The best of both worlds? *General* principles of psychopathology in *personalized* assessment

Merlijn Olthof[1,2], Fred Hasselman[2], Benjamin Aas[3,4], Daniela Lamoth[3], Silvia Scholz[3,5], Nora Daniels-Wredenhagen[3,5], Florens Goldbeck[3], Els Weinans[6], Guido Strunk[7], Günter Schiepek[8], Anna M. T. Bosman[2], Anna Lichtwarck-Aschoff[1]

[1]Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, the Netherlands

[2]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

[3]sysTelios Think Tank, sysTelios Health Center, Siedelsbrunn, Germany

[4]Faculty of Psychology and Educational Sciences, LMU Munich, Munich, Germany

[5]Department of Clinical Psychology and Psychotherapy III, Witten/Herdecke University, Witten, Germany.

[6]Copernicus Institute of Sustainable Development, Environmental Sciences, Faculty of Geoscience, Utrecht University, Utrecht, The Netherlands

[7]Complexity Research, Vienna, Austria; FH Campus Vienna, Vienna, Austria and Technical University Dortmund, Dortmund, Germany.

[8]Institute of Synergetics and Psychotherapy Research, University Hospital of Psychiatry, Psychotherapy and Psychosomatics, Paracelsus Medical University, Salzburg, Austria; Department of Psychology, Ludwig Maximilian University of Munich, Munich, Germany

Note: this manuscript is the final author version that was accepted for publication at the *Journal of Psychopathology and Clinical Science*. This version has not been proof-edited and is not the copy of record. The final article is available, upon publication, at the *Journal of Psychopathology and Clinical Science*.

Supplemental materials are attached at the end of this file

Correspondence should be sent to Merlijn Olthof, merlijn.olthof@ru.nl

Citation:

Olthof, M., Hasselman, F., Aas, B., Lamoth, D., Scholz, S., Daniels-Wredenhagen, N., Goldbeck, F., Weinans, E., Strunk, G., Schiepek, G., Bosman, A., & Lichtwarck-Aschoff, A. (in press). The best of both worlds? General principles of psychopathology in personalized assessment. *Journal of Psychopathology and Clinical Science*. https://doi.org/10.31234/osf.io/jqp7g

Abstract

A complex systems approach to psychopathology proposes that general principles lie in the dynamic patterns of psychopathology, which are not restricted to specific psychological processes like symptoms or affect. Hence, it must be possible to find general change profiles in time series data of fully personalized questionnaires. In the current study, we examined general change profiles in personalized self-ratings and related these to four measures of treatment outcome (International Symptom Rating, Depression Anxiety and Stress Scale-21, daily symptom severity and self-reflective capacity). We analyzed data of 404 patients with mood and/or anxiety disorders who completed daily self-ratings on personalized questionnaires during psychotherapy. For each patient, a principal component analysis was applied to the multivariate time series in order to retrieve an univariate person-specific time series. Then, using classification and regression methods, we examined these time series for the presence of general change profiles. The change profile classification yielded the following distribution of patients: no-shift (n = 55; 14%), gradual-change (n = 52; 13%), one-shift (n = 233; 58%), reversed-shift (n = 39; 10%) and multiple-shifts (n = 25; 6%). The multiple-shifts group had better treatment outcome than the no-shift group on all outcome measures. The one-shift and gradual-change group had better treatment outcome than the no-shift group on respectively two and three outcome measures. Overall, this study illustrates that person-specific (idiographic) and general (nomothetic) aspects of psychopathology can be integrated in a complex systems approach to psychopathology, which may combine 'the best of both worlds'.

Keywords: idiographic, nomothetic, complex systems, sudden gain

General scientific summary

Personalized questionnaires have clinical and intuitive appeal, but the question rises how to generalize if questionnaires measure different constructs in different patients. In this paper, we develop a novel method to abstract general change profiles from personalized daily questionnaire data collected during psychotherapy. We find that the way in which patients change on their personalized questionnaire over the course of treatment is related to treatment outcome as predicted by the complexity theory of psychopathology.

Psychopathology is highly individualized as studies illustrate vast heterogeneity in symptoms (Allsopp et al., 2019; Fried & Nesse, 2015), case formulations (van den Bergh et al., 2022), and change trajectories (Schiepek et al., 2020). This heterogeneity limits group-level research in psychopathology, which often rests on the assumption of homogenous sets of patients with similar diagnosis. It is therefore, that recent research aims at developing personalized approaches to psychopathology and clinical change (Wright & Woods, 2020). A pressing challenge in this field of research is how to balance the tension between doing justice to the individuality of psychopathology while at the same time generalizing across patients.

Recent research on personalized psychopathology often uses time series of self-ratings collected through methods like experience sampling (e.g., Fisher, 2015). Advancements in multilevel modelling have enabled researchers to model both inter- and intra-individual differences in time series in great detail (e.g., Beltz et al., 2016). However, even the most advanced modelling approaches still require standardized questionnaires (i.e., the same questionnaire for every patient) which limits the extent of personalization. Given the massive heterogeneity in psychopathology, personalized questionnaires potentially better capture the person-relevant dynamics of individual patients, for instance by basing questions on clinical case formulation (Kramer, 2020; Schiepek, et al., 2016).

While personalized questionnaires have intuitive and clinical appeal (e.g., Lloyd et al., 2019), the question rises how to generalize across patients if personalized questionnaires measure different constructs in different individuals. A complex systems approach aims to generalize in terms of dynamics, not content, and is therefore a promising theoretical framework for studying personalized questionnaires. Specifically, the complexity theory of psychopathology (Olthof et al., in press) states that general principles lie in the dynamic patterns of psychopathology, which are not restricted to specific psychological processes like symptoms or affect, and should in principle be measurable in any person-relevant *collective variable*[1].

In this paper, we use personalized questionnaires to study two of these general principles – attractors (stable configurations of a complex systems) and transitions (often abrupt changes from one attractor to another) – as 'fine grains' in psychopathology (for detailed introductions see e.g. Hayes & Andrews, 2020; Olthof et al., in press). First, psychopathology may function as an *attractor,* a dynamic pattern that a person keeps 'falling

[1] Collective variables are variables which 'summarize' the macro-level dynamics of a high-dimensional complex system in much lower dimensionality. For example, the macro-level dynamics for an individual with rapid-cycling bipolar disorder may be adequately described by time series measures of mood.

into'. The impossibility to disengage from psychopathology is crucial to the idea of an attractor. For example, it is not considered pathological to react anxious in an unexpected encounter with a spider, while it is considered pathological to be fearful of spiders all the time, even when they are not present. To be 'stuck' in a state of fear (an attractor) is considered pathological, not the fear itself (Kashdan & Rottenberg, 2010).

Second, clinical improvement, in the complex systems framework, is the *transition* from a psychopathological attractor towards a healthier attractor state (Hayes et al., 2007). Such transitions are theorized to often occur suddenly and abruptly at so-called *tipping points* where a new attractor starts to attract the system more strongly than the old one (Scheffer et al., 2009). For psychotherapy and other treatments, transitions between attractors have been hypothesized to be the key vehicle for clinical change (Hayes et al., 2007; Schiepek et al., 2016a). Patients seeks help because they are *stuck* in an undesirable attractor state and successful psychotherapy supports the transition towards a more desirable attractor state, for which 'being stuck in' is often referred to as *resilience*. The content and meaning of the attractor states (e.g., in terms of experienced symptoms) can be completely person-specific in this framework, which may in fact be expected given the massive heterogeneity of psychopathology.

Although often not phrased in complex system terms, abrupt transitions have been studied extensively in clinical change trajectories. Most notably, sudden gains – abrupt and enduring improvements in symptoms – may be understood as transitions from one attractor to another (Hayes et al., 2007). Sudden gains occur frequently (estimates range from ~ 15-60%) and have mostly been studied in weekly and session-based symptom or therapy process questionnaires (for a meta-analysis see Shalom & Aderka, 2020). Patients may also experience sudden losses (i.e., sudden deterioration), as well as combinations of (multiple) gains and losses over the course of treatment (Olthof et al., 2020a). Sudden gains have also been found in time series of daily symptom ratings collected during psychotherapy (Helmich et al., 2020). The same study also found gradual change profiles, albeit in only few patients. Such gradual transitions from one attractor to another are also possible in complex systems and form a special case in which systems are to a lesser extent 'stuck' in attractors (i.e., there is no hysteresis effect; Kèfi et al., 2012).

In the present work, we aim to identify personalized transitions in psychopathology, as measured with fully personalized daily self-ratings over the course of psychotherapeutic treatment. Following the complexity theory of psychopathology (Olthof et al., in press), we propose that these self-ratings can function as personalized collective variables, which

measure the macro-level dynamics of individual change processes of patients. In line with the hypothesis that successful psychotherapy is characterized by a transition from one attractor state to another, we predict that patients who transition towards a different state in their personalized self-ratings have better treatment outcome than patients who do not transition or those who do transition but later fall back to their initial state. As we do not a priori know what types of transitions will be identified, we have no specific hypotheses about the differences between different types of transitions, as long as they reach a different end state compared to the start of treatment.

Methods

Study sample

Data were collected as part of routine care, which adhered to the Declaration of Helsinki. All patients gave consent to use their data for scientific purposes. The original dataset included 653 patients diagnosed with mood and/or anxiety disorders completed personalized self-ratings while receiving psychotherapy treatment at an inpatient clinic between 2013-2020. The clinic provides a treatment routine incorporating elements from various therapeutic orientations. Daily group and individual sessions of conversation-based therapy were provided by psychotherapists who completed training in cognitive-behavioral therapy, psychodynamic therapy or systemic therapy. In addition, creative and body-oriented therapy sessions were provided by trained music-therapists, art-therapists, body-therapists and physiotherapists. The average (*SD*) treatment duration was 53 (16) days.

