k-Centres Functional Clustering

Hofmans, Joeri; Vantilborgh, Tim; Solinger, Omar N.

published in
Organizational Research Methods
2018

DOI (link to publisher)
10.1177/1094428117725793

document version
Publisher's PDF, also known as Version of record

Article

Organizational Research Methods
2018, Vol. 21(4) 905-930

DOI: 10.1177/1094428117725793
journals.sagepub.com/home/orm
SAGE

k-Centres Functional Clustering: A Person-Centered Approach to Modeling Complex Nonlinear Growth Trajectories

Joeri Hofmans[1], Tim Vantilborgh[1] and Omar N. Solinger[2]

Abstract

In the present paper, we introduce *k*-centres functional clustering (*k*-centres FC), a person-centered method that clusters people with similar patterns of complex, highly nonlinear change over time. We review fundamentals of the methodology and argue how it addresses some of the limitations of the traditional approaches to modeling repeated measures data. The usefulness of *k*-centres FC is demonstrated by applying the method to weekly measured commitment data from 109 participants who reported psychological contract breach events. The *k*-centres FC analysis shows two substantively meaningful clusters, the first cluster showing reaction patterns with general growth in commitment after breach and the second cluster showing general decline in commitment after breach. Further, the reaction patterns in the second cluster appear to be the result of a combination of two interesting reaction logics: immediate and delayed reactions. We conclude by outlining how future organizational research can incorporate this methodology.

Keywords

functional data analysis, functional principal component analysis, functional clustering, psychological contract breach, organizational commitment

During recent years, organizational researchers have become increasingly aware that the implicit assumptions of cross-sectional research—that organizational phenomena are stable across time and that fluctuations around the average levels of these phenomena are randomly distributed across occasions—are flawed (Ilies & Judge, 2002). This awareness, together with the emergence of new technologies such as handheld devices and smartphones, has increased the interest and ability of

[1]Vrije Universiteit Brussel, Brussels, Belgium
[2]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Corresponding Author:
Joeri Hofmans, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.
Email: joeri.hofmans@vub.be

organizational scholars to model change in organizational phenomena (Dass & Shropshire, 2012; Mehl & Conner, 2012).

Despite these technological advancements, the statistical models typically applied to such data (e.g., latent growth models) fail to capture complex patterns of change as these methods were developed to model change with only a few waves of measurement (Fok & Ramsay, 2006). With more intensive longitudinal designs and more complex patterns of change, traditional statistical models fall short (Dass & Shropshire, 2012). One of the reasons is that the traditional methods typically make use of polynomials to represent time-based change (e.g., Duncan & Duncan, 2009). However, polynomials only reflect periodic trends and therefore cannot capture complex, nonlinear patterns of change, which are typically characterized by many irregular ups and downs (Collins, 2006; Fok & Ramsay, 2006). As a consequence, substantively relevant information about nonlinear change (e.g., periodically peaking or plateauing, recovering after a dip, or deteriorating after a high) remains underexposed and untested. Second, the traditional models for repeated measurements were developed within the variable-centered tradition, assuming that the population is homogeneous with respect to how phenomena unfold and how predictors impact on outcomes (Laursen & Hoff, 2006). An increasing number of studies, however, demonstrate that this assumption is not always valid (Jung & Wickrama, 2008; Meyer & Morin, 2016). Because of this reason, several researchers have made a call to supplement the large amount of variable-centered research with person-centered research, in which it is acknowledged that the underlying population is potentially heterogeneous in nature (Meyer & Morin, 2016; Weiss & Rupp, 2011).

In the present paper, we start by reviewing the dominant longitudinal data analysis techniques in the field, arguing that they fall short when it comes to modeling complex, nonlinear change over time. In response to this limitation, we introduce functional clustering, a family of methods that is relatively unknown to the field of organizational studies (for an exception, see Dass & Shropshire, 2012) and specifically developed to capture rich temporal information in high-density, longitudinal time-series data by grouping people with similar patterns of complex, nonlinear change over time. Following the introduction of functional clustering, we focus our attention on *k*-centres functional clustering (*k*-centres FC; Chiou & Li, 2007), demonstrating the auspiciousness of this method by applying it to data measuring within-person changes in organizational commitment following psychological contract breach. To facilitate the implementation of the method in the organizational sciences, we provide a step-by-step tutorial and present MATLAB code. By doing this, the present paper offers a methodological-substantive synergy (Marsh & Hau, 2007), showing how sophisticated methodology can be used to address important substantive issues.

The Traditional Way of Modeling Longitudinal Data

In the organizational sciences, change over time is typically tested using latent growth curve modeling (Collins, 2006), with the most common growth curve model fitting a linear growth function to the data:

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij} \quad (1)$$

In this linear growth model, Y_{ij} represents individual i's outcome score at time j, t_{ij} represents the measurement time j for individual i, and β_{0i} and β_{1i} represent the intercept and (linear) slope for individual i. The model parameters (in this case the intercept β_{0i} and linear slope β_{1i}) are random effects, implying that they can differ between individuals.

Whereas the linear growth curve model is simple and therefore an elegant way to model growth, its major downside is that it often oversimplifies reality (Stoel, van den Wittenboer, & Hox, 2004). The first reason for this oversimplification is that it only captures individual differences in the starting position (β_{0i}) and the *amount* of change over time (β_{1i}). Linear growth with a constant rate

of change, however, is the exception rather than the norm. For example, in their paper on newcomer adjustment, Solinger, Van Olffen, Roe, and Hofmans (2013) showed that socialization scenarios are highly nonlinear for a large group of newcomers. While some newcomers show clear signs of initial rise, followed by a steady decline of their affective commitment (i.e., the honeymoon hangover pattern), others were characterized by a pattern of growth that starts off slowly but increases rapidly thereafter (i.e., the learning to love scenario).

One way to test these more complex functional forms while staying within the latent growth curve modeling approach is to consider the family of polynomial functions (Curran, Obeidat, & Losardo, 2010). Using this approach, one expands the polynomial relating time t_{ij} to the outcome by adding higher-order terms. However, the use of polynomials has several limitations (Davidian & Giltinan, 1995). First, whereas polynomials allow capturing relatively simple nonlinear patterns, the approximation of highly complex nonlinear patterns with a set of polynomials is difficult (Collins, 2006). The reason is that polynomials only capture periodic trends, whereas "raw, unadjusted time-series data often contain periodic effects as well as long-term trends or drifts" (Fok & Ramsay, 2006, p. 109). While polynomials are thus well suited to approximate simple growth or change over time in longitudinal series with only a few waves of measurement, they fail to capture more complicated patterns of change such as abrupt nonlinear changes, plateaus, or short-term periods of elevation (Wood & Jackson, 2013). As a result, when patterns of fluctuation are more complex, which is typically the case with more intensive schedules of measurement, time-graded polynomials no longer suffice (Fok & Ramsay, 2006).

A second important limitation of polynomials is that the individual coefficients lack substantive meaning. For example, whereas a cubic polynomial can be used to closely approximate logistic and exponential functions, the coefficients of the polynomial have no direct interpretation in terms of the growth process (Hox, 2002). One of the reasons is that the terms making up the polynomial are typically highly correlated, which makes it difficult to interpret individual coefficients. Whereas the lack of meaningfulness of the individual coefficients is not a major problem per se, it becomes problematic when one wants to predict growth because predicting growth implies relating covariates to individual coefficients. Therefore, if trajectory characteristics are the focus of investigation, one needs to move beyond the use of polynomials, and therefore, "other ways of specifying the change over time may be preferable" (Hox, 2002, p. 85).

The second reason why the latent growth model oversimplifies reality is that it assumes that a single growth trajectory holds for the whole population. Although this assumption might hold in very specific cases, existing theory and research suggests that describing an entire population with a single growth trajectory oversimplifies the complex growth patterns describing continuity and change among members of different groups (Jung & Wickrama, 2008). For example, Adi-Japha, Karni, Parnes, Loewenschuss, and Vakil (2008) showed that population-average representations of learning curves in skilled performance misrepresented the actual learning curves at the individual level. Similar results were presented by J. Li and Roe (2012) on the development of team conflict and Solinger et al. (2013) on the development of organizational commitment during onboarding. In other words, the assumption that the population can be approximated by a single set of growth parameters thus appears to be doubtful in what is presumably a large number of cases. One can relax this assumption by shifting from latent growth modeling to the growth mixture modeling (GMM) framework (Jung & Wickrama, 2008). GMM relaxes the assumption that all individuals in the sample are drawn from a single population with common parameters by allowing for differences in growth parameters across unobserved subpopulations. By making use of latent trajectory classes (i.e., categorical latent variables that serve as an indicator for the different unobserved subpopulations), GMM allows for different growth trajectories for each latent class. The result is a set of separate growth models, each with their unique estimates of variance and covariate influences (Jung & Wickrama, 2008). Yet, whereas switching from latent growth modeling to GMM is an elegant

way to shift from a variable-centered to a person-centered approach to modeling growth, the GMM remains an extension of the simpler latent growth model, and therefore the limitations concerning the usage of polynomials to model growth also apply to GMM.

Because of the limitations associated with latent growth curve modeling and its relatives such as GMM, we follow a different route to modeling change. In the present paper, we argue that functional clustering in general and *k*-centres FC (Chiou & Li, 2007) in particular constitute promising methods for OB researchers interested in modeling high-density repeated measurements data. In what follows, we first discuss functional cluster analysis, a group of statistical techniques specifically developed to capture heterogeneity in highly complex, nonlinear patterns in (repeated measures) data. Next, we describe *k*-centres FC, and finally we apply *k*-centres FC to a data set consisting of 109 participants reporting on their level of organizational commitment over a period of 10 weeks following the experience of a psychological contract breach event (Solinger, Hofmans, Bal, & Jansen, 2016).

