

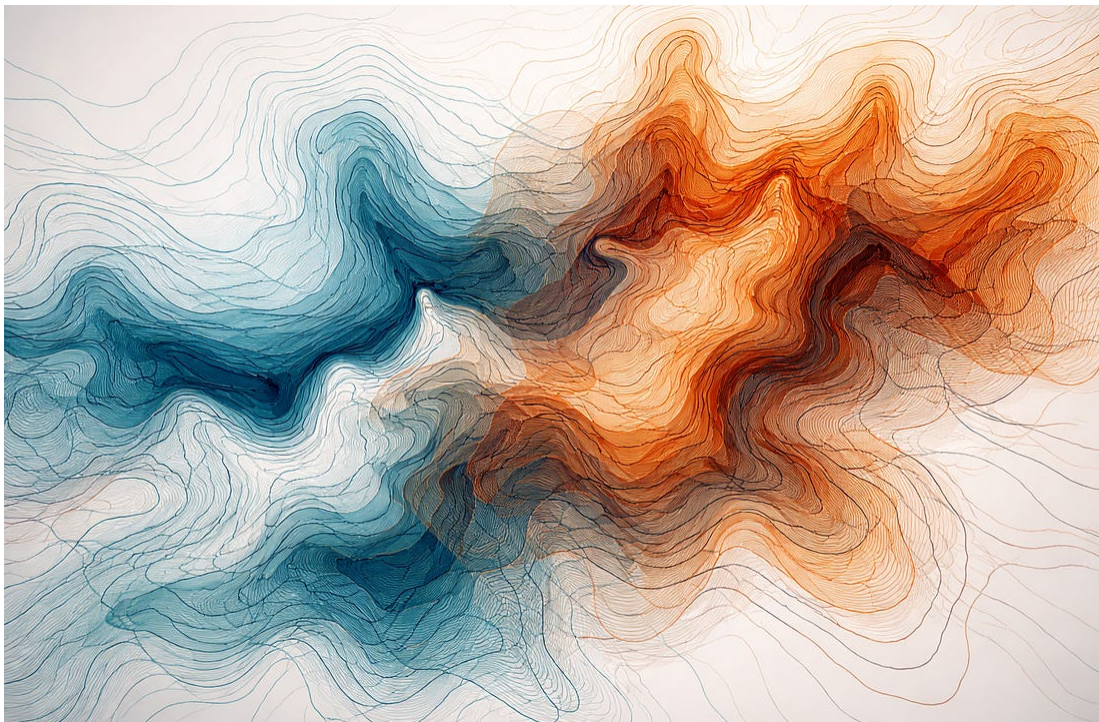
The Convergence Problem

What happens when the map of psychopathology changes depending on who's holding it



KIMBERLY (KIMMIE) GILBERT

MAR 17, 2026



Here is a finding that should be generating far more discussion than it is: when you measure psychopathology every day and factor analyze the data at two levels simultaneously, the structure of how people differ from each other and the structure of how any one person's symptoms move together across days are not the same.

This is not a psychometric nuisance. It is a window into the nature of psychopathology itself. And the specific pattern of where the two

structures converge and where they fracture contains information about mechanism that neither level can provide alone.

I want to make three arguments. First, that the degree of convergence between within-person and between-person structure is itself meaningful data, indexing the match between assessment timescale and the timescale of the underlying process. Second, that the most important tension in the field right now, between the idea that personality pathology has a shared interpersonal core and the finding that within-person processes are wildly heterogeneous, can be resolved by placing these claims at different levels of a single hierarchy. Third, that holding both levels in view simultaneously generates novel hypotheses about specific clinical phenomena, using self-harm as a case study.

