Clinical Psychology Review 129 (2026) 102800

Contents lists available at ScienceDirect

Clinical Psychology Review

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

One size fits none: A call for an idionomic revolution

Steven C. Hayes [a], Joseph Ciarrochi [b], Baljinder K. Sahdra [b,*], Stefan G. Hofmann [c], Clarissa W. Ong [d], Cristóbal Hernández [e]

[a] *Department of Psychology, University of Nevada, Reno, USA*
[b] *Institute for Positive Psychology and Education, Australian Catholic University, Australia*
[c] *Department of Psychology, Philipps-Universität Marburg, Germany*
[d] *Department of Psychological and Brain Sciences, University of Louisville, USA*
[e] *Universidad de los Andes, Chile, Facultad de Ciencias Sociales, Escuela de Psicología*

ARTICLE INFO

Keywords:
Idionomic analysis
Ergodic error
Personalized intervention
Process-based therapy
Eugenics
Intensive longitudinal data

ABSTRACT

Clinical psychology is stalled. Treatment effect sizes have plateaued, syndromal diagnosis has not yielded actionable mechanisms, reliance on WEIRD-centric research limits cultural relevance, and rising mental-health needs continue to outstrip available care. New therapeutic brands keep appearing, yet few new principles or processes emerge, and practitioners still lack scientific guidance for fitting interventions to particular people's lives. These problems have many causes. One is deep and correctable: the ergodic error. Psychologists routinely apply average findings from ensemble-level statistics to particular people without verifying the assumptions that make such applications valid. This error is not incidental to clinical science's history. Population-ranking methods were originally developed for eugenic aims; clinical science retained their logic while abandoning their aims, yielding an analytic system poorly suited to the applied task: predicting and influencing what particular people do, in a particular context, over time. To move beyond this impasse, we propose idionomic analysis, which begins with idiographic modeling of change processes in particular people (e.g., individuals, couples, families), then derives functional groupings and nomothetic principles, retaining only those that deepen idiographic understanding. We discuss supporting algorithms: i-ARIMAX, iBoruta/tsBoruta, GIMME/S-GIMME, GSOM, Multilevel-VAR, idionomic random-effects meta-analysis, and PECAN/PLAN. Matching analytic methods to the demands of intervention is not a technical adjustment. It changes what applied life science can coherently claim, what will replicate, and whether personalized care is built from evidence about the clients being treated, or from averages that may describe no one.

Clinical and applied psychology need a new way forward. For decades, the field has pursued innovation through new diagnostic manuals, newly branded therapies and intervention packages, and a growing number of randomized controlled trials. The APA Presidential Task Force on Evidence-Based Practice found that evidence-based interventions now exist for the major syndromal clinical problems, crossing every major therapy tradition, from cognitive-behavioral to analytic, systems, and humanistic (APA, 2006). Yet this apparent growth masks a deeper stagnation. The core promise of clinical psychology – producing reliable, meaningful change for people who are in distress or pursuing positive life goals – is only partially fulfilled (e.g., Bhattacharya, Goicoechea, Heshmati, Carpenter, & Hofmann, 2023; Cuijpers, Miguel, Harrer, Ciharova, & Karyotaki, 2024; Hofmann, Kasch, & Reis, 2025). Treatment effect sizes remain modest, diagnostic categories remain overwhelmingly heterogeneous, and the proliferation of therapies and intervention methods has far outpaced our understanding of what works, for whom, and under what conditions.

We will argue that this stagnation has several contributing causes, but one root is especially deep, traceable, and correctable, and it has gone largely unrecognized. Clinical psychology's core methods rest on a false assumption: that findings from ensemble-level data can be applied to particular people. This assumption, known as ergodicity, is rarely tested and almost never holds in psychological research. Its roots lie in statistical tools designed for purposes quite different from promoting adaptive change. Until we build methods that start with particular people rather than the aggregate, new therapies, better training, or larger trials will not break the impasse. Later in the paper we will propose idionomic analysis as the basis for a different approach, but first we

* Corresponding author at: IPPE, ACU, Level 9, 33 Berry Street, North Sydney, NSW 2060, Australia.
E-mail address: baljinder.sahdra@acu.edu.au (B.K. Sahdra).

https://doi.org/10.1016/j.cpr.2026.102800
Received 30 March 2026; Received in revised form 2 September 2026; Accepted 3 September 2026
Available online 6 September 2026

need to lay the groundwork for it.

We will not argue that aggregate findings are worthless. Used for the right purpose, they can be valuable. A well-estimated average is a reasonable starting guess when nothing is yet known about a particular person. It can flag candidate processes worth testing for that person or it can serve as a Bayesian prior to be updated by the person's own data. A traditional aggregate-based approach can guide genuine population-level decisions when the data and the decision both concern entire populations or truly random samples drawn from them. An ensemble average cannot stand in for a particular person, couple, or family, however. We describe these legitimate uses, and their limits, more fully below. We flag them here so that the critique that follows is read for what it is: a claim about misapplication rather than a total rejection of traditional aggregate methods.

The ergodic assumption shapes not only how interventions are built but how they travel. A method designed around aggregated averages is already poorly suited to particular cases before it even reaches the practitioner. Behavioral and cognitive science has often not translated well into practice. In the United States, as few as a third of students in training as psychotherapists are adequately exposed to variants of cognitive behavior therapy, and many of the methods that are called "evidence-based" do not reach even minimal standards for that label (Eubanks Fleming, Ernestus, Wenze, & Blomquist, 2025). In practice, use of these methods by psychotherapists appears both to be somewhat limited and to decline over time (Olfson & Marcus, 2010; Wolitzky-Taylor, Zimmermann, Arch, De Guzman, & Lagomasino, 2015). As a result, when particular people seek out psychological help, they may be unlikely to be treated using evidence-based methods (Kazdin, 2008).

Meanwhile those without financial means often cannot access care at all (Funk, Drew, & Knapp, 2012). The worldwide gap between human needs and evidence-based service is shocking. The World Health Organization (WHO) estimates that 75% or more of people with mental disorders in low- and middle-income countries do not have access to the treatment they need (World Health Organization, 2022). Furthermore, when we consider the full range of issues addressed by applied psychology, it becomes clear that behavioral science is often ignored when society faces problems that hinge on human behavior, such as climate-related behavior change, widening economic inequality, and large-scale migration (Clayton, Manning, Krygsman, & Speiser, 2017).

This restricted relevance shows up in the clinic as well. As systems of care have narrowed their gaze to the symptoms of hypothesized latent diseases, the full range of challenges clients face is increasingly unlikely to capture the attention of providers. A "depression clinic" may fail even to ask about chronic pain, tinnitus, financial hardship, or poor nutrition habits; an "anxiety clinic" may leave unexplored issues of loneliness or inadequate exercise (Hayes & Hofmann, 2018). The system discourages clinicians from focusing on prosperity goals such as better relationships, a more cooperative workplace, or a deeper sense of personal meaning, because these concerns fall outside reimbursable treatment for "mental illness" (Hayes, 2019). Meanwhile, only a small fraction of those in distress will receive psychosocial care of any kind, and every year people are more likely to be given only medication when they are distressed. For example, across age ranges, between 15 and 25% of women in the United States take antidepressants (Elgaddal, Weeks, & Mykyta, 2025; Pratt, Brody, & Gu, 2017) – a staggering number that is unwarranted by any fair assessment of appropriate prescribing practice.

The breadth and depth of these trends do not reflect a minor stumble or temporary plateau of progress. Said more directly, applied and clinical psychology are headed in the wrong direction and have been for some time. The field needs a fundamental reconsideration of how we theorize, measure, and intervene in psychological suffering and in the promotion of human prosperity.

A recent umbrella review synthesizing over 100 meta-analyses covering 650,000 patients found average effect sizes of roughly 0.34 for psychotherapy and 0.36 for pharmacotherapy: real but modest benefits (Leichsenring, Steinert, Rabung, & Ioannidis, 2022). Combined treatments performed marginally better, but even that increment may not be valid since it is rare to include a proper control condition for the medication + psychotherapy combination (namely, a placebo + psychotherapy combination). After half a century and massive resource investment, the field appears to have reached its ceiling under current research paradigms.

A common objection to such conclusions is that the science is getting better, and including proper control groups, blinding, training of the therapist and so on, so of course effect sizes are staying flat or are sloping downward because more stringent controls raise the bar. This could be a minor factor, but it seems unlikely to account for all of what is known (Driessen, Hollon, Bockting, Cuijpers, & Turner, 2015). Cuijpers et al. (2024) examined randomized controlled trials (RCTs) on the effects of 125 psychological treatments of depression as compared to clinical care as usual. There was hardly any improvement over the years. The publication year as predictor in the full dataset resulted in a non-significant coefficient of 0.1. Stagnating effects over the years were also found when examining good quality RCTs of one of the best supported methods in all of psychotherapy: CBT for anxiety disorders as compared to psychological or pill control conditions (Hofmann et al., 2025). The result? "Despite 30 years of treatment research, we found no evidence for improvement in CBT effect sizes over time" (p. 8).

The incentives of science admittedly reward new and positive findings over slow, careful follow-up (Munder, Brütsch, Leonhart, Gerger, & Barth, 2013; Smaldino & McElreath, 2016) but that does not explain the poor progress of highly studied areas. Our measures are noisy and our diagnoses broad, so that a single label groups together people with very different problems (Fried & Nesse, 2015; Olbert, Gala, & Tupler, 2014) and we have to consider more deeply why that may be. Something problematic seems to reside within our assumptions, our models, and the way we build treatments. Put plainly, our whole strategy for conceptualizing cases, developing interventions, and testing them has run into a dead end. To understand why people struggle and how to help them, "an as yet unknown paradigm shift may need to occur" (Kupfer, First, & Regier, 2002, p. xix). The purpose of this paper is to propose what that might look like.

A word on the scope and aim of this paper. This is a critical and conceptual review rather than a systematic one. Our goal is to diagnose a foundational problem and argue for a remedy, drawing on the studies that most clearly reveal the problem and the methods that most directly address it. We have updated our reading of the relevant literatures through the first half of 2026 so that the argument reflects current work, including recent independent demonstrations of the problem and recent methodological advances.

This paper synthesizes and extends empirical and methodological work that our group and others have reported elsewhere. Previous pieces have provided specific elements of our present argument such as identifying ways traditional aggregated analysis conceals person-specific effects in individuals and couples (Ciarrochi et al., 2024; Sahdra et al., 2023; Sahdra et al., 2024), or in ways that idiographic algorithms and tools can recover person-level structure and provide avenues of personalization (Li et al., 2024; Li et al., 2026; Ong, Ciarrochi, Hofmann, Karekla, & Hayes, 2024; Sahdra et al., 2026). The present contribution provides the larger historical and conceptual argument that ties these findings together. We will show how an inherited error, the misapplication of ensemble statistics to particular people, came to hide in plain sight and foster a field's stagnation. We will also provide a detailed account of how an idionomic workflow provides the needed correction, in a practical but broadly applicable way.

1. The syndromal model: Aggregate thinking applied to diagnosis

The ergodic assumption runs deeper than our statistical tools. It shapes the very categories we use to understand human suffering. Syndromal classification is built on the premise that averaging symptom patterns across large groups of people will reveal categories that apply meaningfully to each individual within them. But an average across diverse individuals can misrepresent every one of them. That is the ergodic assumption operating at the level of diagnosis. The whole idea behind syndromal classification is that clusters of formal features will eventually lead to the identification of the latent factors that explain the etiological processes, mechanistic course, and response to intervention of specific areas of human suffering (Kupfer et al., 2002). That has not happened – not even one psychiatric disease has emerged from this approach (Insel, 2022; Kupfer et al., 2002).

In lieu of real progress we have created a system of enormous complexity and unknown treatment utility. There are over 10 million combinations of signs and symptoms that can yield a DSM diagnosis, and if specifiers are included, that number soars to 161 septillion (Borgogna et al., 2024). Diagnostic criteria allow for incredible combinatorial heterogeneity – for example, there are 636,120 ways of being diagnosed with Post-Traumatic Stress Disorder (Galatzer-Levy & Bryant, 2013). Despite its aspirations to do so, traditional diagnostic categories apparently do *not* carve nature at its joints, and once we identify these various "disorders" we cannot reliably map effective interventions to them. Most patients receive a diagnosis of "not otherwise specified" (NOS), or have presentations that span multiple categories, rendering category-specific treatment recommendations moot even before the system can be tested (Rajakannan, Bunting, & Dziadkowiec, 2016). The ultimate goal of assessment and diagnosis is treatment utility (Hayes, Nelson, & Jarrett, 1987) but the DSM itself appropriately warns that the current system is not known to have it (APA, 2022). Prediction of individual treatment response remains very weak even in large samples, using contemporary analytic methods. A recent meta-analysis (Curtiss & DiPietro, 2025) showed that machine-learning models can achieve good predictive performance but most studies to date have not tested their models on unseen data, which is the proper standard for evaluating prediction. When we examined their data tables and focused only on models assessed with unseen data (Curtiss & DiPietro, 2025; Table 1),

Table 1
Comparison of key idionomic and process-based analytic methods.

Method	Description	Pros	Cons
i-ARIMAX (Idiographic AutoRegressive Integrated Moving Average with eXogenous variables) (Ciarrochi, Sahdra, Hayes, et al., 2024; Sahdra et al., 2024)	Models individual time-series data by accounting for autoregression, differencing (trends), and moving averages to isolate the strength of an exogenous predictor on an outcome.	• Suitable for modest time-series lengths (e.g., ~20–30 points common in routine clinical care) when signal is strong. • Controls for temporal dependencies and autocorrelation without shrinking individual estimates to a group mean. • Yields easily interpretable betas and standard errors at the person level.	• Restricted to bivariate relationships (one process, one outcome) rather than complex multivariate networks • Does not automatically group individuals; requires secondary meta-analytic or clustering steps • Assumes stability in variance and autocorrelation after detrending.
iBoruta / tsBoruta (Machine Learning Feature Selection) (Kursa & Rudnicki, 2010; Li et al., 2026; Sahdra et al., 2026)	iBoruta is the idiographic application of a supervised machine learning feature selection algorithm (Boruta, a wrapper around Random Forest) that forces features to compete against randomized "shadow features" to identify true predictors; tsBoruta first removes temporal dependencies via ARIMA.	• Highly effective at identifying the most vital variables out of massive candidate pools • Does not require strong distributional assumptions and easily detects non-linear relationships and interactions • tsBoruta significantly reduces false positives caused by autocorrelation.	• Decision trees can be difficult to translate into simple, actionable clinical mechanisms compared to path models • Traditional Boruta (and thus iBoruta) ignores time-series dependencies, which can inflate false positives if not corrected.
GIMME (Group Iterative Multiple Model Estimation) and S-GIMME (Subgrouping-Group Iterative Multiple Model Estimation) (Beltz & Gates, 2017; Gates, Lane, Varangis, Giovanello, & Guskiewicz, 2017; Gates & Molenaar, 2012).	A unified structural equation modeling approach that estimates idiographic dynamic networks first, then iteratively retains shared group/subgroup pathways if they improve individual fit.	• Premier tool for mapping complex multivariate systems • Makes zero assumptions of ergodicity or population homogeneity • • Recovers both contemporaneous and lagged/temporal interdependencies.	• High assessment burden: often requires a minimum of 60 (ideally 100+) longitudinal data points per person • Assumes weak stationarity (relationships do not fundamentally change structure during the measurement window) • Not optimized for focused feature selection of individual process–outcome links.
GSOM (Growing Self-Organizing Maps) (Alahakoon, Halgamuge, & Srinivasan, 2000)	An unsupervised, neural network-based machine learning method that dynamically generates a topology-preserving map of the input space to identify latent structures and subgroups.	• Expands dynamically to match data complexity without forcing individuals into a predefined number of clusters (unlike K-means or PAM) • Captures fine-grained heterogeneity and gradual transitions/continuums between clinical presentations • • "Neighborhood preserving" feature allows for highly interpretable visual mapping of how different response profiles relate to one another.	• Requires careful tuning of hyperparameters (e.g., spread factor) to balance granularity and interpretability; too much granularity leads to fragmentation • Interpreting the final nodes can require a secondary clustering step to establish clear clinical boundaries • Atypical edge cases may remain "unclassified" if they do not fit into any regional topology.
Multilevel-VAR (Multilevel Vector Autoregression) (Bringmann et al., 2013; Epskamp, Deserno, & Bringmann, 2024)	Combines multilevel modeling with vector autoregression to simultaneously model networks of multiple variables across time at both within- and between-person levels.	• Good for exploratory network building when a-priori theory is limited • Captures bidirectional, contemporaneous, and autoregressive associations.	• Relies on "partial pooling" (shrinkage), which pulls individual estimates toward the group mean. • Can underestimate true idiographic heterogeneity, reproducing a form of the ergodic error.
Idionomic RE-MA (Random-Effects Meta-Analysis) (Ciarrochi, Sahdra, Hayes, et al., 2024; Sahdra et al., 2024)	Treats each individual as a "study" to pool idiographic effect sizes via meta-analysis (e.g., from i-ARIMAX) to estimate true heterogeneity and prediction intervals.	• Precisely quantifies how much individual effects deviate from the norm • Wide prediction intervals expose when group averages are statistically unsafe for individual clinical prediction.	• Not a primary time-series estimator; depends entirely on the outputs of base idiographic models.
Perceived Causal Networks (PECAN); Personalized Life Analysis Network (PLAN) Klintwall, Bellander, & Cervin, 2023; Ong et al., 2024.	Networks created by client based on their views of key variables and their inter-relationships.	• Can be done rapidly. • Highly acceptable. • Appears to increase the ability to personalize.	• Based on client memory.

predictive performance dropped to poor discrimination levels.