Functional Cluster Analysis

Functional cluster analysis belongs to the functional data analysis (FDA) family, a family of statistical techniques specifically developed to model functions or curves. In FDA, the curves typically describe trajectories in two-dimensional space with time being one of the dimensions. The main idea underlying FDA is that the (repeated measurements) data are generated by a smooth process,[1] and accordingly the goal of FDA is to obtain insight in this smooth process (Ramsay & Silverman, 2005). To achieve this goal, FDA replaces the original (raw) observations by curves or functions, which are then used as the unit of observation for further analyses. A discussion of the potential of FDA to organizational and managerial science can be found in Dass and Shropshire (2012), which provides an overview of the fundamentals of FDA and briefly introduces and illustrates three FDA techniques: functional principal component analysis, functional regression, and functional clustering. Unlike Dass and Shropshire, the aim of the present paper is not to offer a broad introduction to FDA. Instead, we provide a brief overview of the different approaches to functional clustering, after which we focus on *k*-centres FC (Chiou & Li, 2007), demonstrating the usefulness of this technique for a person-centered analysis of high-density, repeated measurements data.

Functional Clustering Approaches

Very much like in traditional cluster analysis, the goal of functional cluster analysis is to look for homogeneous groups (or clusters) of observations in a data set. Several methods can be used for clustering functional data, and these different functional clustering methods can broadly be categorized in three categories or approaches: two-stage methods, distance-based functional clustering methods, and model-based clustering methods (Jacques & Preda, 2014a). In what follows, we will briefly review the different approaches, each time indicating their unique advantages and disadvantages.

The first category covers functional clustering methods that follow a two-stage approach. The idea underlying these methods is that one first replaces the raw observations by curves (i.e., the filtering stage), after which the curves are clustered (i.e., the clustering stage). In the filtering stage, the raw data are typically represented by a set of basis functions such as Fourier series (Serban & Wasserman, 2005) or B-spline functions (Abraham, Cornillon, Matzner-Lober, & Molinari, 2003), allowing one to summarize the raw data by their coefficients on the basis functions. In the second (clustering) stage, classical clustering algorithms such as the *k*-means algorithm are used to cluster the coefficients on the basis functions. The main advantage of the two-step approach is that it is conceptually simple, while the major downside is that the filtering step is performed separately from

the clustering step, which means that it is done independently of the goal of the clustering (Jacques & Preda, 2014a).

The second approach is the distance-based functional clustering approach. The idea driving this approach is to apply nonparametric clustering techniques, such as *k*-means or hierarchical clustering, in combination with distance measures or proximity measures specifically developed for functional data (see e.g., Ieva, Paganoni, Pigoli, & Vitelli, 2012). In other words, methods belonging to this approach use adaptions of the popular geometric clustering algorithms in the sense that they develop specific measures of similarity or proximity between the curves and then apply nonparametric clustering techniques to these similarity measures. As with the two-stage approach, the advantage of the nonparametric methods is that these methods are easy to understand and implement. Their major downside, however, is that complex cluster structures cannot be efficiently modeled (Jacques & Preda, 2014a).

The third category accommodates model-based functional clustering methods. Similar to the two-stage methods, model-based clustering methods assume a density probability on a finite number of parameters. However, as opposed to the two-stage approach, the filtering stage is not separated from the clustering stage; instead, the approximation of the raw data and the clustering are performed simultaneously. This approach is superior to the two-step approach and the nonparametric approach because model-based clustering methods perform the dimensionality reduction and the clustering simultaneously and because these methods allow modeling complex covariance structures (Jacques & Preda, 2014a).

The category of model-based clustering methods can further be divided in two subcategories. A first subcategory approximates the raw data using basis functions such as cubic splines (James & Sugar, 2003; note that this is the clustering method Dass & Shropshire [2012] refer to when discussing functional clustering), B-splines, or P-splines (Coffey, Hinde, & Holian, 2014). Methods belonging to the second category instead rely on functional principal component analysis, representing the curves by estimating the eigenfunction basis from the data using functional principal component analysis (FPCA) (Yao, Müller, & Wang, 2005a, 2005b). Examples of such methods are the Gaussian mixture model applied to FPCA scores (Jacques & Preda, 2014b) and *k*-centres FC (Chiou & Li, 2007).

In the present paper, we focus on *k*-centres FC, a model-based clustering method using data-adaptive basis functions. The choice for *k*-centres FC is motivated by several reasons. First, unlike the approaches based on the clustering of basis coefficients, *k*-centres FC does not require the same basis function for the clustering. Instead, it relies on an eigenfunction decomposition of the functional data and therefore allows each cluster to have its own functional form, which allows for great flexibility. This assumption also resonates better with the theoretical basis assumption in person-centered research that different subpopulations can have recognizable and qualitatively different pattern characteristics. Second, whereas most clustering methods are designed for clustering around mean functions, *k*-centres FC allows both the means and the shapes of the functions (i.e., the eigenfunctions) to differ between clusters (Chiou & Li, 2007). Therefore, if the clusters differ not only in the mean function but also in the structure of the covariance function, *k*-centres FC improves the quality of the clustering (Wang, Chiou, & Müller, 2016) while it also performs well when the clusters differ in the mean function only (Chiou & Li, 2007). Finally, and as opposed to most other model-based clustering approaches that require Gaussian model assumptions, *k*-centres FC does not rely on distributional assumptions (Chiou & Li, 2007).

k-centres Functional Clustering

As mentioned before, a typical step in functional data analysis is the projection of the data onto a low dimensional space using a set of basis functions. In the *k*-centres FC method, the low-dimensional

projection is performed by approximating the basis functions through functional principal component analysis (fPCA) using the following expansion:

$$Y_i(t) = \mu(t) + \sum_{k=1}^{\infty} A_{ik}\phi_k(t) \tag{2}$$

where $\mu(t)$ is the mean trajectory, A_{ik} are the functional principal components of Y_i and ϕ_k are the eigenfunctions. Unlike in traditional PCA, the functional principal components in fPCA do not reflect common domains in the scores to different variables but common domains in the repeated measures data (i.e., particular pattern shapes). This implies that the component loadings are replaced by *component profiles* over time, capturing the functional (trajectory) characteristics that describe the major forms of variability in the functional data, while the component scores represent to what extent the individual profile is characterized by each of the functional characteristics. Because the first K terms in Equation 2 provide a good approximation of the infinite sum, the following approximation is used:

$$Y_{iK}(t) = \mu(t) + \sum_{k=1}^{K} A_{ik}\phi_k(t). \tag{3}$$

Further, k-centres FC assumes that the random curves or functions are independently sampled from a mixture of nonparametric random effect models (see Equation 4). These nonparametric random effect models consist of a fixed component (i.e., $\mu^c(t)$), and a random component (i.e., $\sum_{k=1}^{K} A_{ik}^c \phi_k^c(t)$) with both components being cluster-specific. As a result, k-centres FC accounts for both the means and modes of variation in the functional data. This property of k-centres FC sets the method apart from other functional coefficient-based clustering methods, which are typically designed for clustering according to mean functions, without considering the cluster covariance structures (Chiou & Li, 2007).

$$\tilde{Y}_i^c(t) = \mu^c(t) + \sum_{k=1}^{K_c} A_{ik}^c \phi_k^c(t). \tag{4}$$

The cluster memberships in the k-centres FC method are determined by minimizing the discrepancy between Y_i and $\tilde{Y}_i^c$ using the following L^2-distance criterion:

$$c^*(Y_i) = \underset{c\in\{1,\ldots,L\}}{\operatorname{argmin}} \sum_{i=1}^{n} d^2(Y_i, \tilde{Y}_i^c). \tag{5}$$

The k-centres FC procedure runs according to the following sequence: In the *initial clustering* step, the marginal mean and covariance structures (i.e., the marginal fPCA scores) are clustered using the k-means clustering method. Next, in the *iterative updating via reclassification* step, the cluster-specific mean and covariance structures are repeatedly reestimated while these means and covariance structures are simultaneously used to reclassify each curve into the best predicted cluster using criterion (Equation 5). Note that this iterative procedure allows each of the different clusters to be described by a different number of principal components, which allows for great flexibility in the description of the subprocesses. Ideally, the reclassification is done using the leave-one-out technique, where estimates of the mean function and eigenfunctions are obtained for each cluster using the observed curves minus one curve that is left out of the sample. Next, this left-out curve is classified into one of the clusters according to Equation 5. This procedure is performed for each single curve of the sample and repeated until no more curves can be reclassified.

Applying *k*-centres FC: How Commitment to the Organization Recovers From Psychological Contract Breach

In what follows, we apply the *k*-centres FC method to commitment data from 109 participants who reported psychological contract breach (PC breach) events (as reported in Solinger et al., 2016). A PC breach event represents an employee's perception that the organization has failed to deliver on an obligation (Bal, Hofmans, & Polat, 2017; Morrison & Robinson, 1997), and such breaches have been shown to carry considerable consequences in terms of within-person fluctuations of commitment and performance (Griep, Vantilborgh, Baillien, & Pepermans, 2016; Solinger et al., 2016). The data were selected from a larger data file including weekly measures of breach events and organizational commitment with up to 25 waves per person. From this larger data file, we selected for each person the first breach event reported in the data and retained the commitment data until 10 weeks after this event (yielding a maximum of 11 observations per individual). In total, 775 observations were retained, with on average 7.11 observations per participant ($SD = 2.46$). Further details on the procedure and instruments can be found in Solinger et al. (2016).

The major difference between this analysis and the analysis reported by Solinger and colleagues (2016) is that the functional principal component analysis applied by Solinger et al. assumes that individual differences in the commitment trajectories of all respondents can be captured by the same set of principal components. As such, it assumes that the growth trajectories of all individuals are driven by the same underlying mechanism, which Solinger et al. assumed to be the breach resolution process. In line with this assumption, Solinger et al. found that the largest portion of the variance in post-breach change could simply be described by successful breach resolution (i.e., bouncing back after breach) versus unsuccessful resolution (i.e., durable impairment of commitment levels after breach), while the second principle component described a set of nonlinear responses. Whereas this might be true, existing theory suggests that the assumption that these two principle components describe everyone's breach repair process is an oversimplification of reality. For example, Skinner, Edge, Altman, and Sherwood (2003) detail that apart from a broad process of (un)successful recovery from breach, there will be subtle differences in the ways individuals cope with a negative event (see also Bankins, 2015). Importantly, these different ways of coping will most likely correspond with a distinct type of breach resolution (Tomprou, Rousseau, & Hansen, 2015) and distinct patterns of development of organizational commitment. By applying *k*-centres FC to these data, we put this assumption to the test. Moreover *k*-centres FC will offer a detailed insight into subprocesses that may give rise to a specific set of trajectories. This is an important contribution because currently the process through which people recover from breach has received scarce research attention (Tomprou et al., 2015), and therefore research on the different mechanisms that drive breach repair is needed.

