Received: 1 December 2016 | Revised: 13 December 2017 | Accepted: 19 December 2017

DOI: 10.1111/peps.12260

WILEY PERSONNEL PSYCHOLOGY

ORIGINAL ARTICLE

Modeling consensus emergence in groups using longitudinal multilevel methods

Jonas W. B. Lang[1] | **Paul D. Bliese**[2] | **Alex de Voogt**[3]

[1]Ghent University

[2]University of South Carolina

[3]Drew University

Correspondence
Jonas W. B. Lang, Department of Personnel Management, Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.
Email: jonas.lang@ugent.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: G019217N

Abstract

Organizational researchers have long been interested in studying bottom-up multilevel processes where lower level units (e.g., employees) in organizations interact to jointly create characteristics of higher level units (e.g., work groups). This article contributes to the literature on bottom-up processes by detailing a statistical approach—the consensus emergence model (CEM)—that allows researchers to study emergence of shared perceptions and feelings or climates in groups over time. The described methodological approach extends standard multilevel methodology by examining residual variances within a growth model to account for dynamic change in group consensus. The CEM provides a formal test for consensus emergence. The approach also allows researchers to test explanatory models of consensus emergence by including person-level, group-level, and observation-level predictors. We illustrate the CEM by applying the method to data from two longitudinal studies of work units. The first study investigated job satisfaction in military companies. Our second study examined professional archeologists working in groups on a field excavation mission and focused on fatigue at the end of the work day. Our analyses demonstrate the CEM's ability to detect and study emergence, and suggest that the CEM may be a valuable tool to help extend the study of emergence in organizational research.

1 | MODELING CONSENSUS EMERGENCE IN GROUPS USING LONGITUDINAL MULTILEVEL METHODS

In bottom-up multilevel processes, lower level units in organizations (e.g., group members) interact to jointly create characteristics of higher level units (Klein & Kozlowski, 2000; Kozlowski & Chao, 2012; Kozlowski, Chao, Grand, Braun, & Kuljanin, 2013). Bottom-up or emergent phenomena are central for understanding social interactions in organizations and form the basis for numerous organizational theories and empirical studies (Cronin & Weingart, 2011; Humphrey & Aime, 2014; Kozlowski et al., 2013). The research literature has identified a variety of bottom-up processes (Chan, 1998; Kozlowski & Klein, 2000); however, much of the emergence literature centers on the idea of consensus. That is, multilevel studies often assume that group members develop shared perceptions or climates over time and that the likelihood of forming shared perceptions is partially a function of attributes of the work context (Ashforth, 1985; Kostopoulos, Spanos, & Prastacos, 2011; Ployhart & Moliterno, 2011). Several examples (see also Kozlowski et al., 2013) include group-consensus emergence in leader–member exchange (LMX) quality (e.g., Schyns & Day, 2010), group conflict (Jehn & Mannix, 2001; Tekleab, Quigley, & Tesluk, 2009), and affect "contagion" (Barsade & Knight, 2015; Kelly & Barsade, 2001). In personnel psychology, similar ideas revolve around concepts such as collective attitudes toward turnover (Felps et al., 2009; Hausknecht, 2017; Hausknecht & Trevor, 2011; Nyberg & Ployhart, 2013) and employee perceptions of human resources management practices (e.g., Bowen & Ostroff, 2004).

The notions of bottom-up processes, emergence, and consensus are fundamentally longitudinal involving patterns of change within groups over time. Importantly, however, empirical research rarely investigates how consensus develops (Cronin & Weingart, 2011; Humphrey & Aime, 2014; Kozlowski et al., 2013). In other words, researchers rarely ask questions such as "Do group members' feelings and perceptions become more similar or less similar over time?"

One potential reason for the lack of temporally focused studies on bottom-up processes is that the methodological tools organizational researchers routinely use for studying multilevel phenomena have not yet been fully extended to capture temporal changes in group consensus (Kozlowski et al., 2013). For instance, the intraclass correlation, type 1 (ICC1) from a multilevel model (also known as mixed-effects models, random coefficient models, and hierarchical linear models) allows researchers to estimate whether members of organizational units are more similar to each other than to members from other work units at discrete snapshots in time. Unfortunately, the discrete snapshots provide incomplete insight into temporally based emergent processes (Kozlowski et al., 2013). As we detail later, change in ICC1 values over time can result from either groups becoming less similar from each other or from individuals within groups becoming more similar. As such, a pattern of increasing ICC1 values at each time in a construct such as climate could either mean that the employees within the groups were developing shared climates, that groups were become increasingly divergent, or a combination of both processes.

We believe organizational theory and research would benefit from being able to (a) formally test and quantify whether consensus in groups emerges over a period in time, (b) understand how group member characteristics (e.g., being in a leadership role) relate to emergent tendencies within groups (Cronin & Weingart, 2011; Humphrey & Aime, 2014; Morgeson & Hofmann, 1999), and (c) study how unit characteristics (e.g., baseline collective efficacy) predict consensus emergence over time. In this article, we detail a statistical approach that expands researchers' ability to study consensus emergence over time within work units. We define consensus emergence as a pattern of increased similarity among unit-member perceptions and describe an extended multilevel model—the consensus emergence model (CEM)—that allows researchers to test for, quantify, and understand the nature of consensus emergence.

2 | EXTANT RESEARCH METHODS AND APPROACHES FOR STUDYING CONSENSUS EMERGENCE

In this section, we review existing approaches for studying consensus emergence in organizational units. We begin by briefly describing the ICC1 and other existing multilevel-based indices and how these indices are currently used in organizational research to examine consensus emergence. We then review approaches utilizing tools other than multilevel modeling and discuss how the CEM approach can complement and extend these existing approaches.

2.1 | Multilevel-based indices

As noted, the ICC1 can be estimated using a basic multilevel model. This multilevel model describes the response Y_{jk} for person j in group k as a function of a common intercept γ_{00}, the group-specific deviation from the intercept u_{0k}, and

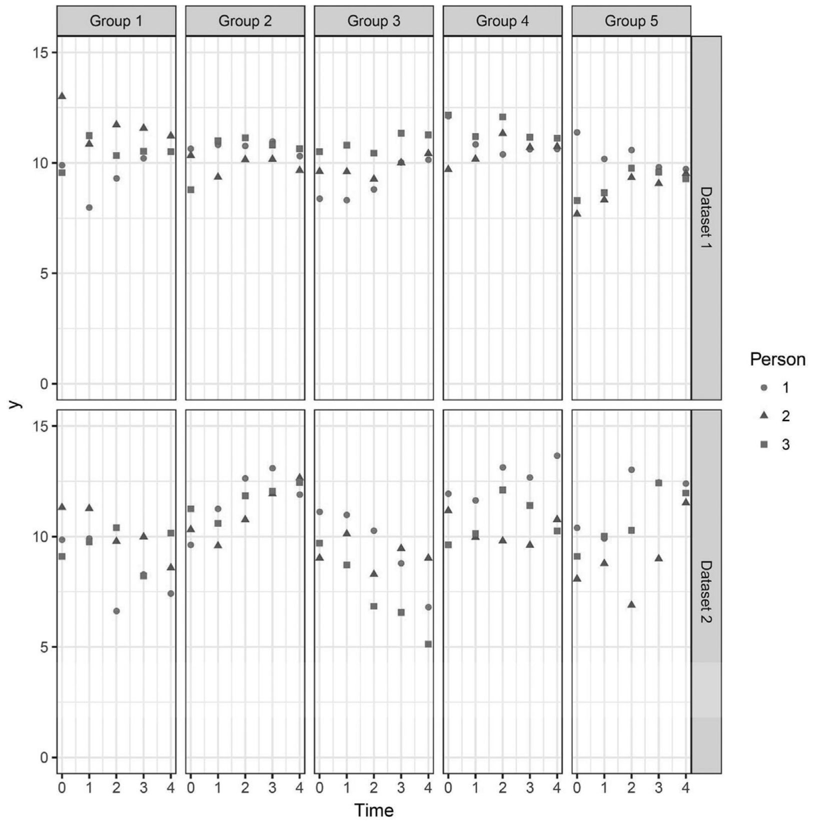

FIGURE 1 The graphs show two hypothetical data sets

the residual error e_{jk}.

$$Level\text{-}1: Y_{jk} = \beta_{0k} + e_{jk} \quad \text{where } e_{jk} \overset{iid}{\sim} N\left(0, \sigma_e^2\right), \tag{1}$$

$$Level\text{-}2: \beta_{0k} = \gamma_{00} + u_{0k} \quad \text{where } u_{0k} \overset{iid}{\sim} N\left(0, \sigma_{\beta 0}^2\right). \tag{2}$$

The group deviation term captures the group's deviation from the intercept, and the residual error describes the deviation of each observation from the intercept and the group deviation. The amount of variance in both terms the—group variance $\sigma_{\beta 0}^2$ and the residual variance σ_e^2—are used to estimate the ICC1 using the formula ICC1 = $\sigma_{\beta 0}^2$ / ($\sigma_{\beta 0}^2$ + σ_e^2). The ICC1 specifies the amount of variance that group membership explains in the overall variance. High ICC1 values consequently indicate that the members of a group are more similar to each other than they are to members of other groups in the data set.

The ICC1 allows researchers to quantify the degree of similarity among group members. Importantly, though, the index is limited when attempting to examine temporal changes associated with bottom-up processes (Cronin & Weingart, 2011; Humphrey & Aime, 2014; Kozlowski et al., 2013). As we noted in the beginning of our article, one major limitation of the ICC1 is that increases in ICC1 values can result from either groups becoming less similar over time (an increase in $\sigma_{\beta 0}^2$) or from individuals within groups becoming more similar over time (a decrease in σ_e^2). Figure 1 demonstrates this phenomenon using hypothetical data. The first five groups in the upper panel show patterns that support the theoretical phenomenon of emergence. In all five groups, group members' perceptions become increasingly similar. In contrast, the second data set in the lower panels show patterns that do not reflect consensus emergence. In the lower panels, no notable trend of increased similarity among group members is apparent.

Despite the obvious differences between the sets of groups, both data sets show a similar pattern of increasing ICC1 values over time. ICC1 values for data set 1 (upper panel) are .03, .00, .25, .40, and .58 from T1 to T5, and ICC1 values for data set 2 (lower panel) are .01, .00, .19, .54, and .72 from T1 to T5. In data set 1, decreases in σ_e^2 lead to an increase in ICC1 values over time. In contrast, in data set 2, increases in $\sigma_{\beta 0}^2$ lead to an increase in ICC1 values.

A second limitation of the ICC1 is the lack of an omnibus statistical test for interpreting ICC1 patterns across measurement occasions. That is, even if one could assume that an increase in ICC1 values was only associated with a decrease in σ_e^2, one would still be faced with trying to determine whether the overall pattern associated with specific values for each measurement occasion (e.g., .03, .00, .25, .40, .58) did or did not provide statistically significant evidence of emergence. For instance, our example ICC1 values from data set 1 appear to support a pattern of emergence, but ideally we would like a single test that summarizes the pattern as a whole. A third limitation with examining ICC1 values in a static manner is that within any applied data set, some groups are likely to show patterns consistent with emergence whereas other groups will not. Therefore, it would be valuable to test hypotheses about factors that potentially explain different emergence patterns. Unfortunately, when ICC1 values are separately estimated for each time period there is no clear way to include predictors of differential patterns among groups.

In short, patterns of change in ICC1 over time produce ambiguous indices of consensus emergence. Later, we return to this important point; however, the field generally recognizes that relying on ICC1 provides incomplete information about consensus emergence. Indeed, a recent review focused on changes in ICC1 values over time found limited support for the theoretical notion that changes in these values occur (Allen & O'Neill, 2015).