The current approach is simply not giving practitioners stable nomothetic knowledge (general principles meant to apply broadly) that can be used reliably in the unique, context-bound lives of the people they serve. Clients themselves sense this flaw and complain of it. A recent meta-analysis of client accounts found that when psychotherapy fails, the most common complaint is that the intervention did not fit the person's actual issues or needs (Vybíral, Ogles, Řiháček, Urbancová, & Gocieková, 2024).

2. Intervention science: Aggregate thinking applied to treatment

The same logic that produced an unwieldy diagnostic system also shaped how treatments were built and tested. Interventions are developed by identifying what works on average across groups, then packaged as uniform protocols to be applied to everyone. A treatment optimized for the average patient may fit no particular patient well, and no amount of rebranding corrects that fundamental mismatch. Paradoxically, in parallel to this failure in diagnostic understanding, the field has witnessed a proliferation of therapies and treatment brands. Every year, new protocols, named therapies, and manualized packages appear. Some real benefits emerge from addressing system problems (e.g., Clark, 2018) or by focusing on behavioral health or social wellness domains that have hitherto not been carefully studied, but in established areas in mental health when researchers look at the behavioral kernels, component mechanisms, or change processes addressed, they have found little increase in innovation. For example, a detailed review of child and adolescent psychosocial treatment trials found that while the number of named protocols rose dramatically, the identification of unique practice elements plateaued decades ago (Okamura et al., 2020). Incremental rebranding, new labels, and minor variations are generating novelty of appearance and faddish blooms of interest, but the novelty is of neither substance nor clinical impact.

Meanwhile, our systems of care are under unprecedented strain. Mental health needs have surged, during but also following the COVID-19 pandemic (Kader, Rahman, Bristi, & Ahmmed, 2024). Mental health and substance use conditions are highly prevalent; yet access to care remains uneven, waitlists are long, insurance coverage spotty, and telemental health expansion appears generous in policy but uneven in practice. High demand, constrained supply, and changes in health care funding make it harder to sustain even existing levels of care, much less to test newer interventions rigorously.

Recent meta-analyses offer cause for guarded optimism in an interesting area we will explore in this paper: personalization. Simply allowing practitioners to adjust protocols to fit individual needs, contexts, and preferences reliably outperforms a standardized, one-size-fits-all approach (Li et al., 2024; Nye, Delgadillo, & Barkham, 2023). Yet even here, the gains are modest and unevenly distributed, perhaps because the underlying science provides few useful guides for *how* to personalize interventions. Our analytic tools in evidence-based intervention are based on the assumption that ensemble averages will suffice, perhaps with the addition of a small set of readily understood demographic moderators. That assumption is extremely unlikely to be progressive, however, due to known and arguably irredeemable flaws in that approach.

3. The ergodic error: Historical roots and inescapable consequences

The field did not always pursue aggregate-level answers to individual-level questions. For decades, clinicians and clinical researchers have seen that people vary widely in their needs, trajectories, and contexts, and early attempts to address this were promising. Early behavior therapy was founded on idiographic methods of assessment (Shapiro, 1961a), functional analysis as a means of diagnosis (Kanfer & Saslow, 1965), and single case experimental designs as a means of evaluating outcomes (Hersen & Barlow, 1976; Shapiro, 1961b). Although promising, our idiographic measurement and analysis methods of the time were not up to the challenge (Hayes & Follette, 1992). The science of evidence-based practice soon shifted toward ensemble research methods, driven in part by the rise of DSM-III and the search for latent mental diseases, and accelerated by federal funding decisions that tied research support to syndromal diagnostic categories (Zachar & Kendler, 2012).

The research paradigm that resulted led to an enormous increase in data on psychoactive medications and psychotherapy, especially CBT, but the analytic approach struggled to provide actionable tools for tailoring care to particular people. The study of processes of change was minimized in favor of outcomes tied to technical protocols. Analogue studies on basic principles languished and major clinical research journals such as *Behavior Research and Therapy* developed policies against publishing them at all. Behavioral assessment began to be dominated by psychometrics (Nelson & Hayes, 1986) and gradually withered as a distinct field. Even applied behavior analysis became more purely technological, and gradually stopped citing idiographic basic research (Hayes, Rincover, & Solnick, 1980). All of this made the early principle-focused idiographic vision of behavior therapy, or indeed the yearnings of the humanistic tradition for a new more personalized form of science (e.g., Maslow, 1962; Rogers, 1961) impossible to mount scientifically.

That is all distant history, but it has arguably contributed to the current impasse. We will attempt to show that problems in personalizing care in the modern era arise from an inescapable fact: Most of the applied psychological science now rests on the provably false assumption that aggregate-level analyses can be applied to particular people (Molenaar, 2004). As the early and hopeful trends toward idiographic methods faded, a sugar high from government money encouraged intervention science to adopt an approach that was inadequate for the actual use case that intervention science sought to address. Much was learned, but in the context of our current stagnation a new direction is now needed. We are caught up in an ergodic illusion – an untested assumption of applicability. When that assumption is empirically falsified and its violation ignored, the illusion becomes an active ergodic error which stymies clinical scientific progress.

4. Ergodicity

To understand why this error is not a minor technical quibble but a foundational problem, it helps to examine what ergodicity actually requires mathematically, and how rarely those requirements are met in psychological research. Ergodicity has been researched in statistical physics for over 140 years (Boltzmann, 1884), and its mathematical foundations have been accepted for nearly a century – ever since two world class mathematicians proved the ergodic theorem in back-to-back issues of the Proceedings of the National Academy of Sciences (Birkhoff, 1931; von Neumann, 1932). The ergodic theorem specifies the conditions under which the behavior of an ensemble coheres with the behavior of elements of that collective over time. We can assume such coherence if and only if the phenomena display both stable statistical properties over time and a radical degree of homogeneity in which the same dynamic model applies to all units or samples of units. If these two features apply, the phenomena are "ergodic" and findings can be generalized from an ensemble at a given point in time to elements of the aggregate over time.

These conditions – stationarity and homogeneity – are the classic route to the ergodic error (Molenaar, 2004), but they do not exhaust it: even a stationary process can be non-ergodic when outcomes compound multiplicatively rather than add, so that a single trajectory's average over time departs systematically from the ensemble average at a moment (Peters, 2019; Peters & Gell-Mann, 2016). Non-ergodicity is thus not a single switch but a family of failures – and the multiplicative, path-dependent forms that dominate human change pathways are among the least forgiving, with deep roots in the formal theory of when such long-run averages are defined at all (Oseledets, 1968).

There are indeed ergodic elements in nature such as some noble gases (Hansen, 2018), and ergodic conditions sometimes exist for a wider range of phenomena, but stationarity and extreme homogeneity across units are relatively rare in all of the life sciences and perhaps especially so in psychology. People are unique; psychological events are historically embedded, contextually sensitive, and they evolve over time. Processes of change are by definition not stationary. All of this means that in psychological science and practice, non-random heterogeneity and the existence of trends and cycles have to be assumed. In some unusual situations ergodicity may still apply, but the assumption that it will do so without formal testing is bizarre.

In psychology, the implications of this problem were first laid out clearly by Molenaar (2004), who demonstrated that unless ergodicity holds, results derived from traditional normative ensemble statistics cannot be reliably generalized to any particular person, couple, or family. As Hamaker (2012) observed, this oversight has led to a disconnect between the questions clinicians face when they need to understand what is happening over time with this client, and how these trajectories can be modified – and the answers scientific analyses based on aggregates can possibly provide. The failure to meet the analytic assumptions necessary to generalize from a collective to particular people (individuals, couples, families, and so on) is a threat to human research when it is conducted and analyzed in a traditional normative fashion (Fisher, Medaglia, & Jeronimus, 2018; see also Sundstrom, Lavefjord, Buhrman, & McCracken, 2025, for an independent demonstration in chronic pain).

The hegemony of the ergodic error has impacted even our ability to describe it. "Group vs individual" misdirects the field since the ergodic error applies equally to couples and families (e.g., Sahdra et al., 2023) and real groups cannot be analyzed by group statistical methods without notable adjustments. The alert reader will have noticed we have so far used terms like aggregate, collective, and ensemble methods, contrasting them with methods focused on particular people. From here we will generally use the term "ensemble methods". Its dictionary definition applies and its common connotations are helpful. In a musical ensemble players are obviously not interchangeable – their particular characteristics matter in a way that terms like "collective" or "group" disguise. A drum sounds different from a trumpet or a violin. The musical image is thus a useful corrective, because it makes the central point visible: the average of an ensemble is never a description of any member of it. The solution on the other side of the dialectic is harder. Idiographic systems exist at all levels of complexity from cells to neighborhoods. "Particular people" applies since the word "people" can be plural, but we have so far been unable to fully avoid the word "individual". We caution the reader going forward.

A recent study of a year's worth of articles in three high citation impact psychology journals found that nearly 90% of the articles clearly assumed *without any checking* that the phenomenon being studied was ergodic (Speelman, Parker, Rapley, & McGann, 2024). The authors concluded that theories and interventions built on ensemble data are unlikely to apply to particular people: a problem that cuts to the heart of how clinical psychology currently operates.

The consequences for psychology and for all the behavioral and life sciences are far from abstract. An example from our research team is characteristic: the assessment of mindfulness. Conventional psychometrics presumes that deviations from ensemble-level factor structures are random error, but intensive longitudinal research shows otherwise. Hernández, Sahdra, Ciarrochi, and Hayes (2025) analyzed the relevance of emotional awareness to sadness and anhedonia assessed longitudinally and idiographically, person by person. Mindfulness processes reliably predicted statistically significant increases in sadness and anhedonia for some people but decreases for other people. Because traditional ensemble statistical analyses link mindfulness to higher well-being (Sahdra, Ciarrochi, & Parker, 2016), a very wide range of professionals – clinicians, physicians, teachers, coaches, and workplace advisers – now recommend it for almost any difficulty, from stress and anxiety to low mood, poor focus, sleep problems, and general dissatisfaction. Yet the idiographic data show that its effects vary so widely across individuals that recommendations based on between-person averages can fail to fit or even actively harm particular people.

Many if not most psychologists are well aware that ensemble averages do not apply to particular people, but they falsely assume that modern normative statistical methods have resolved the problem. Multilevel modeling, for example, appears to address individual variation by modeling individual growth curves while also aggregating data. The appeal is understandable: if the method estimates individual trajectories, surely those estimates can now guide individual clinical care. Unfortunately, this reasoning overlooks a critical feature of how multilevel modeling actually works. Multilevel modeling includes a shrinkage factor in which individual deviations are differentially weighted based on their model-assumed precision. This statistical adjustment can seriously distort individual findings, as we will show.

Authoritative texts on multilevel modeling are explicit about this adjustment. Raudenbush and Bryk (2002; e.g., see pp. 39-46) describe how individual estimates are pulled toward the grand mean, a procedure they call shrinkage. Shrinkage estimators are not simple averages. They are empirical Bayes (EB) estimates that are weighted compromises between the group mean and the overall grand mean, biased toward the latter. Statistically speaking, the shrinkage estimator in multilevel modeling was developed with logical and well-intentioned goals. By pulling individual estimates toward the grand mean, shrinkage serves to prevent overfitting and stabilizes estimates when individual-level data is sparse or unreliable (Raudenbush & Bryk, 2002). If the primary goal of an analysis is to generate a stable population parameter that is less affected by random fluctuations, this algorithm is appropriate (Liu, Moeller, & Kospi, 2019). However, a profound problem arises when these exact same adjusted estimates are repurposed to understand particular people in clinical settings – as might reasonably be concluded when the method supposedly first models individual trajectories. The problem is that by artificially restricting variance for the desired statistical good, the shrinkage factor can have a massive impact on individual trajectories if heterogeneous units are pooled together. It routinely erases or even reverses the direction of true person-specific effects, thus creating the illusion of homogeneity where none actually existed (Sahdra et al., 2024).

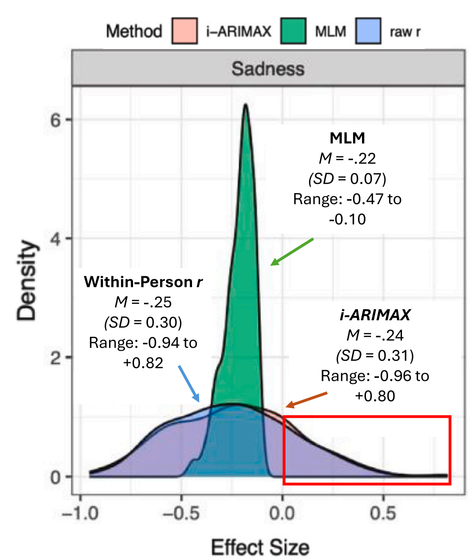

Fig. 1. Comparison of three analyses of the same dataset. Adapted from Sahdra et al. (2024), *Journal of Contextual Behavioral Science*, 32, 100728, under the terms of the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/). Changes were made.

Fig. 1 provides a graphic display of the degree of distortion that can result. In a large experience sampling dataset, Sahdra and colleagues (Sahdra et al., 2024; see also Catts et al., 2025) used idiographic time-series models to examine the relationship between valued action and affect. This figure shows the distribution of idiographic results of three different ways of modeling the degree of relationship between "doing what matters" (a values-based action item) and entanglement with sadness – a one item measure sometimes used as a proxy for "depression" in Ecological Momentary Assessment (EMA; Bolger & Laurenceau, 2013) studies. The three methods were (1) idiographic raw correlations of the sort one might readily calculate in a simple spreadsheet for each individual; (2) idiographic estimates that are corrected for trends, autocorrelation and moving averages, which is a more proper way to calculate within-person effects; and (3) multilevel modeling. All methods had approximately the same central tendency – at the level of the ensemble, valued action is negatively correlated around −0.25 with entanglement with sadness. Said simply, in terms of central tendency all methods agreed that moments when people engage in values-based action are somewhat less likely to be moments in which they struggle with sadness and vice versa. The three distributions differed dramatically, however.

Traditional correlations and those corrected for auto-regression and moving averages showed a similar, broad, and almost entirely overlapping distribution, ranging from nearly −1 to around +0.8. Multilevel modeling (in green) showed a much tighter distribution with a range of about −0.5 to −0.1 eliminating the very information clinicians needed to know about.

Since MLM is a more modern and "sophisticated" method that is said to address individual trajectories, a typical researcher publishing such a study (or a clinician reading it) would likely only be aware of the green results. The MLM analysis conveys a nearly universal nomothetic conclusion: when people engage in values-based action they are somewhat less likely to struggle with sadness. That is false, and misleadingly so for a practitioner who may end up then treating "depressed" people in the red box – for whom values-based action tends to be associated with *more* entanglement with sadness. As we will document later, as these within-person values deviate further from zero, more and more individuals show a statistically significant relation that cannot be accounted for by mere random heterogeneity. Some people truly are more miserable the more values-focused they are. For them, targeting values without other changes occurring could even be iatrogenic. That risk is clear with raw or correctly calculated idiographic correlations, but it disappears when a more "sophisticated" statistical method is used that supposedly models individual trajectories.