In what follows, we will discuss the important steps in the analysis. In the Appendix, we provide Matlab code that can be used to run *k*-centres FC using the kCFC Matlab package.

Step 1: Pre-Processing of the Data

As *k*-centres FC approximates the basis functions by means of fPCA, the clustering in *k*-centres FC is based on those features that capture the majority of the variation in the temporal data. Relevant to this issue is that with raw data, the first principal component typically captures level differences (applied to our example: between-person differences in the average level of commitment), which means that level differences usually strongly impact the cluster solution. Often such level differences are substantively meaningful, and in these circumstances, one can safely proceed with clustering raw data. In some cases, however, level differences should be removed before one starts with the clustering. For example, when—as is the case in our paper—the goal is to identify inter-individual

differences in the reactions to breach, one would typically like the clustering to only reflect differences in patterns of within-person fluctuations following breach and not differences in overall commitment levels (see also Solinger et al., 2016). In this case, person-centering (or group-mean centering) the data before performing the *k*-centres FC analysis is recommended because it removes level differences from the data (the person-centered commitment data are shown in Figure 1). In summary, when it is theoretically important that the clustering only reflects differences in patterns of within-person fluctuations and not differences in overall level, one should first person-center the data. Note that person-centering the data is not commonly done in person-centered research and that centering decisions should be informed by the substantive question at hand.

Step 2: Run the k-centres FC Analysis

The *k*-centres FC procedure consists of an *initial clustering* step and an *iterative updating via reclassification* step. In the initial clustering step, the marginal means and the marginal fPCA scores are clustered using *k*-means. To this end, one needs to specify how many clusters to retain (Step 2a), how many functional principal components to retain (Step 2b), and the method that should be used to estimate the fPCA scores (Step 2c). In the iterative updating via reclassification step, the cluster-specific means and covariance structures are repeatedly reestimated and used to reclassify each curve into the best predicted cluster. Therefore, in addition to the aforementioned steps, one also needs to specify the stopping rule (Step 2d), whether the leave-one-out technique should be used (Step 2e), and the maximal number of iterations (Step 2f). In what follows, we will discuss each substep.

Step 2a: Specify how many clusters to retain. A reasonable approach for deciding on the number of clusters is to perform the *k*-centres FC analysis with different numbers of clusters and select the cluster solution for which the results make sense from a substantive point of view (Chiou & Li, 2007). Recently, however, P.-L. Li and Chiou (2011) developed a formal and systematic method to help identifying the number of clusters in *k*-centres FC. This approach, called *forward functional testing* (FFT), tries to find the maximal number of clusters while retaining significant differences between the different cluster structures. To do so, FFT starts with a small number of clusters and each time tests equalities between the cluster mean functions and between the sets of cluster eigenfunctions using bootstrap resampling methods. The number of clusters is each time increased until the maximum number of distinguishable clusters is reached. Whereas this procedure is promising, no software code is yet available. In our study, we found that when asking for three or more clusters during the reclassification step, the cluster solution converged on a two-cluster solution. This suggests the presence of two important clusters in the data, and because of this reason, we proceeded with the two-cluster solution.

Step 2b: Selection of the number of functional principal components. There are several ways to decide on the number of functional principal components. The first option is to a priori specify this number. Whereas this approach is straightforward, it is clearly suboptimal as the optimal number of functional principal components can differ per cluster. Moreover, in the reclassification step, the cluster-specific covariance structures are repeatedly reestimated, which means that the optimal number of functional principal components can even vary per iteration. Therefore, we suggest to not specify a fixed number of functional principal components but instead to rely on model selection criteria, such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), or the cumulative percentage of total variance accounted for. For the cumulative percentage of total variance accounted for, the set of components that jointly explains more than 90% of variance is typically considered appropriate, while in case of the AIC and BIC, the component solution that minimizes the

AIC (respectively, BIC) is selected (Dass & Shropshire, 2012). Among these model selection criteria, the BIC and the cumulative percentage of total variance accounted for yield the most parsimonious models, which is why we opted for the BIC in the present study.

Step 2c: Specify the method for estimating the fPCA scores. There are two methods for estimating fPCA scores: the classical integration method (IN) and the conditional expectation method (CE). The difference between both methods is that classical integration can only be used with dense, regular data. Conditional expectation in turn borrows information from the data on all subjects when estimating the fPCA scores, which means that it can deal with sparse, irregularly spaced longitudinal data (see Yao et al., 2005a). Because of that reason, we recommend using the conditional expectation method when performing *k*-centres FC.

Step 2d: Specify the stopping rule for the k-centres FC algorithm. In the reclustering step, the cluster means and covariance structures are iteratively updated, which means that with each iteration, a reclustering takes place. To determine whether the *k*-centres FC algorithm should continue iterating (i.e., to check when the *k*-centres FC algorithm has converged), the user can specify the minimum percentage of curves that needs to be reclassified for the reclustering to continue. In the present study, we set this number to 0, meaning that the *k*-centres FC algorithm kept reclassifying curves until no more curves could be reclassified.

Step 2e: Specify whether the leave-one-out technique should be used. One needs to decide whether the *k*-centres FC analysis should be performed with or without the leave-one-out technique. With the leave-one-out technique, estimates of the mean function and eigenfunctions are obtained for each cluster using the observed curves minus one curve, after which the left-out curve is classified into one of the clusters. Although performing *k*-centres FC using the leave-one-out technique typically improves the clustering, it is very time-consuming. Because of that reason, Chiou and Li (2007) suggest to use the leave-one-out technique especially when the number of curves is not large. The reasoning is that for a large number of curves, the cluster structures are less influenced by each individual curve, meaning that the bias of including the to be classified curve in the estimation of the mean function and eigenfunctions is ignorable. Whereas Chiou and Li are not explicit about what a large enough number of curves is, we believe that most studies in the organizational sciences (working with samples sizes up to a few hundred) would be well advised to use the leave-one-out technique. Because of this reason, also in the present paper, the *k*-centres FC analysis was performed using the leave-one-out technique.

Step 2f: Specify the maximum number of iterations. Because *k*-centres FC uses an iterative procedure, one needs to specify the maximum number of iterations. Regarding this issue, Chiou and Li (2007) report that in their simulation study, convergence occurred mostly after 5 or 6 iterations, with some cases requiring up to 10 iterations. In our analysis, we specified the maximum number of iterations to be 50, with convergence being reached after 17 iterations. Because the number of iterations depends among other things on the quality of the initial clustering and the strictness of the stopping rule (see aforementioned), one might try to adopt a less strict stopping rule or increase the number of iterations when the model does not convergence. Note that specifying a large maximum number of iterations typically does not increase computation time because the algorithm stops iterating once convergence is reached.

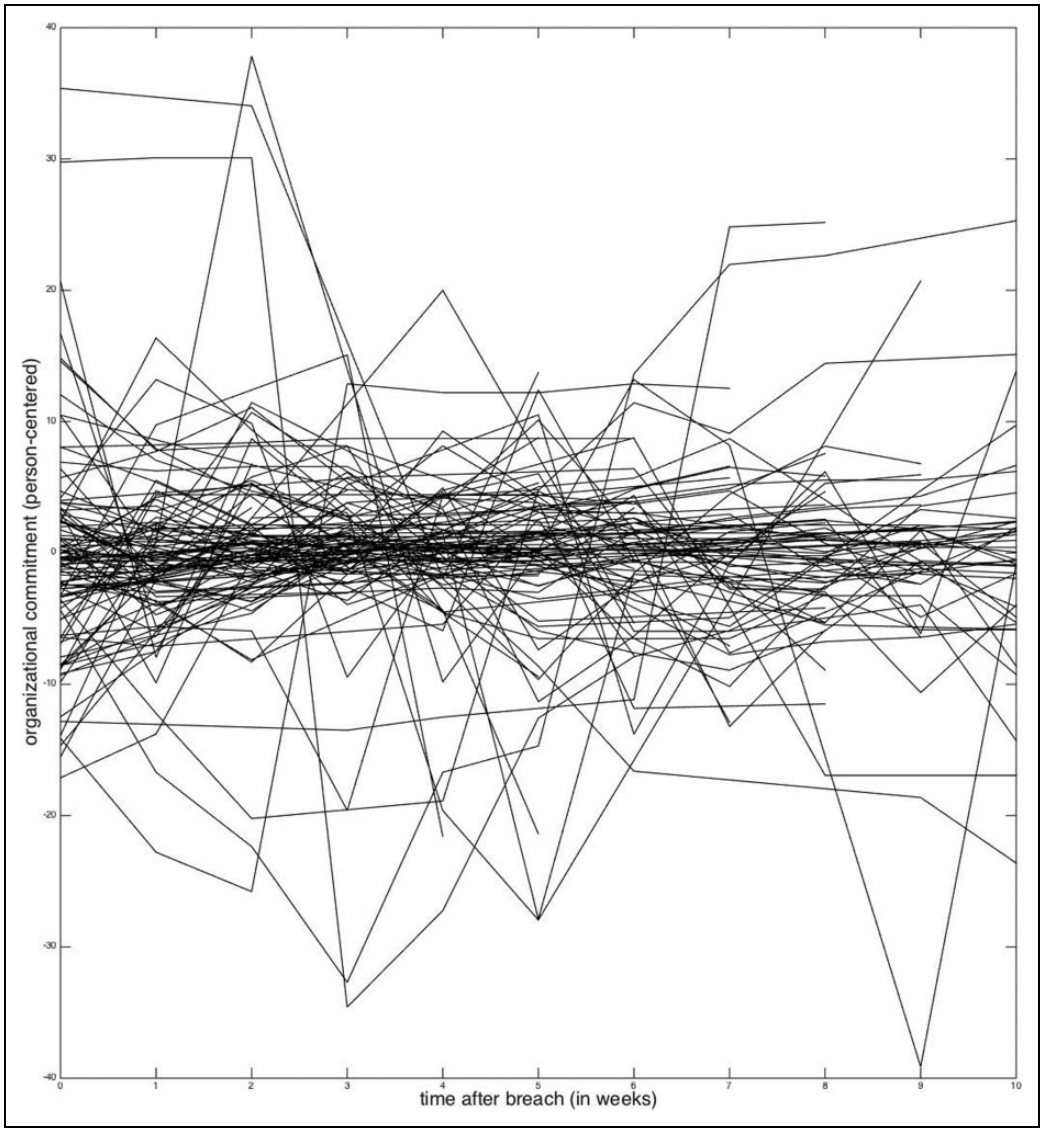

Figure 1. Raw commitment trajectories for all individuals in the sample (person-centered data).