Patients were selected from the sample when their time series were 14 days or longer (n = 520), had at least 3 items (n = 498), 80% compliance (n = 432) and maximum 3 missing values in a sequence, leading to a final sample of 404 patients consisting of 168 (42%) men and 235 (58%) women. Mean age was 48 years (*SD* = 12). Patients had on average 13 items (*SD* = 7) in their personalized questionnaire and a time series length of 36 days (*SD* = 15) with a median of 2% missing values (range 0 – 20 %).

Materials

Personalized questionnaires

Personalized questionnaires were answered daily in the evening using a web application called the Synergetic Navigation System (Schiepek et al., 2016a). The personalized questionnaires were co-created by therapists and patients based on case formulation using idiographic system modelling (ISM; Schiepek, 1986). The ISM procedure has been described in more detail elsewhere (Schiepek et al., 2016a, 2016b; van den Bergh et al., 2022). In short, therapists and patients collaboratively identified all concepts crucial to a patient's current situation and mapped these onto a network visualization. All concepts referred to processes that change over time (e.g., 'relationship with my partner' instead of 'my partner'), which could therefore be well-translated to questionnaire items. Therapists also aimed to balance positive items (e.g., 'Today I felt self-confident') with negative items (e.g., 'Today I felt anxious'). The dataset contained 8184 different items of which 7823 (95.6 %) were entirely person-specific (i.e., mentioned by only one patient). The vast majority of non-unique items (207; 2.5 %) were answered by two individuals; 48 (0.6 %) items were answered by three or more individuals, with the most common item present for 7 individuals.

Outcome questionnaires

We used four outcome measures. Two outcome measures were measured once at the beginning and once at the end of treatment (last or second-last day of stay in the clinic): the ICD-10 symptom rating (ISR; Tritt et al., 2013) and the depression anxiety stress scale (DASS–21; Lovibond & Lovibond, 1995), which we refer to as the 'standard outcomes'. The ISR is based on the international classification system of diseases (ICD-10; World Health Organization, 2004) and measures a broad range of psychopathology symptoms over multiple subscales. Specifically, the ISR consists of 41 items of which 29 items refer to symptoms of depression (4 items), anxiety (4 items), compulsive-obsessive disorder (3 items), somatoform disorders (3 items), eating disorders (3 items), and 12 questions about general psychopathology and daily life functioning. The ISR has a 5-point Likert scale ranging from 0 (does not apply) to 4 (applies extremely). Previous research found high internal consistency for the overall ISR score as well as for the subscales (Fischer et al., 2010). An example item from the ISR depression subscale is: '*I no longer enjoy doing things I used to enjoy*'. The DASS-21 consists of 21 items and is a shorter version of the 42-item DASS by Lovibond and

Lovibond (1995). In the DASS-21, depression, anxiety, and stress are each measured with 7 items using a 4-point Likert scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time). By convention, DASS-21 scores are multiplied by 2 (because the questionnaires is half the length of the original), so the reported range is 0-6. Previous research found high internal consistency for the overall DASS-21 scale as well as for the subscales (Henry & Crawford, 2005). An example item from the DASS-21 is '*I felt scared without any good reason*'.

The other two outcome measures were computed from standardized daily self-ratings which were collected in addition to the personalized self-ratings with a process questionnaire developed by the clinic (adapted from the Therapy Process Questionnaire; Schiepek et al., 2019). These 'process outcomes' were included because of their better temporal resolution compared to the standard outcomes (Schiepek et al., 2016a). The first process outcome was daily symptom severity, measured with one item '*Today, my symptoms were…*' answered on a visual analog scale ranging from 0 (not at all) to 100 (very much). The other process outcome was self-reflective capacity, the ability to actively observe one's thoughts, feelings, motivation, expectancies, and actions from an inner 'bird's eye perspective' (Beitman & Soth, 2006). Self-reflective capacity was included as an outcome based on expert judgement: Therapists working at the clinic identified self-reflective capacity as the main treatment target of the clinic's treatment rather than symptom reduction per se. Also, previous research suggests that frequent assessment may increase self-insight, which makes it an interesting outcome measure in the current study (van Os et al., 2017). Self-reflective capacity was measured by 4 items of the process questionnaire about self-reflection and insight in one's experience, feelings, problems, and process. We examined the 4 self-reflective capacity items with confirmatory factor analysis and reliability analysis of which the result support the use of these 4 items as a scale (supplemental text 1; Table S1; Figure S1). An example item from the self-reflective capacity scale is *'Today, I was able to adopt an observer's perspective on my inner experience'* answered on a visual analog scale ranging from 0 (not at all) to 100 (very much).

In line with previous research, we used the person-specific average of the process outcomes in the first 7 days of data collection as pre-score and the average of the last 7 days as post-score (e.g., Olthof et al., 2020b). Internal consistencies for the ISR, DASS-21 and self-reflective capacity scale in the current sample are reported in Table 1.

Table 1

Descriptive Statistics of Outcome Measures

	Cronbachs α pre-score	Cronbachs α post-score	Pre-score Mean (SD)	Post-score Mean (SD)	Cohen's D [95% CI]
ISR	0.86	0.91	1.5 (0.6), N = 403	0.7 (0.5), N = 334	1.45 [1.28, 1.62]
DASS-21	0.91	0.92	2.7 (1.1), N = 402	1.0 (0.8), N = 337	1.76 [1.58, 1.94]
Daily symptom severity	-	-	45.5 (21.4), N = 389	36.1 (23.9), N = 389	0.42 [0.27, 0.56]
Self-reflective capacity	0.93	0.94	52.2 (17.7), N = 379	65.3 (18.8), N = 379	0.72 [0.57, 0.86]

Note. Daily symptom severity is measured with 1 item, hence internal consistency cannot be calculated. Cohen's *d* is calculated from the subset of patients who had both pre- and post-scores. Cronbach's alpha for self-reflective capacity was computed for patients with ratings on all 4 items (same sample as used for the factor analysis).

Data-analysis

General strategy

The data-analytic strategy can be understood as a two-part process. The first part was an exploratory analysis to identify change profiles in the personalized self-ratings. The second part tested the hypothesis that patients who transition towards a different state over the course of therapy have better treatment outcomes. To safeguard a fair hypothesis test of the second analysis, the outcome variables were only examined and analyzed after the change profiles were decided upon. All performed outcome analyses, including those on all subscales of the ISR and the DASS (not reported in this manuscript) can be found at the open science framework [https://osf.io/ksym9/; DOI 10.17605/OSF.IO/KSYM9].

Time series analysis

For each person, we performed a principal component analysis (PCA) on the personalized items. We then used the projection on the first principal component (hereafter *PC1*) as one univariate time series per person. Prior research has shown that transitions in multivariate systems can be studied in the first principal component (e.g., Held & Kleinen, 2004). The idea is that the PC1 entails the primary dimension in which the system changes, which is most likely to feature a transition if one occurs (Lever et al., 2020). PCA thus abstracts the relevant dynamics out of the multivariate item time series of each individual patient into one person-specific PC1 time series per person, which may be advantageous compared to some of its alternatives, such as analyzing all items separately (e.g., Bos et al., 2022) or averaging dynamic markers over all items (e.g., Olthof et al., 2020a; 2020b). Missing values in the PC1 time series were imputed using a Kalman filter as available in the R-package ImputeTS (Moritz & Bartz-Beielstein, 2017), which was also used for imputation of the process outcome time series.

Next, we applied recursive partitioning, a classification and regression trees algorithm (also known as CART) for data-driven classification of stable mean levels, to the PC1 time series (Breiman et al., 1984; for a detailed treatment see Therneau & Atkinson, 2022). Recursive partitioning tries to separate a vector of data most optimally into two vectors. Then, it tries to separate those two vectors most optimally in two and so on (i.e., recursive). The resulting solution will always have too many separations, which can be tackled by providing a criterion of change that must occur before classifying a new level. By using the ANOVA method for recursive partitioning (as we did), this criterion, which we call changeSensitivity, equals the R^2 increase that must be observed when introducing a new level.

Recursive portioning was performed using the function *ts_levels* in the R-package casnet (Hasselman, 2022a), which uses the function *rpart* from the R-package rpart (Therneau & Atkinson, 2022). We set the parameters minDataSplit and minLevelDuration to 7, meaning that stable levels should at least last one week. We set minChange to 0, since we cannot interpret change in the PC1 in absolute terms, so we do not want to set an absolute change criterion. We set changeSensitivity to 0.1, which yielded the clearest results after visual inspection with different parameter values. To examine the influence of the changeSensivity parameter setting on the results, we performed the same classification procedure and subsequent analyses for two other values: .05 (more sensitive to shifts) and .15 (less sensitive to shifts).

Based on the results from *ts_levels* we initially classified 4 change profiles: no shift, one shift, reversed shift (an equal number of shifts in opposite directions, e.g., 1 shift to a higher level and 1 shift to a lower score), and multiple shifts (multiple shifts in the same direction, either up or down). Visual inspection of this initial classification suggested that the profile of some patients is better described by a gradual change. We examined this by performing a linear regression and a step-function regression to model the change of PC1 over time. The step-function was dummy coded following the *ts_levels* results (i.e., the dummy code indicates to which stable level of the time series each day belongs). Next, we compared the Akaike Information Criterion (AIC) of both models for each person to see which model fitted best. The results not only confirmed that most trajectories can best be described by a step-function, but also that some trajectories may be better described as a gradual change (see Table S2, Table S3). We therefore added a 'gradual change' profile, featuring persons that met two criteria: a significant slope (at $p < .05$) in the linear model and a better fit in the linear model compared to the step-function (Table S4).