An intuitive way to address the fact that the ICC1 takes both the within-group variance and the between group variance into account would be to look at the average of the raw within-group variance or *SD* instead of the ICC1. Focusing on raw within-group variances would not fully capture a potential scenario in which the between-group variance is simultaneously decreasing; nevertheless, changes in within-group variance can provide insights about the within-group variance in isolation. In practice, however, it can be difficult to interpret a pattern of averaged within-group *SD* values over time because there is no statistical test and no clear effect size estimate.

An index that is closely related to the within-person *SD* is the r_{wg} index (James, Demaree, & Wolf, 1993) and several conceptually similar indices (Bliese, 2000; LeBreton & Senter, 2008). In theory, r_{wg} is a direct function of the within-group *SD* as the index compares the variability of response categories (s^2) to a theoretical null distribution (σ^2) for the respective Likert-type scale format ($r_{wg} = 1 - [s^2/\sigma^2]$). However, r_{wg} can substantially differ from the within-group variance because the observed within-group variances can exceed the expected random variance of the theoretical null distribution. In this case, r_{wg} becomes negative and is typically set to 0 for the group (LeBreton & Senter, 2008; LeBreton, James, & Lindell, 2005). This behavior of r_{wg} may be problematic when the index is used in longitudinal research as a potential emergence pattern may be affected by this truncation behavior. As with the ICC1, a researcher would also still lack an omnibus test and could not easily study potential explanations for emergence patterns. Given the described limitations of existing multilevel-based indices of emergence, researchers have suggested other methods to study change in social dynamics over time.

2.2 | Other approaches: Social network analysis, qualitative approaches, and computational modeling

One alternative to multilevel-based indices is social network analysis (Balkundi & Harrison, 2006; Burt, Kilduff, & Tasselli, 2013; Butts, 2008; Carter, DeChurch, Braun, & Contractor, 2015; Fowler & Christakis, 2008; Jones & Shah, 2016; Kenny & Judd, 1996). Social network analysis is a relational approach that estimates how closely one person in a larger sample or group is linked to the other members. The approach focuses on collecting information on connections (ties) between group or organization members (nodes). One possible method for establishing ties among others is to ask organizational members to rate a specific type of relation to each of the members in an organizational unit or larger environment using Likert-type scales. Information on social ties in social network analysis can be analyzed using a variety of different statistical methods. For instance, the strength of group members' ties can be summarized in indices of group centrality (Butts, 2008; Venkataramani, Zhou, Wang, Liao, & Shi, 2016), or changes in ties can be studied over time (Huisman & Snijders, 2003; Kalish & Luria, 2016).

A second approach for studying consensus emergence is based on qualitative research methods (Blee, 2013; Gehman, Trevino, & Garud, 2013; Kozlowski & Chao, 2012). Qualitative methods gather rich data from discussions, documents, and interviews, and focus on detecting evidence of change in climates and shared perceptions over time within these data. For instance, organizational researchers have used qualitative methods to study the emergence of shared ethical values in an educational organization over time (Gehman et al., 2013).

A third approach for studying phenomena related to emergence is computational modeling (Vancouver & Weinhardt, 2012). In computational modeling, a detailed model is built that captures the presumed underlying mechanisms and processes, and hypothetical data are simulated on the basis of this model. The core idea is that the computational model should generate simulation data that are plausible. Computational modeling thus provides an explicit plausibility check for formal theories. Computational modeling is well suited to exploring the emergence of complex patterns over time in a way that can then inform the design of research studies and the collection of primary research data (Kozlowski, 2017; Kozlowski et al., 2013).

The social network approach, the qualitative approach, and computational modeling all have the potential to help elucidate emergent phenomena, and the use of a variety of approaches can facilitate developing a rich understanding of emergence. At the same time, however, each alternative approach has some limitations. For instance, the qualitative approach allows researchers to make unique discoveries that may not be possible using quantitative methods, but the qualitative approach also puts rather heavy demands on researchers and respondents (Gephart, 2004). In contrast, the computation modeling approach, at its core, does not place heavy demands on respondents, but the confirmation of patterns suggested by computational modeling often requires additional empirical tests of the nature we describe (Vancouver, Tamanini, & Yoder, 2010). Finally, a defining characteristic of social network approaches is that these approaches require information on ties between each of the member in a network via a unique data structure.

Social network approaches are also based on a different theoretical foundation than the ICC1 and the models we describe in this paper. The ICC1 and most multilevel research on group-level constructs conceptualize constructs as latent variables (the random effects in multilevel mixed-effects models are latent variables, see, e.g., Skrondal & Rabe-Hesketh, 2007) and, thus, assume that group members are affected by an underlying latent climate in the group. Consensus emergence in this context is consequently the strengthening of the effects of the group climate on its members over time. In contrast, constructs in social network analysis are systems of causally connected members (Borsboom & Cramer, 2013). In this context, consensus emergence could be viewed as the strengthening of connections over time. Both approaches tap different types of constructs that are conceptually distinct and help to answer different types of theoretical questions. Ultimately, we recognize that each approach (to include the CEM) has some limitations; however, we demonstrate that the CEM represents an additional tool in large part due to the CEM's ability to utilize common data structures (e.g., repeated measures survey data).

3 | THE CEM

The conceptual rationale behind the CEM is to systematically model changes in residual variances over time in a way that provides insights into consensus emergence. Trends in residual variances have previously been discussed in organizational research (e.g., Bliese & Ployhart, 2002), and many statistical packages for multilevel analyses include relatively simple command options that allow for tests involving residual variances. For instance, Bliese and Ployhart (2002) have shown that the option "weights = varExp(form = ∼TIME)" in the lme function in R returns a model that allows residual variances to change over time. Similar options exist in other programs, so the described models can be fit in software packages like NLMIXED in SAS, and the multilevel structural equation modeling software Mplus (Muthén & Muthén, 2015). Conceptually similar models can also be estimated in programs like Stata, lme4 in R (Bates, Maechler, Bolker, & Walker, 2015), and MCMCglmm in R (Hadfield, 2010), so the approach we describe is not tied to any specific software package. Notably, organizational researchers have also described and used multilevel approaches that use variability as a predictor variable of other outcomes in longitudinal designs (Griffin, 1997; Stewart & Nandkeolyar, 2007). For instance, Stewart and Nandkeolyar (2007) described a longitudinal multilevel modeling approach in which they used variance in performance in previous years as a predictor for subsequent performance.

Where the CEM builds upon other treatments of residual variance (e.g., Bliese & Ployhart, 2002; Singer & Willett, 2003) and variability (Griffin, 1997; Stewart & Nandkeolyar, 2007) is in its focus on the residual variance as an outcome of substantive interest. The CEM departs from the idea that trends in residual variances represent potential confounds that can compromise inferences surrounding other model components. Instead, the CEM is based on the idea that residual variance patterns can answer relevant research questions under certain model specifications (Hoffman, 2007; Kim & Seltzer, 2011; Raudenbush & Bryk, 1987). Furthermore, what may be less well known is that variants of the command options in multilevel software provide the ability to add group-level and individual-level predictors of change in residual variances. In the CEM, the addition of residual variance predictors allows researchers to answer questions about why (a) emergence may be more pronounced in some groups than others and (b) certain types of individuals might have differential effects on emergence patterns.

Outside of organizational research areas, the notion that modeling residual variances can help researchers address questions of substantive interest has existed for some time (e.g., Raudenbush & Bryk, 1987). Fields including health research (e.g., Hedeker & Mermelstein, 2007; Hoffman, 2007), aging research (Rast, MacDonald, & Hofer, 2012), educational research (Raudenbush & Bryk, 1987), and emotion research (Kuppens & Yzerbyt, 2014) have used variance functions to study predictors of fluctuations in residual variance within persons and within groups. For instance, health researchers (Hedeker & Mermelstein, 2007; Hedeker, Mermelstein, & Demirtas, 2012) modeled residual variance in an experience sampling research design to study fluctuations in positive and negative affect in cigarette smokers. The researchers were interested not only in the general level of positive and negative affect but also in individual variation in affect because smokers presumably experience diminished affect when they have not smoked for an extended period of time and heightened affect immediately after smoking. An examination of residual variance patterns revealed that smokers showed more within-person variability than nonsmokers. As another example, in educational research, predictors of residual variance have been used to study differences between classrooms in terms of within-class heterogeneity in cross-sectional data (Raudenbush & Bryk, 1987). In the education context, one can argue that classrooms with more variability are less egalitarian than classrooms showing a pattern of deceased residual variance regardless of the average class level.

In considering how to adapt the models to study emergence, it is important to recognize that the focal level of analysis is the group rather than the individual. That is, groups demonstrate emergence patterns via responses of the individuals; therefore, at a conceptual level, consensus emergence represents a specific form of intragroup response patterns over time. Stated otherwise, in the context of emergence, individual members represent fluctuations around latent group means. Thus, the CEM model we describe represents a novel combination of a growth model for latent group means along with existing approaches for modeling residual variances over time that goes beyond existing approaches.

3.1 | Two-level CEM

Table 1 provides three basic CEM model specifications. The first CEM model shown in Table 1 is a two-level model and specifies that the response in group k at measurement occasion i is a function of a common intercept γ_{00}, one fixed-effects predictor (TIME$_i$), a group random effect u_{0k}, a group-specific random slope u_{1k}, and the residual e_{ik}. The variance of the residual e_{ik} changes as a function of TIME$_i$. TIME is typically coded 0 at the origin with a one point increases for each occasion (0, 1, 2, 3, etc.) but other coding schemes are possible.

As noted previously, one approach is to view the basic CEM as a combination of a growth model for latent group means with a variance function model that captures fluctuations around these latent group means and models change in these fluctuations over time. The underlying idea is to model two fundamental sources of variance change jointly. The first source is variance change resulting from changes in latent group means over time (group slope variability). Controlling for this source of variance removes the bias produced by group mean change that is problematic when examining ICC1 values over time. The second source of variance change is reflected in the residual variances. In the

TABLE 1 Consensus emergence model (CEM)

Model	Interpretation	Equation	Variance components
Two-level CEM	• δ_1 indicates consensus emergence • observations are influenced by group membership	Level 1: $Y_{ik} = \beta_{0k} + \beta_{1k}\mathrm{TIME}_i + e_{ik}$ Level 2 (group): $\beta_{0k} = \gamma_{00} + u_{0k}$ $\beta_{1k} = \gamma_{10} + u_{1k}$	$e_{ik} \overset{iid}{\sim} N(0, \sigma_e^2 \exp[2\delta_1 \mathrm{TIME}_i])$ $\begin{pmatrix} u_{0k} \\ u_{1k} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_1\beta_0} & \sigma^2_{\beta_1} \end{bmatrix}\right)$
Three-level CEM	• δ_1 indicates consensus emergence • observations are influenced by group membership • prior individual experiences influence the baseline level	Level 1: $Y_{ijk} = \pi_{0jk} + \pi_{1jk}\mathrm{TIME}_i + e_{ijk}$ Level 2 (person): $\pi_{0jk} = \beta_{00k} + r_{0jk}$ $\pi_{1jk} = \beta_{10k}$ Level 3 (group): $\beta_{00k} = \gamma_{000} + u_{00k}$ $\beta_{10k} = \gamma_{100} + u_{10k}$	$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2 \exp[2\delta_1 \mathrm{TIME}_i])$ $r_{0jk} \overset{iid}{\sim} N(0, \sigma^2_{\pi_0})$ $\begin{pmatrix} u_{00k} \\ u_{10k} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\beta_{00}} & \sigma_{\beta_{00}\beta_{10}} \\ \sigma_{\beta_{10}\beta_{00}} & \sigma^2_{\beta_{10}} \end{bmatrix}\right)$

Note: The variable TIME_i in the model specification omits a subscript for the person and the group because the model specification we present here assumes that the data are time structured; that is, there are fixed points in time at which the variables were measured. In situations in which the model is used with unstructured data and the time of measurement occasions vary across groups and individuals, it is necessary to add subscripts for the person and the group to the TIME_i variable (i.e., TIME_{ijk}).