Step 3: Interpret the Cluster Solution

After running the *k*-centres FC analysis, one would typically like to interpret the cluster solution. As *k*-centres FC accounts for both the means and modes of variation in the data, one needs to inspect the cluster-specific mean profiles as well as the cluster-specific principal components. To aid in this interpretation, one can plot the cluster-specific means and modes of variation (Step 3a) and reconstructed curves (Step 3b).

Step 3a: Plot the cluster-specific means and modes of variation. To interpret the cluster-specific mean profiles and the cluster-specific principal components, for each cluster × principal component combination, we made a plot combining (a) the cluster-specific mean profile and (b) predicted

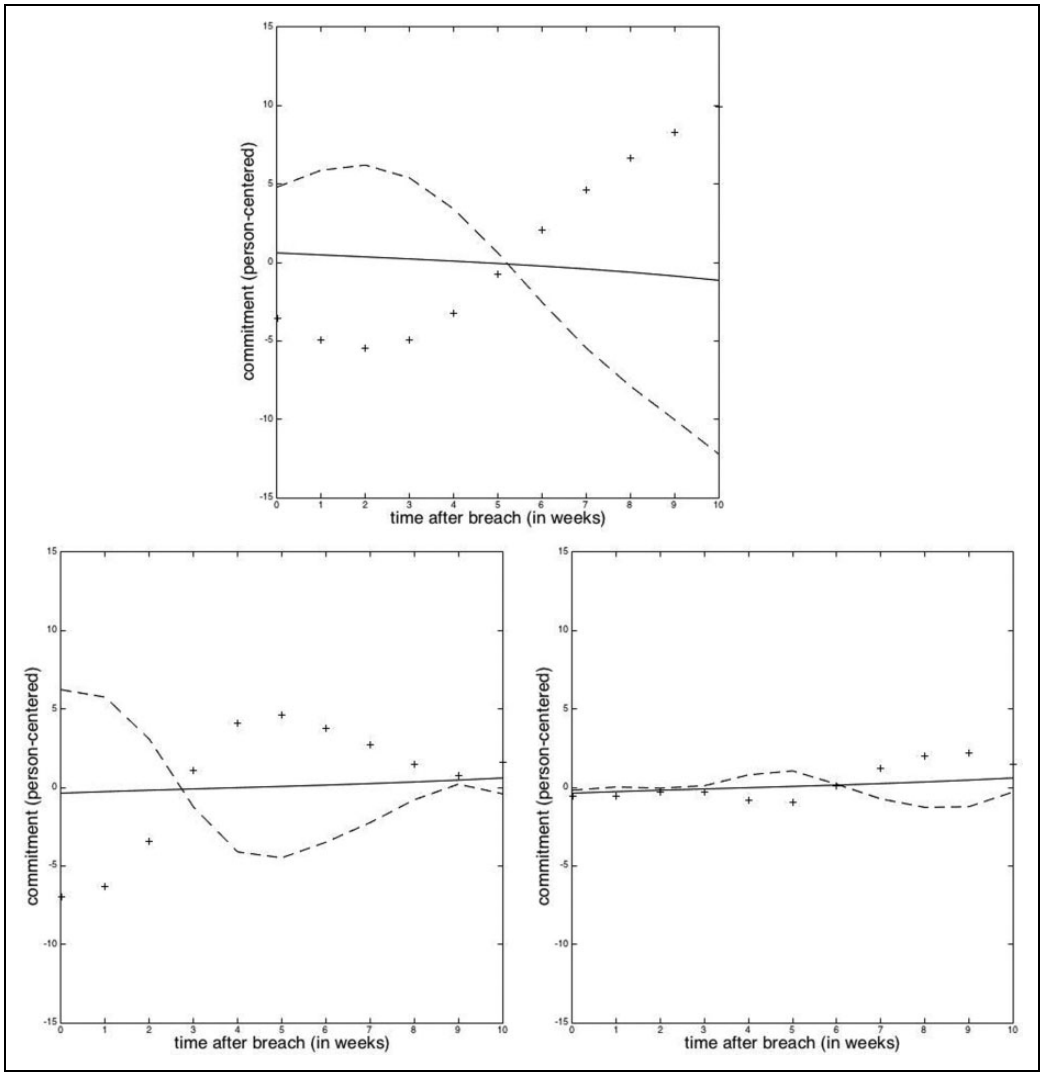

Figure 2. Mean function (full line) and principal components for Cluster 1 (upper panels) and Cluster 2 (lower panels).
Note: Predicted commitment trajectories for people scoring –1 and +1 *SD* on the functional principal component scores are shown by the minuses and plusses, respectively.

commitment trajectories for people scoring –1 *SD* and +1 *SD* on the functional principal component scores while taking an average score on the other cluster-specific functional component(s) (see Figure 2). These plots show how one's organizational commitment trajectory is impacted when one scores higher (respectively lower) on one of the cluster-specific principal components.

Regarding the cluster means (i.e., mean trajectory profiles), we see that individuals belonging to Cluster 1 ($N = 42$; 38.53% of the participants) showed a decreasing trend in organizational commitment after breach, while people belonging to Cluster 2 ($N = 67$; 61.47% of the participants) on average experienced an increase in their commitment to the organization.

Turning to the interpretation of the functional principal components, we see that the commitment trajectories of people belonging to Cluster 1 are captured by one principal component that explains

almost all variation in the data (i.e., 97.50% of the variance). This component captures a subprocess in which there is a short period of relative inertia, after which the commitment level starts to increase or decrease. For people with high principal component scores (represented as +1 *SD* in Figure 2, upper panel), commitment *bounces back to a higher level* than before, a temporal pattern that is indicative of "psychological contract thriving" (Solinger et al., 2016; Tomprou et al., 2015). In a psychological contract thriving scenario, the psychological contract ends up in a *more favorable* state than before the breach event, which is consistent with prior research showing that sometimes negative events may elicit a "cognitive restructuring" coping response in which individuals reinterpret the negative event in a positive light such that individuals feel "stronger" than before (e.g., Bankins, 2015; Park & Folkman, 1997; Skinner et al., 2003). People with low component scores (–1 *SD* in Figure 2, upper panel), instead, de-commit after the breach event, which means that they end up in a *less favorable* state. Such a pattern is indicative of "psychological contract impairment" (Tomprou et al., 2015), a scenario in which the individual accepts a less than ideal employment situation and is less willing to invest effort in the organization than before the breach event. Since there is no evident sign of plateauing for this particular pattern (at least, not in this 11-week timeframe), it may even be the case that an individual will over time completely lose his or her commitment to the organization, something that is associated with a dissolution of the psychological contract (Solinger et al., 2016; Tomprou et al., 2015).

The commitment trajectories in Cluster 2 in turn are captured by two important principal components. The first principal component (explaining 86.86% of the variance) reflects an *immediate reaction* to the breach event, after which the commitment level stabilizes at a higher level of commitment (+1 *SD* in Figure 2, lower left panel) or at a lower level of commitment than before the breach event (–1 *SD* in Figure 2, lower left panel). The second principal component (explaining an additional 10.48% of the variance in the data) represents a *delayed reaction* to the breach event and predominantly captures increases (+1 *SD* at the bottom right panel of Figure 2) or decreases (–1 *SD* at the bottom right panel of Figure 2) in commitment near the end of the 10-week period. In summary, whereas the first component in Cluster 2 captures inter-individual differences in the *immediate* reaction to breach, the second component captures more subtle differences in the *delayed* reaction.

Step 3b: Plot reconstructed curves. To interpret the between-cluster differences further, one can plot reconstructed curves based on the cluster-specific means and the person-specific functional principal component scores (see Figure 3). In fact, this procedure plots for each participant the predicted curve based on the *k*-centres FC estimates. From this figure, it can be seen that the commitment profiles of participants belonging to Cluster 1 (i.e., the left panel in Figure 3) are indeed characterized by a short period of relative inertia, after which the commitment level increases or decreases. In contrast, the commitment profiles of participants belonging to Cluster 2 (i.e., the right panel in Figure 3) are characterized by a mixture of periodic effects and plateauing and therefore show substantially more irregularities.