Outcome analysis

We analyzed the relationship between the change profiles and treatment outcome with linear mixed-effects models using the function *lmer* from the R-package lme4 (Bates et al., 2014). Mixed-effects models can also include patients who only have a pre- or post-score, instead of both, on the outcome variable, which is why they were preferred over the general linear model (for number of observations per measure see Table 1). The models had the following general form (written in 'lme4 language'):

(1) Outcome ~ Time * Change Profile + Time * Treatment Duration + (1 | Patient Identifier)

where *Outcome* is the scores on an outcome variable (e.g. ISR), *Time* is a dummy-coded factor of the measurement occasion (pre = 0, post = 1), *Change Profile* is a dummy-coded factor representing the change profile that each person had (reference group = 'no shift'), *Treatment Duration*, is the duration of treatment in days, which was used as a control variable, *(1 | Patient Identifier)* is the random intercept term which models the between-person differences on the pre-test. The 'no-shift' group was chosen as reference group to compare the different types of transitions with. We also exploratively re-ran all models with

the ‘gradual-change’ as reference group to test for possible differences between this group and the reverse shift, one shift, and multiple-shift group.

The set-up of the model means that its results should be interpreted as follows. The grand intercept is the estimated average pre-score of the ‘no-shift’ group. The main effect of *Time* is the estimated average pre to post change of the ‘no-shift’ group. The main effects for the other change profiles are their deviations from the no-shift group at pre-score. The interaction effect of *Time * Change Profile* is the effect of interest, as this one yields the pre-post difference on the outcome variables for the different change profiles compared to the ‘no-shift’ group.

Transparency and openness statement

Open materials for this study are available at the open science framework [https://osf.io/ksym9/; DOI 10.17605/OSF.IO/KSYM9]. Besides code, the open materials feature the output of all performed outcome analyses and additional figures. This study was not preregistered due to the explorative nature of the change profile classification. To ensure a fair hypothesis test on the treatment outcome variables, the relation between change profiles and outcome was only examined after determining the final change profiles classification.

Results

Change profiles

The change profile classification led to the following distribution of patients over the groups: no-shift (n = 55; 14%), gradual-change (n = 52; 13%), one-shift (n = 233; 58%), reversed-shift (n = 39; 10%) and multiple-shifts (n = 25; 6%). Exemplars of each group are shown in Figure 1. For these exemplar patients, we also examined the loadings of the personalized items on the PC1. In Table 2, we list for each exemplar patient the main themes from the 5 items that loaded highest on their PC1, thereby illustrating what was important in their person-specific change processes. The exact items are more person-specific in phrasing and sometimes contain highly idiosyncratic terms (e.g., mentioning of specific objects) and are therefore not included to protect patients’ privacy.

The average variance explained by the first principal component was 50% (SD = 15%). The groups did not differ in how much variance the PC1 explained (F(4, 399) = 0.469, p =.76) nor in the number of items patients had (F(4, 399) = 1.316, p = .26). The groups differed with respect to the length of the time series (F(4, 399) = 8.4716, p <.001). The

multiple-shifts group had the longest time series (*M* = 43.36, *SD* = 12.03, median = 42, range = 26-71), followed by the reversed-shift group (*M* = 41.77, *SD* = 15.55, median = 36, range = 22-93), the one-shift group (*M* = 36.54, *SD* = 14.48, median = 34, range = 14-88), the no-shift group (*M* = 31.09, *SD* = 14.69, median = 26, range = 14-77) and finally, the gradual-change group (*M* = 28.83, *SD* = 10.37, median = 28, range = 14-53). These differences in time series length follow the same pattern as differences in treatment duration; a control variable in the outcome analysis. The groups also differed in their percentage of missing values in the time series (*F* (4, 399) = 2.61, *p* = .035). The reversed shift group had most missing values (*M* = 6,0%, SD = 6,0%), followed by the no-shift group (*M* = 5,0%, *SD* = 6,3%), the one-shift group (*M* = 3,9%, *SD* = 5,0%), the multiple-shift group (*M* = 3,5%, *SD* = 4,8%) and the gradual-change group (*M* = 2,9%, *SD* = 4,7%). In none of the groups, there was a significant differences between the weekdays in terms of missingness (as computed with χ^2 tests).

Table 2

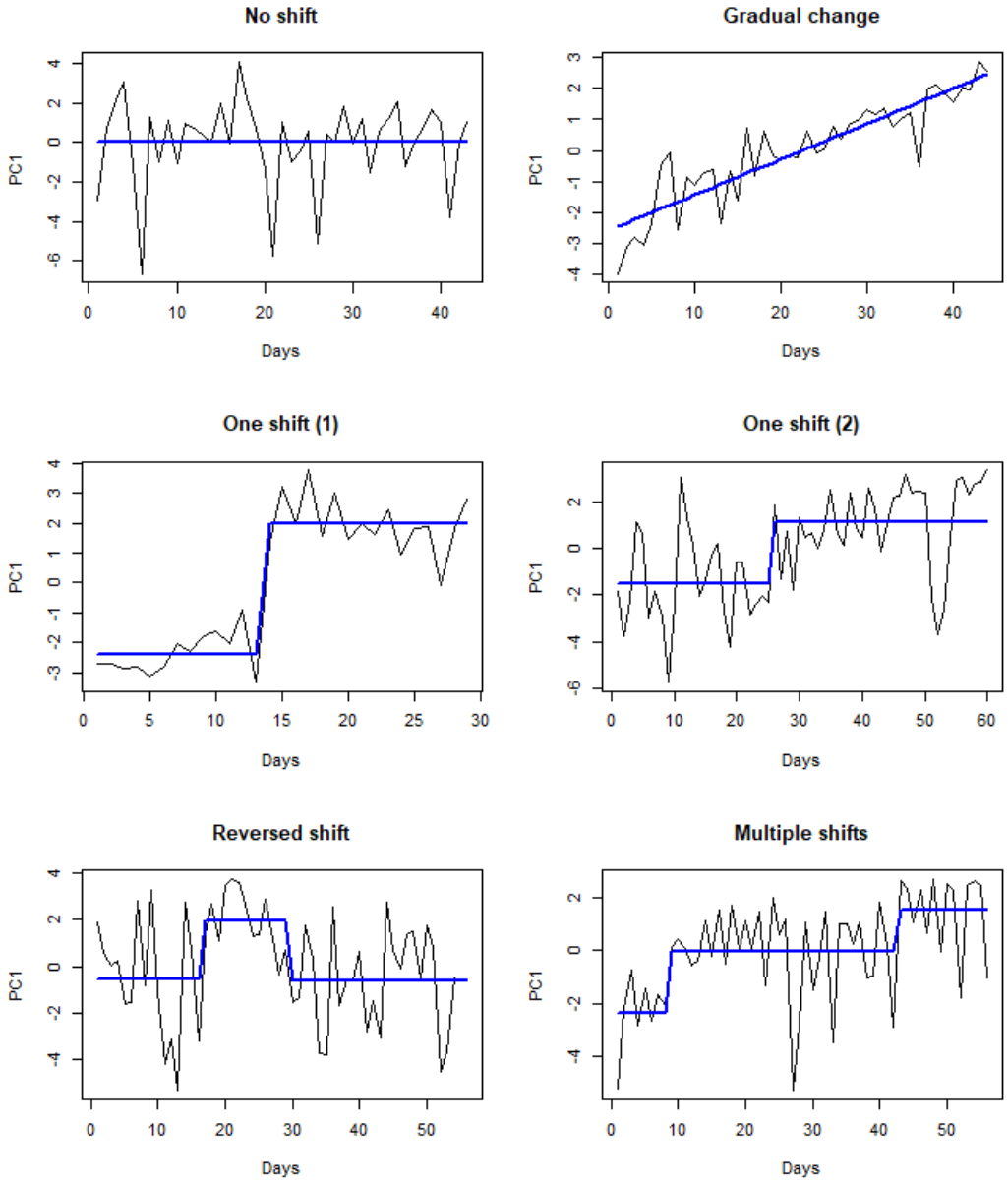
Item themes characterizing the PC1 of the exemplar time series for each change profile from Figure 1.

Exemplar patient	Themes of five highest loading items
No shift	*Patience; Active and healthy life; Allowed current condition; Self-worth; Followed body*
Gradual change	*Distinguished myself; Weakness and strengths; Shown feelings; Slept well*
One shift (1)	*In the here and now; Shown feelings to others; Accepted attention from others; Share what is important; Give space to feelings*
One shift (2)	*Self-confidence; Trust; Felt a little lighter; Felt less guilty; Balance fear*
Reversed shift	*Self-esteem; Outlet for emotions; Experienced creativity; Curious and open; Showed myself*
Multiple shifts	*Not to want; Pushed myself (-); Loaded myself (-); Balance; Marked boundaries*

Note. The exemplar with gradual change only had 4 items. The (-) sign indicates that the item loaded negatively on the PC1.

Figure 1

Exemplar PC1 time series for each change profile



Outcome analyses

Descriptive statistics and correlations

On average, patients decreased on ISR, DASS-21, and daily symptom severity and increased in self-reflective capacity over the course of therapy (Table 1). The means and standard deviations at post-measure for the ISR (range 0-4) and DASS-21 (range 0-6) suggest

a possible floor effect. Assuming normal distributions, 8% of patients for the ISR and 11% for DASS-21 would be attenuated to a score of 0. All outcome measures were significantly correlated with each other at pre- and post-assessment (Table S5; S6). ISR and DASS-21 were strongly correlated at both pre- and post- assessment (between .7 and .8). Daily symptoms and self-reflective capacity were at pre-measure weakly correlated with each other and with the ISR and DASS-21 (absolute *r* values between .19 and .26) and at post-measure moderately correlated with each other and the ISR and DASS-21 (absolute *r* values between .33 and .48).

ISR

For the ISR (range 0-4), the multiple-shift group decreased an additional 0.34 points compared to the no-shift group (Table 3; Figure 2). The other groups did not significantly differ from the no-shift group. Re-analysis with the gradual-change group as reference yielded no significant differences between the gradual-change group and the other groups in how strong they decreased on the ISR.