When confronted with the reality of non-ergodicity, one understandable reflex is to seek salvation in complexity: better machine learning algorithms, larger datasets, and more precise moderators will handle heterogeneity. Such approaches still fundamentally rely on between-person variance. They attempt to answer the question, "What works for a person *like* this?" but without first adequately characterizing anyone in an idiographic sense. In other words, the "this" that is being referred to is not a person at all – it is an analytic abstraction, leaving no empirical foundation for the far more important question "What works for *this* particular person or couple or family?" (Fisher et al., 2018). Unfortunately, such approaches rest on a category error. Slicing a cross-sectional aggregate thinner does not transform it into a longitudinal model of particular human lives – it merely yields smaller pools of non-ergodic data. A science of intentional change cannot be salvaged by sophisticated algorithms if the algorithms are fed data that obfuscate the very phenomenon of within-person change (Molenaar, 2004). You cannot model the dynamic trajectory of a flying bird by endlessly subdividing a photograph of a flock.

5. The fit with the applied use case

Our point is not just that particular human lives are distorted by modern statistics that claim to characterize them. It is also that analyses based on ensembles actively disguise the true nomothetic findings that practitioners need. This requires an additional terminological clarification that cuts to the heart of the problem.

The word "nomothetic" literally means general laws. Practitioners and researchers need general principles that can be reliably applied to particular human beings: individuals, couples, families, and other groups. But ensemble analyses have come to be called "nomothetic methods," as though studying ensembles of people is the same as deriving generally applicable principles. That terminology contains a costly scientific fib that is as common as the nearest introductory psychology textbook. Unless the phenomenon being studied is ergodic, ensemble analyses cannot yield knowledge that is both nomothetic and mathematically valid for the applied use case.

The unit of analysis in applied work may not be a single person. A clinician may need to understand a particular couple, family, or organization, and ensemble statistics cannot reliably guide that work any more than they can guide work with individuals. As we have noted, even researchers aware of the ergodicity problem can fall into this trap when they frame the solution as moving from "group-level" to "individual-level" analysis, implicitly excluding couples, families, and organizations from view.

The consequences of this confusion are not merely terminological. In Sahdra et al.'s (2024) study, nomothetic findings appeared only when idiographic methods were used first, before searching for distinct subgroups. Termed "stoics" by the authors, for this subgroup valued action was unrelated or even negatively related to joy but stoics persisted in the face of challenging circumstances; for another subgroup, ("non-stoics"), the positive relationship held (valued action was associated with joy) but stressful circumstances undermined joy. Building on this work, Catts et al. (2025) replicated the finding in a transdiagnostic clinical sample.

This replicated finding illustrates a one-two punch directed at the heart of the applied use case. Not only was the substantial non-random heterogeneity in the linkage of valued action and mood masked by pooled analyses, these same so-called "nomothetic analyses" failed to yield nomothetic findings practitioners can actually use.

Applied and clinical psychology has entered a reality distortion field, ensnared by assumptions that cannot be met and terms that obscure the true situation. It is hard even to describe how to escape the cul-de-sac clinical science has entered when what has entrapped us has been commonly described as essential features of behavioral science. This is why we have found it necessary to introduce the new terms you will soon encounter in this paper – such as "idionomic" (Hayes & Hofmann, 2021) or "equisyncratic" – not out of fondness for neologisms but because the existing vocabulary makes the problem we are facing nearly impossible to name clearly.

6. Characterizing the ergodic error

Fortunately, there is a well-developed area of behavioral science that allows us to approximate an assessment of the ergodic error. In meta-analyses well-established metrics such as I^2 or Q exist that address how much of the difference in effect sizes between studies are likely due to non-random heterogeneity. If more than 50% of the differences between studies across sites are due to real but unknown site or study features (that is if I^2 is greater than 0.5), researchers are asked by organizations such as the Cochrane Collaboration to be highly cautious about statements of central tendency. If more than 75% of the differences are due to non-random heterogeneity, averaging effects across individual studies should be avoided entirely or interpreted as essentially non-informative. As we discuss below, I^2 is a relative rather than absolute index and these cutoffs are only rough guides; in our own work we rely on τ and the prediction interval, but I^2 is a familiar entry point for readers.

This same approach can be applied to aggregated samples of idiographic data when there is enough longitudinal information to calculate relationships person by person (or couple by couple, etc.), and there are enough people in the study to consider the resulting distribution of idiographic results properly. In the experience of our research team, when process–to–outcome relations are modelled that way, I^2 values generally exceed the Cochrane Handbook's 75% cutoff for "considerable heterogeneity". Levels reaching 80% or 90% are common, and values below 50% are the exception (Hayes, Ciarrochi, Hofmann, Chin, & Sahdra, 2022; Moritz et al., 2024).

In many ways this revelation is familiar. Such findings echo earlier demonstrations by Cattell (1952) with the P-technique, or decades of warnings about the ecological fallacy (Estes, 1956; Nesselroade, 2001). In biology, they parallel concerns that population genetics cannot fully explain evolutionary change because it ignores how path-dependent outcomes are, and how much they are impacted by specific histories and contexts (Wilson & Sober, 1994). What is new is the awareness that these problems can be traced back to the assumption of ergodicity that is built into our normative analytic methods. It then becomes obvious that a fundamental failure to meet necessary assumptions can never be repaired merely by ignoring the failure.

The implications are direct. Ensemble analytic tools such as regression, ANOVA, and multilevel models all rest on assumptions that do not hold for human change processes, and more data or larger samples cannot repair that mismatch. The alternative explored in this paper, idionomic analysis, begins with idiographic data collected at a temporal scale appropriate to the processes of interest, and if the data are non-ergodic, proceeds to model dynamic intraindividual relationships, and then search for nomothetic generalizations that improve idiographic understanding (Ciarrochi, Sahdra, Hayes, et al., 2024; Hayes & Hofmann, 2021; Hayes et al., 2022; Hofmann & Hayes, 2019, Hofmann, Hayes, & Lorscheid, 2021; Hofmann, Curtiss, & McNally, 2016; Hofmann, Curtiss, & Hayes, 2020). Personalization has been promised for decades. Without first addressing the ergodic error, it cannot be delivered.

Before we lay out the details of an idionomic approach, we must understand how an applied field came to adopt methods that cannot serve its actual use case. The failure to recognize the ergodic error is not merely technical; it reflects the historical forces that shaped biostatistics and, in turn, shaped psychology. It is time to confront our own history.

7. The ergodic error has a history: Eugenics, amnesia, and inherited assumptions

The ergodic error did not arise from innocent statistical convenience. The tools psychology treats as scientific bedrock were built by people with explicit ideological goals: to rank, classify, and control populations, not to understand particular human lives. That origin may help to explain why the assumption of ergodicity was never seriously questioned. When tools are designed to characterize aggregates, the question of whether aggregates describe individuals never arises. Statistical physics has understood for nearly 150 years that aggregates generalize to particulars only under strict conditions (Boltzmann, 1884), yet psychology has only begun to confront this reality in the last two decades (Molenaar, 2004). We still have not been able to find a single introductory psychology text that even mentions the word "ergodicity." After World War II, when eugenic atrocities became undeniable, the field chose amnesia over reckoning. The history was scrubbed from textbooks, journals were rebranded, and the underlying analytic assumptions were inherited intact.

Although understanding this history is morally necessary, our primary focus in this paper is scientific. Scientifically speaking, statistics are not invalidated by the purposes of their inventors. A method is not refuted by the aims of those who built it. The point is narrower and more durable: the tools were designed to rank and classify aggregates, and the field inherited the ergodic assumption embedded in them without ever asking whether aggregates describe individuals. We cannot correct a scientific error we have not named.

For almost all of our history as a field, "good science" has been *defined* in psychology as adherence to methods and analytic tools that require ergodicity to fit the use case of applied and clinical psychology, even though ergodicity is a mirage for applied psychology. There is a kind of reality distortion field in psychology that is many decades old. Casting a light on it helps practitioners and researchers alike sense how hard and necessary it is to change our direction in fundamental ways. It has been 75 years since eugenics collapsed in the field, but it is telling that this appears to be the first major paper to draw the link between eugenics, the ergodic error, and our analytic methods. We as a field need to take responsibility and make the needed repairs on behalf of our discipline. The scientific reasons are sufficient. The moral ones remove any excuse for delay.

The reach of that error is wide, touching personality testing, factor analysis, IQ research, mediational analysis, and the application of randomized trial results to particular clients, but its origin is singular. Francis Galton set the stage. His 1869 book *Hereditary Genius* was less a neutral inquiry into human variation than a blueprint for breeding "eminent" lineages while discouraging the "unfit" (Galton, 1869). To achieve this, he and his close colleagues invented tools like the standard deviation and correlation, explicitly for the purpose of ranking humans on anthropometric and intellectual traits (Galton, 1889). These methods were dry-tested in agriculture and beer making, where there was little interest in *particular* seeds, but they were immediately applied to areas where particular human lives were central. There, the larger purpose was far darker. Galton was clear: these tools were to give us "cognizance of all influences that tend in however remote a degree to give more suitable races … a better chance of prevailing speedily over the less suitable" (Galton, 1883, p. 24). In order to translate what he claimed were the scientific implications of objective statistical methods into public policy, Galton created the field of eugenics.

Behavioral science eagerly adopted both normative ensemble statistical methods and their underlying policy purpose, especially in the United States, where they aligned both with post-Darwinian progressivism and with social agendas of exclusion and hierarchy in the post-Civil War era. Galton's disciple Karl Pearson formalized regression and the chi-square test while promoting racial hierarchies in his writings and lectures. He founded the journal *Annals of Eugenics*, subtitled "*A Journal for the Scientific Study of Racial Problems*" and warned for decades against Jewish immigration to Britain, arguing it would dilute the nation's "stock" (Porter, 2004). In that journal's first volume, Pearson wrote that "only the superior stocks should be allowed entrance, not the inferior stocks in the hope – unjustified by any statistical inquiry – that they will rise to the average native level by living in a new atmosphere" (Pearson & Moul, 1925, Part II, p. 126–127). For our argument, the critical phrase is "*unjustified by any statistical inquiry*" because it contains the mathematically false assumption that methods of ensemble statistics could predict the life trajectories of particular people and how they might be impacted by a changed environment. Ronald A. Fisher, Pearson's editorial successor and also a professor of eugenics, extended these ideas into analysis of variance and experimental design, while advocating for policies that rewarded "superior" families with financial incentives for having children (Fisher, 1918; Mazumdar, 1992). The methods and the ideology were, for Fisher as for Pearson, inseparable.

These tools quickly permeated psychology. Charles Spearman applied Pearson's correlation to invent factor analysis, positing a latent "g" factor that could rank individuals hierarchically (Spearman, 1904) – a goal that fit seamlessly with the eugenic agenda. In the United States, Lewis Terman adapted the Binet-Simon scale into the Stanford-Binet IQ test, arguing that low scores marked "feeblemindedness" warranting reproductive restrictions (Terman, 1916).

Terman's involvement with the Iowa Child Welfare Research Station (ICWRS) provides a stark example of how eugenic biases corrupted psychological science (Brookwood, 2021). Although researchers at ICWRS had carefully found in a controlled investigation that environmental factors could significantly boost IQ scores in disadvantaged children (Skeels & Dye, 1939), Terman actively distorted these data to fit genetic determinism, suppressing contrary evidence from other labs, and even threatened to out a leading ICWRS researcher, Harold Skeels, as a gay man if he continued to stand behind his data – a career-ending act at the time (Brookwood, 2021; Cravens, 1992; Gould, 1996). Robert Yerkes' Army Alpha and Beta tests during World War I took this further, testing 1.7 million recruits under rushed, biased conditions, "showing" Black and immigrant soldiers to be intellectually subpar, fueling the Immigration Act of 1924, which curtailed non-Nordic European entry (Kevles, 1985; Yerkes, 1921). Henry Goddard tested immigrants at Ellis Island and concluded that many were "morons" unfit for society (Goddard, 1917) despite unfair testing conditions (Zenderland, 1998). Even John B. Watson, the founder of behaviorism – a tradition that ostensibly emphasized environment over heredity – served alongside Yerkes and Terman on the three-person American Eugenics Society's Committee on the Genetic Basis of Behavior (Cohen, 1979). Without the invisibility of the ergodic error, none of these steps would be scientifically justified or even rational. Group rankings cannot describe individuals unless ergodicity holds, and it did not.

The human cost of eugenics was immense. By the 1920s, 32 U.S. states had eugenic sterilization laws, leading to tens of thousands of forced procedures, disproportionately targeting women, the poor, and minorities (Lombardo, 2008). The Supreme Court's 1927 Buck v. Bell decision upheld this, with Justice Oliver Wendell Holmes famously declaring, "Three generations of imbeciles are enough" (Lombardo, 2008). These laws drew directly on normative psychological evidence. Across the Atlantic, Eugen Bleuler, who coined "schizophrenia," endorsed sterilizing those with mental illnesses (Bleuler, 1911). Nazi lawyers cited U.S. sterilization statutes word for word when drafting the 1933 "Law for the Prevention of Hereditarily Diseased Offspring" (Proctor, 1988). This law led to Aktion T4, which murdered hundreds of thousands of people with disabilities and mental illnesses between 1939 and 1941 in facilities set up for their compassionate "care" (Friedlander, 1995). Physical and psychological tests were required of all citizens, with results reported directly to the government. Facility staff received daily orders of the patients who were to die each night. Families were lied to with explanations made plausible by personal information (e.g., "the wart on your mother's foot became cancerous") gathered in the nationwide professional screening. The tools of early biostatistics, used to identify and classify the undesirable, had organized a trial run for the Holocaust.

Scientific psychology and behavioral science, especially in the United States, contributed to the deaths of millions. The mechanism was the embrace of normative biostatistics as a tool for predicting the trajectories of particular human lives, combined with diagnosis and testing pressed into the service of eugenic policy. It is a common lament today that society has not listened to psychological science. Seen in historical context, that lament needs to be turned on its head. The problem is not that society has not listened. The problem is that it did.

As eugenic atrocities became undeniable after World War II, a collective sense of shame and guilt emerged within the biobehavioral sciences. Such emotions can foster responsible acts of repair in human beings, but in part because of the invisibility of the ergodic error, the field opted for amnesia over acknowledging and changing its analytic assumptions. Journals like *Annals of Eugenics* were eventually rebranded as *Annals of Human Genetics*, while retaining much of the same editorial board, focus, and empirical strategies (Mazumdar, 1992). Textbooks presented correlation, regression, and factor analysis as value-neutral pillars of scientific method (Tucker, 1994). Even today, very few introductory textbooks mention that Terman, Fisher, Pearson, Frank Yates, and many others were eugenicists. The APA itself states that the majority of its Presidents from 1892 to 1947 were eugenic advocates (APA, 2021; see also Yakushko, 2019).

No serious scholar today defends eugenics, and using ensemble statistics of course does not equate to being a racist or antisemite. But our field has not yet grappled with the possibility that some analytic assumptions from that legacy are limiting psychological science. *Scientific errors allowed behavioral science to play this role, and these errors continue to the present day.* Those errors cannot be corrected until the field understands what they were. Of the few textbooks that do actually address psychology's role in eugenics in a major way, they all do so purely as a matter of history. We have been unable to find even one that mentions ergodicity or links that issue to assumptions central to this inherited legacy. To the contrary, every introductory psychology textbook lauds the applicability of methods built on assumptions that can virtually never be met in the actual use case of most interest to our field.

This is not yesterday's news. The late Richard Herrnstein's still-popular *The Bell Curve* (Herrnstein & Murray, 1994), drawing directly on Galton's methods, argued for cutting programs targeted at disadvantaged children "who are also disproportionately low in cognitive ability" (p. 402). The same logic runs through subsequent works treating national IQ as evidence for inherent population differences (Lynn, 2008), arguing genetic divergence explains cross-societal economic inequality (Wade, 2015), or contending that DNA is the dominant shaper of psychological outcomes (Harden, 2021; Plomin, 2019). In each case, ensemble genetic findings are treated as applicable to particular lives without testing whether that generalization holds mathematically – it does not. The ergodic error is equally present across the political spectrum: popular books and programs proliferate suggesting that everyone needs to eat in a particular way (e.g., Inchauspé, 2022), think in a particular way (e.g., Shetty, 2020), parent in a particular way (e.g., Kennedy, 2022), or exercise in a particular way (e.g., Starrett & Starrett, 2023). Each makes the same logical move: ensemble findings are presented as prescriptions for particular lives without testing for ergodicity. That is an error. Progress itself demands a change.