Background: two levels, two structures

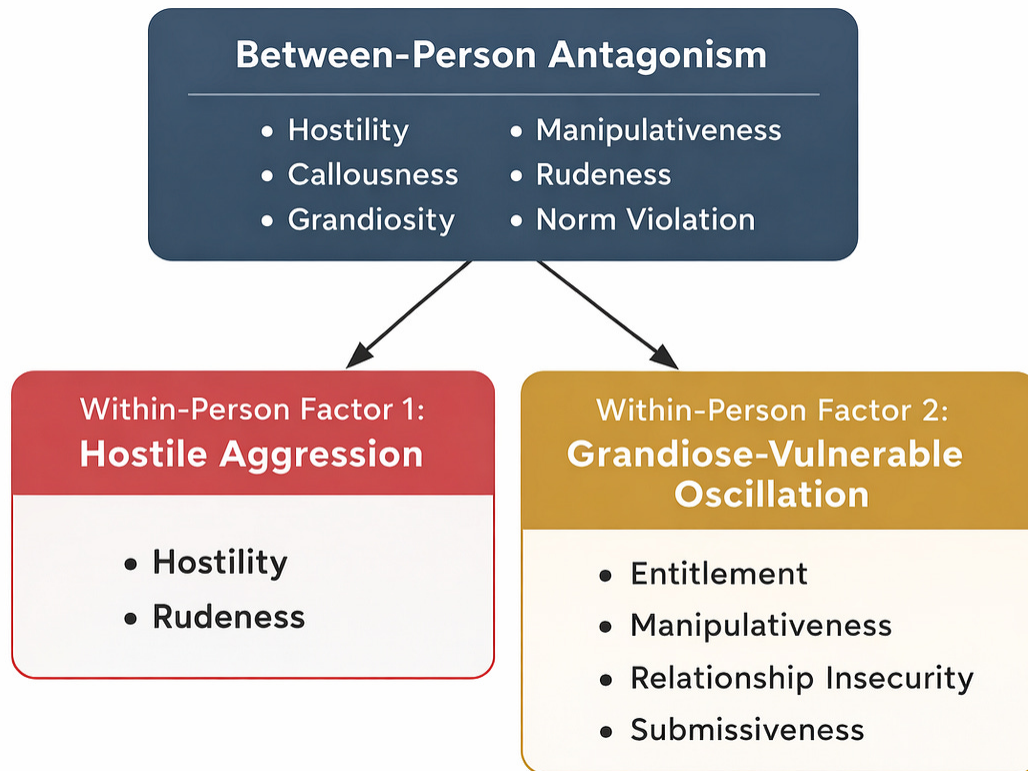
The between-person structure of psychopathology is now well established. Through decades of factor analytic work culminating in systems like HiTOP (Kotov et al., 2017), we have a hierarchical dimensional model organized around broad spectra: Internalizing, Antagonism, Detachment, Disinhibition, Thought Disorder. These domains replicate across cultures, informants, and methods. They align with the Big Five personality structure in ways that integrate normative and pathological individual differences (Markon, Krueger, & Watson, 2005; Widiger & Simonsen, 2005). This is a genuine scientific achievement.

But this structure was derived entirely from between-person data. It tells us how individuals differ from one another in their tendencies. It does not tell us how any given person's symptoms move together from day to day. Molenaar (2004) argued two decades ago that between-

person patterns need not apply to within-person processes, and the accumulating data bear this out.

Recent multilevel factor analyses of 27 daily psychopathology facets (Wright, Ringwald, & Zimmermann, 2025) showed that between-person and within-person structures diverge in systematic, interpretable ways. Some factors have clear counterparts across levels: Internalizing, Detachment, Disinhibition, and Anankastia each emerged at both levels, though narrower and differently composed within persons. But other factors fractured entirely.

The most dramatic fracture involved Antagonism. Between persons, it is a coherent factor: people who are more hostile tend also to be more callous, grandiose, and manipulative than others. Within persons, it split into two factors. One captured hostile aggression straightforwardly. The other combined entitlement and manipulativeness with relationship insecurity and submissiveness. On the same days that individuals felt more entitled and manipulative, they also felt more insecure and submissive. These experiences, which sit on opposite sides of the interpersonal circumplex (Alden, Wiggins, & Pincus, 1990) and have near-zero or negative correlations between persons, rise and fall together within persons across days.



Clinicians will recognize this immediately. Kernberg (1984) described oscillation between idealization and devaluation. Pincus and Lukowitsky (2010) formalized fluctuation between grandiose and vulnerable narcissistic states. Edershire and colleagues (2019) demonstrated that these states are positively correlated between persons but uncoupled within persons at the momentary level. The multilevel factor analysis recovers this clinical reality from raw daily data, without theoretical imposition.