DASS-21

For the DASS-21 (range 0-6), both the multiple-shift group and the gradual-change group decreased with 0.54 points more than the no-shift group (Table 3; Figure 2). Re-analysis with the gradual-change group as reference showed that the gradual-change group also decreased more than the one-shift group (*Estimate* = .46, *95% CI* = [0.13, 0.78]).

Daily symptom severity

For daily symptom severity (range 0-100), the multiple-shift group (15.48 points), the one-shift group (8.47 points), and the gradual-change group (10.34 points) had a significantly stronger symptom reduction compared to the no-shift group (Table 4; Figure 2). The multiple-shifts group started with significantly higher symptom severity (11.04 points) than the no-shift group. Re-analysis with the gradual-change group as reference showed a greater reduction in daily symptoms for the gradual-change group compared to the reversed-shift group (*Estimate* = 9.39, *95% CI* = [1.55, 17.24]), but no significant differences with the one-shift or multiple-shift group. Also, the multiple-shift group had significantly higher daily symptom severity at pre-measure than the gradual-change group (*Estimate* = 13.58, *95% CI* = [2.41, 24.76]).

Self-reflective capacity

For self-reflective capacity (range 0-100), the multiple-shift group (18.28 points), the one-shift group (7.18 points) and the gradual-change group (10.84 points) increased significantly more in self-reflective capacity compared to the no-shift group (Table 4; Figure 2). The multiple-shifts group started with a significantly lower self-reflective capacity than the no-shift group. Re-analysis with the gradual-change group as reference showed that the multiple-shift group increased more in self-reflective capacity than the gradual-change group (*Estimate* = 7.44, *95% CI* = [1.08, 13.81]) and that the gradual-change group increased more than the reversed-shift group (*Estimate* = -9.13, *95% CI* = [-14.60, -3.66]). Also, the multiple-shift group had significantly lower self-reflective capacity at pre-measure than the gradual-change group (*Estimate* = -10.71, *95% CI* = [-19.47, -1.94]).

Sensitivity analysis

The classifications with two other values for changeSensitivity had considerable overlap with the original classification (70% for the more sensitive value of 0.05 and 84% for the more conservative value of 0.15) and led to mostly similar patterns of results on the outcome analysis. Detailed results of the sensitivity analyses are provided in supplemental text 2 and tables S7-S12.

Table 3

Estimates and Confidence Intervals for the Relationship between Change Profiles and Pre- to Post Test Change on ISR & DASS-21

	ISR		DASS-21	
Effect	Estimate	95% CI	Estimate	95% CI
Intercept	1.46	[1.32, 1.60]	2.65	[2.41, 2.89]
Time	-0.81*	[-0.95, -0.67]	-1.47*	[-1.77, -1.18]
One Shift	0.00	[-0.16, 0.16]	-0.04	[-0.31, 0.23]
Multiple Shifts	0.25	[0.00, 0.51]	0.18	[-0.25, 0.61]
Reverse	0.05	[-0.17, 0.27]	0.07	[-0.31, 0.44]
Gradual	0.06	[-0.14, 0.27]	0.24	[-0.11, 0.58]
Treatment Duration	0.19*	[0.14, 0.24]	0.32*	[0.22, 0.41]

Time x One Shift	0.05	[-0.11, 0.20]	-0.08	[-0.41, 0.24]
Time x Multiple Shifts	-0.34*	[-0.58, -0.09]	-0.54*	[-1.04, -0.04]
Time x Reverse	-0.02	[-0.24, 0.20]	-0.19	[-0.65, 0.27]
Time x Gradual	-0.10	[-0.30, 0.10]	-0.54*	[-0.95, -0.13]
Time x Treatment Duration	-0.05	[-0.11, 0.00]	-0.12*	[-0.23, -0.02]

Note. Number of observations is 737 (ISR) and 739 (DASS-21). *statistically significant based on 95% CI.

Table 4

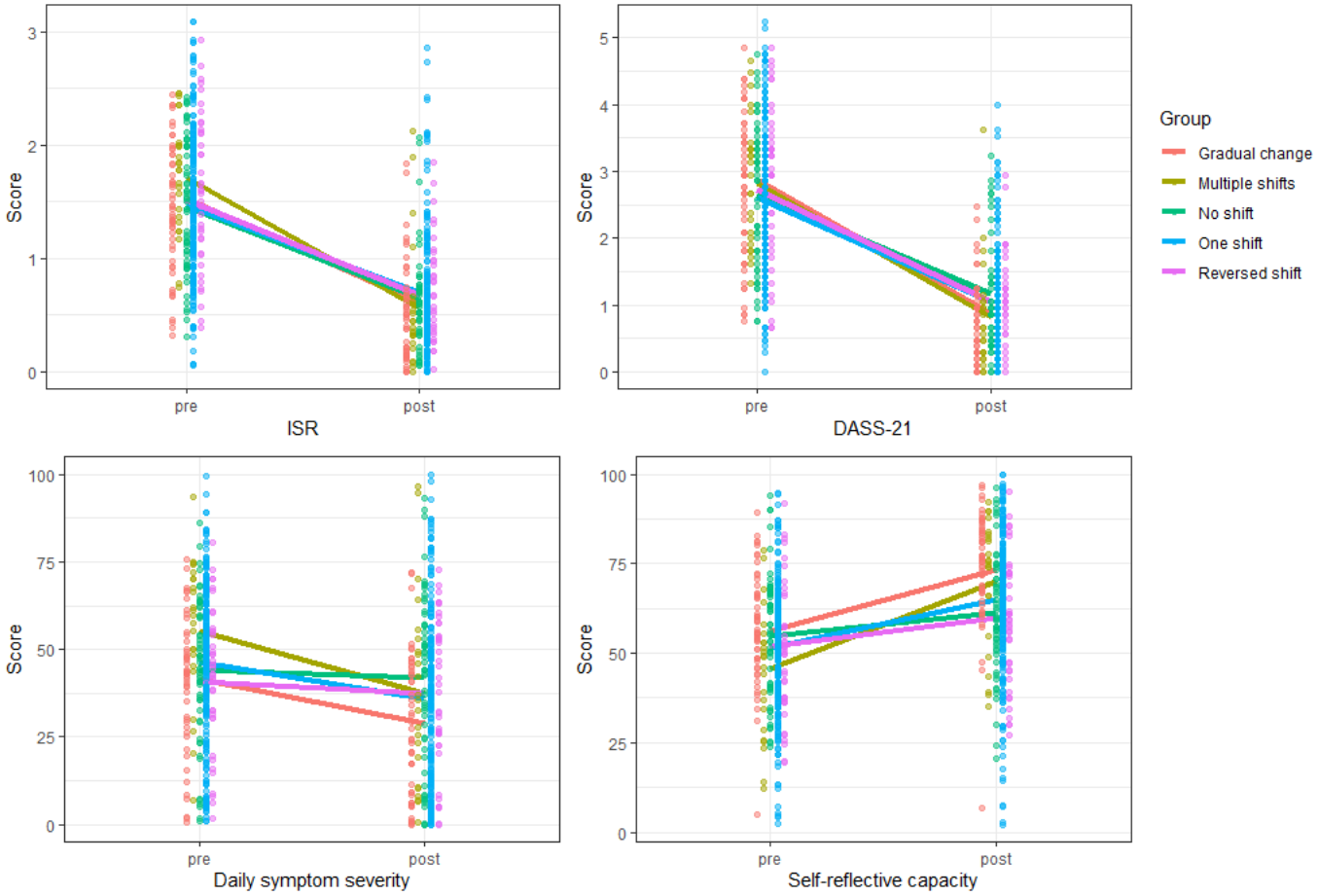
Estimates and Confidence Intervals for the Relationship between Change Profiles and Pre- to Post Test Change on Daily Symptom Severity and Self-reflective Capacity

Effect	Daily symptom severity		Self-reflective capacity	
	Estimate	95% CI	Estimate	95% CI
Intercept	44.08	[38.14, 50.01]	54.84	[50.11, 59.56]
Time	-2.24	[-7.07, 2.59]	6.33*	[2.90, 9.76]
One Shift	2.40	[-4.24, 9.04]	-3.55	[-8.82, 1.72]
Multiple Shifts	11.04*	[0.17, 21.91]	-9.32*	[-17.97, -0.67]
Reverse	-3.43	[-12.65, 5.80]	-3.18	[-10.53, 4.17]
Gradual	-2.54	[-11.29, 6.20]	1.39	[-5.40, 8.17]
Treatment Duration	3.49*	[1.19, 5.78]	-2.44*	[-4.29, -0.59]
Time x One Shift	-8.47*	[-13.87, -3.07]	7.18*	[3.35, 11.01]
Time x Multiple Shifts	-15.48*	[-24.32, -6.64]	18.28*	[12.00, 24.56]
Time x Reverse	-1.03	[-8.54, 6.47]	1.71	[-3.62, 7.05]
Time x Gradual	-10.43*	[-17.54, -3.32]	10.84*	[5.91, 15.77]
Time x Treatment Duration	-2.24*	[-4.11, -0.37]	0.02	[-1.32, 1.37]

Note. Number of observations is 758 (daily symptom severity) and 778 (self-reflective capacity), *statistically significant based on 95% CI.

Figure 2

Estimated Effects (lines) and Raw Scores (points) on the ISR, DASS-21, Daily Symptom Severity and Self-Reflective Capacity Pre- and Post- test Scores per Group



Discussion

We classified *general* change profiles in fully *personalized* self-ratings collected over the course of psychotherapy. We identified five different change profiles: no-shift, gradual-change, one-shift, reversed-shift, and multiple-shifts, which relate to different types of transitions in complex systems (e.g., Kéfi et al., 2013) and to previously identified change profiles in standardized process measures (including sudden gains, Shalom & Aderka, 2020). The highest loading items for the exemplar cases (Table 1, Figure 1) illustrates the strong individuality of each patient's PC1 time series, which is characteristic for the whole sample and logically follows from the massive heterogeneity in items. The classification results thereby corroborate the assumption of a complex systems approach to psychopathology that

general dynamic principles apply to individual change processes which are not specific to certain (psychological) constructs.