8. Why ergodicity has caused our field to fail the applied use case

The historical roots of the ergodic error are clear. What remains to be shown is exactly how it plays out in the moment-to-moment work of clinical practice, and why no refinement of current methods can repair it. When providing an intervention, a practitioner faces multiple decision points: identifying relevant information sources, defining problems or goals, understanding origins and maintenance factors, selecting and sequencing interventions, developing workable explanations, monitoring response, determining termination, and so on. For evidence-based practice, another critical decision emerges at each step: how to integrate the best available evidence.

The most common approach to such integration is the protocol-for-syndromes approach (Ciarrochi et al., 2024), where treatment manuals are implemented based on their efficacy for those with the symptoms of specific syndromes randomized trials carefully assess. A lack of ergodicity fundamentally undermines this approach as a route to needed intervention personalization. Treatment manuals typically define target problems, theoretical bases for intervention methods, and descriptions of principles and techniques (Kiesler, 1994; Wampold & Imel, 2015). These manuals became the gold standard for testing efficacy through RCTs, enabling more standardized treatment delivery (Mohr, 2024) which was thought to be key in disseminating empirically supported treatments (Becker, Smith, & Jensen-Doss, 2013). Empirical support required demonstrating superiority in well-crafted group designs (Chambless & Hollon, 1998; Tolin, McKay, Forman, Klonsky, & Thombs, 2015). While manuals did create a more homogeneous independent variable and delivery, and RCTs did provide needed evidence of efficacy at the ensemble level, the ultimate adequacy of both changes for intervention with particular people rests on the assumption of ergodicity.

Broader approaches such as the evidence-based psychological practices framework (APA, 2006), suggest flexible integration of best available evidence, clinical expertise, and client characteristics. The data on personalization (e.g., Li et al., 2024) recommend this second and more flexible approach, but unfortunately, current psychological science largely cannot provide replicable scientific guidance for *how* practitioners should integrate these sources of information, nor how practitioners can gain the expertise to see problems in processes of change in flight so that personalization is reliably based on actual client needs. The gap is not a failure of clinical intention. It is a failure of scientific foundation. What clients receive is often more determined by the training background of practitioners than the features of the client (Cook, Biyanova, Elhai, Schnurr, & Coyne, 2010). Without a science built around particular people, even the most flexible practitioner has no empirical foundation for personalization.

Non-ergodicity is far more the rule than the exception when dealing with psychological processes (Fisher et al., 2018; Hunter, Fisher, & Geier, 2024; Molenaar, 2008) given that the historical and situational context may alter the many functions a single biopsychosocial event can have. For instance, using EMA methods, Sahdra et al. (2025) found a small average negative contemporaneous association between moments in which participants worried that their positive emotions were fading and their experience of happiness. Especially because these results were based on longitudinal research, a practitioner might be tempted to extrapolate them to the within-persons realm and to apply the "take home" idea that worrying about losing positive emotions modestly decreases happiness to their particular clients. The practice base can perhaps be forgiven for this error since analytic level-switching (from the ensemble to the particular) is extremely common in psychological research (Speelman et al., 2024), and indeed most biopsychosocial research. But that kind of switching is exactly what non-ergodicity prohibits.

Fig. 2 shows a different picture of the Sahdra et al.'s (2025) finding, by providing a plot of the relationship between worrying that positive emotions are fading and happiness when this relationship is examined *one participant at a time* and then summarized without averaging. Individual time series were analyzed with an idiographic Autoregressive Integrated Moving Average with Exogenous Covariate algorithm (i-ARIMAX). Green horizontal lines signal statistically significant positive within person effects, red horizontal lines show statistically significant negative within person effects, and black horizontal lines show indeterminant within-person effects. The dots in each line show the ARIMAX beta estimate, and the line width shows the confidence interval for this estimate. The vertical blue lines represent the pooled effect of a Random Effects Meta-Analysis (RE-MA) with a 95% confidence interval.

The small effect size was not a good representation of the collective, as it was composed of positive and negative effects that cancelled each other out. Not a single person showed a significant idiographic effect within the confidence interval of the pooled effect, while 36% of the analyzed sample (58 of 151 analyzed participants) showed significant

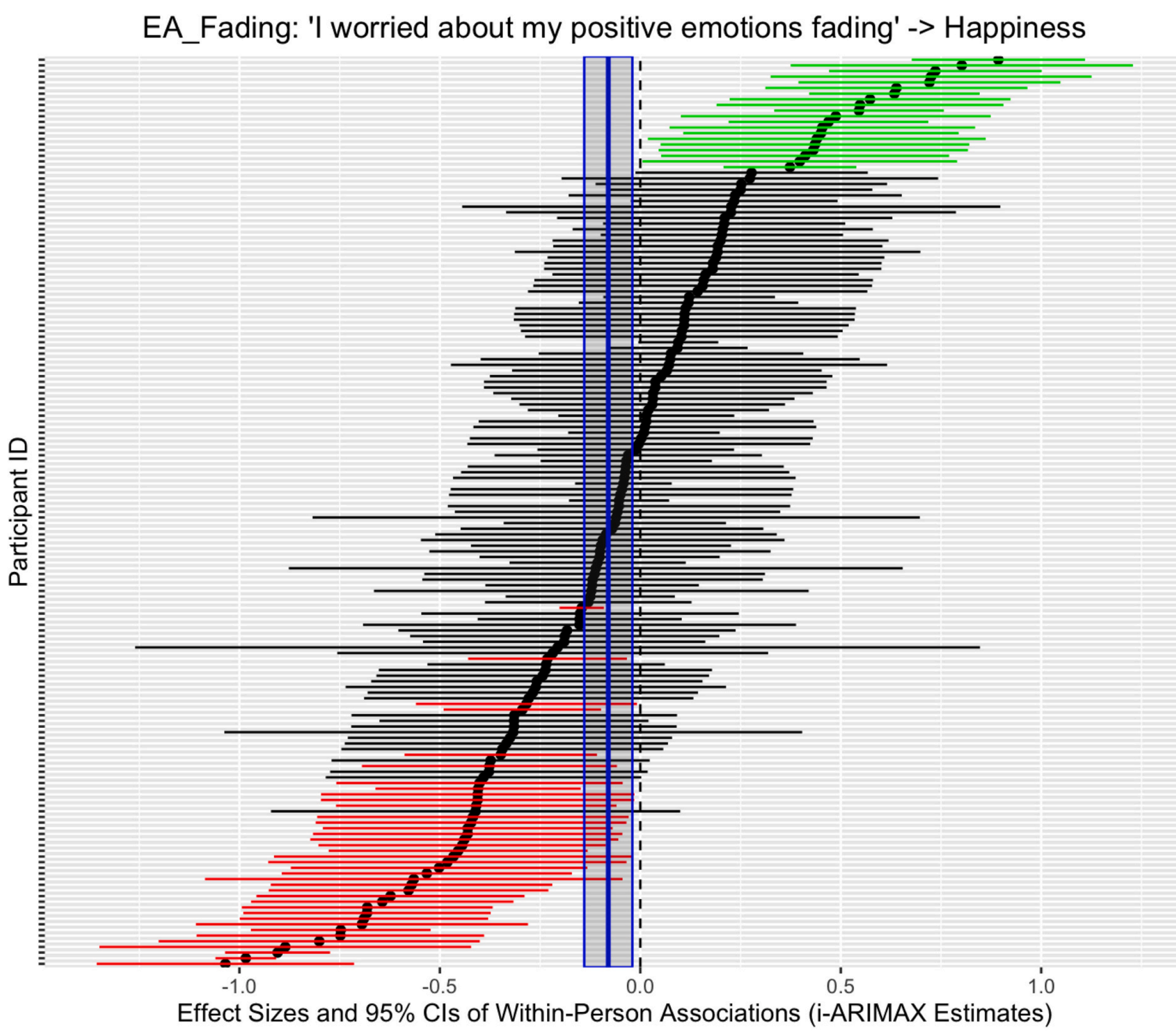

Fig. 2. i-ARIMAX results for the relationship of worrying about positive emotions and happiness, organized by the idiographic estimate. Reproduced from Sahdra et al. (2025), Journal of Happiness Studies, 26, 101, under the terms of the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/). See text for explanation of the figure.

idiographic effects outside of that interval. Of those 58 participants, 20 (34.5%) showed significant negative effects and 38 (65.5%) showed significant positive effects. The small negative average effect simply does not characterize what a practitioner will likely see, based on the size and direction of the relationships observed.

To fully grasp how ensemble statistics actively conceal clinically vital information, consider the phenomenon of "equisyncratic" relationships (Ciarrochi, Hernández, Hill, et al., 2024). Equisyncratic relationships are processes of change that display no significant average effect at the ensemble level but possess significant and highly heterogeneous effects at the individual level. For example, in a recent study social behaviors such as being assertive or expressing feelings had a pooled average effect of around zero on loneliness at the ensemble level (Ciarrochi, Sahdra, Hayes, et al., 2024). If a practitioner relied solely on the ensemble average, these variables appeared to be clinically inert. However, idionomic analyses found that for many individuals, expressing feelings reliably decreased loneliness, whereas for others, it reliably increased loneliness (Ciarrochi, Sahdra, Hayes, et al., 2024). The positive and negative individual realities perfectly cancel each other out at the level of the aggregate – but no practitioner ever sees or treats the mythical aggregate.

The statistical tools that became psychology's methodological bedrock were designed for contexts where particular cases do not matter: crop yields, barley plots, manure treatments and the like (Box, 1987; Fisher & Mackenzie, 1923; Parolini, 2014). Average-based methods suit average-based applications such as these. The problem is that psychological practice is not an average-based application, and when the eugenic era passed, these tools remained, carrying their assumptions intact into a use case they were never designed to serve.

The long-term cost shows up in three issues that have resisted solution despite decades of methodological refinement: stagnating effect sizes, the replication crisis, and the gap between science and practice.

Since Smith and Glass's (1977) first meta-analysis demonstrating psychotherapy efficacy, average treatment effects have remained stable or deteriorated (e.g., Hofmann et al., 2025; Johnsen & Friborg, 2015). A recent meta-analysis found positive average effects across 12 mental health problems, but most prediction intervals (except for OCD) included zero, meaning future studies could plausibly yield negative results (Harrer et al., 2025). At the ensemble level there is modest progress, but current gold-standard treatments (both psychosocial and pharmacological) fail many, with some experiencing deterioration (Cuijpers et al., 2021).

When an RCT pools positive, neutral, and negative individual effects into a single estimate, the resulting intervention is designed around an average that may describe no one in the sample. Larger studies, better control groups, blinding, better training of study therapists, preregistration, and so on cannot rectify a faulty assumption. Some proposed reforms, including the systematic monitoring of negative effects in RCTs, could help (Honkalampi, Urhonen, & Virtanen, 2024). Yet when the phenomena are non-ergodic, such analyses must include high-density idiographic data and idionomic methods to separate random variation from true person-level heterogeneity. The metrics currently collected in intervention science were never designed to capture what matters most: whether this intervention helps this person. The only way to meet the needs of particular people is to create a science of particularity, but that is logically and mathematically impossible if our studies are based on the ergodic error.

The ergodic error also provides an additional lens through which to view the replication crisis. In non-ergodic complex systems, initial conditions matter enormously, history and path-dependence profoundly affect outcomes, and contextual effects compound in ways that simple additive models cannot capture. Since population averages do not represent individual processes, normative research does not have access to the information needed to develop nomothetic principles that may be able to simplify this complexity. When honest and well-crafted studies fail to replicate findings, they may simply be reflecting the very nature of the heterogeneous, path-dependent landscape of human functioning (Fisher et al., 2018; Peters, 2019). That does not reduce concerns over questionable research practices (Tyner et al., 2026) but it does call for the careful disentanglement of these possible sources, especially in areas where direct replication is difficult and populations are poorly characterized. For example, a treatment that reduces depression in one sample may fail in another not because the original finding was false or based on traditional questionable research practices such as "p hacking", but because the participants have different developmental histories, face different contextual pressures, or have different intervention histories, and all of this may interact with therapists' actions – precisely the kind of non-random heterogeneity that idionomic analysis is designed to detect and model. Viewed through this lens, beyond scientific malpractice, the replication crisis may sometimes include a level of analysis problem. Lack of replication is the natural consequence of applying statistical methods that assume ergodicity to massively non-ergodic phenomena. Better methodology within an ensemble framework can never fully solve this problem, but the adoption of idionomic approaches can do so at least in principle, if they help psychologists identify when, for whom, and under what conditions effects reliably occur.

The gap between science and practice is usually blamed on practitioners, including the trend toward psychotherapists and other helping professions having lower-level degrees and more limited scientific training (Baker, McFall, & Shoham, 2008). But this framing may have it backwards. Because applied clinical work operates case-by-case, part of that gap may exist because practitioners are sensibly sensitive to the limitations of group averages and protocols, but the science that informs practice minimizes these factors (Hayes, Barlow, & Nelson-Grey, 1999). Only 7.8% of psychotherapists report using manuals frequently, 51.4% reported using them to some degree and 40.7% reported never using them at all (Becker et al., 2013). Even when manuals are indeed used, therapists often add modifications that are not always consistent with the principles laid down in the protocol, a practice known as *therapist drift* (Speers, Bhullar, Cosh, & Wootton, 2022; Waller, 2009). These may be examples of practitioners trying to make the best of a bad situation, especially since practitioner personalization appears to increase effect sizes (Li et al., 2024).

Effective evidence-based practice requires translating scientific findings into guidance for particular people (Rousseau & Gunia, 2016). In other words, a practitioner needs to develop practical principles based on an empirical understanding of the problems to solve. If that scientific understanding is based on the ergodic fallacy, the confrontation with applied problems can increase rather than decrease the science to practice gap given that inconsistencies will be found. The gap is not a failure of practitioners to apply the science. It is a failure of the science to apply to the intended use case.

A reasonable question is whether ordinary clinicians, working in ordinary settings, can actually do this. We think a usable version is already within reach, and that it can start simply and grow. A clinician does not need hundreds of observations to begin. Brief, repeated self-report a few times a week across a handful of personally relevant variables is enough to support simple idiographic models such as ARIMAX, and rapid client-built tools such as PECAN and PLAN can produce a personalized network in a single session. For example, a clinician might ask a client to record three ratings a day on four personally relevant items for 14 days, using a phone-based EMA app that calculates a series of within-person models. Finding that for this client, experiential avoidance predicts sadness ($\beta = 0.35$) while valued action does not predict improved mood, the practitioner might adjust the treatment plan accordingly. Routine outcome monitoring, which already improves outcomes and reduces deterioration and dropout (de Jong et al., 2021), supplies much of the needed data as a byproduct of good care. Widely available phone-based experience sampling lowers the burden further, and clinical decision support that pairs a group-informed starting plan with the patient's own ongoing data shows how the two levels can be combined in practice (Lutz et al., 2022). The aim is not to ask every clinician to become a time-series analyst, but to build software and training into clinical care in a way that puts person-level feedback in reach, beginning with the simplest case and adding complexity only as the data and the question warrant.

9. An uncomfortable implication: Most process findings require idionomic vetting

We can now state an implication that the field has been reluctant to face. If ensemble averages routinely fail to describe the particular people they are meant to guide, then most of what we currently call a process of change has not actually been shown to be a process of change for individuals. It has been shown to hold on average in an ensemble. Until a putative process has been examined idionomically and found to operate within particular people, it should be treated as a candidate to be tested, not as an established principle to be applied. Stated plainly, most current process-of-change findings require idionomic vetting before they can be fully trusted as guides to individual treatment.

This implication arises from an unforgiving asymmetry. A significant average does not establish that a relationship holds within any given person, because a strongly signed average can coexist with a substantial minority for whom the relationship is null or reversed, as the distribution of valued-action effects in Fig. 1 shows. A null average is even less informative, because equisyncratic processes with large but opposing individual effects can cancel to nearly zero (Ciarrochi, Hernández, Hill, et al., 2024). The aggregate can be significant, null, or anything between, and still mislead about the particular case. Non-random heterogeneity at the level needed for individual application is typical. In our process-to-outcome work the I^2 statistic commonly exceeds 75%, and prediction intervals routinely span zero and reverse sign.