The point: this dynamic is invisible at the between-person level. Average across days and individuals and the grandiose moments and insecure moments cancel out or separate onto different factors. The between-person map says antagonistic people are not submissive. The within-person map says the same person is often both, and that this oscillation may be the pathology itself.

Argument 1: convergence indexes timescale

Why do some factors hold together across levels while others fracture? The standard explanation is that within-person structure is noisier, less reliable, or less well-defined. But the divergences in the Wright, Ringwald, and Zimmermann (2025) data are not random degradation. They are systematic and interpretable. Hostile aggression separates cleanly from grandiose-vulnerable oscillation. Between-person Antagonism is not a noisy version of within-person Antagonism; it is a different construct, composed of different elements.

I want to propose a specific explanation: convergence across levels signals that the trait and its constituent processes share a timescale with the assessment window. Divergence signals a mismatch. Other factors could also contribute to structural divergence across levels, including differences in item sampling, recall distortion in end-of-day reports, context aggregation, and constraints imposed by the factor model itself. The timescale account is not the only possibility, but it is the one that generates the most specific and testable predictions.

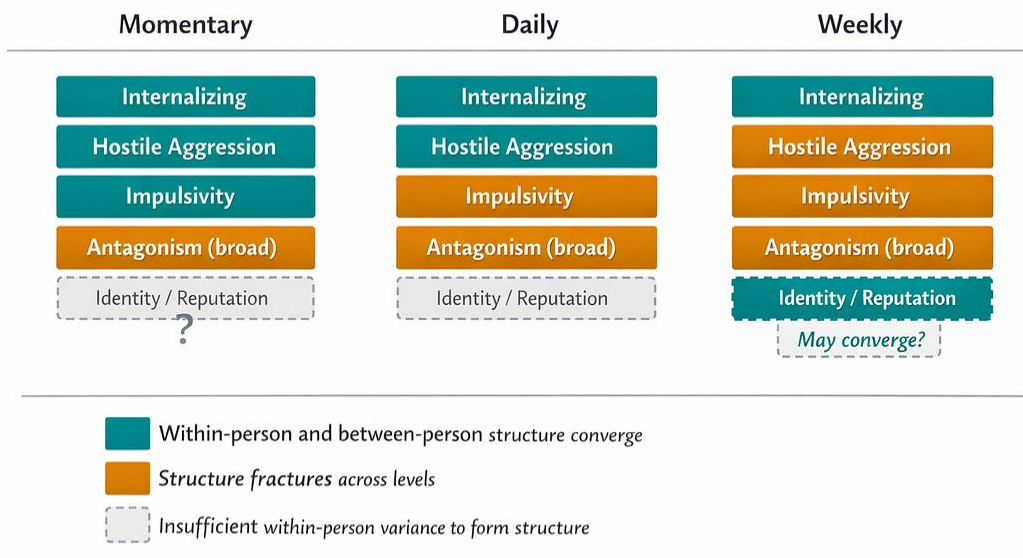
Internalizing features (anxiety, depression, affective lability) converge across levels in daily data because these experiences fluctuate meaningfully on a daily timescale. The between-person and within-person structures align because daily assessment captures the relevant dynamics at the right temporal resolution. Collins (2006) argued generally that the most informative longitudinal designs match measurement timing to process timing. Here, the timescale match produces structural convergence as a byproduct.

Antagonism fractures because its between-person composition reflects accumulated behavioral patterns that differentiate individuals over months and years, while the within-person processes generating those patterns operate faster than daily assessment can cleanly resolve. Rapid state shifts between grandiosity and vulnerability, context-dependent hostile reactivity, momentary regulatory failures: these play out over

hours or minutes, not days. Daily assessment captures the residue of these processes but averages over their internal structure. The between-person factor is the geological sediment of within-person dynamics, and sediment has a different composition from the river that deposited it.

The metaphor is useful but incomplete. The relationship between levels is not one-directional. Traits constrain which situations people enter and how they process them, shaping within-person dynamics, which over time either reinforce or gradually reshape trait levels (Hopwood, Bleidorn, & Wright, 2022). Parallel process growth models have shown that changes in Big Five traits track with changes in borderline personality disorder symptoms over 16 years (Wright, Hopwood, & Zanarini, 2015). The river deposits the sediment, but the canyon also channels the river. Still, the core point holds: even in a bidirectional system, the composition of the sediment and the composition of the river can differ, and the difference is informative.