CEM model, the residual variance is explicitly modeled over time and can be systematically predicted and tested. The residual variance in the CEM is therefore not only error variance but a model component of interest. In other words, the focus in the CEM model is to use the fixed and random effects components of the multilevel mixed-effects model as controls such that changes in the residual variance can be meaningfully interpreted. Although our focus in this article is on modeling consensus emergence as a novel contribution, it is important to recognize that interpreting mean change in latent group means over time (group slope variability) in the CEM is also of substantive interest. Changes in means provide insights into the degree to which groups shift the average of their consensus over time and a scenario where groups do not increase their consensus but shift the location of their consensus (the latent group means) provides relevant information for researchers.

As shown in Table 1, the CEM uses an exponential variance function with the coefficient δ_1 (Pinheiro & Bates, 2000) to model change in the residual variances over time.[1] The exponential variance function ensures that variances estimates for the residual variance σ_e^2 cannot be negative because the exponential function can never be smaller than 0 even when δ_1 reaches high negative values. The model also uses a constant of 2 in the exponential function to bring the delta to the scale of the *SD* of the dependent variable. Notice that $\exp(2 \times \delta_1 \times \mathrm{TIME})$ is mathematically identical to $\exp(\delta_1 \times \mathrm{TIME})^2$, and $\sigma_e^2 \exp(2\delta_1 \mathrm{TIME}_i)$ is identical to $(\sigma_e \times \exp[\delta_1 \times \mathrm{TIME}_i])^2$. The *SD* scale is easier to interpret than the variance scale because a linear increase in the predictor can be interpreted as an approximate linear increase on the *SD* scale. When TIME_i is coded 0 at the origin of time and increases by one with each measurement occasion (0, 1, 2, 3, etc.), the interpretation of δ_1 is relatively straightforward. Negative values of δ_1 can be interpreted as an approximately linear decrease in the amount of residual variance with each measurement occasion (and thus emergence); values at or near zero indicate residual stability over time, and positive values for δ_1 indicate an approximate linear increase (and thus increased residual variability over time). The δ_1 coefficients from the CEM provide a convenient measure of effect size as the percentage of residual *SD* reduction associated with each measurement occasion. For instance, a delta coefficient of –0.03 would mean that an residual variance *SD* of 2 (equivalent to a variance of 4) is reduced by approximately 3% with each measurement occasion, $\sqrt{2^2 \times \exp(2 \times -0.03)} = 1.94$. An option for researchers who want a time-based measure of effect size is to change the TIME variable in the model to this respective time frame. For instance, researchers could use months as the time frame so that TIME in the model would be coded 0, 6, and 12 for a person surveyed at the beginning, after 6 months, and after 12 months of a study. The δ_1 coefficient then approximately refers to the percentage change in the residual *SD* each month.

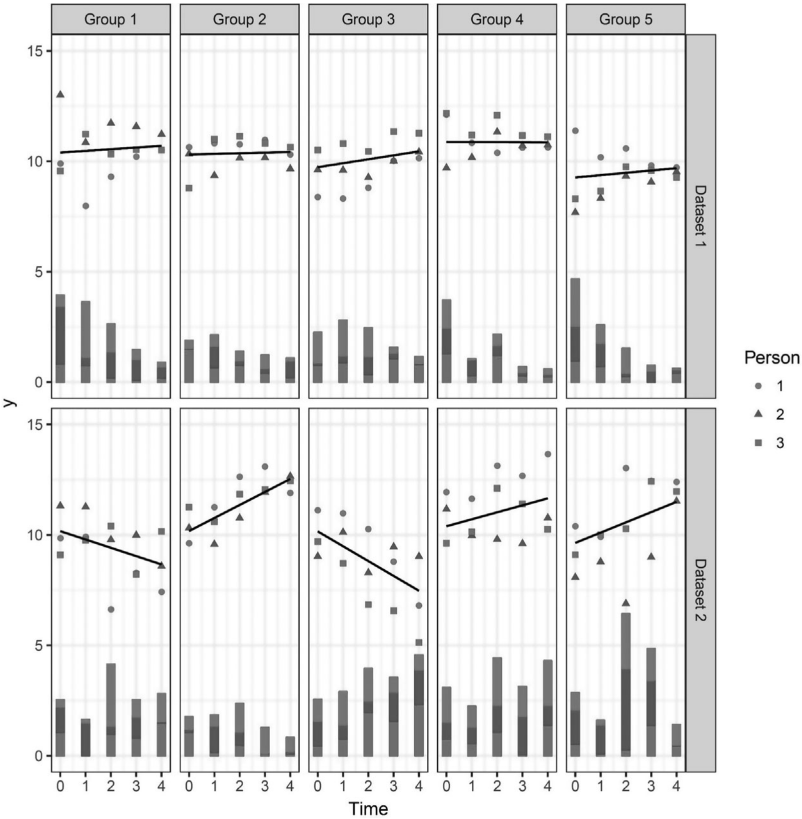

FIGURE 2 The graph shows the same hypothetical data as in Figure 1 and additionally includes the group trends from a two-level consensus emergence model (CEM). The bars in the bottom of the panels show the residuals (the overall amount of differences between the observed and the fitted values)

A key feature of the CEM is that the significance of change in the residual variance can conveniently be tested using log-likelihood ratio test (χ^2—difference test). This test compares the fit of the CEM to a null model that does not include residual variance change (conceptually analogous to a hierarchical *F*-test). Specifically, the null model constrains the residual variance σ_e^2 to be homogenous, $e_{ijk} \sim N(0, \sigma_e^2)$. This log-likelihood ratio test provides a formal overall test of the stability of the residual variance, which, in the case of a negative term, reflects an overall test of consensus emergence.

Figure 2 shows the same data sets as Figure 1 and applies the basic two-level CEM to these data sets. The panels in Figure 2 plot the observed data against the fitted values (lines) from the CEM along with residual histograms in the lower parts of the panels. Recall that the unit of analyses for emergent effects is the group. As a consequence, the group-level latent trends are shown and individual responses are deviations from the group mean pattern. In line with the intuitive interpretation of the data, the panels for data set 1 show that the residuals—the differences between the fitted values (the lines) from the CEM and the observed values (the points)—decrease over time and accordingly indicates that the groups develop consensus because the variance/size of the residuals declines. These differences are most pronounced for Group 1 and Group 5. In contrast, no similar pattern is apparent for data set 2.

The presence of the consensus emergence trend in data set 1 and the absence of a consensus emergence trend in data set 2 is also apparent in the numerical results of the CEM. CEM results for both data sets are provided in Table 2 and Table 3. Table 2 provides the omnibus likelihood test of whether an emergence trend exists. The contrast of M1a and M1b in data set 1 returns $\chi^2(df = 1) = 21.46, p < .001$, suggesting that the model that includes a variance function fits the data significantly better than a model that excludes this term.

TABLE 2 Consensus emergence model (CEM): Model comparisons in two example data sets

Data, model (M)	AIC	BIC	*logLik*	*df*	vs. previous model χ^2
Data set 1					
M1a: Two-level model, no emergence	219.55	233.30	−103.78	6	
M1b: Two-level CEM	200.09	216.12	−93.05	7	21.46**
M2a: Three-level model, no emergence	198.32	214.35	−92.16	7	
M2b: Three-level CEM	170.37	188.69	−77.18	8	29.95**
Data set 2					
M1a: Two-level model, no emergence	277.09	290.83	−132.5	7	
M1b: Two-level CEM	277.67	293.70	−131.8	8	1.43
M2a: Three-level model, no emergence	256.42	272.45	−121.21	7	
M2b: Three-level CEM	257.96	276.28	−120.98	8	0.46

Note: N = 75 observations nested in five groups for both data sets.
**$p < .01$.

TABLE 3 Consensus emergence model: Model estimates in two data sets

Parameters	M1a	M1b	M2a	M2b
Data set 1				
Intercept, γ_{000}	10.06	10.12	10.06	10.20
TIME, γ_{100}	0.10	0.08	0.10	0.05
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.58	0.49	0.46	0.58
Group variance for TIME, $\sigma^2_{\beta_{10}}$	0.01	0.01	0.01	0.03
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	−0.07	−0.06	−0.07	−0.11
Person intercept variance, $\sigma^2_{\pi_0}$			0.47	0.22
Residual variance, σ^2_e	0.79	1.92	0.42	1.72
TIME, δ_1		−0.31		−0.51
Data set 2				
Intercept, γ_{000}	10.11	10.11	10.11	10.11
TIME, γ_{100}	0.06	0.06	0.06	0.06
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.14	0.18	0.05	0.04
Group variance for TIME, $\sigma^2_{\beta_{10}}$	0.35	0.34	0.38	0.37
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	−0.07	−0.07	−0.13	−0.12
Person intercept variance, $\sigma^2_{\pi_0}$		0.34	0.95	0.92
Residual variance, σ^2_e	1.63	1.14	0.85	0.68
TIME, δ_1		0.08		0.06

Note: N = 75 observations nested in five groups for both data sets. M = model.

Table 3 provides the model estimates. As shown in Table 3, the exponential variance function for M1b in data set 1 associated with TIME is $\delta_1 = -0.31$ and the residual variance at the start of the study (when TIME = 0) is $\sigma^2_e = 1.92$. Based on these model parameters for M1b in Table 3, we can estimate that the residual variance changes from $\sigma^2_e \times \exp(2 \times \delta_1 \times [\text{TIME} = 0]) = 1.92 \times \exp(2 \times -0.31 \times 0) = 1.92 \times \exp(0) = 1.92$ at the start of the study at T1 to $\sigma^2_e \times \exp(2 \times \delta_1 \times [\text{TIME} = 4]) = 1.92 \times \exp(2 \times -0.31 \times 4) = 1.92 \times \exp(2 \times -0.31 \times 4) = 0.16$ at the end of the study. Together, the results from Table 2 and Table 3 indicate that within-group residual variance in data set 1 is significantly decreasing, which formally supports the idea of emergence.

Table 2 and Table 3 also provide results for data set 2. In Table 2, the $\chi^2(df = 1) = 1.43$, $p = .23$ provides no evidence to suggest that within-group variance is decreasing over time. Similarly, in Table 3, the δ_1 value of 0.08 would be considered a nonsignificant increase. Again, to emphasize our earlier point, the ICC1 values over time for data set 1 and data set 2 were quite similar, so the ability of the CEM to formally differentiate between the two mechanisms leading to increases in ICC1 values is clearly valuable. In addition, the fact that the CEM provides a formal omnibus test by contrasting the fit of alternative models (Table 2) gives researchers a basis by which to determine whether the pattern differs from chance.

3.2 | Three-level CEM

The basic two-level CEM can be extended by controlling for potential baseline differences between persons making the model a three-level model. That is, it is possible that systematic prestudy differences exist between persons on the outcome of interest. For instance, in a study of job satisfaction, employees may hold certain a priori views of their job or work environments in general before they actually become a part of the group. These individual differences represent a confounding factor that can impact all observations from a particular person. In the basic three-level model, a person-random effect is added to the model specification that accounts for these potential differences.

The resulting model (see Table 1) specifies that the response of person *j* in group *k* at measurement occasion *i* is a function of a common intercept γ_{000}, one fixed-effects predictor (TIME_i), a person random effect r_{0jk} that accounts for baseline differences, a group random effect u_{00k}, a group-specific random slope u_{10k}, and the residual e_{ijk}. The variance of the residual e_{ijk} changes as a function of TIME_i.