Step 4: Add Covariates

In contrast to traditional mixture models, where covariates (i.e., predictors, outcomes, or correlates) can be directly included in the model, adding covariates in *k*-centres FC must be done according to a two-stage strategy. A first way in which covariates can be tested is by relating the cluster memberships to covariates (Step 4a). Second, because *k*-centres FC relies on an eigenfunction decomposition of the repeated measures data, one can—besides predicting cluster memberships—also predict the scores on each of the cluster-specific eigenfunctions. That is, by looking for covariates that predict only a subset of the components, one might gain better insight in the factors that

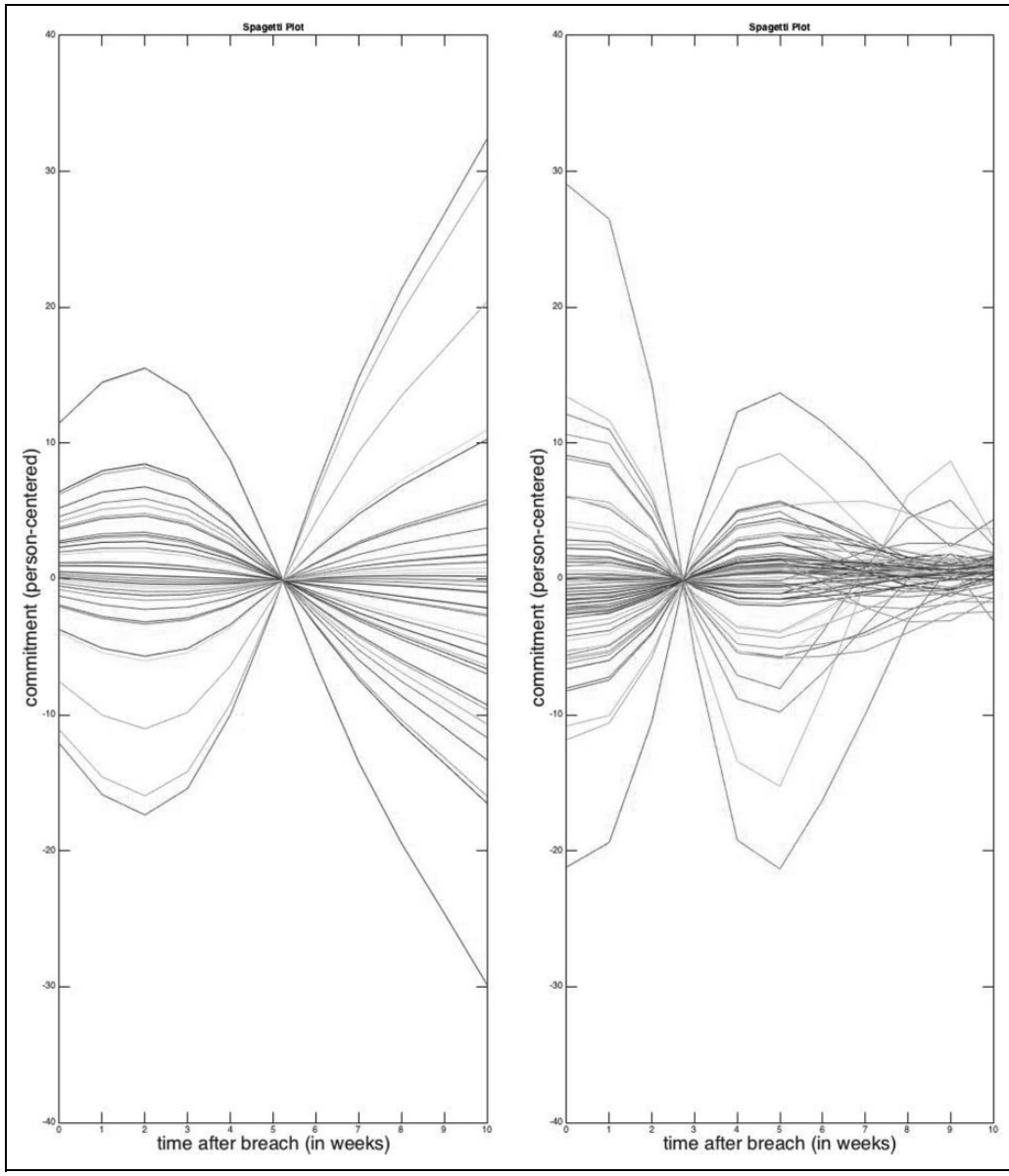

Figure 3. Reconstructed person-centered commitment trajectories based on the principal components for Cluster 1 (left panel) and Cluster 2 (right panel).

differentially drive cluster membership and/or their separate subprocesses. To this end, the component scores per cluster can be saved, after which they can be related to covariates (Step 4b).

In the context of the present paper, commitment level at the time of the breach event and evolution in commitment prior to the breach event were included as covariates. Commitment level was included because individual differences in commitment level were removed by person-centering the data. Reintroducing them as a covariate allowed testing whether inter-individual differences in level relate to inter-individual differences in patterns of change. Evolution in commitment before the breach event in turn allowed testing whether the evolution in commitment before

the breach event impacts the way people react to the breach. Evolution in commitment was measured by testing a linear growth model on the commitment data in weeks $t-5$, $t-4$, $t-3$, $t-2$, and $t-1$ (with t being the week of the breach event) and by saving the linear slope per participant. Important to note here is that because we had less than two pre-breach measurements for 51 participants, there are many missings for this covariate.

Step 4a: Relating cluster membership to the covariates. Relating commitment level and evolution in commitment to cluster membership showed that the average level of commitment of Cluster 1 ($M = 63.25$; $SD = 22.46$) was significantly *lower* than that of Cluster 2 ($M = 77.57$; $SD = 15.91$), $t(107) = -3.60$; $p < .001$. Regarding the evolution of commitment before the breach, there was no difference between people in Cluster 1 ($M = .42$; $SD = 1.01$) and people in Cluster 2 ($M = .36$; $SD = 1.01$), $t(56) = .22$; $p = .824$.

Step 4b: Relating cluster-specific fPCA scores to covariates. Relating the scores on each of the cluster-specific eigenfunctions to the person's average level of commitment and to the individual's pre-breach evolution in commitment showed that for Cluster 1, individual differences on the principle component were (marginally) related to commitment level, $r(N = 42) = -.28$; $p = .075$, but not to pre-breach evolution in commitment, $r(N = 21) = -.04$; $p = .872$. In other words, for individuals belonging to Cluster 1, higher average levels of commitment are predictive of a negative reaction in which commitment descends after the breach event. This finding is in line with the results of Brockner, Tyler, and Cooper-Schneider (1992), showing that people who feel highly committed have the most negative reactions when treated in an unfair way by their organization (see also Bal, Chiaburu, & Jansen, 2010). For people belonging to Cluster 2, the reverse pattern of associations was found. The component scores on Components 1 and 2 related to pre-breach evolution in commitment, $r(N = 37) = -.39$; $p = .019$ and $r(N = 37) = .59$; $p < .001$, respectively, while they were unrelated to the person's average commitment level, $r(N = 67) = -.09$; $p = .464$ and $r(N = 67) = .12$; $p = .345$, respectively. This means that for people in Cluster 2, a stronger increase in commitment before the breach event is predictive of an immediate negative reaction in which commitment descends in the first weeks after breach and then plateaus at a lower level of commitment. At the same time, individuals showing a stronger increase in commitment before the breach event tend to show a slight increase in commitment near the end of the 10-week period. Together, these findings extend the findings of Brockner et al. (1992); while people in a positive flow before a breach incident indeed are more likely to show negative immediate reactions, they also tend to recover more easily at a later point in time. Note that this observation also connects with the notion that path dependency is highly relevant when explaining organizational behavior: Those in a "winning mood" react differently to similar conditions compared to those on demise (Solinger et al., 2013).

In summary, our application of k-centres functional clustering to repeated measurements data following PC breach yielded important insights in the processes driving breach recovery. First, our analysis revealed that breach recovery is not driven by one and the same mechanism for everyone: For some people, breach recovery was captured by a relatively straightforward increase or decrease in commitment resulting from (un)successful resolution. For others, recovery involved a mixture of immediate (positive or negative) and delayed (positive or negative) reactions. As such, our findings extend the findings of Solinger et al. (2016) by showing that there are important inter-individual differences in the processes driving breach repair. Interestingly, a comparison of our findings with the findings of Solinger et al. shows that the first principle component in Solinger et al. strongly resembles the principle component of Cluster 1. The second principle component in Solinger et al. in turn looks like a mixture of the two principle components of our Cluster 2. As such, it appears that functional PCA captures the heterogeneity in the data by means of separate, complex principle components. k-centres functional clustering in turn recognizes that the reaction patterns of

different people are characterized by different driving factors, and it is successful in revealing what these factors are.

Discussion

Throughout the paper, we argued that FDA in general and functional clustering in particular "can often extract information contained in the functions . . . not normally available through traditional methods" (Levitin, Nuzzo, Vines, & Ramsay, 2007, p. 135). Indeed, whereas the typical growth curve model fits a linear growth function to the data (Ghisletta, Cantoni, & Jacot, 2015), there is ample evidence that nonlinear growth may be the rule rather than the exception (Stoel et al., 2004). By modeling nonlinear growth in a linear way, the true growth processes are misrepresented, leading to incomplete or even erroneous conclusions. Within the traditional, nonfunctional framework, there are several ways to accommodate such nonlinear relationships. Perhaps the most well-known and popular approach concerns the inclusion of higher order polynomials in the latent growth curve model. However, this requires the researcher to specify the order of the polynomial in advance, which is problematic because in organizational psychology, there are—to our knowledge—no theories that propose how attitudes, behaviors, or affect change as a function of elapsed time (for a similar criticism in the domain of developmental psychology, see Ram & Grimm, 2007). Instead, dynamic organizational theories often make predictions about growth, stasis, or decline without explicating the precise form of these changes. As a result, applied researchers often recover the form of change in a data-driven way rather than testing it in a theory-driven way (e.g., by testing increases in predictive power when increasing the order of the polynomial). Because polynomials have a relatively simple nonlinear form, highly complex relationships, such as those encountered in our example (and especially the patterns of Cluster 2), are difficult to model. Therefore, polynomials might not be the best choice when one wants to recover the precise form of change.

Of course, polynomials are not the only way to model nonlinear growth. Another option is to test the nonlinear function directly by, for example, fitting logistic or exponential growth curves (Burchinal & Appelbaum, 1991). Although the advantage of this approach is that the parameters of the function have a direct interpretation in terms of the hypothesized growth process, one needs to have strong theoretical expectations regarding the growth function as misspecification of the function leads to misrepresentation and therefore misinterpretation of the growth process. Because strong theoretical expectations about the form of change are lacking in contemporary organizational theories, modeling the nonlinear growth function directly is challenging. Still another option is to test a latent basis growth model (McArdle, 1986). The core idea of the latent basis growth model is that the rate of change is not necessarily constant across time (Berlin, Parra, & Williams, 2014). Therefore, the loadings for the different measurement occasions on the slope factor are not fixed according to the time intervals between the measurement occasions but are estimated from the data. This is typically done by fixing the loading for the first measurement occasion to zero, fixing the last loading to one, and estimating all other loadings freely. The result is that time is rescaled in the model with loadings representing the proportion of the total amount of change that has occurred up to that point in time (Grimm, Ram, & Hamagami, 2011). Because it has no specific functional form, the latent basis growth model is able to capture a wide variety of nonlinear growth patterns. At the same time, because all but two shape factor loadings need to be estimated, latent basis growth models are parameter heavy, and this problem only gets worse when the data have many measurement occasions (Grimm et al., 2011). Moreover, as the nonlinear growth pattern is reflected in the estimated factor loadings, it might be challenging to map the parameters onto theoretical notions of developmental processes (Grimm et al., 2011).