This does not strip ensemble findings of all value for the applied use. It disciplines their use, and as we will show later, their value increases with even limited idiographic information. For example, an ensemble finding earns the status of a clinical heuristic when an aggregate effect is present, and the within-person distribution does not contradict it, meaning that non-random heterogeneity is modest and few individuals run counter to the average. In the absence of any such information an ensemble average is a hypothesis for applied use at best and a trap at worst. Only person-level analysis can tell which. Adopting this discipline does not require discarding the existing literature. It requires reading it differently, as a catalog of candidates awaiting idionomic confirmation rather than a set of established individual truths. That reframing is the practical heart of what follows.

10. Idionomics – Creating a science of the particular

If the central failure of our current paradigm is the assumption that ensemble results can be applied to particular people, the remedy must be a methodological program that puts the particular first and uses aggregates only when they help accomplish that purpose. In practice, that means: (a) observing frequently within the lives of intact analytic units (individuals, couples, families), (b) checking for the ergodic error, (c) modeling unit-specific dynamics idiographically, and (d) seeking nomothetic generalizations that clarify and support the idiographic nature of the applied use case. The idionomic vision has deep roots. Decades ago, early attempts to build an 'idiothetic' psychology demonstrated that fundamental psychological assumptions – such as the presumed synchrony between physiological arousal and verbal report – often held true at the level of the group but reversed or vanished at the level of the individual (Turner & Hayes, 1996). What we lacked in the 1990s were the computational tools to model this reality at scale. Today, that barrier has fallen. In this section we will describe some of the idionomic tools that have emerged.

The first pillar of a toolbox that supports idionomic design logic is low-burden longitudinal measurement that captures processes and outcomes at the time scale in which they unfold. The methods that can be used are growing rapidly but these include ecological momentary assessment, wearable devices, AI scoring of verbal or non-verbal behavior, daily diary measurement, and similar measurement systems. In multiple EMA datasets, single- or few-item process and outcome measures have demonstrated excellent within-person reliability (ICC(2) commonly $\geq$0.90), including in clinical samples, enabling repeated measurement without undue burden (e.g., Ciarrochi, Sahdra, Hayes, et al., 2024; Sahdra et al., 2024, 2025). For example, a trichotillomania EMA battery used brief items grounded in the Extended Evolutionary Meta-Model (EEMM) and standard treatment kernels (e.g., DBT, ACT), achieving ICC(2)s between 0.93 and 0.99 across the study period (Sahdra et al., 2026; Woolley et al., 2025). Similar reliability was observed for mood and valued-action items in transdiagnostic in- and out-patient samples (ICC(2) $\approx$ 0.84–0.96) and for single-item happiness and affect in student samples (Catts et al., 2025; Sahdra et al., 2025). These studies show that high temporal density is critical because pooled associations across people often mask strong, systematic within-person heterogeneity; high-frequency measurement allows that heterogeneity to be modelled rather than averaged away. Assessment burden in routine clinical use remains a practical challenge, but emerging solutions are promising: items can be reduced to a critical few of personal relevance, wearables can replace self-report for some processes, and AI-assisted scoring of transcripts or natural speech can automate what once required extensive human coding.

The second pillar is method complementarity. An idionomic science of intentional change cannot rely on a single algorithm; it requires an ecosystem of analytic tools tailored to varying data densities, theoretical constraints, and clinical goals. Method complementarity acknowledges that identifying a functionally important process of change requires different statistical lenses: from evaluating simple bivariate temporal links, to selecting features from massive variable pools, to mapping highly complex, multivariate dynamic networks, and discovering the topological structure of individual differences that traditional clustering methods obscure.

Table 1 outlines the premier idionomic methods currently available, detailing their specific use cases, strengths, and methodological vulnerabilities. By applying these tools complementarily, we can rigorously model particular people without falling prey to the ergodic error.

Idionomic methods carry their own risks, and we do not want to trade one set of blind spots for another. Idiographic models require dense, repeated measurement, and when time series are short or measurement is sparse their estimates can be unstable. They can overfit, mistaking noise for signal, and they can generate false positives when temporal dependencies are not handled correctly. Variance and autocorrelation are assumed to be reasonably stable within a measurement window, and that assumption can fail. Person-level models can also be harder to interpret and to teach than a single group estimate.

These risks are real, but they are manageable, and several are addressed by the tools themselves. tsBoruta removes temporal dependencies before selecting features, which controls false positives; idionomic random-effects meta-analysis reports heterogeneity and prediction intervals, which makes overconfident generalization visible; and methods that shrink individual estimates toward the mean are flagged precisely because that shrinkage can reintroduce the ergodic error. Beyond the tools, the ordinary safeguards of good measurement apply: enough observations for the question at hand, transparent and shared analysis pipelines, explicit reporting of within-person heterogeneity, and replication of person-level findings across occasions and samples. Reproducibility in this framework is not the reproducibility of a single average but the reliable recovery of person-specific structure, and it has to be demonstrated rather than assumed.

ARIMAX is well suited to estimate within-person, bivariate contemporaneous links while controlling trends and autocorrelation, delivering interpretable betas and standard errors at the person level that can be meta-analyzed across individuals (i-ARIMAX when both steps are used with an ensemble; Ciarrochi, Sahdra, Hayes, et al., 2024). Across extensive Monte Carlo simulations parameterized from real EMA data, ARIMAX achieved the highest or near-highest F1 scores for linear effects and maintained good specificity (often near or above 0.80), outperforming multivariate network methods when the aim is to detect specific predictor–outcome links (Li et al., 2026). In the same simulations, sensitivity generally rose with more time points but specificity and precision tended to fall beyond roughly 100 observations, suggesting a practical balance for many clinical EMA designs around 50–70 observations per person (Li et al., 2026). These patterns align with an empirical study using a trichotillomania sample, where i-ARIMAX models delivered convergent nomothetic signals (e.g., generalized positive effects of fixation on pulling) and revealed broad heterogeneity with respect to other processes (Sahdra et al., 2026).

Tree-based feature selection adds a second capability: identifying all-relevant features, including nonlinear or interaction-driven effects, without strong distributional assumptions. A powerful example is idiographic Boruta (iBoruta) a wrapper around the Boruta feature selection algorithm (Kursa & Rudnicki, 2010), which generates Random Forest–based importance tests per person, benchmarking each predictor against its "shadow" permutations to guard against false positives, and has shown utility in psychological and clinical data to uncover meaningful predictors in complex, high-dimensional contexts (Li et al., 2026; Sahdra et al., 2026). One limitation of iBoruta is that it ignores time-series dependencies (autocorrelation, trends), which can inflate false positives in EMA applications. Time Series Boruta (tsBoruta) addresses that limitation by first removing temporal dependencies from each predictor via ARIMA and then applying Boruta to the residuals (Li et al., 2026; Sahdra et al., 2026). Across diverse simulated conditions, tsBoruta delivered substantially higher specificity and precision than iBoruta, especially as autocorrelation or trends increased, and surpassed all linear time-series or network methods in detecting nonlinearities and interactions when those were present (Li et al., 2026). In a trichotillomania sample, tsBoruta was more conservative than iBoruta, rejecting more spurious predictors among effects that ARIMAX judged non-significant, but converging with ARIMAX and iBoruta in confirming fixation as a robust predictor nomothetically (Sahdra et al., 2026). This conservative profile makes tsBoruta particularly valuable when clinical time and attention are scarce and chasing false positives is costly.

While i-ARIMAX and iBoruta or tsBoruta are excellent for estimating within-person links, system-level network approaches such as Group Iterative Multiple Model Estimation (GIMME) and its variants such as S-GIMME offer a third capability: recovering contemporaneous and temporal interdependencies among multiple processes at the person level, while also identifying shared group- or subgroup-level structure via an extended unified SEM approach (Beltz & Gates, 2017). GIMME works at the strictly idiographic level first. Only after the dynamic structure of the individual is established does it search for shared group- or subgroup-level edges, retaining them if and only if they improve the fit for most of the individual models (Gates & Molenaar, 2012). By refusing to assume homogeneity a priori, tools like GIMME allow us to identify complex network structures of psychopathology and change without forcing particular people onto a Procrustean bed of normative assumptions.

Clinical applications of GIMME (e.g., Sanford et al., 2022) have sometimes found only a few subgroups after idiographic patterns are modelled. GIMME currently demands a great deal of longitudinal data, and it is not optimized for focused feature selection of individual process–outcome links, lagging behind ARIMAX and tree-based approaches on that task (Li et al., 2026). In a trichotillomania sample studied by Woolley et al. (2025), GIMME revealed group-level links between sensory seeking, cognitive fixation, and pulling while preserving person-specific networks, but it required manual item curation and assumed linearity, equal spacing (also an assumption of ARIMAX models), and detrending outside the model – constraints that motivated Sahdra et al.'s (2026) comparison with automated feature selection.

Four additional idionomic methods are described in Table 1. When the goal is to discover topological continuums of individuals rather than forcing people into arbitrary, predefined categories, Growing Self-Organizing Maps (GSOM; Alahakoon et al., 2000) provide a solution. GSOM is an unsupervised machine learning method based on artificial neural networks, that dynamically expands its structure to group individuals into a "neighborhood" while maintaining idiographic complexity of the data. It yields interpretable visual mapping of how different individual response profiles relate to one another, enabling practitioners to see gradual transitions between clinical presentations that traditional clustering algorithms often obscure.

For exploratory network building when a priori theory is limited, Multilevel Vector Autoregression (Multilevel-VAR; Bringmann et al., 2013) can be a valuable method. It combines multilevel modeling with vector autoregression to simultaneously model networks of multiple variables across time at both the within-person and between-person levels, capturing bidirectional, contemporaneous, and autoregressive associations simultaneously. The multilevel component relies on partial pooling to pull individual estimates toward the group mean, however, and thus users must apply this tool cautiously to avoid underestimating true idiographic heterogeneity.

Idionomic Random-Effects Meta-Analysis (RE-MA; Sahdra et al., 2024) is the second step of the i-ARIMAX algorithm, and offers a useful way to evaluate whether individual effects can be safely generalized. This approach treats each individual as a separate study to pool idiographic effect sizes derived from base estimators like ARIMAX. By examining I^2 statistics and prediction intervals, RE-MA quantifies how much individual effects deviate from the normative average, identifying when group averages are statistically inappropriate for individual clinical prediction.

Finally, Perceived Causal Networks (PECANs; Klintwall et al., 2023), such as the Personalized Life Analysis Network (PLAN) we have been using in our work (Ong et al., 2024) are known to be of some use in personalization (Burger et al., 2024). They can lay down a foundation for EMA and personalized work.

10.1. *The role of traditional methods and the evolution of idionomic analysis*

An idionomic revolution does not require throwing 50 years of randomized controlled trials (RCTs) or 150 years of aggregate statistics into the scientific dustbin. Ensemble statistics are not useless; they have simply been miscategorized as individual mandates rather than what they are in the primary use cases of psychology: starting possibilities typically based on convenience samples.

When a client walks into a clinic, and the practitioner has zero time-series data available relevant to their unique dynamic network, aggregate averages based on any plausible set of participants are a more sensible place to begin to form hypotheses than random guesswork. They can reasonably serve as Bayesian priors (van de Schoot et al., 2021), especially if at least some idiographic information suggests there is a very low rate of sign heterogeneity. For example, based on ensemble statistics, we might assume that facilitating values-based action will likely improve the mood of a client who is entering treatment entangled with feelings of sadness (Sahdra et al., 2024). The prior belief must be held lightly, however, until at least some information exists on idiographic sign heterogeneity. For some people, values-based action may initially increase distress (Sahdra et al., 2024). If prediction intervals are wide that is a clear signal to treat the ensemble average with caution rather than as a reliable individual prescription.

The Bayesian priors framework will quickly become far more powerful, however, once high temporal density idiographic data are more routinely collected inside RCTs. Processes of change are now widely viewed as important (Hayes & Hofmann, 2021) – the present paper shows why the inclusion of idionomic analysis inside RCTs is a logical next step. The needed "idionomic revolution" will be needlessly delayed, however, if it is assumed that such inclusion means using self-report EMA methods with massive numbers of often poorly understood assessment items scored by clients repeatedly over many weeks if not months in order to pass a statistical hurdle.

Examples of better ways forward abound. Researchers and practitioners might limit assessment to only the key processes indicated by functional analysis of particular cases. Techniques such as network-based case conceptualization (Hofmann et al., 2021; Ong et al., 2024) can quickly narrow the need to collect process information to a small set of clinically relevant items.

In another direction that echoes the earliest days of evidence-based therapy (Shapiro, 1961a), AI methods can be used to score entirely idiographic self-report items in terms of process models such as the EEMM, with recent data showing that EEMM dimensions can be reliably classified by both humans and AI (Ciarrochi, Hernández, Hill, et al., 2024). This should allow small numbers of relevant idiographic self-report items to be used that are meaningful and heartfelt across very diverse populations, reducing or perhaps even eliminating a sense of "burden" and replacing it with client curiosity about their experience.

In the modern day, people are quite used to the use of wearables and the scoring of passive data in such areas as sleep patterns, heart rate variability, or movement. AI tools already allow transcription and process scoring of routine speech or daily qualitative reports – perhaps in lieu of "standardized" methods. This very paper mentions analytic routines our team has developed (e.g., iBoruta; tsBoruta) that become analytically stable in a fraction of the time of previous methods. Such innovations could rapidly self-amplify. For example, as idionomic databases grow, the identification of functionally distinct subgroups seems likely to increase quickly – and recognizing a known within-person pattern is far simpler and faster analytically than discovering such patterns. At this point we can only say "let's see," but pouring more years into traditional methods without taking such steps does not seem likely to change the tragic trajectory of our field.

All of this does not mean that idiographic modeling is always required or that aggregate evidence is invalid, once its limitations are understood. A clear case arises when the prediction interval for person-level effects is narrow and falls entirely on one side of zero. In that situation, the pooled effect can be treated as a reasonable approximation for most individuals in the population from which the sample was drawn, at least as a plausible Bayesian prior. We state this criterion in terms of the prediction interval and the between-person standard deviation τ rather than I^2, and the distinction matters for how our argument should be applied. I^2 is a proportion of observed variance, not an absolute measure of dispersion; it rises and falls with the precision of the estimates that enter it even when the true spread of effects is unchanged, so a given I^2 value does not tell a reader how widely effects actually range (Borenstein, Higgins, Hedges, & Rothstein, 2017). The familiar 25/50/75% rules of thumb hold only where within-study error is relatively constant, and idionomic data routinely violate that condition: per-person series lengths differ widely, and lagged effects are estimated with substantially more error than contemporaneous ones, which lowers I^2 without lowering τ. Empirically, I^2 varies considerably across replication projects at identical levels of true heterogeneity, driven largely by the sample sizes of the constituent studies (Olsson-Collentine, Wicherts, & van Assen, 2020). Two sets of person-level effects with the same τ can therefore return very different I^2 values. τ and the prediction interval do not behave this way, and they are the quantities that bear on whether an average can stand in for a person.

The most mentioned situation is when population-level policy is the goal, such as designing a public health campaign or allocating system-wide resources. This case is more complex. If the underlying data and the policy application both focus on random samples from a known population, ensemble effect sizes are a defensible unit of inference (Glasgow, Lichtenstein, & Marcus, 2003). Often, however, the input data are samples of convenience and policies will be applied differentially to subpopulations depending on socioeconomic status or racioethnic background. An even more troublesome worry occurs when I^2 values exceed 75%, and prediction intervals span zero and reverse the sign. In that case, without idionomic analyses, it may never be known that a policy reliably harms particular people. This toxic outcome can be especially invisible when the policy has equisyncratic effects in some areas because aggregate statistics will falsely reassure policy makers that there is no need for further investigation.

11. From ensemble findings to the individual: Lessons from adjacent fields

Moving from the ensemble to the person is a separate step, one the average cannot take and one that can even point the wrong way (Ciarrochi, Sahdra, Hayes, et al., 2024). This problem is not unique to therapy. Several other fields have tried to turn ensemble data into advice for one person, and their experience is worth a close look.

The approach those fields tried most often is the one our field favors too: matching people to treatments by between-person differences such as symptom type or severity. Given our critique, it would be easy to assume such matching simply fails. The truth is more mixed. It helps, but the gains are small and uneven, and their pattern is telling. The first warning came from education, where Cronbach and Snow tried to match teaching styles to student aptitudes, giving weaker students more structure and confident students more freedom. Matches looked good in single studies but rarely held up, and as interacting traits and settings multiplied, the picture dissolved into what Cronbach called a hall of mirrors that could not guide teaching (Cronbach, 1975).