This framing generates a falsifiable prediction: the degree of structural convergence between within-person and between-person factors should be a function of assessment timescale. Specifically, constructs driven by processes that fluctuate on a daily scale should converge in daily data. Constructs driven by faster processes (impulsive episodes, hostile outbursts, affective cascades within social interactions) should converge in momentary assessment but fracture in daily data. And constructs that accumulate slowly (interpersonal reputation, identity coherence, relational patterns) may never show meaningful within-person structure in any intensive longitudinal design, because the relevant variability exists across developmental timescales.



Predicted pattern: convergence tracks the match between assessment timescale and process timescale

Hypothesized pattern

There is already preliminary evidence consistent with this. Daily impulsivity, hostility, and manipulativeness have the lowest intraclass correlations in daily diary data, meaning the highest proportion of within-person variance, and they show episodic spikes rather than smooth oscillation (cf. Trull et al., 2008; Ebner-Priemer et al., 2007). These are processes whose natural timescale is faster than daily assessment, and daily measurement captures their echo but not their structure. When within-person antagonism narrows to hostile aggression specifically, this may be because hostile aggression is the component that actually fluctuates at a daily resolution, while the rest of between-person antagonism (callousness, grandiosity, manipulativeness as dispositions) varies meaningfully only at longer intervals.

The strong test would be a study measuring the same constructs at momentary, daily, and weekly timescales in the same individuals, then comparing the factor structures that emerge at each level. The timescale hypothesis predicts specific, differential patterns of

convergence and fracture across assessment windows. If hostile aggression converges at the momentary level but fractures at the daily level, that tells you something precise about its temporal grain. If callousness never shows within-person structure at any timescale, that suggests it is a genuinely dispositional construct without meaningful short-term fluctuation, at least in the populations studied.

Argument 2: domain versus architecture

There is an unacknowledged contradiction in the field right now, and resolving it requires the kind of multilevel thinking the convergence data invite.

One body of evidence says personality pathology has a shared interpersonal core. The two explicitly interpersonal dimensions of psychopathology, Antagonism and Detachment, are uniquely saturated by personality disorder diagnoses in meta-analytic structural models (Ringwald, Forbes, & Wright, 2021). Wilson, Stroud, and Durbin (2017) confirmed that interpersonal dysfunction is pervasive across personality disorder categories. Clinically, what makes personality disorders difficult to treat is that the therapeutic relationship itself becomes a site of difficulty (Bender, 2005).

But the personalized modeling literature says something very different: within-person processes are wildly heterogeneous, even among individuals sharing the same diagnosis. In samples analyzed with automated model-search algorithms like GIMME (Gates & Molenaar, 2012), almost no dynamic paths are shared by a majority of individuals. In p-technique studies, factor structures range from one to five factors. Fisher and colleagues (2017) found comparable heterogeneity in internalizing disorders.