Considering the limitations of the more traditional models, our offering introduces a different route to modeling such data by introducing functional clustering. The main advantage of functional

methods is that because of their flexibility in modeling nonlinear and heterogeneous data, they outperform static models when it comes to capturing processes and dynamics over time (Dass & Shropshire, 2012; Jank & Shmueli, 2006). Moreover, as for clustering functional data, we have argued that *k*-centres FC has some unique advantages over other functional clustering methods: It does not require the same basis function for the clustering, it accounts for both the means and the shapes of the functions between clusters, and it does not rely on distributional assumptions. At the same time, functional methods in general and *k*-centres FC in particular also come with some limitations. A first limitation is that because the functional methods use the raw repeated measurements data to estimate complex functions, they typically need data with a higher resolution than the more traditional techniques, which means that data with a high signal-to-noise ratio and with sufficient repeated measurements per individual are required (Levitin et al., 2007; Ramsay & Silverman, 2005). However, recently developed functional methods (i.e., the conditional expectation method, Yao et al., 2005a) allow for the modeling of noisy, irregularly spaced, sparse data as well, and we take advantage of these methods in the current study. A related issue is that one needs to understand that the cluster quality of *k*-centres FC is affected by the number of individuals (or number of curves) and the number of observations per individual. This is due to the fact that the clustering criterion of *k*-centres FC depends on the estimated fPCA scores, which have asymptotic properties under large sample size. When the number of observations per subject or when the number of subjects is small, the clustering result may be suboptimal, and in such a situation, functional clustering methods that are specifically developed for sparse data might be preferred (e.g., James & Sugar, 2003). Second, although *k*-centres FC makes no distributional assumptions, the classification according to minimal distances (see Equation 5) implies equal within-cluster variations, an assumption that is often not true in field data (Morin et al., 2011). Third, *k*-centres FC performs hard clustering, meaning that the growth curves have a probability of one to belong to one of the clusters and a probability of zero to belong to all other clusters. Although this nonprobabilistic clustering is an important limitation of traditional *k*-means clustering (where because of the hard clustering it is assumed that each individual corresponds entirely to the prototypical cluster profile; Meyer & Morin, 2016), it is less of a problem in *k*-centres FC. The reason is that in *k*-centres FC, each cluster is characterized by a cluster-specific mean and cluster-specific eigenfunctions, meaning that curves belonging to the same cluster are not assumed to be identical to the prototypical curve but are only assumed to have the same basis functions. In other words, curves belonging to the same cluster can still widely differ because they take different scores on the principal functional components. A final limitation is that whereas the traditional statistical techniques are included in most standard software packages, functional methods are still limited to tailor-made packages for Matlab or R (see Graves, Hooker, & Ramsay, 2009). Specifically, apart from the kCFC Matlab package that is used in the present paper, there are two R packages that allow one to perform *k*-centres FC. The first package, *fdapace*, mirrors the possibilities of the kCFC Matlab package but currently does not work when one has data with less than 10 observations per individual. However, when one works with rather dense data, we can recommend the *fdapace* package to R users. The second package, *funcy*, also performs *k*-centres FC but currently only allows users to predefine the number of functional principle components. As the optimal number of functional principal components can vary per cluster and even per iteration (see Step 2b), this is in our opinion a major limitation.

Choosing a Method for One's Data

Because there are multiple person-centered methods that can accommodate nonlinear growth—each with their individual strengths and weaknesses—several considerations should be taken into account when choosing a method for one's data. In particular, when one has strong theoretical expectations

about the form of the growth function, modeling the nonlinear function directly is probably the preferred option because in that case, the parameters of the function have a direct interpretation in terms of the hypothesized growth process. Instead, if strong theoretical expectations about the form of growth are missing—which is true for most organizational theories—the exact growth function has to be inferred from the data. In case the researcher collected only a few waves of measurement, the raw data can probably be approximated reasonably using polynomials, which allows for the modeling of simple growth or change over time. With more intensive schedules of measurement, polynomials typically no longer suffice, and one has to go for methods that can flexibly model complex nonlinear growth patterns. In such a situation, functional clustering methods in general and *k*-centres FC in particular have a lot to offer.

Unique advantages of *k*-centres FC—as compared to other functional clustering methods—are that *k*-centres FC (a) allows each cluster to have its own functional form, (b) does not rely on distributional assumptions, and (c) allows both the means and the shapes of the functions (i.e., the eigenfunctions) to differ between clusters (Chiou & Li, 2007). Of particular importance is that *k*-centres FC also works well in case there are no between-cluster differences in the eigenfunctions (see Chiou & Li, 2007), making it an appropriate choice for modeling a wide variety of intensive longitudinal data.

Does this mean that *k*-centres FC should always be used? Adhering to the principle of model parsimony, we strongly believe that less complex models should be preferred above more complex ones when they fit the data equally well. The problem is, of course, that this information is unknown to the researcher before doing the analysis. In these situations, however, *k*-centres FC can be a useful tool for data exploration. For example, when the results of a *k*-centres FC analysis show that the first eigenfunction of each cluster is nearly linear in time and explains a considerable amount of the total variance of the data (e.g., >80%), a latent class linear growth model might be appropriate for the data. Thus, when used in an exploratory fashion, the outcome of *k*-centres FC might suggest viable parametric models that are more parsimonious for the data at hand (Wang et al., 2016).

k-centres FC as a Tool for Inductive Research

We believe that *k*-centres FC offers a lot of opportunities for organizational researchers. Of particular importance is its ability to build typologies of change over time by identifying clusters of change (Dass & Shropshire, 2012). For example, *k*-centres FC provides researchers with the tools to test and extend recent dynamic theories such as Tomprou and colleagues' (2015) post-violation theory on the recovery from psychological contract breach or Lynall, Golden, and Hillman's (2003) theory on the evolution of board composition over time. Moreover, it allows scientists to study in a more exploratory way how, for example, commitment trajectories evolve over time (Meyer & Morin, 2016) or careers and career success unfold over the life span (Judge, Klinger, & Simon, 2010). In other words, beyond greater accuracy of description, we believe that *k*-centres FC is a useful tool for inductive theorizing based on the data. The state of temporal-dynamic research in the management and organizational sciences is such that there are only few theories available that would be informative about predicting particular types of temporal change in employees (e.g., predicting a cyclical movement or a peak or a dip in performance or well-being followed by a period of restoration; Langley, Smallman, Tsoukas, & Van de Ven, 2013). In a situation where suitable theory is scant, sticking to the straightjacket of deductive logic (i.e., having a *theory*, assuming a general *rule*, and then inferring whether it applies to the present *data*) is restrictive (Hambrick, 2007; Ketokivi & Mantere, 2010; Locke, 2007). Instead, *k*-centres FC allows a researcher more leeway to discover new (and temporal-dynamic) aspects of phenomena. Particularly the cluster-specific eigenfunction decomposition sets *k*-centres FC apart from other methods as it provides insights in the processes driving change by revealing functional (trajectory) characteristics that describe the major forms of

variability in the repeated measurements data. This property of *k*-centres FC is key to inductive reasoning and has the potential to help develop new theories about the subject matter.

Inductive reasoning is more easily said than done, though, as it involves an accurate representation of the *data* (i.e., a *k*-centres FC analysis), a possible *explanation* (e.g., the level of commitment prior to breach), and then making inferences about a *rule* (e.g., other cases where one might find similar findings). However, in full absence of theory (which is not uncommon when studying nonlinear phenomena in the management and organizational sciences), induction is not possible (for lack of an explanation), and one should revert to abduction instead. In abduction, where one has the data (a *k*-centres FC analysis), the observation of a *rule* (e.g., a pattern occurs in 55% of the cases), or a conspicuous anomaly (e.g., commitment increasing immediately after breach), the researcher is then tasked to come up with an *explanation* (Mantere & Ketokivi, 2013).

We have two suggestions that an inductive usage of *k*-centres FC could focus on. First, if a researcher encounters an interesting, unanticipated pattern characteristic with k-centres FC, he or she could ask the question "What is this a case of?" (Langley et al., 2013), which will most likely involve a combination of inductive and abductive thinking. For instance, we found two clusters of individuals in terms of their reactions following a breach event. Following such an observation, inductive reasoning could follow two distinct strategies: contextual or theoretical induction (Ketokivi & Mantere, 2010). In contextual strategies of induction, we would look for particular clues in the research context that might explain why we found our two subpopulations. For instance, we would have to look for clues for induction in the specific context of job changes in the realm academia (our sample). Alternatively, we might opt for theory-based induction, assuming that observing two subpopulations in reactions to breach events is not a one-off event, neither the result of a particular sample setting but rather a *theoretically* meaningful observation. Any observation gains substantive meaning if it can be coupled with a *theoretical* understanding of the subject matter (involving abductive thinking from our part). In our case, it would make sense to combine contextual with theoretical strategies for induction. For instance, one could make a case that the dearth of long-term employment contracts for young academics, and especially for those with unfavorable levels of commitment and match (Cluster 1), there is no "safe" middle ground (no plateauing at lower levels); one either has to completely de-commit and look for employment elsewhere or recover quickly and get on with relevant tasks in the quest for a more permanent work relationship. It follows that recovering from breach is largely a binary decision (i.e., Do I stay or do I go?) that will be decided by either problem-focused ways of coping (staying) or behavioral disengagement (leaving). In contrast, for those with higher levels of match and commitment (Cluster 2), recovery from breach involves more complex resolution processes that can be classified in the category of *relationship* repair. In relationship repair, breach resolution is not only an individual matter but also conditional on the response of the other party. Both parties in the relationship have to show a willingness to repair the wrongdoing (Tomprou et al., 2015). For instance, a complete dissolution of the psychological contract is not the first thing that comes to mind; plateauing at higher or lower levels of commitment seems to be the rule. However, this releveling can be disturbed at a later moment, such as when it becomes clear that the organization is unwilling to repair the wrongdoing (a late decline in commitment) or when the organization shows unexpected signs of post-breach support (a late increase in commitment).