The three-level version of the model may frequently be more realistic because people likely differ before they interact in the group. Furthermore, the model may also increase the power of CEM analyses when group members' a priori views are considerable in magnitude. In these situations, the power can increase analogously to adding a covariate in a multiple regression analysis or including a Level 2 error term when investigating Level 1 effects (Bliese, Maltarich, & Hendricks, 2018). By reducing error variance, the power to detect effects increases.

Notice in Table 2 that the contrast between M2a and M2b produces a larger χ^2 value (data set 1) than the contrast between M1a and M1b. Similarly, in Table 3, the estimate of the residual variance (M2b) is –0.51 instead of –0.31 (M1b). These increases reflect the fact that a large source of variance surrounding individual differences are accounted for by including an additional level that captures individual mean differences in response patterns.

The two-level CEM and the three-level CEM models that we described are appropriate models for most organizational and also most basic social psychology studies on consensus emergence. In typical organizational studies or classical social psychology laboratory experiments (Sherif, 1935), researchers attribute changes in consensus within groups over time to the nature of the group context. That is, researchers can assume that changes in consensus within groups over time are a function of group processes that occur during the study. For instance, in a classic study by Sherif, individuals made estimates of autokinetic light movement and there was a strong tendency for consensus over time. Sherif attributed the change pattern in consensus to the group situation in which group members saw other group members' estimates and then felt a need to move their own estimates into the direction of the others.[2]

In specific situations, it may make sense to further extend the basic three-level CEM model by also controlling for individual-level random slope variability. This type of model deviates from the logic behind the ICC1 and adds considerable computational complexity. Although the pattern of results in our empirical examples we present later in this paper were similar using this type of model, we would argue that this model is useful only in cases where individuals are likely to have strong change trajectories that are not driven by the group environment such as a study of children and reading skills over time. The individual slope random effects act as a control variable. Like control variables in other models, it can potentially decrease the amount of residual variance and group-level random slope variability. Change effects may become more or less significant as a result. The inherent challenge in this model, however, is its interpretation. The model assumes that individual change is driven by processes outside the group and that these individual change processes should be controlled before one makes inferences on group change. In most cases, and in line with much of the existing organizational literature and the use of the ICC1 in this literature, researchers often assume that changes in consensus result from group processes and not from individual factors external to the group. Therefore, we focus on the basic three-level model as the most common form of the CEM in the remainder of this paper.

In summary, the basic rationale behind the CEM is to remove the ambiguity in changes in ICC1 values over time by modeling changes in the within-person residual variance over time. In the context of the CEM, significant decreases in residual changes can then be modeled and interpreted as providing evidence of emergence.

3.3 | Demonstration of δ_1 recovery and the temporal ambiguity of ICC1

In our basic examples in Figure 1 and Figure 2, we demonstrated how the CEM can be applied to hypothetical data and yield statistical values that are congruent with the conceptual definition of consensus emergence. However, our analyses of these example data sets do not offer insights into the degree to which the model can recover true parameter estimates. To examine whether the CEM effectively recovers known change in residual variance, and to study the relationship between change in residual variances from the CEM (δ_1) and ICC1 more systematically, we therefore generated simulated data sets.

The data sets were generated on the basis of the three-level CEM and varying combinations of residual variance change with change in the between-group variance ($\sigma^2_{\beta_{10}}$ and $\sigma_{\beta_{00}\beta_{10}}$ in the CEM). We calculated ICC1 values on the basis of our simulated data at each point in time. For model fitting, we used restricted maximum likelihood estimation (REML), the nlme package (Pinheiro & Bates, 2000), and the Nelder-Mead optimization method in the R environment (R Core Team, 2014).

The top portion of Table 4 provides true values for the simulated data. Underneath the true values are the values obtained from the CEM. For instance, Model 2 data were generated with an intercept value of 1.00 and a –0.20 change in residual variance, and the models estimated values of 1.01 and –0.20. The lower portion of Table 4 also includes estimates of ICC1 values for each measurement occasion produced by the simulated data structure. For instance, had ICC1 been estimated at each occasion, the ICC1 values for Model 1 would have been 0.08, 0.11, 0.09, 0.08, and 0.10 for TIME = 0 through TIME = 4, respectively.

Notice in Table 4 how different trends in σ^2_e, different values of $\sigma^2_{\beta_{10}}$, and different values $\sigma_{\beta_{00}\beta_{10}}$ lead to similar patterns of change in ICC1 values over time. In other words, ICC1 values fail to reveal the underlying nature of the change process in consensus. Importantly, however, despite different values for σ^2_e, $\sigma^2_{\beta_{10}}$, and $\sigma_{\beta_{00}\beta_{10}}$, the CEM model recovers the values for δ_1 (change in residual variance) across different conditions. These results demonstrate that values of δ_1 provide unambiguous evidence for the presence or absence of emergence within the CEM.

4 | APPLICATION IN ORGANIZATIONAL DATA SETS

To illustrate the use of the models on nonsimulated data, we analyze consensus emergence in affect-related and motivational variables in two data sets that differ in terms of group sizes and number of groups. Research on the development of consensus in affect-related and motivational variables in groups over time is frequently motivated by the theoretical idea that motivation and affect spreads from the group environment to the individual group member so that group members are increasingly in tune with the affective or motivational tone of the group over time (Barsade, 2002; Damen, van Knippenberg, & van Knippenberg, 2008; Kelly & Barsade, 2001; Sy, Côté, & Saavedra, 2005).

Our two longitudinal field studies involve newly formed work units where changes and dynamics in group consensus are likely to occur. Table 5 provides details on both data sets. The Army data are based on a previously published data set (Bliese & Ployhart, 2002) and investigate job satisfaction among groups of Army soldiers in units chosen to test new equipment. The Nile valley data center on a sample of professional archeologists working in groups on field excavation missions. The archeologists included in the sample were studying Nubian sites in the Nile valley of present-day Sudan. Archeological excavation missions are often intense and stressful (Berggren & Hodder, 2003). The sample size at the group level in this latter example is limited with just six teams; however, the number of observations in this example is relatively large with up to 34 measurement occasions for each team ($N = 815$). The CEM analyses focus on change in

TABLE 4 Demonstration of the ambiguity of intraclass correlation, type 1 (ICC1) values for making inferences on consensus emergence and the successful recovery of changes in residual variances (δ_1) under varying conditions

	Simulated data set						
	Model 1: No change	Model 2: Residual variance decrease	Model 3: Slope variability	Model 4: Slope variability + negative covariance	Model 5: Slope variability + negative covariance + residual variance decrease	Model 6: Strong slope variability	Model 7: Slope variability + positive covariance + residual variance decrease
Underlying true values							
N ($n_{\text{level-3}} \times n_{\text{level-2}} \times n_{\text{level-1}}$)	2,500	[a]	[a]	[a]	[a]	[a]	[a]
$n_{\text{level-3}}$ (groups)	100	[a]	[a]	[a]	[a]	[a]	[a]
$n_{\text{level-2}}$ (persons/group)	5	[a]	[a]	[a]	[a]	[a]	[a]
$n_{\text{level-1}}$ (measurements/person)	5	[a]	[a]	[a]	[a]	[a]	[a]
Intercept, γ_{000}	1.00	[a]	[a]	[a]	[a]	[a]	[a]
TIME, γ_{100}	0.10	[a]	[a]	[a]	[a]	[a]	[a]
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.50	[a]	0.40	[a]	[a]	[a]	[a]
Group slope variance, $\sigma^2_{\beta_{10}}$	0.00	[a]	0.05	[a]	[a]	0.20	0.10
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	0.00	[a]	[a]	−0.10[b]	[a]	0.00	0.05
Person intercept variance, $\sigma^2_{\pi_0}$	2.00	[a]	[a]	[a]	[a]	[a]	[a]
Residual variance, σ^2_e	3.00	[a]	[a]	[a]	[a]	[a]	[a]
Change in residual variance, δ_1	0.00	−0.20	0.00	[a]	−0.20	0.00	−0.20

(Continues)

TABLE 4 *(Continued)*

	Simulated data set						
	Model 1: No change	Model 2: Residual variance decrease	Model 3: Slope variability	Model 4: Slope variability + negative covariance	Model 5: Slope variability + negative covariance + residual variance decrease	Model 6: Strong slope variability	Model 7: Slope variability + positive covariance + residual variance decrease
Consensus emergence models fitted to simulated data							
Intercept, γ_{000}	1.00	1.01	1.01	1.00	1.08	1.03	0.98
TIME, γ_{100}	0.11	0.10	0.09	0.09	0.08	0.10	0.10
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.53	0.51	0.38	0.34	0.47	0.46	0.35
Group slope variance, $\sigma^2_{\beta_{10}}$	0.00	0.00	0.06	0.05	0.05	0.22	0.10
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	−0.01	0.01	0.01	−0.09	−0.10	−0.02	0.04
Person intercept variance, $\sigma^2_{\pi_0}$	2.03	2.02	2.05	2.01	2.00	2.04	2.09
Residual variance, σ^2_e	3.01	3.00	3.03	3.10	3.08	3.00	3.01
Change in residual variance, δ_1	0.00	−0.20	0.00	0.00	−0.20	0.00	−0.20
ICC1 from standard multilevel models for each measurement occasion							
$ICC1_{TIME=0}$	0.08	0.10	0.06	0.07	0.10	0.07	0.07
$ICC1_{TIME=1}$	0.11	0.11	0.09	0.03	0.06	0.12	0.12
$ICC1_{TIME=2}$	0.09	0.15	0.12	0.03	0.07	0.22	0.20
$ICC1_{TIME=3}$	0.08	0.15	0.16	0.04	0.09	0.31	0.33
$ICC1_{TIME=4}$	0.10	0.17	0.22	0.08	0.14	0.42	0.46

[a]Value is identical to the value in the previous column.
[b]This covariance value corresponds to a correlation of $r = -0.10 / (\sqrt{0.40}\ \sqrt{0.05}) \approx -.71$.
[c]This covariance value corresponds to a correlation of $r = 0.10 / (\sqrt{0.40}\ \sqrt{0.05}) \approx .71$.

TABLE 5 Study characteristics for two longitudinal field data sets

Model	Army data	Nile valley data
Context	U.S. Army companies surveyed during a period of significant technological change in their work	Archeological excavation missions on a Nubian site in the Nile valley of present-day Sudan.
Participants	Officers and enlisted soldiers	• Archeologists working in groups • Participants were mostly experts from France and the United States • The questionnaires were in English and included French translations of the items in parentheses
Time period	3 times (0, 6, and 12 month)	Team members worked at the sites for an average of $M = 21.64$ days and filled in daily surveys. Fridays were work free.
Sample size	Companies for which data from at least three members with at least two observations each was available (34 companies with 471 soldiers and a total of 1,351 observations)	Two groups had six group members, one group had five group members, one group had seven group members, one group had 10, and one group 12 group members. The participants filled out a total of 815 daily questionnaires.
Variables	• Job satisfaction: Three items modified from Hackman and Oldham (1975) and a 5-point Likert-type scale • Unit combat readiness: Four item measure (Jex & Bliese, 1999) and a 5-point Likert-type scale • Leadership status: We treated all senior leadership ranks (officers and noncommissioned ranks above staff sergeant) as leaders. Army companies are led by a leadership team that includes more than one senior leader	• After-work fatigue using a two adjectives (exhausted and tired) from the Profiles of mood scales (McNair, Lorr, & Droppelman, 1971; also see Sonnentag, Binnewies, & Mojza, 2008) and a 5-point Likert scale (mean daily Cronbach's $\alpha = .78$)

Note: The Army data were originally reported in Bliese and Ployhart (2002) and are available in the multilevel package (Bliese, 2016) in R (R Core Team, 2014).

