person levels, without sacrificing the voices of the minority of individuals in the data who clearly deviated from the norm. These conclusions would simply have been invisible if a purely normative method had been used.

Group-level estimates of the effect of valued action on affect failed to accurately reflect the individual experiences of a subgroup of people in the data. This violation of the ergodic assumption highlights the need for caution in over-relying on the nomothetic effects of valued action to clinical decision-making. Given the current normative methods that dominate in our field, an empirical clinician might understandably target values clarity as a way of, say, reducing entanglement with sadness (or its closely allied normative label, "depression") in new clients from whom the clinician may not yet have any systematic idiographic data. The present results show how risky that decision could be since the prediction interval of the values → sadness relation included both positive and negative effects. Yes, the 'normal' effect can be helpful in orienting practitioners to potentially useful psychological processes of change, but our study suggests that norms may poorly describe how some people experience values in daily life (including the Stoic subgroup). In contrast, our idionomic findings (while tentative and initial) can potentially better serve clinicians by orienting them to monitor the specific relationship of valuing with negative affect in their individual client, and guiding clinicians to consider whether their client demonstrates a pattern of responses (e.g., stoic or non-stoic) that may be meaningful at the individual level.

To be clear, we are not claiming that short-term increase in sadness as a result of enhancing valued activity is necessarily an unhealthy outcome. The 7-day data collection window of our study with undergraduates cannot speak to the effects of valued action beyond a week, nor its potential effects on overall well-being or quality of life in a clinical context where the client is well supported by a skilled therapist during the transition of short-term sadness in the service of long-term gains in meaning and overall well-being. Examining the long-term effects of valued activity was not a goal of our study. Our key objective was to test the potential utility of an idionomic approach, using the bare minimum time series data needed to examine both the idiographic and nomothetic effects.

We have shown that idionomic methods can be applied even to relatively short time series data from clients to yield insights that can potentially inform personalized interventions. Idionomic methods are already being implemented in user-friendly software applications to automate high temporal density data collection, to analyze idiographic data in an idionomic fashion, and to generate reports for clinicians and their clients to facilitate interventions based on data-driven case conceptualizations (e.g., Hayes, Ciarrochi, & Jansen, 2023). Thus, an empirical clinician might better serve their clients by being aware that the analyses in much of what they read in psychology violates the ergodic assumptions of these very analyses, and they should consider using high temporal density measures in their own practice, analyzing them idionomically, to supplement their current knowledge as based on the research literature. Even a simple spreadsheet can yield within-person correlations and the tools needed for more sophisticated analyses that already exist. Thus, the long-held belief (e.g., Hayes, Barlow, & Nelson-Grey, 1999) that empirical clinicians need to produce and not merely consume research in order for the field to progress may be more applicable today than ever. Indeed, clinical testing of idionomic methods will be key to providing a strong scientific foundation for measurement and analyses that can empower clinicians to effectively personalize their interventions.

Of course, if an idionomic approach were expanded to include a wider variety of processes of change and the contextual features impacting on them, it might be found that the different values → affect relations were part of a larger network of process → process and process → outcome relations over time. If so, the narrow clinical implications of our study for values-focused interventions may change. But that is our larger point. The present study is a proof of concept. It is an early test of the limits of our current normative methods and whether an idionomic alternative is worth pursuing. Results of this study suggest that it is, and that focusing on the between-person effects alone is not sufficient for understanding the moment-to-moment dynamics of processes of change that practitioners need to understand to guide interventions for specific people. Conversely, within-person processes examined from an idionomic lens better honored the diversity of the different voices in the data, which tended to be smothered in "error" terms of group-level effects of purely nomothetic methods. Practitioners do not treat "error" terms—they treat people. If this finding of smothering of individual voices is replicated across a variety of empirical associations and samples, an expansion into idionomic analysis may help alleviate current concerns over the tendency of behavioral science methods to minimize the importance of human diversity and individuality (Mertens, 2019).

4.1. Limitations and future directions

An idionomic method refers to an analytic strategy, not to a particular statistical tool. It can be performed using a variety of existing techniques, as we demonstrated in this study. Future research on valued action may expand on our results through the use of additional variables and longer time series, which will lend themselves to other statistical techniques, which are also suitable for idionomic analysis but require more within-person observations (e.g., see Gates & Molenaar, 2012; Gates et al., 2023; Gates et al., 2010 for uSEM and GIMME methods). Longer time series would also allow systematic examination of potential causal directions using different lags in the lagged temporal networks, which was not possible in the current sample. There is an implied causal direction from values to affect in the current study because participants were asked to rate their valued activity 'since the last prompt,' and affect was measured 'right now.' However, future research is needed to tease apart the consequences of different temporal anchoring of responses.