A second suggestion for the inductive use of *k*-centres FC results would be to look for *counterfactuals*, that is, findings that defy one's set of theoretical assumptions but that are nevertheless interesting (Cornelissen & Durand, 2014; Davis, 1971). The reason to look for counterfactuals is that deductive logic has the pernicious problem that it cannot lead to new theory by itself (Ketokivi & Mantere, 2010; Locke, 2007; Mantere & Ketokivi, 2013). For instance, many longitudinal studies on the OB constructs in the management and organizational sciences are used simply to confirm preexisting ideas (more data but not more understanding; cf. Alvesson & Sandberg, 2013). More

imaginative and innovative research could come from interesting counterfactuals in the data, which is followed by a problematization of assumptions from the part of the researcher and the presentation of an alternative (Alvesson & Sandberg, 2011; Cornelissen & Durand, 2014). Several counterfactuals came out of the *k*-centres FC analysis in the present paper. For instance, we found that commitment could end up at a *higher* level compared the initial level, which clearly defies variance-based studies showing a consistent negative effect of breach (Zhao, Wayne, Glibkowski, & Bravo, 2007). Our data showed that there are in fact two explanations, or routes, by which one could arrive at a higher level: One is bouncing back after a short period of inertia, and the other is quick, immediate recovery to the initial level, followed by a positive delayed response. All in all, *k*-centres FC proves very useful for the use of inductive reasoning when making sense of research findings, which in turn has the potential of making our research imaginative, innovative, and theoretically important.

Conclusion

Given the technological advances to track human affect, cognitions, and behavior in real time, existing statistical methods for the analysis of change—which were developed for the modeling of only a few waves of measurement—are starting to show their limitations. Their main limitations are the fact that many of these models are constrained by the use of polynomials, which are limited in their ability to model true change over time. Moreover, these models obscure the fact that there may be meaningfully different subpopulations in the data. *k*-centres FC is a technique that circumvents these limitations and (as we have shown) is quite flexible in representing complex nonlinear data of different subpopulations. Our example shows that *k*-centres FC has great potential in exposing substantively important aspects of nonlinear data, in our case, the dynamic reactions to psychological contract breach incidents. This application of *k*-centres has yielded three insights that are of critical substantive importance: First, it has shown that there are different subpopulations that can be meaningfully interpreted, revealing two ways in which a person's pattern of development associates with the way a person reacts to a (negative) event. Second, it has shown that some people may do so by mixing two rather distinct reaction logics: immediate and delayed reactions to breach. Third, *k*-centres FC allows for the prediction of both cluster memberships and specific trajectory dimensions *within* these clusters. For instance, it allowed us to do separate tests for immediate versus delayed reactions to a breach event. Such knowledge is of tremendous substantive value for those interested in building and testing temporal theory on the processes governing the recovery after a breach incidence (cf. Tomprou et al., 2015). Beyond the study of commitment and the psychological contract, such analyses might boost temporal theorizing in the area of nonlinear dynamics of recovery after work, motivation and goal striving, emotion regulation, socialization patterns, reactions to organizational change, performance dynamics, team work, follower reactions to leaders, and so on as in all of these areas there is still much to be learned and discovered. Thanks to an unparalleled level of detail in the description of temporal characteristics, *k*-centres FC is uniquely equipped to support inductive theorizing in order to unravel new temporal characteristics of these phenomena.

Appendix

Matlab Code for k*-centres Functional Clustering*

To perform *k*-centres functional clustering (FC) in Matlab, one needs to download the kCFC package. This package, together with several other functional data analysis packages, can be downloaded from the following webpage: http://www.stat.ucdavis.edu/PACE/.

To run the analysis, the data need to be stored in a Matlab cell array. A Matlab cell array contains indexed data containers called cells. In the context of kCFC, two numeric 1 × n cell arrays are needed; one 1 × n numeric cell array y{i} with the i cells containing the vectors of measurements for the i=1, . . . , n subjects, and a corresponding 1 × n numeric cell array t{i} containing the vectors of measurement moments. In our example, two 1 × 109 cell arrays were created with cell y{i} containing the vector of commitment scores for subject i, and cell t{i} containing the vector of weeks in which these commitment scores were measured. Note that the length of the vectors can differ between cells (i.e., different subjects can have a different number of observations), whereas the length of y{i} and t{i} should evidently be equal.

Assuming the data matrix (Data) is in long format (i.e., one row per measurement moment per subject) with the first column being the subject indicator, the second column containing the measurements, and the third containing the measurement moments, one can transform this data matrix into two cell arrays y and t using the following Matlab code:

```
subjects=unique(Data(:,1));
for i=1: length(subjects)
    index=find(Data(:,1)==subjects(i));
    y{i}=Data(index,2)';
    t{i}=Data(index,3)';
end
```

The kCFC package only accepts numerical values. Thus, NAs need to be removed from the data matrix before performing the analysis. Note that this does not mean that kCFC cannot deal with missing data. In fact, the conditional expectation (CE) method is specifically developed for sparse, irregularly spaced longitudinal data (see Yao, Müller, & Wang, 2005a).

```
%%% STEP 1: PRE-PROCESSING OF THE DATA%%%
% the following code can be used to person-center the data
for i=1: length(y)
    y{i}=y{i}-mean(y{i});
end

%%% STEP 2: RUN THE K-CENTRES FC ANALYSIS %%%
% specify the directories containing the program files
cd('/Users/PACE/PACE-kcfc') % this should be the folder containing
the kCFC.m file
addpath ('/Users/PACE/PACE-kcfc/kcfc') % this allows adding other
folders containing program files

% STEP 2A: SPECIFY HOW MANY CLUSTERS TO RETAIN
clustopt.nc = 2;

% STEP 2B: SELECTION OF THE NUMBER OF PRINCIPAL COMPONENTS
clustopt.P = setopts('selection_k','BIC1'); % for the initial
clustering step
clustopt.M=setopts('selection_k','BIC1');%forthereclusteringstep

% STEP 2C: SPECIFY THE METHOD FOR ESTIMATING THE FPCA SCORES
method = 'CE';
```

```
% STEP 2D: SPECIFY THE STOPPING RULE FOR THE K-CENTRES FC ALGORITHM
clustopt.moveprop = 0; % stop if the number of moving curves is less
than moveprop*n

% STEP 2E: SPECIFY WHETHER THE LEAVE-ONE-OUT TECHNIQUE SHOULD BE USED
clustopt.op_cv = 1; % = 0, perform kcfc without; = 1, perform kcfc
with leave-one-curve-out

% STEP 2E: SPECIFY THE MAXIMUM NUMBER OF ITERATIONS
clustopt.iter = 50;

% ADDITIONAL SETTINGS
clustopt.M.newdata = unique(cell2mat(t)); % specify at which time
points the curves are evaluated
clustopt.seed = 232316; % set seed for kmeans algorithm that is used
in the initial clustering
clustopt.initD = 'correlation'; % distance measure for kmeans
algorithm used in the initial clustering step: 'sqEuclidean',
'cityblock', 'cosine', 'correlation', 'Hamming'
clustopt.P = setopts ('regular',0); % tell that the data are
irregular (sparse) -- initial clustering
clustopt.M = setopts ('regular',0); % tell that the data are
irregular (sparse) -- reclustering

% PERFORM THE CLUSTERING USING THE KCFC PACKAGE
[nc_kcfc, initCluster, reCluster, newCluster, idselect, SSE,
groupM, groupSSE] = kCFC(y, t, clustopt);
```

%%% STEP 3: INTERPRET THE CLUSTER SOLUTION %%%

```
%STEP 3A: PLOT THE CLUSTER-SPECIFIC MEANS AND MODES OF VARIATION
% The following code performs a FPCA per cluster and saves the
output
param_X = setOptions('selection_k','BIC1');
unique_cl=unique(newCluster);

for i = 1: length(unique_cl)
    idx{i} = find(newCluster == unique_cl(i));
end

for i = 1: length(unique_cl)
    for j = 1: length(idx{i})
        y_c{i}{j}=y{idx{i}(j)};
        t_c{i}{j}=t{idx{i}(j)};
    end
end

for i = 1: length(unique_cl)
    [X] = FPCA(y_c{i}, t_c{i}, param_X);
    scores_c{i}=getVal(X,'xi_est');
    phi_c{i}=getVal(X,'phi');
    out1_c{i}=getVal(X,'out1');mu_c{i}=getVal(X,'mu');
    est_c{i}=getVal(X,'y_pred');
end
```

```
% The following code computes low (-1 SD) and high (+ 1 SD) scores on
all principal components per cluster
for i = 1: length(unique_cl)
    for j=1: size(scores_c{i},2)
        mu_high{i}{j}=mu_c{i}+std(scores_c{i}(:,j))*phi_c{i}(:,j)';
        mu_low{i}{j}=mu_c{i}-std(scores_c{i}(:, j))*phi_c{i}(:, j)';
    end
end

% The following code creates the plots of the means and modes of
variation per cluster
[nrows, ncols]=cellfun(@size, scores_c);
for i = 1: length(unique_cl)
    for j=1: size(scores_c{i},2)
        subplot(length(unique_cl),max(ncols),((max(ncols)*(i-1))+j))
        plot(out1_c{i}, mu_c{i},'k')
        hold
        plot(out1_c{i}, mu_high{i}{j},'k--')
        plot(out1_c{i}, mu_low{i}{j},'k+')
        xlabel('time after breach (in weeks)','FontSize',18)
        ylabel('commitment (person-centered)','FontSize',18)
        axis([0 10 -15 15])
    end
end

%STEP 3B: PLOT RECONSTRUCTED CURVES
% The following code creates a plot of the reconstructed curves per
cluster
for i = 1: length(unique_cl)
    for j = 1: length(idx{i})
        t_est_c1{i}{j}=out1_c{i};
    end
end

for i = 1: length(unique_cl)
    subplot(1, length(unique_cl), i)
    createSpaghettiPlot(est_c{i}, t_est_c1{i})
    axis([0 10 -40 40])
    xlabel('time after breach (in weeks)','FontSize',20)
    ylabel('commitment (person-centered)','FontSize',20)
end
```

%%% STEP 4: ADD COVARIATES %%%

```
%STEP 4A: RELATING CLUSTER MEMBERSHIP TO THE COVARIATES
% One can obtain the cluster memberships by asking for the
newCluster vector
newCluster
```

```
%STEP 4B: RELATING CLUSTER-SPECIFIC FPCA SCORES TO THE COVARIATES
% One can obtain the cluster-specific fPCA scores by asking for the
scores_c cell object
scores_c{1} % for cluster 1
scores_c{2} % for cluster 2
```

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. Smoothness of a function means that "a pair of adjacent data values, y_j and y_{j+1} are necessarily linked together to some extent and unlikely to be too different from each other" (Ramsay & Silverman, 2005, p. 38).