TABLE 6 Consensus emergence model (CEM): Model comparisons for two longitudinal field data sets

Data, model (M)	AIC	BIC	*logLik*	*df*	vs. M2a χ^2
Army data, DV = Job satisfaction					
M2a: Three-level model, no emergence	3,333.41	3,369.86	−1,659.70	7	
M2b: Three-level CEM	3,327.90	3,369.56	−1,655.95	8	7.51**
Nile valley data, DV = after work fatigue					
M2a: Three-level model, no emergence	1,932.84	1,965.74	−959.42	7	
M2b: Three-level CEM	1,930.84	1,968.44	−957.42	8	4.00*

Note: For the Army data, $N = 1{,}351$ observations nested in 471 unit members and 34 units. For the Nile valley data, $N = 815$ observations nested in six groups with 46 groups members.
$^{*}p < .05$. $^{**}p < .01$.

observations over time within teams so the models are informative even with the limited number of teams (although obviously questions about whether results from six teams generalize to a larger population are valid concerns).

Table 6 and Table 7 include CEM-based results from the two studies. We also ran the analyses described in this and the next section in Mplus. The findings were highly similar in the two programs and became identical when we switched to maximum likelihood estimation in nlme (Mplus uses maximum likelihood) instead of REML. To help facilitate model

TABLE 7 Consensus emergence model: model estimates for two longitudinal field data sets

Parameters	M2a	M2b
Army data, DV = Job satisfaction		
Intercept, γ_{000}	3.26	3.26
TIME, γ_{100}	0.05	0.05
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.11	0.10
Group variance for TIME, $\sigma^2_{\beta_{10}}$	0.01	0.01
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	–0.03	–0.02
Person intercept variance, $\sigma^2_{\pi_0}$	0.38	0.38
Residual variance, σ^2_e	0.43	0.51
TIME, δ_1		–0.10
Nile valley data, DV = After work fatigue		
Intercept, γ_{000}	2.40	2.39
TIME, γ_{100}	–0.0002	0.001
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.001	0.002
Group variance for TIME, $\sigma^2_{\beta_{10}}$	0.0005	0.0004
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	0.001	0.001
Person intercept variance, $\sigma^2_{\pi_0}$	0.36	0.37
Residual variance, σ^2_e	0.52	0.60
TIME, δ_1		–0.006

Note: For the Army data, N = 1,351 observations nested in 471 unit members and 34 units. For the Nile valley data, N = 815 observations nested in six groups with 46 group members. M = model.

estimation, we provide syntax for both nlme and Mplus in the Appendix (see Appendix A and Appendix B). The examples involve the Army data because these data are easily available as part of the multilevel package in R (Bliese, 2016). We only report the REML results because REML estimation is preferred when the focus is on the estimation of variance components (Snijders & Bosker, 1999).

We conducted the analyses using the three-level CEM model shown in the second row of Table 1 because after-work fatigue and job satisfaction are both variables on which group members likely differ prior to our data collection, so controlling for individual baseline values on these variables makes conceptual sense. Table 6 compares –2log likelihood values and suggests that models allowing for change in the residual error variance fit better than models that assume equal error variance. The –2log likelihood tests provide a statistical test of consensus emergence patterns over time.

Table 6 shows that including an exponential variance function for the Army data increased fit (Model 1.2 vs. Model 1.1), $\chi^2(df = 1) = 7.51, p < .01$, and Table 7 shows that the exponential variance function weight for time was $\delta_1 = -0.10$. This effects equals a decrease in residual variance from 0.51 at baseline (TIME = 0) to $0.51 \times \exp(2 \times -0.10 \times 2) = 0.34$ at TIME = 2. To evaluate how these findings would compare to conclusions we would have reached relying on ICC1 values, we estimated the ICC1 at each time point. The values suggested a pattern of decreasing consensus. ICC1 values were .08, .02, and .01, at TIME = 0, TIME = 1, and TIME = 2, respectively. These values indicate that the conclusion one would draw using ICC1 values fundamentally differs from the conclusion based on the basis of the CEM.

Tables 6 and 7 show that the Nile valley data displayed a significant decrease in variance and thus consensus emergence in after-work fatigue, $\delta_1 = -0.006$, $\chi^2(df = 1) = 4.00, p < .05$. The information criterion shown in Table 6 provides a less clear picture than for the Army data. Although AIC favored the emergence model, BIC preferred the null model. Given that our focus is more on detecting the presence of a pattern that best describes the data and less on identifying a true model, the use of AIC and the use of the CEM is likely justified from an information criterion perspective (see Burnham & Anderson, 2004; Vrieze, 2012, for discussions of AIC vs. BIC).

Our results suggest that team members showed increasingly similar levels of fatigue after each work day over the course of the mission. Note that the δ_1 coefficient in the Nile valley data is smaller than in the other data set. Importantly though, δ_1 captures linear change with each measurement point and the density of measurement occasions in the Nile valley data goes up to 33 daily measurements, which partially explains why the estimate for a one-point change is small. To illustrate this point, we divided the time variable by 30. The resulting larger δ_1 of –0.168 then captures the amount of change over the course of a month.

We again evaluated how the findings from the CEM analysis would compare to an analysis on the basis of ICC1. The pattern of ICC1 values was difficult to interpret because the number of observations is relatively small for each specific time point ($M = 24.19$) in this data set. Observed ICC1 values ranged from .00 to .78. These findings again illustrate our previous point that a formal test for consensus emergence on the basis of a longitudinal model has important advantages over ICC1 values estimated at each specific time point.

In summary, the results for the Army and the Nile valley data show that the CEM approach can be used to test consensus emergence in longitudinal field data.[3] In the next section, we provide examples of how the models can be extended.

5 | EXTENDING THE CEM APPROACH

As noted, in addition to providing a useful omnibus test of patterns of consensus emergence over time, an appealing characteristic of the CEM approach is that it allows researchers to test more complex theoretical models surrounding consensus emergence. In many cases, organizational researchers may not only be interested in testing for the presence of emergence over time but may also want to test for predictors of consensus in addition to time or seek to explain why change occurs by testing interactions between time and predictors (Bliese & Ployhart, 2002; Ployhart & Vandenberg, 2010). In this section, we describe more advanced CEM models that extend the basic CEM and include predictors of consensus emergence. Predictors and interactions between predictors and time can be tested in the CEM by contrasting the fit of different CEMs with or without the predictors. We provide two examples for models of this type with predictors at the group level and also at the person level on the basis of the Army data set (see Appendix C for model specifications).

Our first example incorporates a group-level predictor and tests whether Army companies' combat readiness at the start of the study is related to the degree to which the companies show consensus emergence in job satisfaction over the course of study. Combat readiness reflects members' confidence in their military unit and can be considered a measure of collective efficacy (Jex & Bliese, 1999). Combat readiness was standardized at the sample mean and standard deviation.

Tables 8 and 9 provide results. Table 8 contrasts three different models to formally test whether group readiness predicted the occurrence of consensus emergence. The baseline model (M3a) includes an interaction between time and group readiness for mean differences (fixed effects) but only includes time as a predictor in the variance (random effects) part of the model. The second model (M3b) adds group readiness as an additional predictor of differences in within-group variance in addition to time. Finally, the third model (M3c) adds the interaction effect between time and group readiness to the within-group variance part of the model. Contrasting M3b and M3c provides a formal test for group readiness as a predictor of group-level change in consensus over time, $\chi^2(df = 1) = 6.24$, $p < .05$. Table 9 shows that the residual variance for Model 3c is $\sigma_e^2 = 0.51$. This residual variance estimate refers to a hypothetical group at the start of the study (TIME = 0) with average group readiness. The variance function for time for Model 3c is –0.09 (see Table 9), and the model thus predicts that a hypothetical group with average group readiness would change from 0.51 at baseline to 0.36 ($0.51 \times \exp[2 \times 2 \times -0.09] = 0.36$) at TIME = 2. As indicated by Table 9, the estimated variance function term for group readiness is $\delta_2 = -0.15$, and the variance function interaction term between readiness and time is $\delta_3 = 0.09$. These values can be used to estimate the residual variance for hypothetical groups with group readiness scores 1 *SD* below and above the sample mean. For a group with a group readiness score 1 *SD* above the sample mean, the residual variance would be $0.51 \times \exp(2 \times -0.09 \times 0) \times \exp(2 \times -0.15 \times 1) \times \exp(2 \times 0.09 \times 0 \times 1) = 0.38$ at the start

TABLE 8 Extended consensus emergence models applied to the army data set: Model comparisons

Model (M)	AIC	BIC	*logLik*	*df*	vs. previous model χ^2
Group-level predictor					
M3a: Three-level CEM + group readiness (T1) and a readiness × time interaction as predictors	3,320.42	3,372.47	−1,650.21	10	
M3b: M3a + variance function for readiness	3,317.07	3,374.34	−1,647.54	11	5.34*
M3c: M3b + variance function for readiness × time interaction	3,312.83	3,375.30	−1,644.42	12	6.24*
Person-level predictor					
M4a: CEM + leader status (yes/no) and a leader status × interaction as predictors	3,088.29	3,139.64	−1,534.14	10	
M4b: M4a + variance function for leader status	3,082.90	3,139.39	−1,530.45	11	7.39**
M4c: M4b + variance function for leader status × time interaction	3,081.28	3,142.90	−1,528.64	12	3.62†

Note: For models 3a, 3b, and 3c, $N = 1{,}351$ observations nested in 471 unit members and 34 units. For models 4a, 4b, and 4c, $N = 1{,}260$ observations nested in 438 unit members and 31 units.
†$p < .10$. *$p < .05$. **$p < .01$.

TABLE 9 Extended consensus emergence models applied to the army data set: Model estimates

Parameters	M3c	M4b	M4c
Intercept, γ_{000}	3.22	3.13	3.13
TIME, γ_{100}	0.06	0.07	0.07
GROUP READINESS, γ_{010}	0.26		
TIME × READINESS, γ_{110}	–0.06		
LEADER, β_{01k}		0.48	0.48
TIME × LEADER, β_{11k}		–0.09	–0.09
Group intercept variance, $\sigma^2_{\beta_{00}}$	0.00	0.05	0.05
Group variance for TIME, $\sigma^2_{\beta_{10}}$	0.00	0.003	0.003
Covariance, $\sigma_{\beta_{00}\beta_{10}}$	0.00	–0.01	–0.01
Person intercept variance, $\sigma^2_{\pi_0}$	0.37	0.37	0.38
Residual variance, σ^2_e	0.51	0.56	0.52
TIME, δ_1	–0.09	–0.11	–0.07
GROUP READINESS, δ_2	–0.15		
TIME × READINESS, δ_3	0.09		
LEADER, δ_2		–0.15	–0.02
TIME × LEADER, δ_3			–0.16

Note: For model 3c, $N = 1{,}351$ observations nested in 471 unit members and 34 units. For model 4b and 4c, $N = 1{,}260$ observations nested in 438 unit members and 31 units. M = Model.

of the study and $0.51 \times \exp(2 \times -0.09 \times 2) \times \exp(2 \times -0.15 \times 1) \times \exp(2 \times 0.09 \times 2 \times 1) = 0.38$ at TIME = 2. In other words, the residual variance estimate does not change. In contrast, for a group with a group readiness score that is 1 *SD* below the sample mean, the residual variance would be $0.51 \times \exp(2 \times -0.09 \times 0) \times \exp(2 \times -0.15 \times -1) \times \exp(2 \times 0 \times 0.09 \times -1) = 0.69$ at the start of the study, and $0.51 \times \exp(2 \times -0.09 \times 2) \times \exp(2 \times -0.15 \times -1) \times \exp(2 \times 0.09 \times 2 \times -1) = 0.34$ at TIME = 2. Overall, these results suggest that companies with higher readiness had higher consensus at the start of the study and consensus did not change much over time. Conversely, companies with low readiness started with low consensus and showed a pattern of consensus emergence over the course of the study.