Future research would also benefit from collecting additional quantitative and qualitative data to get deeper insight into the psychological processes related to valued action and affect, especially in those who deviate from the norm. Recent research using the network analysis approach similar to the one we used proved helpful in separating the between-vs. within-person associations of specific kinds of values (e.g., self-transcendence or self-enhancement) and well-being (Fischer & Karl, 2023).

Some might see secondary data analysis as a potential weakness of our study. We took several steps to ensure research transparency (Johnson, 2021) and avoid the issues related to 'false-positive psychology' (Simmons, Nelson, & Simonsohn, 2011). We have disclosed all variables that we selected for this investigation; we chose the valued action and affect variables that were common across the three samples that were pooled for this study to maximize the power and reliability of the results. The design of the studies from which these samples came had at least 20 within-person measurements per person, which is consistent with the minimum sample guidelines (Simmons et al., 2011). The other variables of those studies, that were not relevant for our study, are reported elsewhere (Klimczak et al., 2023; Krafft, Klimczak, & Levin, 2021; Levin et al., 2018). Our research team consists of researchers in independent labs in different countries. All analytical steps were transparent within the team, which meant that the results were verified by different researchers with complementary skills. There were no experimental conditions in the samples we used, thus there is no issue related to under-reporting of failed manipulations. We did not eliminate any observations from the samples, unless there was little variance in a particular time series, which prevents model convergence; and we reported when that happened. We provided clear justifications for the analytical methods we used and have disclosed all analyses we conducted. We have reported all results, including the ones that yielded null or 'imperfect' findings. Finally, we ran many sensitivity tests to ensure that our results did not hinge on idiosyncratic analytical decisions.

Still, it would be important for future research to replicate the findings of our study by explicitly testing the heterogeneity of the effects of valued action linked with affect in daily life using the methods we introduced here. And to further our understanding of the reasons for the heterogeneity of the links between valued action and affect, future research is needed to examine potential moderators, such as: valued activities that are hedonically difficult such as caring for a sick relative; highly challenging valued activities that require a great degree of skill and concentration which may not be pleasant in the moment; and valued activities performed due to external pressures rather than intrinsic motives (Bradshaw et al., 2021). We hasten to add that we need to be wary of the psychological homogeneity assumption that can creep into moderation tests. For instance, if sex is a moderator, the psychological homogeneity assumption would imply that the effect of being male or female is the same for all males and females, which seems too strong of an assumption.

Similarly, it would be a mistake to conclude that we should now treat Stoics differently than Non-Stoics in terms of values, or run out to create a set of new personality tests to detect these subpopulations and to write popular articles on "are you a Stoic or Non-Stoic?" Such an extension would go in precisely the wrong direction. Stoicism is a pattern of behavior in a particular context, not a thing. We caution against any temptation to reify the Stoic versus Non-Stoic sub-grouping or make one sound better or worse. It would be more productive to seek to understand the intervention implications of the Stoic pattern in the context of the person's full life assessed in an idionomic fashion. For example, we might ask, what other physiological, psychological, and social processes are influencing their hedonic experience? How does this larger set of processes link to their values over time (See Hayes et al., 2022, for discussion of networks-informed case conceptualization.)?

Finally, there are potential limitations of our dataset. Steps were taken to prevent data quality issues (e.g., random responding, not paying attention, etc.), as outlined in the method section, but these were limited to Samples 2 and 3. Further, given that the samples consisted of undergraduates, it remains to be seen whether the methods used in this paper would show similar levels of heterogeneity in the links of valued action with affect in clinical samples. Replication is important for scientific progress, especially for nomothetic goals. However, from an idionomic standpoint, the key objective is to make every attempt to represent all 'voices' in the data, especially those who deviate from the norm, even if future data from a different sample may not contain similar voices. While making generalizations that apply broadly is an important aim of research, the job of clinicians is to treat an individual in their specific context. There is already emerging evidence that individually-tailored clinical interventions are more effective than one-size-fits-all approaches that have historically been promoted through the use of randomized controlled trials of complex packages in psychology (e.g., see Nye, Delgadillo, & Barkham, 2023, for a systematic review).