Therapy shows a more hopeful but still limited version of the same story. Matching the level of care to a patient's likely prognosis improves outcomes (Delgadillo et al., 2022; Fletcher et al., 2021), and when a prediction tool, the Personalized Advantage Index, marked which treatment should best suit each patient, those who had received their marked-best option did better (DeRubeis et al., 2014; Huibers et al., 2015). Yet across many trials the average benefit of personalized over standard care is small and varies widely from study to study (Li et al., 2024; Nye et al., 2023), matching the amount of treatment works better than matching its type (Nye et al., 2023), and a rule that predicts the best treatment in one study often fails in the next (DeRubeis et al., 2014).

Cancer medicine shows the same shape. Matching drugs to a patient's biomarkers yields real but narrow wins, such as trastuzumab for HER2-positive breast cancer, yet only a few percent of patients with advanced cancer benefit from gene-targeted drugs overall (Marquart, Chen, & Prasad, 2018; Tannock & Hickman, 2016). The lesson is not that between-person matching is wrong, but that it hits a ceiling when it is fixed in advance, broad, and based only on how people differ from one another. The spread of results is the clue: an approach that helps the average patient can still leave many people unchanged.

We have argued that ensemble findings give us hypotheses for candidates to try with a particular person, which we then keep if they help that person. Medicine's clearest example is the n-of-1 trial, in which a single patient is measured repeatedly as two treatments alternate in random, blinded order, so the better option emerges from that person's own data rather than the ensemble average (Lillie et al., 2011; Schork, 2015). A standard two-group trial cannot do this, because each patient receives only one treatment; separating a real difference in response from chance requires giving the same person both and comparing them (Senn, 2016, 2018). Recent trials in depression, for example, have tested medications rather than psychotherapy (Kronish, Hampsey, Falzon, Konrad, & Davidson, 2018). A larger literature exists in several different fields using that design under the label of an "alternating treatments design" (Barlow & Hayes, 1979).

Generalizing from here does not mean borrowing from the average to steady a person's estimate, which would assume the average applies to them. It means grouping people in whom a process works the same way, so that what helps one becomes a candidate to try with another in that subgroup. We take each person's result at face value, even when the process does nothing or reverses, rather than treating it as noise to be nudged back toward what is typical. Developmental science has the same habit: we judge a child's growth not by where they rank against other children at one moment, but by whether they are drifting off their own line over time (Molenaar, 2004).

Therapy is moving the same way. When clinicians track patients repeatedly during treatment and use the results to guide care, outcomes improve and fewer patients drop out or deteriorate (de Jong et al., 2021), and frequent measurement reveals patterns a single snapshot would miss (Hamaker & Wichers, 2017). Fisher and colleagues mapped how each patient's symptoms fed into one another, chose treatment components to fit that personal map, and saw symptoms drop sharply in an open trial (Fisher et al., 2019). In a controlled trial, an app that selected coping skills from each user's moment-by-moment reports outperformed the same app giving random skills or none (Levin, Haeger, & Cruz, 2019). Other systems combine the two levels: an algorithm trained on ensemble data suggests a plan, and the clinician adjusts it from the patient's own ongoing data (Lutz et al., 2022). A recent review confirms that fully person-specific methods are now used across many conditions, while noting the work still needed on data quality and reporting (Andreoli, Rafanelli, Hofmann, & Casu, 2025).

Between-person data, then, is a fair starting point, most useful when the marker is strong, as in matching the level of care to prognosis. But it takes us only so far. The next gains will likely come from testing each ensemble-derived candidate in the individual, measuring repeatedly before and during treatment, and keeping only what truly helps.

We raised the possibility earlier of using ensemble information to form Bayesian priors. That approach could rapidly take evidence-based practice into the needed direction of practice-based evidence if idionomic information collection were added to systems of care and routine clinical practice. Process to outcome information of that kind would allow kernelized interventions to be applied based on rapidly updated prior expectations. In this way, it is not hard to imagine a world in which our existing ensemble-level mediational data (Hayes et al., 2022) provides the scaffolding for a more process-based approach to practice, with idiographic data holding the ultimate power. The clients and practitioners of applied and clinical psychology would then be creating a world together in which every client's voice mattered empirically; measuring or modeling was synonymous with listening; and personalizing intervention was synonymous with treating. In broad terms that was always part of the vitalizing vision of evidence-based practice and the "scientist-practitioner model" (Barlow, Hayes, & Nelson, 1984; Hayes et al., 1999), but with the advancement of AI methods and idionomic analysis, it has become relevant in a new way.

The argument of this paper is not confined to clinical psychology. Any applied science that draws conclusions about particular cases from group-level regularities inherits the same ergodic hazard. Personalized and precision medicine, psychometrics and assessment, education, organizational science, and developmental science all move, at some point, from an average to an individual, and at that step the same question applies: does the aggregate describe the case in front of us? Wherever the answer is assumed rather than tested, idionomic analysis offers the needed correction.

12. A practical idionomic workflow

When idionomic tools such as those we have mentioned are used together, a transparent workflow naturally emerges. Measurement comes first, guided by a process model that focuses on theories of change relevant to the targeted outcome such as the EEMM (Hayes, Hofmann, & Ciarrochi, 2020). Frequent, brief EMA sampling (e.g., two to four prompts daily for 2–4 weeks) or another adequate longitudinal measurement approach (e.g., wearables; transcription of natural speech) yields sufficient within-person density for a range of idiographic modeling and subgroup detection. In the case of self-report, items span the dimensions and levels of interest: cognition, affect, attention, motivation/values, self, overt behavior, and social and biophysiological levels of analysis (Ciarrochi, Hernández, Hill, et al., 2024; Sahdra et al., 2024; Hernández et al., 2025; Westhoff et al., 2024). The length and frequency of measurement depend on analytic methods used. Modeling proceeds idiographically. For example, ARIMAX estimates might be calculated per-person yielding process→outcome betas with auto.arima (Hyndman et al., 2025) handling temporal structure; those betas can then be meta-analyzed to characterize pooled effects, and non-random heterogeneity can be quantified as the between-person standard deviation τ, together with its associated prediction interval, which is the current i-ARIMAX workflow. A prediction interval that spans zero, or a τ that is large relative to the magnitude of the pooled effect, should be treated as a caution flag against over-interpreting that pooled effect (e.g., Sahdra et al., 2024). I^2 may be reported alongside these as a familiar descriptive companion, but it should not carry the decision, because it is a relative rather than an absolute index and is not comparable across sets of effects that differ in precision (Borenstein et al., 2017, Borenstein, 2024). Where appropriate, multivariate idiographic networks can then be estimated within relatively homogeneous subgroups, either defined a priori (e.g., transdiagnostic subtypes) or empirically via clustering or trajectory modeling and interpreted at the within- and between-person levels separately (e.g., Catts et al., 2025; Sahdra et al., 2024; Sahdra et al., 2025).

Feature selection complements this modeling. iBoruta can be used for exploratory breadth, identifying candidate predictors for a given unit including nonlinear effects; tsBoruta can then narrow the set by removing time series artifacts and prioritizing high-precision positives (Li et al., 2026). In Sahdra et al.'s (2026) trichotillomania analysis, for example, tsBoruta confirmed smaller, higher-confidence predictor sets per person than iBoruta and sometimes disagreed with ARIMAX on marginal effects, surfacing exactly the kinds of decisions that clinicians must make when prioritizing targets. Critically, convergence across methods increases confidence. In that study, all three methods – i-ARIMAX, iBoruta, tsBoruta – converged on cognitive fixation as a nomothetic anchor; such convergence warrants stronger weight when recommending targets in the absence of richer contextual data (Sahdra et al., 2026).

In couples work (Ciarrochi et al., 2024), idiographic estimates and dyadic networks identified "synergistic" couples, in whom compassion robustly increased attraction, and "independent" couples, in whom compassion was largely inert; actor–partner effects also proved asymmetric by gender and context, again undermining the psychological homogeneity assumption. The broader compassion literature replicated the same lesson: the benefits of self- and other-compassion for well-being depended on self–other harmony within persons; when harmony was lacking, compassion's benefits often disappeared, despite positive pooled associations (Sahdra et al., 2023).

In happiness research, i-ARIMAX and multivariate network modeling showed that "prioritizing positivity" and "experiential attachment" to enjoyment diverged at the person level, with very high total nonrandom heterogeneity ($I^2 \approx 94\%$) and prediction intervals spanning negative to positive effects; subgroup networks revealed that experiential attachment dampened happiness contemporaneously within-person even when between-person associations were null, and that prioritizing positivity helped only in a subgroup that was not evident at the aggregate level (Sahdra et al., 2025). These exemplars align with the trichotillomania findings (Sahdra et al., 2026): some nomothetic signals exist and can be useful anchors, but clinical utility arises from modeling particular people and respecting how subgroups and contexts change dynamics.

Performance metrics matter because clinicians must balance missing true targets against pursuing false ones. Simulations provide guidance. For linear effects, ARIMAX consistently yielded the highest F1 scores with good specificity, outperforming network methods for detecting focused process–outcome links and matching or exceeding raw correlations while controlling temporal artifacts (Li et al., 2026). Under strong autocorrelation and trends, tsBoruta retains high specificity and precision thanks to its ARIMA preprocessing, whereas Boruta's sensitivity advantage can come with more false positives; in our trichotillomania data, both methods made a similar number of unique errors but of different kinds, with tsBoruta more conservative and iBoruta more inclusive. Network methods such as GIMME excel at recovering system-level structures but were not competitive for focused feature selection in the conditions tested here (Li et al., 2026).

Training and feasibility are no longer barriers. Open-source R packages implement the key methods: *forecast* for ARIMA and auto.arima (Hyndman et al., 2025), *idionomics* for i-ARIMAX, preprocessing, and a growing number of idionomic analytical routines (Hernández, Ciarrochi, Hayes, & Sahdra, 2026), *Boruta* for iBoruta (Kursa & Rudnicki, 2010), a combination of the functions auto.arima and Boruta for tsBoruta, the package *gimme* for GIMME and indSEM (Lane et al., 2024), and *mlVAR* for within−/between-person networks (Epskamp et al., 2024). Tutorials and example code are publicly available, including step-by-step instructions for conducting tsBoruta (Sahdra et al., 2026) and OSF repositories include code and simulated data where necessary for reproducibility under IRB constraints (e.g.,Li et al., 2026; Sahdra et al., 2026). Practitioners can adopt a "good enough" first pass – e.g., three to six prompts per day for 10–14 days – analyze with ARIMAX and tsBoruta, and iterate as needed, adding network analysis for subgroups where system-level insights are needed and where data density permits.

A final design element in an idionomic workflow is applying this approach to clinical decision-making. A practical sequence is to (a) measure frequently with compact, reliable, and personally relevant items or passive methods; (b) model idiographically using ARIMAX to estimate linear effects, applying tsBoruta to detect nonlinear or interaction-driven predictors; (c) meta-analyze across persons to identify nomothetic anchors and to quantify heterogeneity and prediction intervals; (d) identify relatively homogeneous subgroups where pooled effects and networks make sense; and (e) choose targets that are supported by convergent evidence, link to outcomes in idiographic networks, and align with client goals and context, updating targets and processes as treatment proceeds (Hayes et al., 2026; Sahdra et al., 2026). Consider how this workflow applies to some of the findings we have previously described. In the trichotillomania study (Sahdra et al., 2026), following this sequence preserved the clear nomothetic signal of cognitive fixation and simultaneously revealed that just three processes – fixation, valued action, and anxiety – covered the majority of cases, while leaving room to tailor adjuncts for the sizable minority whose idiographic patterns diverge. In the idionomic study on valued action, this approach would prevent over-promising hedonic lift to Stoics and suggests different sequencing of interventions for Non-Stoics; in couples, it tells the clinician when to focus on compassion and when to look elsewhere; in happiness research, it cautions against blanket prescriptions to "prioritize positivity" and surfaces the hidden cost of experiential attachment, which may need to be addressed first (Sahdra et al., 2025).

Across domains, the message is consistent: an idionomic science is possible – it is already reshaping how we align methods to particular people. Ergodicity violations and shrinkage artifacts are the rule, not the exception; high I^2 values and wide prediction intervals are a signal to model the person and to generalize only when warranted; and small, complementary toolkits can make this feasible in everyday clinical science and practice (Li et al., 2026; Sahdra et al., 2026). By measuring often, modeling idiographically, generalizing only when the data warrant it, and deciding transparently, we can honor the heterogeneity that matters while still building the nomothetic principles needed to serve particular people reliably.

13. Conclusion

The field has hit a barrier and only a change in our assumptions and actions is likely to break through it. We have tried to show that a single and fundamental mistaken assumption, inherited from a troubled history, has prevented our field from serving the people who need care most. When disempowering assumptions are embedded in the very terms and analytic methods used to find a better way forward, it is particularly hard to create transformational change. That is true for the clients who come to ask for help from practitioners. It is also true for us as psychologists and for the field we are creating together. Once seen, however, these assumptions can never again be fully hidden. Fortunately, we now have the insight and tools to advance in resolving this error and correcting the problems it produces.

In the evidence-based field of applied and clinical psychology that practitioners and researchers seek to create, our concepts and methods need to focus us on the particular lives we serve. We need to listen more carefully and to use measurement and analysis to amplify the voices of our clients so that these human beings can be heard. We need to collaborate with our clients to empower them with methods that serve them as unique people, no longer handing out the ill-fitting suits of "one size fits all" protocols. We need to generate truly "nomothetic principles" from the bottom up, so that these principles can create real and lasting progress toward a psychology more worthy of the human condition. All of these needs call for an idionomic revolution that allows such a journey to be taken by clients, practitioners, and researchers – together.

Author's note

Dr. Hofmavnn receives financial support by the Alexander von Humboldt Foundation (as part of the Alexander von Humboldt Professur), the Hessische Ministerium für Wissenschaft und Kunst (as part of the LOEWE Spitzenprofessur), the DYNAMIC center, funded by the LOEWE program of the Hessian Ministry of Science and Arts (Grant Number: LOEWE1/16/519/03/09.001(0009)/98), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 521379614 – TRR 393SFB/Transregio 393 and the Excellence Cluster EXC3066 "The Adaptive Mind".(Project No. 533717223). He also receives compensation for his work as editor from the American Psychological Association and royalties and payments for his work from various publishers.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude and ChatGPT for routine editing, figure modification, and generation of graphical highlights. The authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Steven C. Hayes: Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Joseph Ciarrochi:** Conceptualization, Writing – original draft, Writing – review & editing. **Baljinder K. Sahdra:** Conceptualization, Writing – original draft, Writing – review & editing. **Stefan G. Hofmann:** Conceptualization, Writing – original draft, Writing – review & editing. **Clarissa W. Ong:** Conceptualization, Writing – original draft, Writing – review & editing. **Cristóbal Hernández:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Alahakoon, D., Halgamuge, S. K., & Srinivasan, B. (2000). Dynamic self-organizing maps with controlled growth for knowledge discovery. *IEEE Transactions on Neural Networks, 11*(3), 601–614. https://doi.org/10.1109/72.846732-

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

American Psychological Association. (2021). Historical chronology. In *APA'S role in addressing racism*. Retrieved September 17, 2025, from https://www.apa.org/about/apa/addressing-racism/historical-chronology.

American Psychological Association, Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

Andreoli, G., Rafanelli, C., Hofmann, S. G., & Casu, G. (2025). A systematic scoping review of fully idiographic network analysis in mental health. *Cognitive Therapy and Research*. https://doi.org/10.1007/s10608-025-10674-2

Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest, 9*(2), 67–103. https://doi.org/10.1111/j.1539-6053.2009.01036.x

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*, 199–210. https://doi.org/10.1901/jaba.1979.12-199

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings*. Pergamon.

Becker, E. M., Smith, A. M., & Jensen-Doss, A. (2013). Who's using treatment manuals? A national survey of practicing therapists. *Behaviour Research and Therapy, 51*(10), 706–710. https://doi.org/10.1016/j.brat.2013.07.008

Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research, 52*(6), 789–804. https://doi.org/10.1080/00273171.2017.1373014

Bhattacharya, S., Goicoechea, C., Heshmati, S., Carpenter, J. K., & Hofmann, S. G. (2023). Efficacy of cognitive behavioral therapy for anxiety-related disorders: A meta-analysis of recent literature. *Current Psychiatry Reports, 25*(1), 19–30. https://doi.org/10.1007/s11920-022-01402-8

Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences, 17*(12), 656–660. https://doi.org/10.1073/pnas.17.12.656

Bleuler, E. (1911). *Dementia praecox oder Gruppe der Schizophrenien*. Leipzig: Franz Deuticke.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. The Guilford Press.