Our second example of an explanatory variable focuses on leadership status as a person-level predictor. Leadership theories are frequently based on the notion that leaders are more influential to the development of group consensus/climate than other group members (Barsade & Knight, 2015; Bass, 1985; House, 1977). CEM models with leader status (yes/no) allow one to test whether leaders show different patterns of consensus than other group members. These differential patterns can help support inferences about how central (close to the group average) leaders' positions are in the group and whether leaders become more central over time. A central position implies influence; however, it is also theoretically possible that the group influences leaders more strongly than other group members. We contend, though, that leader influence is generally the theoretically more plausible model especially when leader centrality increases over time.

The theoretical idea behind CEM models with leader status can be illustrated by reexamining Figure 2. As shown in this graph, the groups and some team members show a tendency to move closer together over time (e.g., Person 1 in Group 5 in data set 1), whereas other group members (e.g., Person 3 in Group 3 in data set 2) show a tendency to move away from the overall consensus pattern. Tables 8 and 9 include results from a model testing differential patterns for leaders. As indicated in Tables 8 and 9, leaders (leader status = 1) showed less fluctuations from the group consensus than other group members (leader status = 0), $\delta_2 = -0.15$ (M4b in Table 9), $\chi^2(df = 1) = 7.39$, $p < .01$ (M4b vs. M4a in Table 8). Leaders are thus more central than other members with respect to consensus across the entire time period. Further evidence of the central role of leaders may be found in examining the leader status × time interaction. As noted in Tables 8 and 9, the interaction is significant at the 10% level, $\delta_3 = -0.16$ (see Table 9), $\chi^2(df = 1) = 3.62$, $p < .10$ (M4c vs. M4b in Table 8). This interaction suggests that leaders became differentially more central to the ultimate pattern of consensus over time relative to nonleaders.

In interpreting the results of the two examples with predictors, it is important to note that the analyses are variants of the growth model and thus include the assumption that the predictors of growth are time invariant (Singer & Willett, 2003; Willett, 1997). Specifically, combat readiness measured at the start of the study in the first example is thought to be stable over the course of the data collection. Similarly, leadership status should not change during the measurement period (likely a plausible assumption in this context). In addition, combat readiness in the first example is also measured using ratings by group members and thus the analysis also makes the assumption that combat readiness is a group-level construct over the course of the study (i.e., does not show a pattern of decreasing consensus). Using time-invariant predictors in the basic growth model makes the interpretation of the models more convenient and allows us to demonstrate how predictors can be included, but the use of such variables does require the assumptions of invariance. That is, the two examples we present are in no way exhaustive. For instance, researchers may decide to focus on research questions that can be studied with time-varying predictors other than time (see Singer & Willett, 2003, for a detailed discussion) and add these types of predictors in the context of the CEM or combine elements of the CEM with other advanced longitudinal data analysis approaches (e.g., Hoffman, 2007; Rast et al., 2012).

6 | DISCUSSION

This paper described a statistical approach—the CEM—that allows organizational researchers to study common bottom-up processes over time within work units. Our approach builds on multilevel models by examining patterns in the residual variance. The CEM can be expressed as either a two-level or three-level model that allows one to interpret residual variances as indices of group consensus emergence. The approach also models systematic change in group means over time and slope variability in this change. Systematic change in group means and variability in group slopes are both additionally important sources of information and provide insights into the degree and direction of change in group perceptions/climates.

Extensions of the basic model allow researchers to study more complex research questions and use the CEM as a general framework to study emergent processes in organizations. We believe that the CEM used in conjunction with other approaches such as computational modeling, qualitative research, and social network analyses can enhance our ability to study emergent processes. We illustrated the use of CEM in a simulation and in two organizational data sets.

6.1 | Methodological and theoretical implications

Even the basic CEM approach appears to provide a valuable tool for researchers interested in testing consensus-related bottom-up emergent processes. The framework is a direct extension of multilevel methods already common in organizational research and builds off of variance models used in other disciplines. Organizational researchers can readily use the described approach to gain insights into bottom-up processes and develop more dynamic views of shared perceptions in organizations (Barsade & Knight, 2015; Cronin & Weingart, 2011; Humphrey & Aime, 2014; Kozlowski & Chao, 2012; Kozlowski et al., 2013).

We have also shown how the basic CEM can be expanded, and in several examples, we discussed group readiness, leadership status, and daily events as potentially meaningful group-, person-level, and time-varying predictors of consensus emergence. Clearly other predictors of consensus emergence at the group-, person-, and measurement level are of interest; however, our examples illustrate the potential application and use of the CEM approach in organizational research. Many of the theoretical ideas that we described in this paper have been introduced and discussed in the broader literature; however, we have demonstrated how these ideas might be built upon and tested. For instance, the notion that some members are more closely related to shared ideas and thus to emergent phenomena than others has been studied by examining the centrality of leaders using social network approaches (Carter et al., 2015). We have shown, though, that the basic idea of centrality can be incorporated using typical organizational data (survey responses over time) within a multilevel model framework. The two types of approaches use different data structures, build on different theoretical assumptions, and have unique advantages (for a broader discussion of conceptual differences between multilevel/latent variable models and network approaches, see Borsboom & Cramer, 2013); nonetheless, we see opportunities for the CEM approach to complement existing theory and methodology.

6.2 | Practical implications

One practical implication of the CEM is that research using this tool may provide useful new temporal insights for organizations. For instance, the CEM could be used to generate insights on how long the development of perceptions and climates takes for specific constructs and organizational contexts. Questions of this type are currently rarely studied in empirical research. Relatedly, organizations may consider whether different on-boarding procedures, team-development procedures (e.g., developing a team charter), or different leadership development practices facilitate the emergence of consensus in teams. For instance, an organization may develop an intervention to decrease LMX differentiation/increase LMX group-consensus by training their leaders to more equally distribute their attention to followers (Henderson, Liden, Glibkowski, & Chaudhry, 2009; Schyns & Day, 2010). The CEM could then be used to track how the intervention affects LMX consensus over time.

Another practical implication of the CEM is that organizations could directly use the approach on their data archives. For example, the CEM could be applied to annual survey data gathered by most large organizations and used to gain insights into team and unit development. In this context, organizations might note that teams in certain functional areas or with managers with specific characteristics (e.g., lack of experience) never seem to come together or "gel." Finally, organizations could use the CEM to monitor how work groups react to the implementation of new policies, initiatives, or programs. For instance, it might be valuable to see whether certain types of employees (e.g., older or younger or longer or shorter tenure) play a particularly influential role in how teams come together with respect to viewing new policies or initiatives.

6.3 | Limitations

One strength of the CEM approach is that it readily builds off of existing multilevel methodology. However, there are also potential limitations. One limitation is that the CEM is less suited for explorative data analyses. Unlike qualitative approaches (e.g., Gehman et al., 2013) and some descriptive social network analysis approaches (e.g., Borsboom & Cramer, 2013; Butts, 2008), CEM-based research requires specific theoretical ideas on what type of organizational structures, attitudes, and perceptions are expected to have a key role in group social dynamics. The CEM approach is thus primarily useful for testing existing theories and may be less useful in exploratory theory development.

A second limitation of the application of the CEM approach is that the approach does not provide insights into the underlying causal mechanisms. The CEM is designed to evaluate and study the presence of one (commonly discussed) bottom-up process in collective groups. In this context, the term bottom-up process is used to describe a pattern of change in collective groups. This use of the term does not provide evidence of the specific causal processes that underlie the pattern. Although alternative explanations are commonly seen as less likely for longitudinal growth models (Ployhart & Vandenberg, 2010; Singer & Willett, 2003) than for cross-sectional evidence, it is important to keep in mind that longitudinal growth models do not directly provide evidence for causality. Causal conclusions are typically only possible using experimental manipulations (Bodner & Bliese, 2018). A potential strategy for future research could be to revive experimental work on group climate (e.g., Sherif, 1935) to gain more insights into causal mechanisms.

A third limitation of the approach is that researchers should thoroughly check their data for potential floor or ceiling effects. The CEM approach models mean change and variability jointly and thus controls for mean change in variability estimates. However, heavily skewed data with strong floor and ceiling effects can violate the assumption of normally distributed residuals that underlies mixed-effects models. We therefore recommend checks using descriptive graphical approaches (Pinheiro & Bates, 2000). We also recommend using long scales (i.e., many items) and/or fine-grained rating scales, and choosing a set of items that cover the entire scale continuum well.

6.4 | Future directions

We see several important avenues for future research. First, we suggest studying the degree to which the CEM can be used to detect consensus emergence in ordinary organizational data. The analyses we presented provide some important insights regarding the amount of consensus emergence that researchers can expect from field data. An important question for future research is how large typical δ_1 coefficients in organizational data are and how sensitive the CEM approach is with respect to detecting typical coefficients. As a first step, we conducted a small power simulation (cf. Mathieu, Aguinis, Culpepper, & Chen, 2012) using values for the CEM of $\gamma_{000} = 3, \gamma_{100} = 0.01, \sigma^2_{\beta_{00}} = 0.02, \sigma_{\beta_{00}\beta_{10}} = 0.01,$ $\sigma^2_{\beta_{10}} = -0.01, \sigma^2_{\pi_0} = 0.30, \sigma^2_e = 0.40$, and $\delta_1 = -0.08$. The sample size was 10, 20, or 30 eight-person groups measured five times. Results from 1,000 simulation runs revealed power values of .72, .95, and .99 for 10, 20, and 30 groups, respectively. These initial findings suggest that a relatively limited number of teams allow researchers to detect consensus emergence effects equal to a 40% decrease in residual group variance. As a preliminary rule of thumb, we suggest using at least 20 groups when no effect size information on the basis of previous research is available.

Another topic for future research could be how the CEM can effectively be expanded to account for measurement error at the observation level. The CEM approach already corrects for measurement error at the group level because the random effects for the group means in mixed-effects models are latent variables. However, it could help to additionally incorporate observation-level measurement error. As a first step, we specified a four-level version of the CEM. In this model, an additional level is included to account for the nesting of ratings in individual observations and the model uses a slightly different parameterization of the change in within-group variance over time (Goldstein, 2005, 2011; also see our Footnote 1 for detail). We applied this more complex model to the Nile-Valley data example. Like in the original analysis (see Table 6), the results revealed a significant consensus emergence effect, $\chi^2(df = 1) = 8.63$, $p = .003$. The p-value was similar to that found in an analysis using the alternative parameterization without observation-level measurement error, $\chi^2(df = 1) = 9.08$, $p = .003$. In this specific example, the substantive conclusion did not change after the inclusion of observation-level measurement error. However, it is possible that observation-level measurement error can have an impact on the findings and the described four-level model provides a basis for studying this impact.

Finally, a question for future research is how forms of emergence other than consensus can be captured in future modeling efforts. The focus in our article was on modeling how groups transition from a situation with considerable variability in group perceptions, behavior, or affect to a situation in which groups are developing patterns consistent with consensus. As noted, team researchers have theoretically argued that the nature of variability in a team can be more complex and have described three different forms of variability within teams (DeRue, Hollenbeck, Ilgen, & Feltz, 2010; also see Harrison & Klein, 2007). Minority belief teams have one person that differs from the rest, bimodal teams are split into two fractions, and fragmented teams simply show a range of different opinions. Mixed-effects models like the ones we discussed commonly assume that error variance is randomly and unsystematically distributed. As we have shown, though, when researchers have assumptions about which types of group members differ, the other two types of variability can readily be incorporated into the CEM through dichotomous predictor variables. The extended CEM with leader status as a predictor for the Army data that we described is an example of such a model that incorporates and tests a minority difference (leaders vs. nonleaders) and also allows researchers to study the degree of this minority difference over time. Hypotheses on bimodal team diversity can be studied within the CEM in similar ways when the predictor variable is known that should theoretically split the team into two fractions.