In line with our hypothesis that successful psychotherapy entails the transition from one attractor state to another over the course of therapy, the change profiles multiple-shift, one-shift and gradual-change were generally related to better treatment outcome compared to patients who did not transition. Patients with multiple-shifts had better treatment outcomes than patients with no-shifts on all four outcome measures. Patients with one-shift or gradual-change also had better treatment outcomes than the no-shift group on the process outcomes, but not on all standard outcomes. The fact that the one-shift group and the gradual-change group did not always improve more on the standard outcomes than the no-shift group may be related to the strong floor effect on the post measures. The reverse-shift group did not differ from the no-shift group on any outcome. These patients may have experienced a transition towards a different state, but later fell back to their initial state. Another explanation is that some of these patients did not actually experience a transition and that fluctuations within the same state were wrongly classified as a reversed shift profile.

For self-reflective capacity, which is an important treatment goal for the clinic where the data were collected, we found similar patterns of improvement as on daily symptom severity, although these two measures were only weakly to moderately correlated. Person-specific transitions may thus not only reduce symptom severity but also increase self-reflective capacity, sometimes even in the absence of symptom reduction. Overall, the relationship between the change profiles and outcome measures supports our hypothesis that patients who transitions from one person-specific attractor to another over the course of therapy have better treatment outcomes. Person-specific transitions may thus be a general mechanism of psychotherapeutic change (Kazdin, 2007).

In line with previous suggestions that gradual changes in psychotherapy may indicate a different change process than abrupt transitions (Gelo & Salvatore, 2016; Hayes et al., 2007), our results cautiously suggest that the gradual-change group indeed may be a special case, particularly when compared to the multiple-shift group. Formal complex system theories, such as synergetics (Haken, 1983) or catastrophe theory (Zeeman, 1976), state that for abrupt transitions there is a space ‘in between’ attractors that is *inaccessible*. Such inaccessibility is accompanied with a *hysteresis* effect: it requires a disproportionally large effort to change from one attractor to another (formally, the system state lags behind the parameter change; e.g., van Rooij et al., 2013). On the contrary, in gradual transitions this ‘direction of change’ is open: the system can freely move through different states without

hysteresis (formally, the system state changes proportionally to the parameter change; e.g., Figure 1 in Kéfi et al., 2013).

Following this reasoning, patients with a gradual change may have been 'less stuck' in their psychopathological attractor at the start of therapy compared to the multiple-shift group, which allowed their smooth transition towards a different state. This may also explain why the gradual-change group scored better on the pre-measures of certain outcomes (lower daily symptom severity, higher self-reflective capacity) than the multiple-shift group. The gradual-change group had the shortest average treatment duration while the multiple-shift group had the longest average treatment duration, which also suggests that the multiple-shift group may have had more severe problems to begin with (therapy duration in the clinic is in part related to severity). Another interesting difference is that the multiple-shift group increased more in self-reflective capacity than the gradual-change group, which suggests that experiencing multiple abrupt transitions from one attractor to another lead to more insights (and therefore more self-reflective capacity) than experiencing a gradual transition. This possible explanation is supported by findings in cognitive science that the emergence of insight itself may be understood as an abrupt transition (Stephen et al., 2009; Wiltshire et al., 2018). Although these possible explanations are worth considering, they are only indirectly supported by the current results. Future research could further examine underlying processes that may differentiate these change profiles.

Sensitivity analysis (supplemental text 2) showed that the results are relatively stable to changes in this parameter and suggests that our original setting was appropriate in not classifying too many shifts but still classifying enough shifts to identify the multiple-shift group, which was validated by the regression results and seems to be of high clinical relevance. The two alternative parameters for changeSensitivity that we examined (.05 and .15) are rather extreme, with only 5,2% of patients in the no-shift group for the sensitive setting and with the multiple-shift group being almost unidentifiable in the conservative setting. Given that, the overlap in classification of 70% (for the more sensitive setting) and 84% (for the more conservative setting) can be interpreted as signaling a certain robustness. Future research should further investigate the influence of parameter settings by applying recursive partitioning to simulated transitions.

With regards to the current special issue of the *Journal of Psychopathology and Clinical Science*, the present study highlights general principles from complex systems (e.g., attractors and transitions) as fine-grained clinical phenomena that can contribute to our understanding of transdiagnostic and personalized psychopathology. Notably, because

attractors and transitions are general features of complex systems, they should not be regarded as specifically *clinical* phenomena. Instead, they generally apply to human development (Thelen & Smith, 1994). In the case of psychological development, they have for instance been studied extensively in identity development (Lichtwarck-Aschoff et al., 2008) and self-esteem development (de Ruiter et al., 2017). The onset and remission of psychopathology may thus be explained through the same mechanisms as other changes in psychological development, which is in line with a developmental (in contrast to a disease) perspective on mental health (e.g., Lewis, 2018; Sameroff et al., 2000).

Our complex systems approach (for details see: Olthof et al., in press) has certain commonalities and differences with the frameworks discussed in the call for this special issue. Just like the hierarchical taxonomy of psychopathology (HiTOP), the research domain criteria (RDOC) framework and the network approach to psychopathology, the complex systems approach aims to move beyond current diagnostic categories and focus on transdiagnostic psychopathology. Also, just like these frameworks, it seeks to understand psychopathology in a 'bottom-up' manner by invoking general principles. The main differences in comparison to HiTOP, RDOC and the network approach is that our complex systems approach assumes strict interdependence, and therefore inseparability of biopsychosocial processes in the person-environment system[2]. As these processes are ever evolving on multiple timescales, the contributions of specific processes are not stable and cannot be isolated. General mechanisms in the complex systems perspective are therefore not expected at the level of specific biological or psychological constructs, but at the level of dynamics, where general laws dictate pattern formation in all biological systems (Thelen & Smith, 1994).

Strengths, limitations and future suggestions

A main strength of this study is that we used data collected as part of routine clinical care from a large group of patients, speaking to the ecological validity of our results. This large dataset enabled us to study a fundamental hypothesis derived from theory, as well as gain insight in the content and dynamics of personalized questionnaires as applied in a clinical setting. The study also has limitations. First, the general applicability of PCA for calculating univariate time series with fully personalized meaning is besides a key strength also a key limitation of the present work as the valence and direction of change could not be taken into

[2] Notably, one could argue that some versions of the network theory of mental disorders share this assumption (e.g., Borsboom, et al., 2018), but the empirical approaches generally do not (Haslbeck et al., 2022; Olthof et al., 2020c).

account. Hence, we do not know whether all patients who transitioned towards a different state also transitioned towards a more healthy state. It is possible that in some scarce cases patients deteriorated during psychotherapy, in which case we would not expect the transition to be related to better treatment outcome. Further exploration (for instance of the PCA loadings) can help to determine valence and direction of individual change processes and is a promising direction for applied clinical research. Second, in determining the change profiles, we limited the identification of stable states (i.e., attractors) to stable mean levels. In this way, we were likely to identify point attractors, but may have missed other forms of stability such as stable cycles (periodic attractors), or more complex attractor dynamics (strange attractors; e.g., Schöller et al., 2018). Thirds, based on the complex systems literature (e.g., Lever et al., 2020) we searched for transitions only in the PC1, while there are circumstances imaginable in which patients would show transitions on the second principal component (PC2). This could happen for patients where a subset of items has high variance, but no transition (e.g., a cyclical dynamic with high amplitude), which then would constitute the PC1, whereas another subset of items perhaps would show a transition that then ends up in the PC2. Notably, visual inspection of the PC1 and PC2 time series suggests that this situation is extremely rare in the current sample. Still, it is important to realize that dimension reduction as employed in this study may have many advantages in selecting relevant information, but could in principle also miss relevant information in some cases.

Future research could further develop the use of per-person PCA, by for instance also examining the PC2 and PC3 and examining the item loadings to interpret what happens on these different dimensions. Also, dynamic factor analysis could be examined in the present context to derive dimensions that can be interpreted as personalized latent variables which can then be studied to identify transitions (Fisher, 2015). Multidimensional recurrence quantification analysis could be used to study transitions in personalized self-ratings without any dimension reduction, thereby leading to even more fine-grained insights at the person level (Hasselman, 2022b). Last, moving window PCA may be used to study changes over time in the PC1; increases over time in the explained variance of the PC1 can be early-warning signals for upcoming transitions, for which the item loadings and distribution of the data could reveal the direction of change (Lever et al., 2020; Schreuder et al., 2022).

Implications

The central implication of the current work is that person-specific (idiographic) and general (nomothetic) aspects of psychopathology can be integrated in a complex systems

approach to psychopathology. The general aspects of psychopathology then lie in the principles of dynamic patterns, while the content of these dynamic patterns is fully person specific. This study also has more specific implications for personalized psychopathology research. First, our findings show that personalized questionnaires to study clinical change can be feasible and insightful, which until now had only been shown in single-case studies (e.g., Schiepek et al., 2016b). Second, the present results illustrate that non-stationarity is the rule rather than the exception in self-rating time series, which is a limitation for many commonly used statistical models in personalized psychopathology research (Wright & Woods, 2020). Lastly, since a considerable number of patients had complex patterns of shifts (reverse shifts and multiple shifts), the findings show that clinical change can be more complex than a singular transition from a pathological attractor to a healthy attractor (i.e., a bi-stable dynamical system; cf., Wichers et al., 2015).