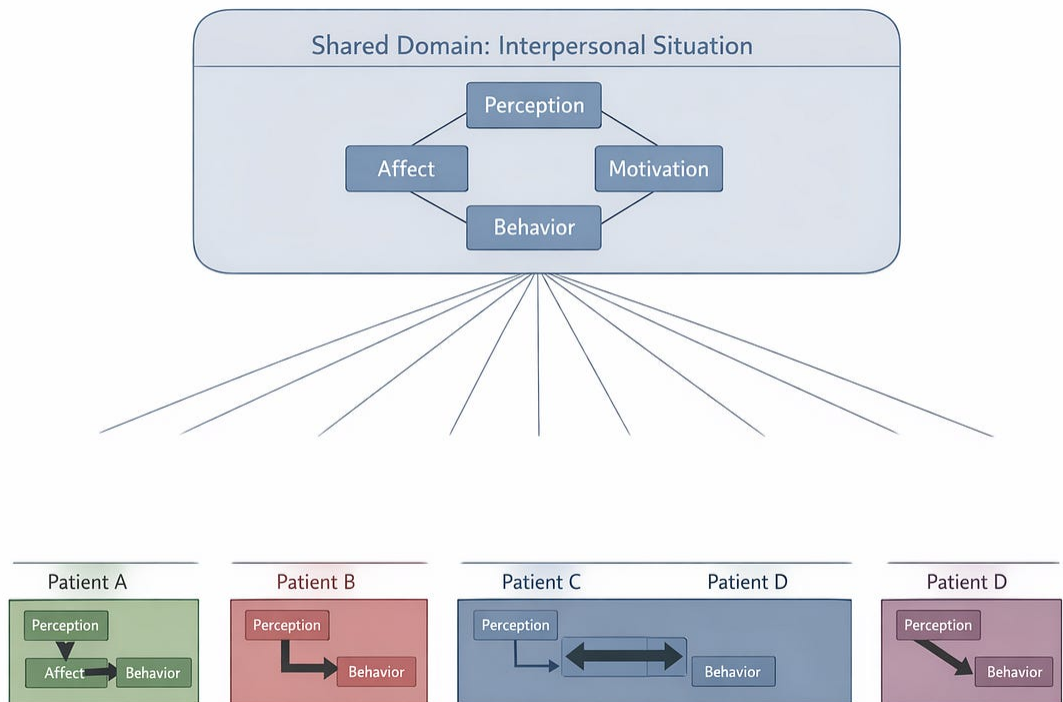
The shared-core proposal says there is a commonality. The personalized models say there is not. Both are well supported. How do

we reconcile them?

By recognizing that they are claims at different levels. The interpersonal core is a claim about content domain: the dysfunction, whatever its specific form, plays out in social transactions involving perception, behavior, affect, and motivation in relation to others. The heterogeneity finding is a claim about process architecture: the specific configuration of which perceptual biases, behavioral rigidities, affective contingencies, and motivational priorities are operative is unique to each individual.

These are not contradictory once you place them in a hierarchy. The interpersonal situation, as elaborated by contemporary integrative interpersonal theory (Sullivan, 1953; Pincus, 2005; Pincus & Hopwood, 2012), is the shared stage on which person-specific dramas play out. Every patient with an interpersonal disorder has dysfunction manifesting in interpersonal situations. But the specific component of the situation that breaks down (perception versus behavior versus affect versus motivation), the specific interpersonal content involved (agency versus communion), and the specific dynamic pattern (rigidity versus instability versus amplification) are all person-specific.

This framing is only as good as the interpersonal domain's empirical track record. If non-interpersonal variables, such as inflammatory processes, sleep architecture, or circadian disruption, consistently proved to be stronger within-person predictors of personality disorder dynamics than interpersonal variables, the domain constraint would need revision. The reason to privilege the interpersonal domain currently is not theoretical preference but the repeated finding that interpersonal variables carry substantial unique predictive weight in personality pathology samples whenever they are measured alongside alternatives. But this is an empirical wager, not an axiom, and it should be treated as such.



The distinction matters practically, not just theoretically. The domain-level claim constrains the model search. If you know the dysfunction is interpersonal, you know to measure interpersonal variables (perceptions of others, behavioral responses, social affect, relational motives) rather than, say, sleep architecture or dietary patterns. You know to look for dysfunction in the transaction cycle between self and other, not in isolated cognitive processes. You know to expect that the therapeutic relationship will become activated as a site of the pathology, and to plan accordingly. The domain is not the formulation, but it tells you where to look for the formulation.

The architecture-level claim then specifies the formulation. Within the interpersonal domain, is this particular patient's hostility driven by distorted perception of others' intentions? Overvalued agentic motives? Maladaptive behavioral selection in response to accurately perceived provocation? These are competing hypotheses that can be pitted against each other using person-specific time-series models, but only if a theory generates them. Platt (1964) called this strong inference, and it

requires theoretical constraint on the model space. The interpersonal tradition provides exactly this constraint, with falsifiable predictions about complementarity, motive satisfaction contingencies, and normative transactional patterns that serve as the reference against which individual deviation can be identified.

Without the domain, personalized models are unconstrained and underdetermined. Bastiaansen and colleagues (2019) demonstrated this directly: twelve teams, same data, twelve different models, twelve different treatment recommendations. Without the architecture, domain labels are clinically inert. Telling a clinician “this patient has interpersonal dysfunction” is about as useful as telling a mechanic “something is wrong with the engine.” The combination of constrained domain and person-specific architecture is what makes formulation actionable.