References

Abraham, C., Cornillon, P. A., Matzner-Lober, E., & Molinari, N. (2003). Unsupervised curve clustering using b-splines. *Scandinavian Journal of Statistics. Theory and Applications*, *30*, 581-595.

Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., & Vakil, E. (2008). A shift in task routines during the learning of a motor skill: Group-averaged data may mask critical phases in the individuals' acquisition of skilled performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1544-1551.

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*, 247-271.

Alvesson, M., & Sandberg, J. (2013). Has management studies lost its way? Ideas for more imaginative and innovative research. *Journal of Management Studies*, *50*, 128-152.

Bal, P. M., Chiaburu, D., & Jansen, P. G. W. (2010). Contract breach and work performance: Is social exchange a buffer or intensifier? *Journal of Managerial Psychology*, *25*, 252-272.

Bal, P. M., Hofmans, J., & Polat, T. (2017). Breaking psychological contracts with the burden of workload: A weekly study of job resources as moderators. *Applied Psychology*, *66*, 143-167.

Bankins, S. (2015). A process perspective on psychological contract change: Making sense of, and repairing, psychological contract breach and violation through employee coping actions. *Journal of Organizational Behavior*, *36*, 1071-1095.

Berlin, K. S., Parra, G. R., & Williams, N. A. (2014). An Introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth and growth mixture models. *Journal of Pediatric Psychology*, *39*, *188-203*.

Brockner, J., Tyler, T. R., & Cooper-Schneider, R. (1992). The influence of prior commitment to an institution on reactions to perceived unfairness: The higher they are, the harder they fall. *Administrative Science Quarterly*, *37*, 241-261.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*, *62*, 23-43.

Chiou, J. M., & Li, P. L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*, 679-699.

Coffey, N., Hinde, J., & Holian, E. (2014). Clustering longitudinal profiles using P-splines and mixed effects models applied to time-course gene expression data. *Computational Statistics & Data Analysis*, *71*, 14-29.

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.

Cornelissen, J. P., & Durand, R. (2014). Moving forward: Developing theoretical contributions in management studies. *Journal of Management Studies*, *51*, 995-1022.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121-136.

Dass, M., & Shropshire, C. (2012). Introducing functional data analysis to managerial science. *Organizational Research Methods*, *15*, 693-721.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. New York, NY: Chapman & Hall.

Davis, M. S. (1971). That's interesting! *Towards a phenomenology of sociology and a sociology of phenomenology.* Philosophy of the Social Sciences, *1*, 309-344.

Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass*, *3*, 979-991.

Fok, C. C. T., & Ramsay, J. O. (2006). Fitting curves with periodic and nonperiodic trends and their interactions with intensive longitudinal data. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 109-123). New York, NY: Oxford University Press.

Ghisletta, P., Cantoni, E., & Jacot, N. (2015). Nonlinear growth curve models. In M. Stemmler, A. von Eye, & W. Wiedermann (Eds.), *Dependent data in social sciences research* (pp. 47-66). New York, NY: Springer.

Graves, S., Hooker, G., & Ramsay, J. (2009) *Functional data analysis with R and MATLAB*. New York, NY: Springer.

Griep, Y., Vantilborgh, T., Baillien, E., & Pepermans, R. (2016). The mitigating role of leader-member exchange in reaction to psychological contract violation: A diary study among volunteers. *European Journal of Work and Organizational Psychology*, *25*, 254-271.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357-1371.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*, 1346-1352.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ieva, F., Paganoni, A. M., Pigoli, D., & Vitelli, V. (2012). Multivariate functional clustering for the morphological analysis of ECG Curves . *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*, 401-418.

Ilies, R., & Judge, T. A. (2002). Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes*, *89*, 1119-1139.

Jacques, J., & Preda, C. (2014a). Functional data clustering: A survey. *Advances in Data Analysis and Classification*, *8*, 231-255.

Jacques, J., & Preda, C. (2014b). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, *71*, 92-106.

James, G., & Sugar, C. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, *98*, 397-408.

Jank, W., & Shmueli, G. (2006). Functional data analysis in electronic commerce research. *Statistical Science*, *21*, 155-166.

Judge, T. A., Klinger, R. L., & Simon, L. S. (2010). Time is on my side: Time, general mental ability, human capital, and extrinsic career success. *Journal of Applied Psychology*, *95*, 92-107.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302-317.

Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review*, *35*, 315-333.

Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, *56*, 1-13.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill Palmer Quarterly*, *52*, 377-389.

Levitin, D. J., Nuzzo, R. L., Vines, B. W., & Ramsay, J. O. (2007). Introduction to functional data analysis. *Canadian Psychology*, *48*, 135-155.

Li, J., & Roe, R. A. (2012). Introducing an intrateam longitudinal approach to the study of team process dynamics. *European Journal of Work and Organizational Psychology*, *21*, 718-748.

Li, P.-L., & Chiou, J.-M. (2011). Identifying cluster numbers for subspace projected functional data clustering. *Computational Statistics and Data Analysis*, *55*, 2090-2103.

Locke, E. (2007). The case for inductive theory building. *Journal of Management*, *33*, 867-890.

Lynall, M. D., Golden, B. R., & Hillman, A. J. (2003). Board composition from adolescence to maturity: A multitheoretic view. *Academy of Management Review*, *28*, 418-431.

Mantere, S., & Ketokivi, M. (2013). Reasoning in organization science. *Academy of Management Review*, *38*, 70-89.

Marsh, H. W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, *32*, 151-171.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, *16*, 163-200.

Mehl, M. R., & Conner, T. S. (2012). *(Eds.).* Handbook of research methods for studying daily life. London: The Guilford Press.

Meyer, J. P., & Morin, A. J. S. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*, *37*, 584-612.

Morin, A., Mainano, C., Nagengast, B., Marsh, H., Morizot, J., & Janosc, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling*, *18*, 613-648.

Morrison, E. W., & Robinson, S. L. (1997). When employees feel betrayed: A model of how psychological contract violation develops. *Academy of Management Review*, *22*, 226-256.

Park, C. L., & Folkman, S. (1997). Meaning in the context of stress and coping. *Review of General Psychology*, *1*, 115-144.

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31*, 303-316.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York, NY: Springer.

Serban, N., & Wasserman, L. (2005). CATS: Cluster analysis by transformation and smoothing. *Journal of the American Statistical Association*, *100*, 990-999.

Skinner, E. A., Edge, K., Altman, J., & Sherwood, H. (2003). Searching for the structure of coping: A review and critique of category systems for classifying ways of coping. *Psychological Bulletin*, *129*, 126-169.

Solinger, O. N., Hofmans, J., Bal, P. M., & Jansen, P. G. W. (2016). Bouncing back from psychological contract breach: How commitment recovers over time. *Journal of Organizational Behavior*, *37*, 494-514.

Solinger, O. N., Van Olffen, W., Roe, R. A., & Hofmans, J. (2013). On becoming (un)committed: A taxonomy and test of newcomer on-boarding scenarios. *Organization Science*, *24*, 1640-1661.

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Methodological issues in the application of the latent growth curve model. In K. van Montfoort, A. Satorra, & H. Oud (Eds.), *Recent developments in structural equation models* (pp. 241-261). Amsterdam: Kluwer.

Tomprou, M., Rousseau, D. M., & Hansen, S. D. (2015). The psychological contracts of violation victims: A post-violation model. *Journal of Organizational Behavior*, *36*, 561-581.

Wang, J. L., Chiou, J., & Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*, 257-295.

Weiss, H. M., & Rupp, D. E. (2011). Experiencing work; An essay on a person-centric work psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *4*, 83-97.

Wood, P. K., & Jackson, K. M. (2013). Escaping the snare of chronological growth and launching a free curve alternative: General deviance as latent growth model. *Development and Psychopathology*, *25*, 739-754.

Yao, F., Müller, H. G., & Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*, 577-590.

Yao, F., Müller, H. G., & Wang, J. L. (2005b). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, *33*, 2873-2903.

Zhao, H. A. O., Wayne, S. J., Glibkowski, B. C., & Bravo, J. (2007). The impact of psychological contract breach on work-related outcomes: a meta-analysis. *Personnel Psychology*, *60*, 647-680.

Author Biographies

Joeri Hofmans is an associate professor of organizational behavior in the Department of Experimental and Applied Psychology at the Vrije Universiteit Brussel. His research focuses on the role of personality, emotions, and motivation at work, with particular emphasis on their temporal dynamics.

Tim Vantilborgh is an assistant professor of research methods in the Department of Experimental and Applied Psychology at the Vrije Universiteit Brussel. His research focuses on processes underlying the development of and reactions to psychological contracts. Other interests are personality, emotions, stress at work, and temporal dynamics in organizational behavior.

Omar N. Solinger is an assistant professor of organizational behavior in the Department of Business Administration at the VU University Amsterdam. He is broadly interested in developing theory about change in organizational behavior–related concepts. Current interests involve the impact of shocks and involuntary transitions (e.g., planned change) and leadership (e.g., charisma).