Our results support the use of personalized questionnaires in clinical practice as they seem to be able to capture person-specific transitions that are related treatment outcome. A possible advantage of monitoring person-specific transitions compared to symptom transitions is that they give more precise information about what has changed for a specific person and also capture relevant changes that are not related to symptoms. Moreover, the concept of person-specific transitions can be used to integrate case formulation with process monitoring, thereby stimulating collaboration between researchers, practitioners and patients (Kramer, 2020). For instance, Schiepek et al., (2016) propose an integrative approach to psychotherapy in which patient and therapist together generate a personalized case formulation, which then is translated into a personalized questionnaire for daily process monitoring. Such data can then be used for data-feedback sessions of patients and therapists, in which for instance possible personalized transitions may be discussed.

Conclusion

This study found that fully personalized self-rating time series collected during psychotherapy allow for the classification of general change profiles that can be interpreted from a complex systems perspective. In line with the hypothesis that person-specific transitions are a general mechanism of change, we found that patients who transitioned towards a different state over the course of treatment had in most cases better treatment outcome than patients who did not. Overall, this study illustrates that person-specific (idiographic) and general (nomothetic) aspects of psychopathology can be integrated in a complex systems approach to psychopathology, which may combine ‘the best of both worlds’

Funding

Merlijn Olthof and Anna Lichtwarck-Aschoff were supported by a NWO VIDI grant, Grant No. VI.Vidi.191.178.

Disclosure of interest

Günter Schiepek is director of the Center for Complex Systems, publisher and developer of the SNS monitoring system that was used for data collection.

Acknowledgements

We thank all past and current employees of the sysTelios Health Center team for the collection of data, with a special thanks to Florian Pommerien-Becht, Konstantin Hegewald, Patricia Gill-Schultz, Mechthild Reinhard and Gunther Schmidt. We thank David Leeftink and Jens Barenbrügge for their advice on data-analysis. We are grateful to the patients who gave consent for scientific use of their data.

Author note

Correspondence should be sent to Merlijn Olthof Merlijn.olthof@ru.nl

Results in this manuscript were presented at various conferences (VNOP research days, VGCt, summerschool Human Change Processes) and included in the doctoral dissertation of Merlijn Olthof. Data were collected as part of routine clinical practice. All participants gave informed consent for research use of their data. Raw data from this study are not publicly available to protect the privacy of patients. Preprocessed (unidentifiable) data can be obtained upon reasonable request. This study was not preregistered. Open materials, including model outputs and scripts, are available at the open science framework [https://osf.io/ksym9/; DOI 10.17605/OSF.IO/KSYM9].

Reference list

Allsopp, K., Read, J., Corcoran, R., & Kinderman, P. (2019). Heterogeneity in psychiatric diagnostic classification. *Psychiatry research*, *279*, 15-22.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beitman, B. D. & Soth, A. M. (2006). Activation of self-observation: A core process among the psychotherapies. *Journal of Psychotherapy Integration*, *16*(4), 383– 397. https://doi.org/10.1037/1053-0479.16.4.383

Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. Assessment, 23(4), 447-458.

Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, *42*, e2.

Bos, F. M., Schreuder, M. J., George, S. V., Doornbos, B., Bruggeman, R., van der Krieke, L., ... & Snippe, E. (2022). Anticipating manic and depressive transitions in patients with bipolar disorder using early warning signals. *International Journal of Bipolar Disorders*, *10*(1), 12.

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) *Classification and Regression Trees.* Wadsworth.

De Ruiter, N. M., Van Geert, P. L., & Kunnen, E. S. (2017). Explaining the "how" of self-esteem development: The self-organizing self-esteem model. *Review of General Psychology*, *21*(1), 49-68.

Fischer, H. F., Tritt, K., Klapp, B. F., & Fliege, H. (2010). Faktorstruktur und psychometrische Eigenschaften des ICD-10 Symptom-Rating (ISR) an Stichproben psychosomatischer Patienten [Factor structure and psychometric properties in the ICD-10 Symptom-Rating (ISR) in samples of psychosomatic patients]. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie*, *60*(8), 307–315. https://doi.org/10.1055/s-0029-1214419

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*(4), 825–836. https://doi.org/10.1037/ccp0000026

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*(February), 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Gelo, O. C. G., & Salvatore, S. (2016). A dynamic systems approach to psychotherapy: A meta-theoretical framework for explaining psychotherapy change processes. *Journal of counseling psychology*, *63*(4), 379.

Haken, H. (1983). *Synergetics: an introduction. Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology.* Springer Verlag.

Haslbeck, J., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930.

Hasselman, F. (2022a). casnet: A toolbox for studying Complex Adaptive Systems and NETworks (0.2.2) URL https://github.com/FredHasselman/casnet

Hasselman, F. (2022b). Early warning signals in phase space: geometric resilience loss indicators from multiplex cumulative recurrence networks. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2022.859127

Hayes, A. M., & Andrews, L. A. (2020). A complex systems approach to the study of change in psychotherapy. *BMC Medicine*, *18*(1), 1–13. https://doi.org/10.1186/s12916-020-01662-2

Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, *27*, 715–723. https://doi.org/10.1016/j.cpr.2007.01.008.Change

Held, H., & Kleinen, T. (2004). Detection of climate system bifurcations by degenerate fingerprinting. *Geophysical Research Letters*, *31*(23), 1–4. https://doi.org/10.1029/2004GL020972

Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, *88*(2), 119–127. https://doi.org/10.1037/ccp0000469

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239. https://doi.org/10.1348/014466505X29657

Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/ 10705519909540118

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical psychology review*, *30*(7), 865-878.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*(1), 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kéfi, S., Dakos, V., Scheffer, M., Van Nes, E. H., & Rietkerk, M. (2013). Early warning signals also precede non-catastrophic transitions. *Oikos*, *122*(5), 641–648. https://doi.org/10.1111/j.1600-0706.2012.20838.x

Kramer, U. (2020). Individualizing psychotherapy research designs. *Journal of Psychotherapy Integration*, *30*(3), 440.

Lever, J. J., van de Leemput, I. A., Weinans, E., Quax, R., Dakos, V., van Nes, E. H., Bascompte, J., & Scheffer, M. (2020). Foreseeing the future of mutualistic communities beyond collapse. *Ecology Letters*, *23*(1), 2–15. https://doi.org/10.1111/ele.13401

Lewis, M. (2018). Brain change in addiction as learning, not disease. *New England Journal of Medicine*, *379*(16), 1551-1560.

Lichtwarck-Aschoff, A., van Geert, P., Bosma, H., & Kunnen, S. (2008). Time and identity: A framework for research and theory formation. *Developmental Review*, *28*(3), 370-400.

Lloyd, C. E. M., Duncan, C., & Cooper, M. (2019). Goal measures for psychotherapy: A systematic review of self-report, idiographic instruments. *Clinical Psychology: Science and Practice*, *26*(3), 1–16. https://doi.org/10.1111/cpsp.12281

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

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *R Journal*, *9*(1), 207–218. https://doi.org/10.32614/rj-2017-009

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M. T. & Lichtwarck-Aschoff, A. (*in press*). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science*. https://doi.org/10.31234/osf.io/f68ej

Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020c). Complexity in psychological self-ratings: Implications for research and practice. *BMC medicine*, *18*(1), 1-16.

Olthof, M., Hasselman, F., Strunk, G., Aas, B., Schiepek, G., & Lichtwarck-Aschoff, A. (2020b). Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders. *Psychotherapy Research*, *30*(4), 520–531. https://doi.org/10.1080/10503307.2019.1633484

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020a). Critical fluctuations as an early-warning signal for sudden Gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, *8*(1), 25–35. https://doi.org/10.1177/2167702619865969

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of statistical software*, *48*, 1-36.

Sameroff, A. J. (2000). Developmental systems and psychopathology. *Development and psychopathology*, *12*(3), 297-312.

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

Schiepek, G., Aichhorn, W., & Strunk, G. (2012). Therapy Process Questionnaire (TPQ)-a combined explorative and confirmatory factor analysis and psychometric properties. *Zeitschrift fur Psychosomatische Medizin und Psychotherapie*, *58*(3), 257-266.

Schiepek, G, Eckert, H., Aas, B., Wallot, S., & Wallot, A. (2016a). *Integrative psychotherapy: A feedback-driven dynamic systems approach*. Hogrefe.

Schiepek, G., Gelo, O., Viol, K., Kratzer, L., Orsucci, F., de Felice, G., Stöger-Schmidinger, B., Sammet, I., Aichhorn, W., & Schöller, H. (2020). Complex individual pathways or standard tracks? A data-based discussion on the trajectories of change in psychotherapy. *Counselling and Psychotherapy Research*, *20*(4), 689–702. https://doi.org/10.1002/capr.12300

Schiepek, G., Stöger-Schmidinger, B., Kronberger, H., Aichhorn, W., Kratzer, L., Heinz, P., Viol, K., Lichtwarck-Aschoff, A., & Schöller, H. (2019). The Therapy Process Questionnaire - Factor analysis and psychometric properties of a multidimensional self-rating scale for high-frequency monitoring of psychotherapeutic processes. *Clinical Psychology & Psychotherapy*, *26*(5), 586–602. https://doi.org/10.1002/cpp.2384

Schiepek, G., Stöger-Schmidinger, B., Aichhorn, W., Schöller, H., & Aas, B. (2016b). Systemic case formulation, individualized process monitoring, and state dynamics in a case of dissociative identity disorder. *Frontiers in Psychology*, *7*(10), 1–11. https://doi.org/10.3389/fpsyg.2016.01545

Schöller, H., Viol, K., Aichhorn, W., Hütt, M. T., & Schiepek, G. (2018). Personality development in psychotherapy: a synergetic model of state-trait dynamics. *Cognitive Neurodynamics*, *12*(5), 441-459. https://doi.org/10.1007/s11571-018-9488-y

Schreuder, M. J., Wigman, J. T. W., Groen, R. N., Weinans, E., Wichers, M., & Hartman, C. A. (2022). Anticipating the direction of symptom progression using critical slowing down: a proof-of-concept study. *BMC Psychiatry*, *22*(1), 1–13. https://doi.org/10.1186/s12888-022-03686-9

Shalom, J. G., & Aderka, I. M. (2020). A meta-analysis of sudden gains in psychotherapy: Outcome and moderators. *Clinical Psychology Review*, 76, 101827. https://doi.org/10.1016/j.cpr.2020.101827

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1811–1832. https://doi.org/10.1037/a0014510

Takens, F. (1981). Detecting strange attractors in turbulence. In. D. Rand & L.S. Young (Eds.), *Dynamical systems and turbulence, Warwick 1980. Lecture Notes in Mathematics* (pp. 366-381). Springer. https://doi.org/10.1007/BFb00919

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT press.