5. Conclusion

We began this discussion with an etymological reminder that methods are means—they are ways to pursue pathways toward goals. Over the last 40 years, top-down normative methods have yielded an intervention science that is producing effect sizes that are either flat or falling (Hayes, Hofmann, & Ciarrochi, 2023) and that have often left practitioners to fend for themselves without statistical guidance for extending normative findings to individual clients in a fashion that is known to fulfill the statistical assumptions that led to these findings. That cannot be allowed to continue if our science and the practice based on it are to prosper.

At least since Kuhn (1962) many philosophers of science have agreed that scientific fields tend to stagnate when the assumptions contained within mainstream perspectives become invisible and no longer foster progressive research. The 150-year-old assumption that within-person variability can be addressed in ways that are broadly similar to between-personal variability is based on the ergodic error (Molenaar, 2004). The present study is a proof of concept that confirms recent concerns over the extent, severity, and negative impact of that error (Hayes et al., 2022), adds to recent empirical studies documenting violations of the ergodic assumption (e.g., Ciarrochi et al., 2023; Sahdra et al., 2020; Sahdra et al., 2023; Sanford et al., 2022), and provides support for the possible practical value of an idionomic analytical alternative (e.g., Ciarrochi, Hayes, Oades, & Hofmann, 2022; Ong, Hayes, & Hofmann, 2022; Sahdra et al., 2023). Idionomic methods, such as the ones showcased in this paper, are needed to foster research and development of personalized evidence-based interventions. Developing such intervention and case conceptualization methods appears to be a logical next step in helping our science improve the lives of those we serve.

Declaration of competing interest

Given their role as Editor-in-Chief, Dr. Michael Levin and Editorial Board member, J. Krafft, had no involvement in the peer-review of this article and had no access to information regarding its peer-review. All other authors have declared no conflicts of interest.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcbs.2024.100728.

References

Aurelius, M. (2002). *The meditations*. New York, NY: Random House (2nd Century AD.

Basagaña, X., Barrera-Gómez, J., Benet, M., Antó, J. M., & Garcia-Aymerich, J. (2013). A framework for multiple imputation in cluster analysis. *American Journal of Epidemiology, 177*(7), 718–725. https://doi.org/10.1093/aje/kws289

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and Analysis. In K. J. Klein, & S. W. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations* (pp. 349–381). San Francisco, CA: Jossey-Bass, Inc.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., … Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1*(1), 58. https://doi.org/10.1038/s43586-021-00055-w

Bradshaw, E. L., Sahdra, B. K., Ciarrochi, J., Parker, P. D., Martos, T.s., & Ryan, R. M. (2021). A configural approach to aspirations: The social breadth of aspiration profiles predicts well-being over and above the intrinsic and extrinsic aspirations that comprise the profiles. *Journal of Personality and Social Psychology, 120*(1), 226–256. https://doi.org/10.1037/pspp0000374

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One, 8*(4), Article e60188. https://doi.org/10.1371/journal.pone.0060188

Ciarrochi, J., Hayes, S. C., Oades, L. G., & Hofmann, S. G. (2022). Toward a unified framework for positive psychology interventions: Evidence-based processes of change in coaching, prevention, and training. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.809362

Ciarrochi, J., Sahdra, B. K., Hayes, S. C., Hofmann, S. G., Sanford, B. T., Stanton, C. E., … Gloster, A. (2023). A personalised approach to identifying important determinants of well-being. *PsyArXiv Preprints*. https://doi.org/10.31219/osf.io/m4zhw

Cumming, G. (2013). The new statistics. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Finkelstein-Fox, L., Pavlacic, J. M., Buchanan, E. M., Schulenberg, S. E., & Park, C. L. (2019). Valued living in daily experience: Relations with mindfulness, meaning, psychological flexibility, and stressors. *Cognitive Therapy and Research, 44*(2), 300–310. https://doi.org/10.1007/s10608-019-10062-7

Fischer, R., & Karl, J. A. (2023). Unraveling values and well-being—disentangling within- and between-person dynamics via a psychometric network perspective. *Journal of Personality and Social Psychology, 124*(6), 1338–1355. https://doi.org/10.1037/pspp0000449

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Galbraith, R. F. (1994). Some applications of radial plots. *Journal of the American Statistical Association, 89*(428). https://doi.org/10.1080/01621459.1994.10476864

Gates, K. M., Chow, S., & Molenaar, P. C. M. (2023). *Intensive longitudinal analysis of human processes*. Boca Raton: Chapman & Hall/CRC Press.