Boltzmann, L. (1884). Ueber eine von Hrn. Bartoli entdeckte Beziehung der Warmestrahlung zum zweiten Hauptsatze. *Annalen der Physik, 258*(5), 31–39. https://doi.org/10.1002/andp.18842580503

Borenstein, M. (2024). Avoiding common mistakes in meta-analysis: Understanding the distinct roles of Q, I-squared, tau-squared, and the prediction interval in reporting heterogeneity. *Research Synthesis Methods, 15*(2), 354–368. https://doi.org/10.1002/jrsm.1678

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Borgogna, N. C., McDermott, R. C., Berry, A., Browning, B. R., Lathan, E. C., & Gonzalez, J. (2024). The absurdity of the latent disease model in mental health: 10,130,814 ways to have a DSM-5-TR psychological disorder. *Journal of Mental Health, 33*(4), 451–459. https://doi.org/10.1080/09638237.2023.2278107

Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science, 2*(1). https://doi.org/10.1214/ss/1177013437

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

Brookwood, M. (2021). *The orphans of Davenport: Eugenics, the great depression, and the war over children's intelligence*. Liveright.

Burger, J., Andikkhash, V., Jäger, N., Anderbro, T., Blanken, T. F., & Klintwall, L. (2024). A novel approach for constructing personalized networks from longitudinal perceived causal relations. *Behaviour Research and Therapy, 173*, Article 104456. https://doi.org/10.1016/j.brat.2023.104456

Cattell, R. B. (1952). P-technique factorization and the determination of individual dynamic structure. *Journal of Clinical Psychology, 8*(1), 5–10. https://doi.org/10.1002/1097-4679(195201)8:1<5::AID-JCLP2270080102>3.0.CO;2-Q

Catts, N. V., Sahdra, B. K., Ciarrochi, J., Fraser, M. I., Hernández, C., Hayes, S. C., & Gloster, A. T. (2025). Distinct pathways to well-being: Exploring valued action and mood among stoics and non-stoics. *Journal of Contextual Behavioral Science, 36*, Article 100898. https://doi.org/10.1016/j.jcbs.2025.100898

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7–18. https://doi.org/10.1037/0022-006X.66.1.7

Ciarrochi, J., Hernández, C., Hill, D., Ong, C., Gloster, A. T., Levin, M. E., … Hayes, S. C. (2024). Process-based therapy: A common ground for understanding and utilizing therapeutic practices. *Journal of Psychotherapy Integration, 34*(3), 265–290. https://doi.org/10.1037/int0000348

Ciarrochi, J., Sahdra, B., Hayes, S. C., Hofmann, S. G., Sanford, B., Stanton, C., Yap, K., Fraser, M. I., Gates, K., & Gloster, A. T. (2024). A personalised approach to identifying important determinants of well-being. *Cognitive Therapy and Research, 48*(3), 552–573. https://doi.org/10.1007/s10608-024-10486-w

Ciarrochi, J., Sahdra, B. K., Fraser, M., Hayes, S. C., Yap, K., & Gloster, A. (2024). The compassion connection: Experience sampling insights into romantic attraction. *Journal of Contextual Behavioral Science*. https://doi.org/10.1016/j.jcbs.2024.100749

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology, 14*, 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Clayton, S., Manning, C. M., Krygsman, K., & Speiser, M. (2017). *Mental health and our changing climate: Impacts, implications, and guidance*. Washington, D.C.: American Psychological Association, and ecoAmerica.

Cohen, D. (1979). *J. B. Watson: The founder of behaviourism*. Routledge & Kegan Paul.

Cook, J. M., Biyanova, T., Elhai, J., Schnurr, P. P., & Coyne, J. C. (2010). What do psychotherapists really do in practice? An internet study of over 2,000 practitioners. *Psychotherapy: Theory, Research, Practice, Training, 47*(2), 260–267. https://doi.org/10.1037/a0019788

Cravens, H. (1992). A scientific project locked in time: The Terman genetic studies of genius, 1920s–1950s. *American Psychologist, 47*(2), 183–189. https://doi.org/10.1037/0003-066X.47.2.183

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116–127.

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*, 288–299. https://doi.org/10.1111/acps.13335

Cuijpers, P., Miguel, C., Harrer, M., Ciharova, M., & Karyotaki, E. (2024). The outcomes of mental health care for depression over time: A meta-regression analysis of response rates in usual care. *Journal of Affective Disorders, 358*, 89–96. https://doi.org/10.1016/j.jad.2024.05.019

Curtiss, J., & DiPietro, C. (2025). Machine learning in the prediction of treatment response for emotional disorders: A systematic review and meta-analysis. *Clinical Psychology Review, 120*, Article 102593. https://doi.org/10.1016/j.cpr.2025.102593

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., … Barkham, M. (2022). Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry, 79*(2), 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One, 9*(1), Article e83875. https://doi.org/10.1371/journal.pone.0083875

Driessen, E., Hollon, S. D., Bockting, C. L. H., Cuijpers, P., & Turner, E. H. (2015). Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US National Institutes of Health-funded trials. *PLoS One, 10*(9), Article e0137864. https://doi.org/10.1371/journal.pone.0137864

Elgaddal, N., Weeks, J. D., & Mykyta, L. (2025). Characteristics of adults age 18 and older who took prescription medication for depression: United States, 2023. *NCHS Data Brief, 528*, 1–9. https://dx.doi.org/10.15620/cdc/174589.

Epskamp, S., Deserno, M., & Bringmann, L. (2024). mlVAR: Multi-level vector autoregression. R package version 0.5.2, https://CRAN.R-project.org/package=mlVAR.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*(2), 134–140. https://doi.org/10.1037/h0045156

Eubanks Fleming, C. J., Ernestus, S. M., Wenze, S. J., & Blomquist, K. K. (2025). Faculty reports about training in evidence-based practice and empirically-supported treatments across mental health graduate programs in the United States. *Journal of Clinical Psychology, 81*(11), 1178–1186. https://doi.org/10.1002/jclp.70026

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy, 116*, 69–79. https://doi.org/10.1016/j.brat.2019.01.010

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Fisher, R. A., & Mackenzie, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *The Journal of Agricultural Science, 13*(3), 311–320. https://doi.org/10.1017/s0021859600003592

Fletcher, S., Chondros, P., Densley, K., Murray, E., Dowrick, C., Coe, A., … Gunn, J. (2021). Matching depression management to severity prognosis in primary care: Results of the target-D randomised controlled trial. *British Journal of General Practice, 71*(703), e85–e94. https://doi.org/10.3399/BJGP.2020.0783

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Friedlander, H. (1995). *The origins of Nazi genocide: From euthanasia to the final solution*. University of North Carolina Press.

Funk, M., Drew, N., & Knapp, M. (2012). Mental health, poverty and development. *Journal of Public Mental Health, 11*(4), 166–185.

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science, 8*(6), 651–662. https://doi.org/10.1177/1745691613504115

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. Macmillan. https://doi.org/10.1037/13474-000

Galton, F. (1883). *Inquiries into human faculty and its development*. Macmillan.

Galton, F. (1889). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*, 135–145. https://doi.org/10.1098/rspl.1888.0108

Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guskiewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research, 52*(2), 129–148. https://doi.org/10.1080/00273171.2016.1256187

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

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93*(8), 1261–1267. https://doi.org/10.2105/AJPH.93.8.1261

Goddard, H. H. (1917). Mental tests and the immigrant. *Journal of Delinquency, 2*, 243–277.

Gould, S. J. (1996). *The mismeasure of man* (Rev. ed.). W. W. Norton & Company.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl, & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). Guilford Press.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Hansen, K. (2018). *Statistical physics of nanoparticles in the gas phase* (2nd ed.). Springer.

Harden, K. P. (2021). *The genetic lottery*. Princeton University Press.

Harrer, M., Miguel, C., van Ballegooijen, W., Ciharova, M., Plessen, C. Y., Kuper, P., … Cuijpers, P. (2025). Effectiveness of psychotherapy: Synthesis of a "meta-analytic research domain" across world regions and 12 mental health problems. *Psychological Bulletin, 151*(5), 600–667. https://doi.org/10.1037/bul0000465

Hayes, S. C. (2019). *A liberated mind: How to pivot toward what matters*. Avery.

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.). 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*, Article 104155. https://doi.org/10.1016/j.brat.2022.104155

Hayes, S. C., & Follette, W. C. (1992). Can functional analysis provide a substitute for syndromal classification? *Behavioral Assessment, 14*, 345–365.

Hayes, S. C., Hernández, C., Ciarrochi, J., Hofmann, S. G., Ong, C., & Sahdra, B. K. (2026). Idionomic analysis in translational research in behavior analysis: Creating a two-way relationship between basic and applied behavioral science. In G. J. Madden, & F. D. Digennaro Reed (Eds.), *APA handbooks in psychology: APA Handbook of behavior analysis, 2nd Edition, Volume 2: Translational research, applied principles and values, and child applications*. Washington, DC: American Psychological Association. in press.

Hayes, S. C., & Hofmann, S. G. (Eds.). (2018). *Process-based CBT: The science and core clinical competencies of cognitive behavioral therapy*. New Harbinger Publications.

Hayes, S. C., & Hofmann, S. G. (2021). "Third-wave" cognitive and behavioral therapies and the emergence of a process-based approach to intervention in psychiatry. *World Psychiatry, 20*(3), 363–375. https://doi.org/10.1002/wps.20884

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020). A process-based approach to psychological diagnosis and treatment: The conceptual and treatment utility of an extended evolutionary meta-model. *Clinical Psychology Review, 82*, Article 101908. https://doi.org/10.1016/j.cpr.2020.101908

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*(11), 963–974. https://doi.org/10.1037/0003-066X.42.11.963

Hayes, S. C., Rincover, A., & Solnick, J. (1980). The technical drift of applied behavior analysis. *Journal of Applied Behavior Analysis, 13*, 275–285. https://doi.org/10.1901/jaba.1980.13-275

Hernández, C., Ciarrochi, J., Hayes, S., & Sahdra, B. K. (2026). *Idionomics: Making idionomic analyses accessible and easy to use [R package]*. GitHub. https://github.com/cristobalehc/idionomics.

Hernández, C., Sahdra, B. K., Ciarrochi, J., & Hayes, S. C. (2025). Idionomic assessment of mindfulness. In O. N. Medvedev, et al. (Eds.), *Handbook of assessment in mindfulness research*. Springer. https://doi.org/10.1007/978-3-030-77644-2_136-1.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. Free Press.

Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change*. Pergamon Press.

Hofmann, S. G., Curtiss, J., & McNally, R. J. (2016). A complex network perspective on clinical science. *Perspectives on Psychological Science, 11*, 597–605. https://doi.org/10.1177/1745691616639283

Hofmann, S. G., Curtiss, J. E., & Hayes, S. C. (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review, 76*. https://doi.org/10.1016/j.cpr.2020.101824

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science, 7*, 37–50. https://doi.org/10.1177/2167702618772296

Hofmann, S. G., Hayes, S. C., & Lorscheid, D. (2021). *Learning process-based therapy: A skills training manual for targeting the core processes of psychological change in clinical practice*. Oakland, CA: Context Press / New Harbinger Publications.

Hofmann, S. G., Kasch, C., & Reis, A. (2025). Effect sizes of randomized-controlled studies of cognitive behavioral therapy for anxiety disorders over the past 30 years. *Clinical Psychology Review, 117*, Article 102553. https://doi.org/10.1016/j.cpr.2025.102553

Honkalampi, K., Urhonen, H.-R., & Virtanen, M. (2024). Negative effects in randomized controlled trials of psychotherapies and psychological interventions: A systematic review. *Psychotherapy Research, 35*(1), 100–111. https://doi.org/10.1080/10503307.2024.2301972

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., & DeRubeis, R. J. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLoS One, 10*(11), Article e0140771. https://doi.org/10.1371/journal.pone.0140771

Hunter, M. D., Fisher, Z. F., & Geier, C. F. (2024). What ergodicity means for you. *Developmental Cognitive Neuroscience, 68*, Article 101406. https://doi.org/10.1016/j.dcn.2024.101406

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeen, F. (2025). Forecast: Forecasting functions for time series and linear approaches. R package version 8.24.0 https://pkg.robjhyndman.com/forecast/.

Inchauspé, J. (2022). *Glucose revolution: The life-changing power of balancing your blood sugar*. Simon & Schuster.

Insel, T. (2022). *Healing: Our path from mental illness to mental health*. Penguin.

Johnsen, T., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141*(4), 747–768. https://doi.org/10.1037/bul0000015

de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review, 85*, Article 102002. https://doi.org/10.1016/j.cpr.2021.102002

Kader, M. R., Rahman, M. M., Bristi, P. D., & Ahmmed, F. (2024). Change in mental health service utilization from pre-to post-COVID-19 period in the United States. *Heliyon, 10*(22), Article e40454. https://doi.org/10.1016/j.heliyon.2024.e40454.

Kanfer, F. H., & Saslow, G. (1965). Behavioral analysis: An alternative to diagnostic classification. *Archives of General Psychiatry, 12*(6), 529–538. https://doi.org/10.1001/archpsyc.1965.01750360009002

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*(3), 146–159. https://doi.org/10.1037/0003-066X.63.3.146

Kennedy, B. (2022). *Good inside: A guide to becoming the parent you want to be*. Harper Wave.

Kevles, D. J. (1985). *In the name of eugenics: Genetics and the uses of human heredity*. Knopf.

Kiesler, D. J. (1994). Standardization of intervention: The tie that binds psychotherapy research and practice. In P. F. Talley, H. H. Strupp, & S. F. Butler (Eds.), *Psychotherapy research and practice: Bridging the gap* (pp. 143–153). Basic Books.

Klintwall, L., Bellander, M., & Cervin, M. (2023). Perceived Causal Problem Networks: Reliability, central problems, and clinical utility for depression. *Assessment, 30*(1), 73–83. https://doi.org/10.1177/10731911211039281

Kronish, I. M., Hampsey, M., Falzon, L., Konrad, B., & Davidson, K. W. (2018). Personalized (N-of-1) trials for depression: A systematic review. *Journal of Clinical Psychopharmacology, 38*(3), 218–225.

Kupfer, D. J., First, M. B., & Regier, D. A. (Eds.). (2002). *A research agenda for DSM-V*. American Psychiatric Association.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software, 36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Lane, S. T., Gates, K. M., Fisher, Z., Arizmendi, C., Molenaar, P. C. M., Hallquist, M., … Merkle, E. (2024). *Gimme: Group iterative multiple approach estimation (version 0.7-18) [R package]*. CRAN. https://CRAN.R-project.org/package=gimme.

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry, 21*(1), 133–145. https://doi.org/10.1002/wps.20941

Levin, M. E., Haeger, J., & Cruz, R. A. (2019). Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: Results from a randomized controlled trial. *Mindfulness, 10*(4), 689–699. https://doi.org/10.1007/s12671-018-1004-2

Li, W., Gleeson, J., Fraser, M. I., Ciarrochi, J., Hofmann, S. G., Hayes, S. C., & Sahdra, B. K. (2024). The efficacy of personalized psychological interventions in adolescents: A scoping review and meta-analysis. *Frontiers in Psychology, 15*, Article 1470817. https://doi.org/10.3389/fpsyg.2024.1470817

Li, W., Hernández, C. E., Ciarrochi, J., Fraser, M. I., Hayes, S. C., Ong, C. W., & Sahdra, B. K. (2026). Advancing idiographic methodology for data-driven treatment personalization: Comparing ARIMAX, network, and tree-based approaches. Retrieved from https://osf.io/preprints/psyarxiv/7mxn3_v2.

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized Medicine, 8*(2), 161–173. https://doi.org/10.2217/pme.11.7

Liu, Y., Moeller, J., & Kospi, R. (2019). Bias in the estimation of within-person variance and its impact on the performance of multilevel models. *Methodology, 15*(3), 115–128.

Lombardo, P. A. (2008). *Three generations, no imbeciles: Eugenics, the supreme court, and Buck v. Bell*. Johns Hopkins University Press.

Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology, 90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lynn, R. (2008). *The global bell curve: Race, IQ, and inequality worldwide*. Washington Summit Publishers.

Marquart, J., Chen, E. Y., & Prasad, V. (2018). Estimation of the percentage of US patients with cancer who benefit from genome-driven oncology. *JAMA Oncology, 4* (8), 1093–1098.