Therneau, T.M., & Atkinson, E.J. (2022). *An Introduction to Recursive Partitioning Using the RPART Routines*. Retrieved from https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Tritt, K., Heymann, F. Von, Zaudig, M., Lynch, S., Probst, T., Söllner, W., Loew, T., & Bühner, M. (2013). *The ICD - 10 - Symptom - Rating (ISR) Brief Description.* www.iqp-online.de/index.php?page=download

Van den Bergh, R., Olthof, M., Goldbeck, F., Hegewald, K., Pommerien-Becht, F., Daniels-Wredenhagen, N., Weggemans, R. J., Daalmans, S., Otten, R., Aas, B. G., & Lichtwarck-Aschoff, A. (2022). *The content of personalised network-based case formulations*. PsyArxiv. 10.31234/osf.io/yan4k.

Van Geert, P. L. C. (2019). Dynamic systems, process and development. *Human Development*, *63*(3–4), 153–179. https://doi.org/10.1159/000503825

van Os, J., Verhagen, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., Drukker, M., Reininghaus, U., Jacobs, N., Lataster, T., Simons, C., Lousberg, R., Gülöksüz, S., Leue, C., Groot, P. C., Viechtbauer, W., & Delespaul, P. (2017). The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized

health care in clinical practice. *Depression and Anxiety*, *34*(6), 481–493. https://doi.org/10.1002/da.22647

van Rooij, M. M., Favela, L. H., Malone, M., & Richardson, M. J. (2013). Modeling the dynamics of risky choice. *Ecological Psychology*, *25*(3), 293-303.

Wichers, M., Wigman, J. T. W., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, *7*(4), 363–367. https://doi.org/10.1177/1754073915590623

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive science*, *42*(1), 129-167.

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, *16*, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, *234*(4), 65-83.

Supplemental materials for 'The best of both worlds? *General* principles of psychopathology in *personalized* assessment'

Supplemental text 1: Results of the Factor Analysis and Reliability Analysis of the Self-Reflective Capacity Scale.

We performed confirmatory factor analysis on $N = 374$ subjects, who had ratings on all four items assumed to belong to the self-reflective capacity scale (see Table S1). Factor analysis was computed using R package lavaan (Rosseel, 2012) with MLM estimation and NLMINB optimization method. For both pre- and post-measure, we tested a congeneric model with one latent factor against a congeneric model with two correlated latent factors (see figure S1). The fit indexes of the one factor model show a high goodness of fit for the pre-measure ($X^2(2) = 3.520$, $p = .17$; $RMSEA = .05$; $CFI = 1.00$; $SRMR = .01$; $AIC = 12016.04$) and a moderate goodness of fit for the post-measure ($X^2(2) = 14.877$, $p = .001$; $RMSEA = .131$; $CFI = .99$; $SRMR = .01$; $AIC = 11887.69$). For a discussion of the fit indexes, see Hu and Bentler (1999). The two factor models for pre- and post-measure contained an invalid parameter estimation, namely a correlation of the latent factors higher than $r = 1.00$, and were therefore scrapped. Further analysis concerning self-reflective capacity use the total mean of all four items of the self-reflective capacity scale.

Table S1.

Descriptive statistics of self-reflective capacity for pre- and post-measure.

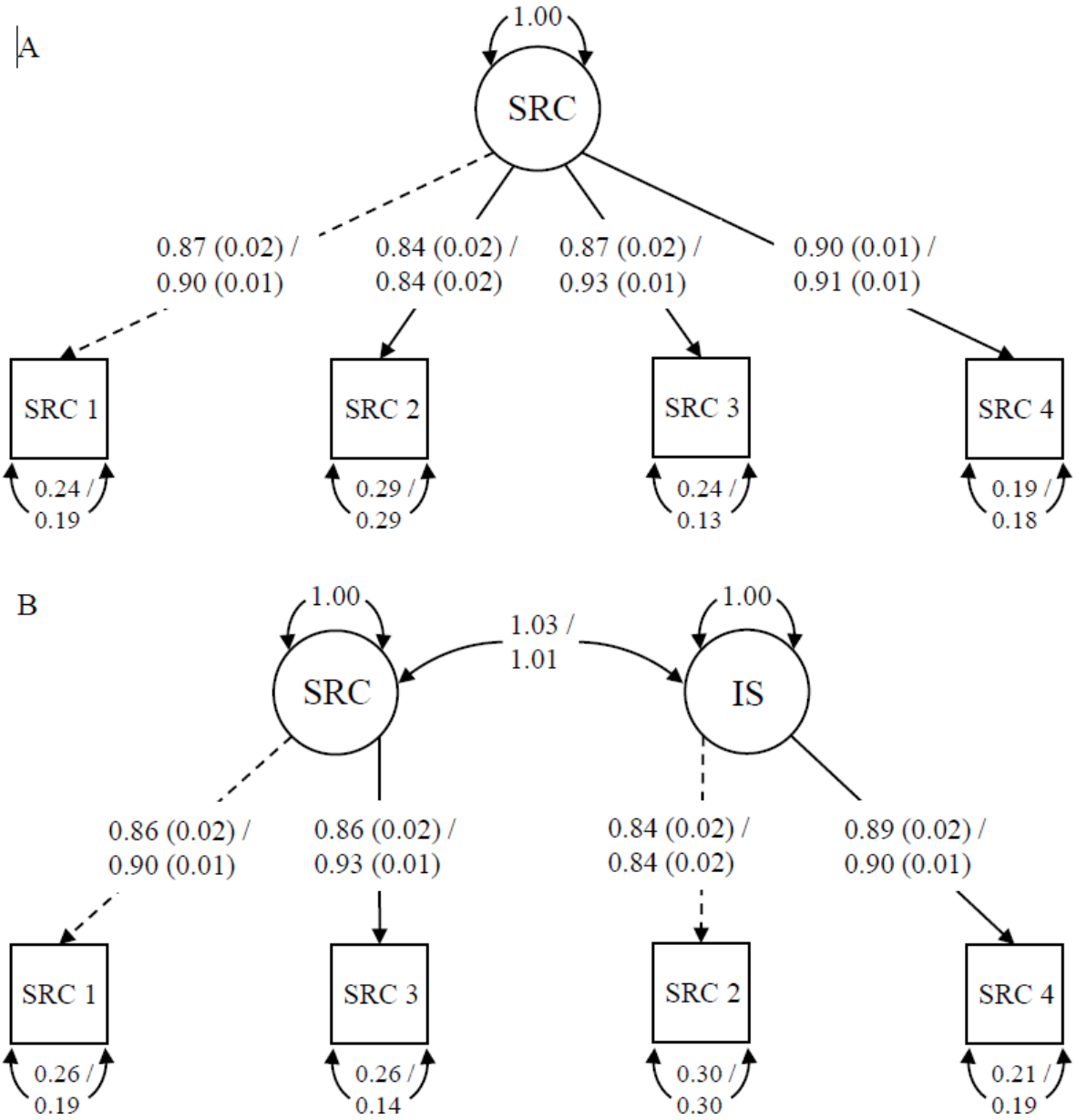
Item	M	SD	Skewness	Kurtosis
SRC_1	58.58 / 68.94	20.32 / 19.22	-0.38 / -0.93	-0.12 / 0.87
SRC_2	51.26 / 64.89	19.95 / 20.22	-0.20 / -0.75	-0.28 / 0.46
SRC_3	47.85 / 61.19	18.46 / 20.64	0.04 / -0.06	-0.27 / -0.06
SRC_4	52.58 / 65.22	19.65 / 21.05	-0.19 / -0.74	-0.25 / 0.13

Note. SRC_1 = "Today, I was able to adopt an observer's perspective on my inner experience.", SRC_2 = "Today, my feelings make sense to me. (I was able to attribute them to events.)", SRC_3 = "Today, I have experienced choices in dealing with my problems.",

SRC_4 = “Today, I was able to look benevolently at myself and my process.”; Scores before slash (“/”) refer to pre-measure. Scores after slash (“/”) refer to post-measure; *N* = 374.

Figure S1.

Tested one-factor and two-factor models of the self-reflective capacity subscale for pre- and post-measure.



Note. A: One-factor model with self-reflective capacity as a single latent factor; B: Two-factor model with self-reflective capacity and insight (Schiepek et al., 2012) as latent factors. The note of table S7 contains the phrases of the four items of the self-reflective capacity subscale used in both models. Scores before slash (“/”) refer to standardized factor loadings at pre-measure with standard errors in brackets. In case of the items, scores before slash (“/”)

refer to standard errors at pre-measure. Scores after slash ("/") refer to the respective values at post-measure; $N = 374$.

Table S2.

Number of patients per change profile without the gradual change group

No shift	One shift	Multiple shift	Reversed shift
56	279	30	39

Table S3.

Best fitting model per change profile without the gradual change group

	No shift	One shift	Multiple shift	Reversed shift	Total (%)
N linear model fits better	26	47	5	0	78 (19%)
N step model fits better	30	232	25	39	326 (81%)

Note. Mean (*SD*) AIC of linear model: 147.73 (63.4); and step model: 146.01 (64.82). Absolute *Mean* (*SD*) beta weights of linear model: 1.05 (0.73); and step model: 2.24 (1.59).

.

Table S4.