Gates, K. M., Henry, T., Steinley, D., & Fair, D. A. (2016). A Monte Carlo evaluation of weighted community detection algorithms. *Frontiers in Neuroinformatics, 10*. https://doi.org/10.3389/fninf.2016.00045

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*(1), 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage, 50*(3), 1118–1125. https://doi.org/10.1016/j.neuroimage.2009.12.117

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Genolini, C., Écochard, R., & Jacqmin-Gadda, H.l. n. (2013). Copy mean: A new method to impute intermittent missing values in longitudinal studies. *Open Journal of Statistics, 3*(4), 26–40. https://doi.org/10.4236/ojs.2013.34A004

Grégoire, S., Doucerain, M., Morin, L., & Finkelstein-Fox, L. (2021). The relationship between value-based actions, psychological distress and well-being: A multilevel diary study. *Journal of Contextual Behavioral Science, 20*, 79–88. https://doi.org/10.1016/j.jcbs.2021.03.006

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Boca Raton, FL and London: Chapman & Hall/CRC Press.

Hayes, S. C., Barlow, D. H., & Nelson-Grey, R. O. (1999). *The Scientist-Practitioner: Research and accountability in the age of managed care* (2nd ed.). New York: Allyn & Bacon.

Hayes, S. C., Ciarrochi, J., Hofmann, S. G., Chin, F., & Sahdra, B. (2022). Evolving an idionomic approach to processes of change: Towards a unified personalized science of human improvement. *Behaviour Research and Therapy, 156*, 1–23. https://doi.org/10.1016/j.brat.2022.104155

Hayes, S. C., Ciarrochi, J., Jansen, G., & July. (2023). The next ACT: How 40 Years of development has prepared CBS for what is coming. In *Two-day invited workshop presented at the meeting of the association for contextual behavioral science*. Cyprus: Nicosia.

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2023). The idionomic future of cognitive behavioral therapy: What stands out from criticisms of ACT development. *Behavior Therapy*. https://doi.org/10.1016/j.beth.2023.07.011

Hayes, S. C., Merwin, R. M., McHugh, L., Sandoz, E. K., A-Tjak, J. G. L., Ruiz, F. J., et al. (2021). Report of the ACBS Task Force on the strategies and tactics of contextual behavioral science research. *Journal of Contextual Behavioral Science, 20*, 172–183. https://doi.org/10.1016/j.jcbs.2021.03.007

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change* (2nd ed.). Guilford Press.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology, 7*(1), 141–168. https://doi.org/10.1146/annurev-clinpsy-032210-104449

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., et al. (2022). *Cochrane handbook for systematic reviews of interventions. version 6.3* (updated February 2022).

von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology, 15*(1). https://doi.org/10.1186/s12874-015-0024-z

Honaker, J., King, G., & Blackwell, M. (2011). AmeliaII: A program for missing data. *Journal of Statistical Software, 45*(7). https://doi.org/10.18637/jss.v045.i07

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://doi.org/10.18637/jss.v027.i03

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7). https://doi.org/10.1136/bmjopen-2015-010247

James, G., Witten, D., & Hastie, T. (2014). *An introduction to statistical learning: With applications in R*. New York: Springer.

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin, 147*(1), 1–15. https://doi.org/10.1037/bul0000317

Klimczak, K. S., Schwartz, S. E., Donahue, M. L., Capel, L. K., Snow, J. L., & Levin, M. E. (2023). Disentangling trait and state psychological inflexibility: A longitudinal multilevel approach. *Journal of Contextual Behavioral Science, 29*, 13–22. https://doi.org/10.1016/j.jcbs.2023.05.006

Krafft, J., Klimczak, K. S., & Levin, M. E. (2021). Effects of cognitive restructuring and defusion for coping with difficult thoughts in a predominantly white female college student sample. *Cognitive Therapy and Research, 46*(1), 86–94. https://doi.org/10.1007/s10608-021-10242-4

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist, 36*(3), 276–289. https://doi.org/10.1037/0003-066x.36.3.276

Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy, 43*(4), 741–756. https://doi.org/10.1016/j.beth.2012.05.003

Levin, M. E., Krafft, J., Hicks, E. T., Pierce, B., & Twohig, M. P. (2020). A randomized dismantling trial of the open and engaged components of acceptance and commitment therapy in an online intervention for distressed college students. *Behaviour Research and Therapy, 126*. https://doi.org/10.1016/j.brat.2020.103557

Levin, M. E., Krafft, J., Pierce, B., & Potts, S. (2018). When is experiential avoidance harmful in the moment? Examining global experiential avoidance as a moderator. *Journal of Behavior Therapy and Experimental Psychiatry, 61*, 158–163. https://doi.org/10.1016/j.jbtep.2018.07.005

Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. New York: Sage.