Maslow, A. (1962). *Toward a psychology of being*. Van Nostrand.

Mazumdar, P. M. H. (1992). *Eugenics, human genetics and human failings: The eugenics society, its sources and its critics in Britain*. Routledge.

Mohr, D. C. (2024). Standards for randomized controlled trials of efficacy of psychological treatments. *World Psychiatry, 23*(2), 286–287. https://doi.org/10.1002/wps.21207

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 and Perspectives, 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. Portico https://doi.org/10.1002/dev.20262.

Moritz, S., Ciarrochi, J., Fraser, M., Krafft, J., Klimczak, K., Levin, M., … Sahdra, B. K. (2024). *An idionomic analysis of mindfulness, experiential avoidance, and affect among individuals exhibiting high and low-hoarding behaviours*. Manuscript in preparation.

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review, 33*(4), 501–511. https://doi.org/10.1016/j.cpr.2013.02.002

Nelson, R. O., & Hayes, S. C. (Eds.). (1986). *Conceptual foundations of behavioral assessment*. Guilford Press.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist, 6*(3), 187–193. https://doi.org/10.1027/1016-9040.6.3.187

von Neumann, J. (1932). Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences, 18*(1), 70–82. https://doi.org/10.1073/pnas.18.1.70

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

Okamura, K. H., Orimoto, T. E., Nakamura, B. J., Chang, B., Chorpita, B. F., & Beidas, R. S. (2020). A history of child and adolescent treatment through a distillation lens: Looking back to move forward. *The Journal of Behavioral Health Services & Research, 47*(1), 70–85. https://doi.org/10.1007/s11414-019-09659-3

Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: Theoretical framework and empirical application. *Journal of Abnormal Psychology, 123*(2), 452–462. https://doi.org/10.1037/a0036068

Olfson, M., & Marcus, S. C. (2010). National trends in outpatient psychotherapy. *American Journal of Psychiatry, 167*(12), 1456–1463.

Olsson-Collentine, A., Wicherts, J. M., & van Assen, M. A. L. M. (2020). Heterogeneity in direct replications in psychology and its association with effect size. *Psychological Bulletin, 146*(10), 922–940. https://doi.org/10.1037/bul0000294

Ong, C. W., Ciarrochi, J., Hofmann, S. G., Karekla, M., & Hayes, S. C. (2024). Through the extended evolutionary meta-model, and what ACT found there: ACT as a process-based therapy. *Journal of Contextual Behavioral Science, 32*, Article 100734. https://doi.org/10.1016/j.jcbs.2024.100734

Oseledets, V. I. (1968). A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society, 19*, 197–231.

Parolini, G. (2014). The emergence of modern statistics in agricultural science: Analysis of variance, experimental design and the reshaping of research at Rothamsted Experimental Station, 1919–1933. *Journal of the History of Biology, 48*(2), 301–335. https://doi.org/10.1007/s10739-014-9394-z

Pearson, K., & Moul, M. (1925). The problem of alien immigration into Great Britain illustrated by an examination of Russian and Polish children: Part II. *Annals of Eugenics, 1*(1), 56–128.

Peters, O. (2019). The ergodicity problem in economics. *Nature Physics, 15*(12), 1216–1221. https://doi.org/10.1038/s41567-019-0732-0

Peters, O., & Gell-Mann, M. (2016). Evaluating gambles using dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 26*(2), Article 023103. https://doi.org/10.1063/1.4940236

Plomin, R. (2019). *Blueprint: How DNA makes us who we are*. MIT Press.

Porter, T. M. (2004). *Karl Pearson: The scientific life in a statistical age*. Princeton University Press.

Pratt, L. A., Brody, D. J., & Gu, Q. (2017). Antidepressant use among persons aged 12 and over: United States, 2011–2014. In *, 283. NCHS Data Brief* (pp. 1–8). https://www.cdc.gov/nchs/data/databriefs/db283.pdf.

Proctor, R. N. (1988). *Racial hygiene: Medicine under the Nazis*. Harvard University Press.

Rajakannan, T., Bunting, M., & Dziadkowiec, O. (2016). National trends in psychiatric not otherwise specified (NOS) diagnosis and medication use among adults in outpatient settings: 1999 to 2014. *Psychiatric Services, 67*(3), 289–295. https://doi.org/10.1176/appi.ps.201500045

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage. https://doi.org/10.4135/9781412985037

Rogers, C. (1961). *On becoming a person. A therapist's view of psychotherapy*. Houghton Mifflin.

Rousseau, D. M., & Gunia, B. C. (2016). Evidence-based practice: The psychology of EBP implementation. *Annual Review of Psychology, 67*, 667–692. https://doi.org/10.1146/annurev-psych-122414-033336

Sahdra, B., Ciarrochi, J., & Parker, P. (2016). Nonattachment and mindfulness: Related but distinct constructs. *Psychological Assessment, 28*(7), 819–829. https://doi.org/10.1037/pas0000264

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, 14*(8), 1997–2013. https://doi.org/10.1007/s12671-023-02187-4

Sahdra, B. K., Ciarrochi, J., Klimczak, K., Krafft, J., Hayes, S. C., & Levin, M. (2024). Testing the applicability of idionomic statistics in longitudinal studies: The example of 'doing what matters.'. *Journal of Contextual Behavioral Science, 32*. https://doi.org/10.1016/j.jcbs.2024.100728

Sahdra, B. K., Shin, A., Fraser, M., Levin, M. E., Klimczak, K. S., Krafft, J., … Ciarrochi, J. (2025). One size does not make all happy: Idionomic links between striving for positive states and happiness in experience sampling. *Journal of Happiness Studies, 26*, 101. https://doi.org/10.1007/s10902-025-00933-0

Sahdra, B. K., Woolley, M. G., Hernández, C., Li, W., Hayes, S. C., Ciarrochi, J., … Levin, M. (2026). Time series machine learning for idionomic process-based treatment planning: A tutorial on tsBoruta. *Journal of Contextual Behavioral Science*. , Article 100983. https://doi.org/10.1016/j.jcbs.2026.100983

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

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., … Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers, 1* (1), 1. https://doi.org/10.1038/s43586-020-00001-2

Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature, 520* (7549), 609–611. https://doi.org/10.1038/520609a

Senn, S. (2016). Mastering variation: Variance components and personalised medicine. *Statistics in Medicine, 35*(7), 966–977. https://doi.org/10.1002/sim.6739

Senn, S. (2018). Statistical pitfalls of personalised medicine. *Nature, 563*(7733), 619–621. https://doi.org/10.1038/d41586-018-07535-2

Shapiro, M. B. (1961a). The single case in clinical-psychological research. *Journal of General Psychology, 64*(1), 8–20.

Shapiro, M. B. (1961b). A method of measuring psychological changes specific to the individual psychiatric patient. *British Journal of Medical Psychology, 34*(2), 151–155. https://doi.org/10.1111/j.2044-8341.1961.tb00940.x

Shetty, J. (2020). *Think like a monk: Train your mind for peace and purpose every day*. Simon & Schuster.

Skeels, H. M., & Dye, H. B. (1939). A study of the effects of differential stimulation on mentally retarded children. *Proceedings of the American Association on Mental Deficiency, 44*, 114–136. https://doi.org/10.1037/h0057113

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*(9), Article 160384. https://doi.org/10.1098/rsos.160384

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*(9), 752–760. https://doi.org/10.1037/0003-066x.32.9.752

Spearman, C. (1904). "general intelligence," objectively determined and measured. *American Journal of Psychology, 15*(2), 201–293. https://doi.org/10.2307/1412107

Speelman, C. P., Parker, L., Rapley, B. J., & McGann, M. (2024). Most psychological researchers assume their samples are ergodic: Evidence from a year of articles in three major journals. *Collabra: Psychology, 10*(1). https://doi.org/10.1525/collabra.92888

Speers, A. J. H., Bhullar, N., Cosh, S., & Wootton, B. M. (2022). Correlates of therapist drift in psychological practice: A systematic review of therapist characteristics. *Clinical Psychology Review, 93*, Article 102132. https://doi.org/10.1016/j.cpr.2022.102132

Starrett, K., & Starrett, J. (2023). *Built to move: The ten essential habits to help you move freely and live fully*. Knopf.

Sundstrom, F. T. A., Lavefjord, A., Buhrman, M., & McCracken, L. M. (2025). Are people with chronic pain more diverse than we think? An investigation of ergodicity. *Pain, 166*(8), 1859–1870. https://doi.org/10.1097/j.pain.0000000000003573

Tannock, I. F., & Hickman, J. A. (2016). Limits to personalized cancer medicine. *New England Journal of Medicine, 375*(13), 1289–1294. https://doi.org/10.1056/NEJMsb1607705

Terman, L. M. (1916). *The measurement of intelligence*. Houghton Mifflin.

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice, 22*(4), 317–338. https://doi.org/10.1037/h0101729

Tucker, W. H. (1994). *The science and politics of racial research*. University of Illinois Press.

Turner, A. E., & Hayes, S. C. (1996). Una comparacion de la covariacion de respuesta vista desde una perspectiva ideotetica y nomotetica (A comparison of response covariation viewed idiothetically and nomothetically). *Psicología Conductual, 4*, 231–250.

Tyner, A. H., Abatayo, A. L., Daley, M., Field, S., Fox, N., Haber, N. A., … Errington, T. M. (2026). Investigating the replicability of the social and behavioural sciences. *Nature, 652*(8108), 143–150. https://doi.org/10.1038/s41586-025-10078-y

Vybíral, Z., Ogles, B. M., Řiháček, T., Urbancová, B., & Gocieková, V. (2024). Negative experiences in psychotherapy from clients' perspective: A qualitative meta-analysis. *Psychotherapy Research, 34*(3), 279–292. https://doi.org/10.1080/10503307.2023.2226813

Wade, N. (2015). *A troublesome inheritance: Genes, race and human history*. Penguin.

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy, 47*(2), 119–127. https://doi.org/10.1016/j.brat.2008.12.005

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). Routledge/Taylor & Francis Group.

Westhoff, M., Heshmati, S., Siepe, B., Vogelbacher, C., Ciarrochi, J., Hayes, S. C., & Hofmann, S. G. (2024). Psychological flexibility and cognitive-affective processes in the daily lives of young adults. *Scientific Reports, 14*, 8182. https://doi.org/10.1038/s41598-024-58598-3

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences, 17*(4), 585–608. https://doi.org/10.1017/S0140525X00031474

Wolitzky-Taylor, K., Zimmermann, M., Arch, J. J., De Guzman, E., & Lagomasino, I. (2015). Has evidence-based psychosocial treatment for anxiety disorders permeated usual care in community mental health settings? *Behaviour Research and Therapy, 72*, 9–17. https://doi.org/10.1016/j.brat.2015.06.010

Woolley, M. G., Sahdra, B. K., Bowers, E. M., Capel, L. K., Twohig, M. P., & Levin, M. E. (2025). An idionomic network analysis of trichotillomania treatment targets: Gathering group-level and individual-level insights to inform a personalized therapeutic framework. *Behavior Therapy*. https://doi.org/10.1016/j.beth.2025.01.006

World Health Organization. (2022). *World mental health report: Transforming mental health for all*. World Health Organization.

Yakushko, O. (2019). Eugenics and its evolution in the history of western psychology: A critical archival review. *European Journal of Counselling Psychology, 8*(1), 13–36. https://doi.org/10.5964/ejcop.v8i1.247

Yerkes, R. M. (1921). *Psychological examining in the United States Army*. Washington, DC: Government Printing Office (National Academy of Sciences).

Zachar, P., & Kendler, K. S. (2012). A brief historicity of the diagnostic and statistical manual of mental disorders: Issues and implications for the future of psychiatric canon and practice. *History of Psychiatry, 23*(1), 3–19. https://doi.org/10.1177/0957154X11427727

Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. Cambridge University Press.

Corrigendum to "One size fits none: A call for an idionomic revolution" [Clinical Psychology Review, 129 (2026) 102800]

Steven C. Hayes, Joseph Ciarrochi, Baljinder K. Sahdra, Stefan G. Hofmann, Clarissa W. Ong, and Cristóbal Hernández

The authors regret that the above article contained errors in referencing. Three entries contain errors that also affect the citation in the body of the text; thirteen entries contain errors in the reference alone. None of the corrections alters any claim made in the article. The corrections are given below. The authors apologize for any inconvenience caused.

1. References requiring a change in the body of the text

On page 3 the text refers to "Rajakannan, Bunting, & Dziadkowiec, 2016." It should read "Rajakannan, Safer, Burcu, & Zito, 2016." The published reference had an incorrect author list and article subtitle. The correct reference is: Rajakannan, T., Safer, D. J., Burcu, M., & Zito, J. M. (2016). National trends in psychiatric not otherwise specified (NOS) diagnosis and medication use among adults in outpatient treatment. *Psychiatric Services, 67*(3), 289–295. https://doi.org/10.1176/appi.ps.201500045

On page 4 the text refers to "Zachar & Kendler, 2012." It should read "Kawa & Giordano, 2012." The published reference does not correspond to a published work. The correct reference is: Kawa, S., & Giordano, J. (2012). A brief historicity of the Diagnostic and Statistical Manual of Mental Disorders: Issues and implications for the future of psychiatric canon and practice. *Philosophy, Ethics, and Humanities in Medicine, 7*(1), Article 2. https://doi.org/10.1186/1747-5341-7-2

On page 5 the text refers to "Liu, Moeller, & Kospi, 2019." It should read "Liu, Kuppens, & Bringmann, 2021." The published reference does not correspond to a published work. The correct reference is: Liu, S., Kuppens, P., & Bringmann, L. (2021). On the use of empirical Bayes estimates as measures of individual traits. *Assessment, 28*(3), 845–857. https://doi.org/10.1177/1073191119885019

2. References requiring a change in the reference list only

The following thirteen entries contained errors in the reference itself; the body of the text does not require change. Six had DOIs that were incorrect. The other elements of the published entry are unchanged. The correct references with correct DOIs are:

Cattell, R. B. (1952). P-technique factorization and the determination of individual dynamic structure. *Journal of Clinical Psychology, 8*(1), 5–10. https://doi.org/10.1002/1097-4679(195201)8:1<5::AID-JCLP2270080103>3.0.CO;2-S

Galton, F. (1889). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*, 135–145. https://doi.org/10.1098/rspl.1888.0082

Kanfer, F. H., & Saslow, G. (1965). Behavioral analysis: An alternative to diagnostic classification. *Archives of General Psychiatry, 12*(6), 529–538. https://doi.org/10.1001/archpsyc.1965.01720360001001

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist, 6*(3), 187–193. https://doi.org/10.1027//1016-9040.6.3.187

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy, 47*(2), 119–127. https://doi.org/10.1016/j.brat.2008.10.018

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences, 17*(4), 585–608. https://doi.org/10.1017/S0140525X00036104

Two entries should not have DOIs. The DOI given should be removed. The other elements of the published entry are unchanged.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Skeels, H. M., & Dye, H. B. (1939). A study of the effects of differential stimulation on mentally retarded children. *Proceedings of the American Association on Mental Deficiency, 44*, 114–136.

Five entries had additional features of the reference that were incorrect. The corrected entries and the error corrected are:

Borgogna, N. C., Owen, T., & Aita, S. L. (2024). The absurdity of the latent disease model in mental health: 10,130,814 ways to have a DSM-5-TR psychological disorder. *Journal of Mental Health, 33*(4), 451–459. https://doi.org/10.1080/09638237.2023.2278107 [author list corrected]

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & 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 [author list corrected]

Lane, S. T., Gates, K. M., Fisher, Z., Arizmendi, C., & Molenaar, P. C. M., (2024). Gimme: Group iterative multiple estimation (version 0.7-18) [R package]. CRAN. https://CRAN.R-project.org/package=gimme. [Two collaborators were mistakenly listed as authors; the title contained an extra word]

Shapiro, M. B. (1961a). The single case in fundamental clinical psychological research. *British Journal of Medical Psychology, 34*(3-4), 255–262. https://doi.org/10.1111/j.2044-8341.1961.tb00949.x [title, journal, volume, and pages corrected; DOI was missing]

Yakushko, O. (2019). Eugenics and its evolution in the history of western psychology: A critical archival review. *Psychotherapy and Politics International, 17*(2), Article e1495. https://doi.org/10.1002/ppi.1495 [journal, article number, and DOI corrected]