Best fitting model per change profile including the gradual change group

	Gradual change	No shift	One shift	Multiple shift	Reversed shift
N linear model fits better	52	25	1	0	0
N step model fits better	0	30	232	25	39

Table S5.

Correlation between outcome measures at pre-score

	ISR	DASS-21	Self-reflective capacity	Daily symptom severity
ISR	1	0.73*	-0.26*	0.24*
DASS-21	0.73*	1	-0.19*	0.23*
Self-reflective capacity	-0.26*	-0.19*	1	-0.24*
Daily symptom severity	0.24*	0.23*	-0.24*	1

Note. * = p <.001

Table S6.

Correlation between outcome measures at post-score

	ISR	DASS-21	Self-reflective capacity	Daily symptom severity
ISR	1	0.79*	-0.45*	0.33*
DASS-21	0.79*	1	-0.48*	0.43*
Self-reflective capacity	-0.45*	-0.48*	1	-0.41*
Daily symptom severity	0.33*	0.43*	-0.41*	1

Note. * = p <.001

Supplemental text 2: Sensitivity analysis

The original classification was performed with a changeSensitivity parameter of .10 for detecting shifts in the *ts_levels()* function. The classification with a more sensitive parameter setting, .05 (which is more sensitive to shifts), had an overlap of 70%. The main difference lies in the classification of patients from the no-shift group who were often classified as having shifts with the more sensitive setting, as could be expected (for the crosstab with the original setting see table S7).

The classification with a more conservative changeSensitivity parameter of .15 (which is less sensitive to shifts), had an overlap of 84%. Here, the main differences lie in the

classification of patients who were in the multiple-shift group in the original classification, which were now classified in the gradual-change group (for the crosstab with the original setting see table S8). This is also not surprising: as the more conservative setting barely defined multiple shifts in a time series, patients who originally were in the multiple-shift group now often only had one shift as a result from recursive portioning. This one shift solution was less adequate than the multiple shift solution in the original parameter setting, because now the linear regression model very often outperformed the step regression, leading to an inflated gradual-change group.

In relation to outcome, the classifications with changeSensitivity of .05 showed no significant differences between no-shift and the other groups on the DASS-21 and the ISR, while they did show better outcome for the multiple-shift group on daily symptom severity and for the one-shift, multiple-shift and gradual-change group on self-reflective capacity (see tables S9, S10). The classifications with changeSensitivity of .15 resulted in only 6 patients with multiple shifts. These were therefore included in the one-shift group for the outcome analysis. Compared to the no-shift group, the gradual-change group had better treatment outcome on all outcome measures and the one-shift group had better outcome on daily symptoms and self-reflective capacity (see tables S11, S12). Note that the gradual-change group in this classification included many patients whose time series were actually better described with the multiple-shift solution in the original parameter setting.

Table S7.

Crosstabulation of groups with changeSenstivity = 0.10 and changeSensitivity = 0.05

.10 / .05	Gradual-change	Multiple-shift	No-shift	One-shift	Reversed-shift	Total
Gradual-change	37	0	0	0	0	37
Multiple-shift	15	22	0	35	0	72
No-shift	0	0	21	0	0	21
One-shift	0	3	20	167	5	195

Reversed-shift	0	0	14	31	34	79
Total	52	25	55	233	39	404

Note. Percentage of agreement is 70%.

Table S8.

Crosstabulation of groups with changeSenstivity = 0.10 and changeSensitivity = 0.15

.10 / .15	Gradual-change	Multiple-shift	No-shift	One-shift	Reversed-shift	Total
Gradual-change	52	12	0	0	0	64
Multiple-shift	0	6	0	0	0	6
No-shift	0	0	55	27	6	88
One-shift	0	7	0	205	10	222
Reversed-shift	0	0	0	1	23	24
Total	52	25	55	233	39	404

Note. Percentage of agreement is 84%.

Table S9.

Estimates and Confidence Intervals for the Relationship between Change Profiles defined with changeSensitivity = 0.05 and Pre- to Post Test Change on ISR & DASS-21

Effect	ISR		DASS-21	
	Estimate	95% CI	Estimate	95% CI
Intercept	1.44*	[1.21, 1.67]	2.71*	[2.32, 3.10]
Time	-0.83*	[-1.05, -0.61]	-1.66*	[-2.12, -1.20]
One Shift	0.01	[-0.23, 0.26]	-0.07	[-0.48, 0.34]
Multiple Shifts	0.17	[-0.09, 0.43]	0.04	[-0.40, 0.48]

Reverse	0.07	[-0.19, 0.33]	-0.12	[-0.56, 0.32]
Gradual	0.04	[-0.25, 0.33]	0.12	[-0.37, 0.61]
Treatment Duration	0.18*	[0.13, 0.24]	0.32*	[0.23, 0.41]
Time x One Shift	0.05	[-0.18, 0.29]	0.08	[-0.41, 0.56]
Time x Multiple Shifts	-0.17	[-0.42, 0.09]	-0.25	[-0.77, 0.26]
Time x Reverse	0.10	[-0.15, 0.35]	0.32	[-0.20, 0.84]
Time x Gradual	-0.09	[-0.37, 0.19]	-0.35	[-0.92, 0.23]
Time x Treatment Duration	-0.06	[-0.11, 0.00]	-0.14*	[-0.25, -0.03]

Note. Number of observations is 737 (ISR) and 739 (DASS-21). *statistically significant based on 95% CI.

Table S10.

Estimates and Confidence Intervals for the Relationship between Change Profiles defined with changeSensitivity = 0.05 and Pre- to Post Test Change on Daily Symptom Severity and Self-reflective Capacity

	Daily symptom severity		Self-reflective capacity	
Effect	Estimate	95% CI	Estimate	95% CI
Intercept	48.78*	[39.28, 58.29]	59.76*	[52.25, 67.27]
Time	-3.25	[-10.95, 4.44]	5.74*	[0.38, 11.09]
One Shift	-3.41	[-13.45, 6.63]	-7.73	[-15.64, 0.19]
Multiple Shifts	-0.90	[-11.48, 10.05]	-10.83*	[-19.44, -2.22]
Reverse	-3.41	[-14.18, 7.36]	-9.14*	[-17.66, -0.63]
Gradual	-9.55	[-21.70, 2.59]	-2.32	[-11.77, 7.13]
Treatment Duration	3.13*	[0.80, 5.46]	-2.16*	[-4.02, -0.30]
Time x One Shift	-5.83	[-13.96, 2.29]	6.27*	[0.62, 11.91]
Time x Multiple Shifts	-15.29*	[-24.15, -6.43]	17.15*	[11.01, 23.29]
Time x Reverse	-1.21	[-9.93, 7.51]	1.68	[-4.39, 7.76]
Time x Gradual	-5.17	[-15.00, 4.66]	10.82*	[4.08, 17.56]
Time x Treatment Duration	-1.73	[-3.62, 0.16]	-0.23	[-1.55, 1.10]

Note. Number of observations is 758 (daily symptom severity) and 778 (self-reflective capacity), *statistically significant based on 95% CI.

Table S11.

Estimates and Confidence Intervals for the Relationship between Change Profiles defined with changeSensitivity = 0.15 and Pre- to Post Test Change on ISR & DASS-21

Effect	ISR		DASS-21	
	Estimate	95% CI	Estimate	95% CI
Intercept	1.50*	[1.38, 1.61]	2.69*	[2.49, 2.88]
Time	-0.79*	[-0.90, -0.68]	-1.58*	[-1.81, -1.35]
One Shift	-0.03	[-0.16, 0.10]	-0.05	[-0.28, 0.17]
Reverse	0.02	[-0.22, 0.27]	-0.12	[-0.54, 0.29]
Gradual	0.06	[-0.11, 0.24]	0.17	[-0.13, 0.47]
Treatment Duration	0.19*	[0.14, 0.25]	0.32*	[0.23, 0.41]
Time x One shift	-0.02	[-0.15, 0.11]	-0.03	[-0.30, 0.24]
Time x Reverse	0.05	[-0.20, 0.30]	0.36	[-0.16, 0.88]
Time x Gradual	-0.18*	[-0.35, -0.01]	-0.44*	[-0.79, -0.09]
Time x Duration	-0.06*	[-0.12, -0.01]	-0.15*	[-0.26, -0.04]

Note. Number of observations is 737 (ISR) and 739 (DASS-21). Patients in the multiple-shift group (N=6) were included in the one-shift group as the group size was too small.
*statistically significant based on 95% CI.

Table S12.

Estimates and Confidence Intervals for the Relationship between Change Profiles defined with changeSensitivity = 0.15 and Pre- to Post Test Change on Daily Symptom Severity and Self-reflective Capacity

Effect	Daily symptom severity		Self-reflective capacity	
	Estimate	95% CI	Estimate	95% CI
Intercept	44.73*	[40.01, 49.46]	52.81	[49.05, 56.56]
Time	-4.16	[-8.02, -0.31]	7.39	[4.65, 10.13]
One Shift	1.13	[-4.47, 6.74]	-1.45	[-5.88, 2.99]
Reverse	2.32	[-7.88, 12.52]	-0.86	[-8.96, 7.24]
Gradual	-0.44	[-7.95, 7.07]	1.14	[-4.66, 6.94]
Treatment Duration	3.43*	[1.31, 5.74]	-2.64	[-4.49, -0.80]
Time x One shift	-6.50*	[-11.08, -1.93]	6.93	[3.69, 10.17]

Time x Reverse	-0.72	[-9.05, 7.61]	-0.76	[-6.67, 5.16]
Time x Gradual	-10.39*	[-16.53, -4.26]	11.86	[7.62, 16.09]
Time x Duration	-2.25*	[-4.13, -0.37]	0.23	[-1.12, 1.58]

Note. Number of observations is 758 (daily symptom severity) and 778 (self-reflective capacity). Patients in the multiple-shift group (N=6) were included in the one-shift group as the group size was too small. *statistically significant based on 95% CI.