7 | CONCLUSION

Bottom-up emergent processes are theoretically important for understanding social dynamics in organizations. In this article, we introduce a methodological approach that addresses limitations of the standard multilevel model and the ICC1. We believe the CEM approach allows researchers to (a) statistically test whether shared perceptions in groups converge over time and (b) examine how shared perceptions differ between types of work-group members and as a function of work-group characteristics. The CEM can be applied to longitudinal rating data on groups that includes scores for individual group members who either rate their own behavior or a group characteristic. The model is not designed for typical network studies in which all group members rate all other group members. As we discussed, we recommend that researchers carefully check for potential ceiling effects using graphical analyses and use at least 20 groups as an initial rule of thumb when no effect size information on the basis of previous research is available. Although the CEM is not designed to replace existing methods for studying emergence, we see opportunities for using the methodology to help advance the study of consensus emergence and related bottom-up processes in organizations.

ACKNOWLEDGMENTS

We thank the National Museums of Khartoum, Sudan, the American Museum of Natural History, New York, Vincent Francigny, and Tosha Dupras for their permission to collect the Nile Valley data and their assistance with this data collection. Jonas Lang's work on this research was partly supported by a grant of the Fund for Scientific Research Flanders (G019217N).

NOTES

[1] An alternative parameterization can be based on the observation-level slope approach (Goldstein, 2005, 2011). Goldstein's approach models change in the residual variance by directly adding predictors to the residual variance and either uses an intercept, $e_{0ijk} \sim N(0, \sigma^2_{e0})$, and an uncorrelated linear slope with a time variable that is centered at the end of the observation period, $e_{1ijk} \sim N(0, \sigma^2_{e1}[\text{TIME}_i - max(\text{TIME})]^2)$, or an intercept and a covariance (with the slope fixed to 0). The exponential variance function approach and the observation-level slope approach are not equivalent models but yield substantively similar results in practice.

[2] A recent book chapter (Lang & Bliese, in press) provides a brief introduction of the CEM and shows how the model can be applied to data from Sherif (1935).

[3] We also fit the two-level model to both datasets. For the Army data, the two-level CEM yielded a significant exponential variance function weight for time (two-level CEM: $\delta_1 = -0.05$, $\chi^2[df = 1] = 4.94$, $p < .05$). In the Nile valley data, the exponential variance function weight for time was positive but not significant (two-level CEM: $\delta_1 = -0.001$, $\chi^2[df = 1] = 0.19$, $p = .67$). These results illustrate the value in controlling for initial differences via a three-level model.

ORCID

Jonas W. B. Lang http://orcid.org/0000-0003-1115-3443

REFERENCES

Allen, N. J., & O'Neill, T. A. (2015). The trajectory of emergence of shared group-level constructs. *Small Group Research*, *46*, 352–390. https://doi.org/10.1177/1046496415584973

Ashforth, B. E. (1985). Climate formation: Issues and extensions. *Academy of Management Review*, *10*, 837–847. https://doi.org/10.2307/258051

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, *49*, 630. https://doi.org/10.5465/AMJ.2006.20785500

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*, 644–675. https://doi.org/10.2307/3094912

Barsade, S. G., & Knight, A. P. (2015). Group affect. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 21–46. https://doi.org/10.1146/annurev-orgpsych-032414-111316

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, NY: Free Press.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*. https://doi.org/10.18637/jss.v067.i01

Berggren, Å., & Hodder, I. (2003). Social practice, method, and some problems of field archaeology. *American Antiquity*, *68*, 421–434. https://doi.org/10.2307/3557102

Blee, K. (2013). How options disappear: Causality and emergence in grassroots activist groups. *American Journal of Sociology*, *119*, 655–681. https://doi.org/10.1086/675347

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bliese, P. D. (2016). *Multilevel modeling in R (2.6): A brief introduction to R, the multilevel package and nlme*. Vienna, Austria: The R Foundation for Statistical Computing. Retrieved from https://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf

Bliese, P. D., Maltarich, M. A., & Hendricks, J. L. (2018). Back to basics with mixed-effects models: Nine take-away points. *Journal of Business and Psychology*, *33*, 1–23. https://doi.org/10.1007/s10869-017-9491-z

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*, 362–387. https://doi.org/10.1177/109442802237116

Bodner, T. E., & Bliese, P. D. (2018). Detecting and differentiating the direction of change and intervention effects in randomized trials. *Journal of Applied Psychology*, *103*, 37–53. https://doi.org/10.1037/apl0000251

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bowen, D. E., & Ostroff, C. (2004). Understanding HRM-firm performance linkages: The role of the "strength" of the HRM system. *Academy of Management Review*, *29*, 203–221. https://doi.org/10.5465/AMR.2004.12736076

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference. *Sociological Methods & Research*, *33*, 261–304. https://doi.org/10.1177/0049124104268644

Burt, R. S., Kilduff, M., & Tasselli, S. (2013). Social network analysis: Foundations and frontiers on advantage. *Annual Review of Psychology*, *64*, 527–547. https://doi.org/10.1146/annurev-psych-113011-143828

Butts, C. T. (2008). Social network analysis with SNA. *Journal of Statistical Software*, *24*, 1–51. https://doi.org/10.1016/j.socnet.2009.09.002

Carter, D. R., DeChurch, L. A., Braun, M. T., & Contractor, N. S. (2015). Social network approaches to leadership: An integrative conceptual review. *Journal of Applied Psychology*, *100*, 597–622. https://doi.org/10.1037/a0038922

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*, 234–246. https://doi.org/10.1037/0021-9010.83.2.234

Cronin, M. A., & Weingart, L. R. (2011). Dynamics in groups: Are we there yet? *The Academy of Management Annals*, *5*, 37–41. https://doi.org/10.1080/19416520.2011.590297

Damen, F., van Knippenberg, B., & van Knippenberg, D. (2008). Affective match in leadership: Leader emotional displays, follower positive affect, and follower performance. *Journal of Applied Social Psychology*, *38*, 868–902. https://doi.org/10.1111/j.1559-1816.2008.00330.x

DeRue, D. S., Hollenbeck, J., Ilgen, D., & Feltz, D. (2010). Efficacy dispersion in teams: Moving beyond agreement and aggregation. *PERSONNEL PSYCHOLOGY*, *63*, 1–40. https://doi.org/10.1111/j.1744-6570.2009.01161.x

Felps, W., Mitchell, T., Hekman, D., Lee, T., Holtom, B., & Harman, W. (2009). Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, *52*, 545–561. https://doi.org/10.5465/AMJ.2009.41331075

Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ (Clinical Research Ed.)*, *337*, a2338. https://doi.org/10.1136/bmj.a2338

Gehman, J., Trevino, L., & Garud, R. (2013). Values work: A process study of the emergence and performance of organizational values practices. *Academy of Management Journal*, *56*, 84–112. https://doi.org/10.5465/amj.2010.0628

Gephart, R. P. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal*, *47*, 454–462. https://doi.org/10.2307/20159596

Goldstein, H. (2005). Heteroscedasticity and complex variation. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia for statistics in behavioral science* (Vol. *2*, pp. 790–795). Chichester, UK: John Wiley & Sons.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: Wiley. https://doi.org/10.2307/1534624

Griffin, M. A. (1997). Interaction between individuals and situations: Using HLM procedures to estimate reciprocal relationships. *Journal of Management*, *23*, 759–773. https://doi.org/10.1177/014920639702300604

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, *60*, 159–170. https://doi.org/10.1037/h0076546

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software*, *33*, 1–22.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, *32*, 1199–1228. https://doi.org/10.5465/AMR.2007.26586096

Hausknecht, J. P. (2017). Collective turnover. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 527–544. https://doi.org/10.1146/annurev-orgpsych-032516-113139

Hausknecht, J. P., & Trevor, C. O. (2011). Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management*, *37*, 352–388. https://doi.org/10.1177/0149206310383910

Hedeker, D., & Mermelstein, R. (2007). Mixed-effects regression models with heterogenous variance: Analyzing ecological momentary assessment (EMA) data of smoking. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 183–206). Mahwah, NJ: Lawrence Erlbaum.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, *31*, 3328–3336. https://doi.org/10.1002/sim.5338

Henderson, D. J., Liden, R. C., Glibkowski, B. C., & Chaudhry, A. (2009). LMX differentiation: A multilevel review and examination of its antecedents and outcomes. *Leadership Quarterly*, *20*, 517–534. https://doi.org/10.1016/j.leaqua.2009.04.003

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*, 609–629. https://doi.org/10.1080/00273170701710072

House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *Leadership: The cutting edge* (pp. 189–207). Carbondale, IL: Southern Illinois University Press.

Huisman, M., & Snijders, T. A. B. (2003). Statistical analysis of longitudinal network data with changing composition. *Sociological Methods & Research*, *32*, 253–287. https://doi.org/10.1177/0049124103256096

Humphrey, S. E., & Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals*, *8*, 443–503. https://doi.org/10.1080/19416520.2014.904140

James, L. R., Demaree, R. G., & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, *78*(2), 306. https://doi.org/10.1037/0021-9010.78.2.306

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, *44*, 238–251. https://doi.org/10.2307/3069453

Jex, S. M., & Bliese, P. D. (1999). Efficacy beliefs as a moderator of the impact of work-related stressors: A multilevel study. *Journal of Applied Psychology*, *84*, 349–361. https://doi.org/10.1037/0021-9010.84.3.349

Jones, S. L., & Shah, P. P. (2016). Diagnosing the locus of trust: A temporal perspective for trustor, trustee and dyadic influences on perceived trustworthiness. *Journal of Applied Psychology*, *101*, 392–414. https://doi.org/10.1037/apl0000041

Kalish, Y., & Luria, G. (2016). Leadership emergence over time in short-lived groups: Integrating expectations states theory with temporal person-perception and self-serving bias. *Journal of Applied Psychology*, *101*, 1474–1486. https://doi.org/10.1037/apl0000126

Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes*, *86*, 99–130. https://doi.org/10.1006/obhd.2001.2974

Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin*, *119*, 138–148. https://doi.org/10.1037/0033-2909.119.1.138

Kim, J., & Seltzer, M. (2011). Examining heterogeneity in residual variance to detect differential response to treatments. *Psychological Methods*, *16*, 192–208. https://doi.org/10.1037/a0022656

Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*, 211–236. https://doi.org/10.1177/109442810033001

Kostopoulos, K. C., Spanos, Y. E., & Prastacos, G. P. (2011). Structure and function of team learning emergence: A multilevel empirical validation. *Journal of Management*, *39*, 1430–1461. https://doi.org/10.1177/0149206311419366

Kozlowski, S. W. J. (2017). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, https://doi.org/10.1177/2041386610376255

Kozlowski, S. W. J., & Chao, G. T. (2012). The dynamics of emergence: Cognition and cohesion in work teams. *Managerial and Decision Economics*, *33*, 335–354. https://doi.org/10.1002/mde.2552

Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, *16*, 581–615. https://doi.org/10.1177/1094428113493119

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Kuppens, T., & Yzerbyt, V. Y. (2014). Predicting variability: Using multilevel modeling to assess differences in variance. *European Journal of Social Psychology*, *44*, 691–700. https://doi.org/10.1002/ejsp.2028

Lang, J. W. B., & Bliese, P. D. (in press). A temporal perspective on emergence: Using 3-level mixed effects models to track consensus emergence in groups. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook for multilevel theory, measurement, and analysis*. Washington, DC: APA.