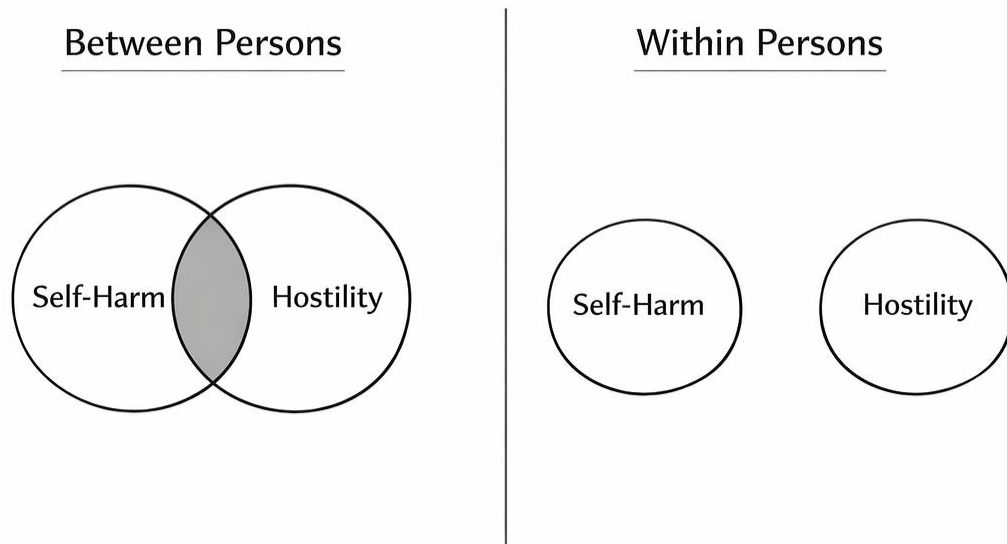
Argument 3: the self-harm puzzle

Holding both levels of structure in view simultaneously generates novel hypotheses that neither level alone supports. Self-harm provides a striking example.

In the Wright, Ringwald, and Zimmermann (2025) multilevel analysis, self-harm loaded on a weak, poorly saturated factor at the between-person level, partially associated with Antagonism and Internalizing but mostly orphaned. Self-harm scales have behaved anomalously in factor analyses for years (cf. Markon, Krueger, & Watson, 2005), and the usual interpretation has been that self-harm is either a method factor, a severity indicator, or simply hard to model structurally.

But the multilevel analysis adds a critical new piece of information: at the within-person level, self-harm showed no association with hostile aggression. On a day when someone harms themselves, they are not more hostile toward others than usual. The processes generating self-

directed harm and other-directed harm are dynamically distinct within persons, even though the people who engage in self-harm tend also to be people elevated on hostility between persons.



Same people. Different days. Different processes.

This dissociation between levels is exactly the kind of signal the convergence framework predicts when a between-person association is driven by shared liability (both self-harm and hostility are associated with personality pathology severity) rather than shared process (the daily dynamics of self-harm and other-directed hostility are independent). The between-person correlation between self-harm and hostility reflects the fact that the same people tend to do both. The within-person independence reveals that they do them on different days, in response to different triggers, through different mechanisms.

The dissociation is the finding. What drives self-harm within persons, if not the hostile-aggressive system, remains an open question with multiple candidates. Some are interpersonal: perceived burdensomeness and thwarted belonging, as Joiner's interpersonal theory of suicide proposes (Hames, Hagan, & Joiner, 2013). Others are intrapersonal:

pain-offset relief, dissociative regulation, shame-driven self-punishment. Kaurin and colleagues (2022) have shown that momentary suicidal surges in borderline personality disorder follow proximal interpersonal trajectories distinct from general negative affect, suggesting that at least some of the time the drivers are social. But whether any given episode of self-harm is driven by interpersonal or intrapersonal processes for a given individual is exactly the kind of question that person-specific modeling within a theoretically constrained framework should answer. The multilevel structural data tell you where to stop looking (the hostile-aggressive system). They do not by themselves tell you where to start. That requires both theory and individual-level data.