Michell, J. (2019). The fashionable scientific fraud: Collingwood's critique of psychometrics. *History of the Human Sciences, 33*(2), 3–21. https://doi.org/10.1177/0952695119872638

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective, 2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology, 50*(1), 60–69. https://doi.org/10.1002/dev.20262

Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 91*(7), 389–397. https://doi.org/10.1037/ccp0000820

Olkin, I., Dahabreh, I. J., & Trikalinos, T. A. (2012). Gosh - a graphical display of study heterogeneity. *Research Synthesis Methods, 3*(3), 214–223. https://doi.org/10.1002/jrsm.1053

Ong, C. W., Hayes, S. C., & Hofmann, S. G. (2022). A process-based approach to cognitive behavioral therapy: A theory-based case illustration. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1002849

Online Etymology Dictionary. (n.d.). *Method*. Retrieved September 3, 2023, from https://www.etymonline.com/word/method.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Richters, J. E. (2021). Incredible utility: The lost causes and causal debris of psychological science. *Basic and Applied Social Psychology, 43*(6), 366–405. https://doi.org/10.1080/01973533.2021.1979003

Sahdra, B. K., Brockman, R., Hayes, S. C., Hofmann, S. G., Kashdan, T. B., & Ciarrochi, J. (2020). Idionomic links of reappraisal and emotion suppression with positive and negative affect in daily life: Group means often fail to apply to individuals. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/w4u7r

Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Yap, K., Haller, E., Hayes, S. C., … Gloster, A. T. (2023). The compassion balance: Understanding the interrelation of self- and other-compassion for optimal well-being. *Mindfulness*. https://doi.org/10.1007/s12671-023-02187-4

Sahdra, B. K., Ciarrochi, J., Parker, P. D., Basarkod, G., Bradshaw, E. L., & Baer, R. (2017). Are people mindful in different ways? Disentangling the quantity and quality of mindfulness in latent profiles and exploring their links to mental health and life effectiveness. *European Journal of Personality, 31*(4), 347–365. https://doi.org/10.1002/per.2108

Sahdra, B. K., Ciarrochi, J., Parker, P., & Scrucca, L. (2016). Using genetic algorithms in a large nationally representative American sample to abbreviate the Multidimensional Experiential Avoidance Questionnaire. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00189

Sanford, B. T., Ciarrochi, J., Hofmann, S. G., Chin, F., Gates, K. M., & Hayes, S. C. (2022). Toward empirical process-based case conceptualization: An idionomic network examination of the process-based assessment tool. *Journal of Contextual Behavioral Science, 25*, 10–25. https://doi.org/10.1016/j.jcbs.2022.05.006

Schmidhuber, J.r. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schubert, E., & Rousseeuw, P. J. (2019). Faster k-medoids clustering: Improving the PAM, CLARA, and CLARANS algorithms. In *Similarity search and applications* (pp. 171–187). https://doi.org/10.1007/978-3-030-32047-8_16

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Tunç, H., Morris, P. G., Kyranides, M. N., McArdle, A., McConachie, D., & Williams, J. (2023). The relationships between valued living and depression and anxiety: A systematic review, meta-analysis, and meta-regression. *Journal of Contextual Behavioral Science, 28*, 102–126. https://doi.org/10.1016/j.jcbs.2023.02.004

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Villatte, J. L., Vilardaga, R., Villatte, M., Plumb Vilardaga, J. C., Atkins, D. C., & Hayes, S. C. (2016). Acceptance and Commitment Therapy modules: Differential impact on treatment processes and outcomes. *Behaviour Research and Therapy, 77*, 52–61. https://doi.org/10.1016/j.brat.2015.12.001

Viroli, C., & McLachlan, G. J. (2019). Deep Gaussian mixture models. *Statistics and Computing, 29*(1), 43–51. https://doi.org/10.1007/s11222-017-9793-z

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods, 20*(1), 63–83. https://doi.org/10.1037/met0000030

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule - expanded form*. https://doi.org/10.17077/48vt-m4t2

Wilson, K., & Dufrene, T. (2008). *Mindfulness for two: An Acceptance and Commitment Therapy Approach to Mindfulness in Psychotherapy*. New Harbinger. .ISBN-10 : ‎ 1608822664ISBN-13 978-1608822669 ‎ : .