LeBreton, J. M., James, L. R., & Lindell, M. K. (2005). Recent issues regarding rWG, rWG, rWG(J), and rWG(J). *Organizational Research Methods*, *8*, 128–138. https://doi.org/10.1177/1094428104272181

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions and interrater agreement. *Organizational Research Methods*, *11*, 815–852. https://doi.org/10.1177/1094428106296642

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*, 951–966. https://doi.org/10.1037/a0028380

McNair, D. M., Lorr, M., & Droppelman, L. F. (1971). *Edits manual for the profile of mood states.* San Diego, CA: Educational and Industrial Testing Service.

Morgeson, F., & Hofmann, D. (1999). The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, *24*, 249–265. https://doi.org/10.5465/AMR.1999.1893935.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide. Seventh edition.* Los Angeles, CA: Muthén & Muthén.

Nyberg, A. J., & Ployhart, R. E. (2013). Context-emergent turnover (CET) theory: A theory of collective turnover. *Academy of Management Review*, *38*, 109–131. https://doi.org/10.5465/amr.2011.0201

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.

Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review*, *36*, 127–150. https://doi.org/10.5465/amr.2011.55662569

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*, 94–120. https://doi.org/10.1177/0149206309352110

R Core Team (2014). R: A language and environment for statistical computing [Version 3.1.1]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rast, P., MacDonald, S. W. S., & Hofer, S. M. (2012). Intensive measurement designs for research on aging. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, *25*, 45–55. https://doi.org/10.1024/1662-9647/a000054

Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational and Behavioral Statistics*, *12*, 241–269. https://doi.org/10.3102/10769986012003241

Schyns, B., & Day, D. (2010). Critique and review of leader–member exchange theory: Issues of agreement, consensus, and excellence. *European Journal of Work and Organizational Psychology*, *19*, 1–29. https://doi.org/10.1080/13594320903024922

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, *27*, 1–60.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Skrondal, A., & Rabe-Hesketh, S. (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, *34*, 712–745. https://doi.org/10.1111/j.1467-9469.2007.00573.x

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London, UK: Sage.

Sonnentag, S., Binnewies, C., & Mojza, E. J. E. (2008). "Did you have a nice evening?" A day-level study on recovery experiences, sleep, and affect. *Journal of Applied Psychology*, *93*, 674–684. https://doi.org/10.1037/0021-9010.93.3.674

Stewart, G. L., & Nandkeolyar, A. K. (2007). Exploring how constraints created by other people influence intraindividual variation in objective performance measures. *Journal of Applied Psychology*, *92*, 1149–1158. https://doi.org/10.1037/0021-9010.92.4.1149

Sy, T., Côté, S., & Saavedra, R. (2005). The contagious leader: Impact of the leader's mood on the mood of group members, group affective tone, and group processes. *Journal of Applied Psychology*, *90*, 295–305. https://doi.org/10.1037/0021-9010.90.2.295

Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group & Organization Management*, *34*, 170–205. https://doi.org/10.1177/1059601108331218

Vancouver, J. B., Tamanini, K. B., & Yoder, R. J. (2010). Using dynamic computational models to reconnect theory and research: Socialization by the proactive newcomer as example. *Journal of Management*, *36*, 764–793. https://doi.org/10.1177/0149206308321550

Vancouver, J. B., & Weinhardt, J. M. (2012). Modeling the mind and the milieu: Computational modeling for micro-level organizational researchers. *Organizational Research Methods*, *15*, 602–623. https://doi.org/10.1177/1094428112449655

Venkataramani, V., Zhou, L., Wang, M., Liao, H., & Shi, J. (2016). Social networks and employee voice: The influence of team members' and team leaders' social network positions on employee voice. *Organizational Behavior and Human Decision Processes*, *132*, 37–48. https://doi.org/10.1016/j.obhdp.2015.12.001

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228–243. https://doi.org/10.1037/a0027127

Willett, J. B. (1997). Measuring change: What individual growth modeling buys you. In E. Amsel & K. A. Renninger (Eds.), *The Jean Piaget symposium series. Change and development: Issues of theory, method, and application* (pp. 213–243). Mahwah, NJ: Lawrence Erlbaum.

How to cite this article: Lang JWB, Bliese PD, de Voogt A. Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*. 2018;71:255–281. https://doi.org/10.1111/peps.12260

Appendix A

R code for the army data

```
library(nlme)

################### PREPARE DATA ######################

## get and prepare data
library(multilevel)
data(univbct)

#prepare variables
univbct2<-univbct
univbct2$UNIT<-paste(univbct2$BTN,univbct2$COMPANY,sep="")

RMEANS<-aggregate(READY1 ~ UNIT,univbct2[univbct2$TIME==0,],mean)
names(RMEANS)<-c("UNIT","CREAD")
univbct2<-merge(univbct2,RMEANS,by="UNIT")

#delete persons with single observations
univbct2<-univbct2[rowSums(sapply(subset(univbct2,select=c(JOBSAT1,JOBSAT2,JOBSAT3)),is.na))<2,]

# at least three
members<-table(univbct2$UNIT)/3
univbct2<-univbct2[univbct2$UNIT %in% names(members[members>2]), ]

################## ANALYSIS #############################

## control settings for lme
csettings=lmeControl(opt="nlminb") # default setting
#alternative
#csettings=lmeControl(maxIter=3000,msMaxIter=3000,opt="optim",optimMethod="Nelder-Mead")

# 3-Level null model. The higher-level unit identifier is UNIT, the person identifier is SUBNUM
# The identifier used first in the list command refers to the highest level (here level-3).
# pdSymm specifies the nature of the random effects covariance matrix and refers to a symmetric
# matrix
# univbct2 is the dataset

M2a<-lme(JSAT ~ TIME, random = list(UNIT=pdSymm(~TIME),SUBNUM=pdSymm(~1)),
        data = univbct2,na.action=na.omit,control=csettings)

# 3-level CEM model. Change in the residual variance over time is added to the null model

M2b<-update(M2a,weights=varExp( form = ~ TIME))
anova(M2a,M2b)

#view additional model details with summary()
summary(M2b)

# 3-level with standardized group-level predictor
# scale(..) centers and standardizes the predictor
# scale(..,scale=FALSE) centers but does not standardize the predictor
# each predictor in the residual variance part of the model needs a separate varExp command
# otherwise R just estimates one coefficient for all predictors

m3a<-lme(JSAT ~ TIME*scale(CREAD), random = list(UNIT=pdSymm(~TIME),SUBNUM=pdSymm(~1)),
        data = univbct2,na.action=na.omit,control=csettings,weights=varExp( form = ~ TIME))
m3b<-update(m3a,weights=varComb(varExp( form = ~ TIME),varExp( form = ~ scale(CREAD))))
m3c<-update(m3b,weights=varComb(varExp( form = ~ TIME),varExp( form = ~ scale(CREAD)),
        varExp( form = ~ scale(CREAD)*TIME)))
anova(m3a,m3b,m3c)

# 2-level with unstandardized group-level predictor
# scale(..,scale=FALSE) centers but does not standardize the predictor

m3a2<-lme(JSAT ~ TIME*scale(CREAD,scale=FALSE), random = list(UNIT=pdSymm(~TIME)),
        data = univbct2,na.action=na.omit,control=csettings,weights=varExp( form = ~ TIME))
m3b2<-update(m3a2,weights=varComb(varExp( form = ~ TIME),varExp( form = ~
scale(CREAD,scale=FALSE))))
m3c2<-update(m3b2,weights=varComb(varExp( form = ~ TIME),varExp( form = ~
scale(CREAD,scale=FALSE)),
        varExp( form = ~ scale(CREAD,scale=FALSE)*TIME)))
anova(m3a2,m3b2,m3c2)

############### prepare data for MPLUS #####################

widedat<-reshape(subset(univbct2,select=c(UNIT,SUBNUM,TIME,JSAT,CREAD)),
        idvar = "SUBNUM",timevar = "TIME", direction = "wide",v.names="JSAT")
widedat$UNIT<-as.numeric(as.factor(widedat$UNIT))

widedat<-subset(widedat,select=c(UNIT,CREAD,JSAT.0,JSAT.1,JSAT.2))
write.table(widedat,file.choose(),
        col.names=FALSE,sep="\t",quote=FALSE,row.names=FALSE,na="9999")
```

Appendix B

MPLUS code for the army data

```
TITLE:  emergence model
DATA:   FILE IS rawdata.dat;
VARIABLE:       NAMES ARE clus cread js1-js3;
        USEVARIABLES = clus js1-js3;
                CLUSTER = clus;
                Missing are all (9999);
ANALYSIS:       TYPE = TWOLEVEL;
MODEL:
        %WITHIN%
        iw BY js1-js3@1;
        js1 (a21);
        js2 (a22);
        js3 (a23);
        %BETWEEN%
        ib sb | js1@0 js2@1 js3@2;
        js1-js3@0;
model constraint:
new (rvar);
new (delta);
a21 = rvar*exp(2*delta*0);
a22 = rvar*exp(2*delta*1);
a23 = rvar*exp(2*delta*2);
```

```
TITLE:  emergence model
DATA:   FILE IS rawdata.dat;
VARIABLE:       NAMES ARE clus cread js1-js3;
    USEVARIABLES = cread js1-js3;
    Missing are all (9999);
    CONSTRAINT = cread;
  DEFINE:       CENTER cread (GRANDMEAN);
MODEL:
    ib sb | js1@0 js2@1 js3@2;
    ib sb on cread;
    js1 (a21);
        js2 (a22);
        js3 (a23);
model constraint:
new (rvar);
new (delta);
new (delta2);
new (delta3);
a21 = rvar*exp(2*delta*0)*exp(2*delta2*cread)*exp(2*delta3*cread*0);
a22 = rvar*exp(2*delta*1)*exp(2*delta2*cread)*exp(2*delta3*cread*1);
a23 = rvar*exp(2*delta*2)*exp(2*delta2*cread)*exp(2*delta3*cread*2);
```

Note: The first script is for the 3-level CEM and the second script is for the 2-level CEM with a group-level predictor. Current versions of Mplus do not allow one to add predictors to the MODEL CONSTRAINT command when TYPE = TWOLEVEL so that three-level consensus emergence models with predictors cannot yet conveniently be fitted in Mplus.

Appendix C

Formulas for the extensions of the basic CEM

Model	Equation	Variance Components
Group or person-level predictor and predictor × time interaction (only changes from basic model in the second row of Table 1)	When at group-level: Level 3: $\beta_{00k} = \gamma_{000} + \gamma_{010}(\text{PRED}_k) + u_{00k}$ $\beta_{10k} = \gamma_{100} + \gamma_{110}(\text{PRED}_k) + u_{10k}$ When at person level: Level 2: $\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{PRED}_{jk}) + r_{0jk}$ $\pi_{1jk} = \beta_{10k} + \beta_{11k}(\text{PRED}_{jk})$ Level 3: $\beta_{00k} = \gamma_{000} + u_{00k}$ $\beta_{01k} = \gamma_{010}$ $\beta_{10k} = \gamma_{100} + u_{10k}$ $\beta_{11k} = \gamma_{110}$	When at group-level: $e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2 \exp[2\delta_1 \text{TIME}_i]$ $\exp[2\delta_2 \text{PRED}_k]$ $\exp[2\delta_3 \text{TIME}_i \text{PRED}_k])$ When at person-level: $e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2 \exp[2\delta_1 \text{TIME}_i]$ $\exp[2\delta_2 \text{PRED}_{jk}]$ $\exp[2\delta_3 \text{TIME}_i \text{PRED}_{jk}])$