Where this leaves us

Between-person structure tells us about liability: the enduring characteristics that make individuals differently vulnerable. Within-person structure tells us about process: the dynamics through which dysfunction manifests and maintains. Where they converge, liability and process share a timescale. Where they diverge, the construct we have reified at one level is actually a composite of different processes at the other.

For diagnosis, this means identifying the domain of dysfunction is a between-person question, while formulating what maintains it is a within-person question. Diagnosis identifies the stage. Formulation identifies the play.

For treatment, it means that manualized therapies typically assume that process models derived from group-level research apply uniformly to individual patients. But the strength and even the direction of key therapeutic targets, such as the link between cognition and affect, may vary substantially across individuals and across time within the same individual. Hayes and colleagues (2019) have argued for exactly this

shift toward identifying individual-level processes as the basis for process-based therapy. If you are treating a patient for whom a core therapeutic target is weak or absent, you are not applying a wrong theory in general. You are applying the right theory to the wrong person.

For research, the field still lacks the decisive comparative test: a trial comparing process-matched treatment, where interventions target each individual's empirically identified dysfunctional process, against standard protocol treatment. Fernandez, Fisher, and Chi (2017) developed an algorithm for this, and the first open trial (Fisher et al., 2019) produced large average improvements, but without a comparison condition. That comparison is the test that would tell us whether personalized models actually improve outcomes. Everything else is promissory.

The anatomy has been mapped. It is time to learn the physiology. And the first step is to stop assuming they will look the same.

References

- Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, 55, 521–536.
- Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F., Boker, S. M., Ceulemans, E., et al. (2019). Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology. *PsyArXiv*.
- Bender, D. S. (2005). The therapeutic alliance in the treatment of personality disorders. *Journal of Psychiatric Practice*, 11, 73–87.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, 57, 505–528.

Ebner-Priemer, U. W., Kuo, J., Kleindienst, N., Welch, S. S., Reisch, T., Reinhard, I., et al. (2007). State affective instability in borderline personality disorder assessed by ambulatory monitoring. *Psychological Medicine*, 37, 961–970.

Edershile, E. A., Woods, W. C., Sharpe, B. M., Crowe, M. L., Miller, J. D., & Wright, A. G. C. (2019). A day in the life of Narcissus: Measuring narcissistic grandiosity and vulnerability in daily life. *Psychological Assessment*, 31, 913–924.

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, 126, 1044–1056.

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63, 310–319.

Hames, J. L., Hagan, C. R., & Joiner, T. E. (2013). Interpersonal processes in depression. *Annual Review of Clinical Psychology*, 9, 355–377.

Hopwood, C. J., Bleidorn, W., & Wright, A. G. C. (2022). Connecting theory to methods in longitudinal research. *Perspectives on Psychological Science*, 17, 884–894.

Kaurin, A., Hallquist, M. N., Dombrovski, A. Y., & Wright, A. G. C. (2022). Momentary interpersonal processes of suicidal surges in borderline personality disorder. *Psychological Medicine*, 52, 2702–2712.

Kernberg, O. F. (1984). *Severe personality disorders: Psychotherapeutic strategies*. Yale University Press.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., et al. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*, 454–477.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139–157.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality. *Psychological Review, 102*, 246–268.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science. *Measurement, 2*, 201–218.

Pincus, A. L. (2005). A contemporary integrative interpersonal theory of personality disorders. In M. Lenzenweger & J. F. Clarkin (Eds.), *Major theories of personality disorder* (2nd ed., pp. 282–331). Guilford Press.

Pincus, A. L., & Hopwood, C. J. (2012). A contemporary interpersonal model of personality pathology and personality disorder. In T. A. Widiger (Ed.), *Oxford handbook of personality disorders* (pp. 372–398). Oxford University Press.

Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology, 6*, 421–446.

Platt, J. R. (1964). Strong inference. *Science, 146*, 347–353.

Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2021). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*. Advance online publication.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. Norton.

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

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*, 110–130.

Wilson, S., Stroud, C. B., & Durbin, C. E. (2017). Interpersonal dysfunction in personality disorders: A meta-analytic review. *Psychological Bulletin, 143*, 677–734.

Wright, A. G. C., Ringwald, W. R., & Zimmermann, J. (2025). Measuring psychopathology in daily life. *Clinical Psychological Science, 13*(3), 649–663.