Published in final edited form as:

Multivariate Behav Res. 2016 ; 51(2-3): 419–423. doi:10.1080/00273171.2015.1086955.

The data box and within subject analyses: A comment on Nesselroade and Molenaar (2016)

William Revelle* and Joshua Wilt**

*Northwestern University

**Case Western Reserve University

Abstract

Nesselroade and Molenaar (2015:in press) suggest that it is the relationships between latent variables within subjects that are invariant across subjects and thus the appropriate unit of analysis. We disagree and take the view that between factor correlations may differ systematically across subjects. And further, individual differences in these correlations may be an important source of information about each unique individual. Following from this premise, analyses of consistencies and differences between subjects of the within subject pattern of inter-factor correlations is a step towards an integrative science of behavior. We give several examples demonstrating how this approach has the potential to yield novel insights into personality functioning.

Keywords

Within-subjects covariation; affect; diurnal rhythm

Nesselroade and Molenaar (2015:in press) make a strong case for the need to focus on the individual as the unit of analysis and to understand how individuals vary over time and situations or both. We agree. Indeed, we have argued previously (Ortony, Norman, & Revelle, 2005) that the systematic patterning of one's feelings, thoughts, motives and behavior over time show a person's unique temporal signature that is one of three independent levels of analysis of personality (within individual, between individuals, between groups of individuals) (Revelle & Condon, 2015).

Nesselroade and Molenaar (2015:in press) consider the three-way analysis of persons, variables, and occasions first introduced as the data box by Raymond Cattell (1946) and subsequently elaborated by him (Cattell, 1966). They propose a variation of Cattell's P technique (the correlation of variables over occasions for single individuals) with the introduction of different variables for different individuals.¹ That is, they want to "explode" the data box such that each Variable $\times$ Occasion slice may be unique for each individual (see their Figure 5). This leads to their claiming that the variables may not have the same meaning for each individual. They speak favorably of the computer generated Item

Correspondence concerning this article should be addressed to William Revelle, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208. Phone: (847)-491-7700. FAX: (847) 491-7859. revelle@northwestern.edu.

¹Unfortunately, to confuse the matter, Cattell (1978) reversed the labels for O and P techniques from his prior and subsequent work.

Response Theory (IRT) approach of Embretson (1998) which considers how to design unique ability items for each individual. This technique of computer generated parallel ability items instantiates the item sampling technique proposed by Lord (1955, 1977). Because the items may be unique to each subject, such item generation methods do not require, indeed, do not allow strict factorial invariance. These techniques do, however, emphasize structural invariance at the level of the cognitive model. Such use of randomly parallel items is appropriate if the items all come from one domain or are sampled from domains with a known structure. This is clearly appropriate for developmental studies of ability but is it appropriate for the more general case? We think not.

At first glance, the emphasis upon unique items for each subject is similar to the position taken by George Kelly (1955) who argued for understanding each individual by understanding the unique structure of their construal of the world. However, Kelly's completely idiographic model did not allow for comparisons across people, whereas Nesselroade and Molenaar (2015:in press) propose that, although the measures might be idiosyncratic, the structural relationships between constructs are the same and allow for theory building.

That is, Nesselroade and Molenaar (2015:in press) suggest that it is the relationships between latent variables within subjects that are invariant across subjects and thus the appropriate unit of analysis. They dismiss the notion of examining the invariance across subjects of the structure of the factor loadings on particular observed variables within subjects and look for invariance of the higher order between factor structure (see Nesselroade, Gerstorf, Hardy, & Ram, 2007, for a similar argument). This reflects the earlier finding of Lebo and Nesselroade (1978) that individuals endorse different items to reflect the presumptively same latent variable. This is an important point and leads to their basic argument. However, here is where we disagree. By emphasizing the variability across subjects of item loadings but the invariance of factor intercorrelations, we believe that Nesselroade and Molenaar (2015:in press) have thrown the baby out with the bathwater. For, as we have previously shown, and will review in the following pages, there are times in which the item loadings are invariant but the between factor correlations differ systematically across subjects. These correlations may be an important source of information about each unique individual.² The between factor correlations are not only stable across time but also relate to other external criteria.

Within subject variation in mood

Some of the studies considered by Nesselroade and Molenaar (2015:in press) involved repeated measurements over years as part of their emphasis upon understanding people over their lifespan; others (e.g., Lebo & Nesselroade, 1978; Zevon & Tellegen, 1982) involve repeated measurements over a few weeks. We do not dismiss the power and elegance of the long term longitudinal studies, but want to consider the alternative example of short term, intense studies of individuals over the period of a few days to a few weeks. These are not

²The examples we will present focus on the difference in correlations between constructs across individuals. For a more general framework by which to approach individual differences in within-person parameters, see Oravecz, Tuerlinckx, and Vandekerckhove (2011).

studies of changes in ability, but rather of fluctuations in mood. Lebo and Nesselroade (1978) was a pioneering study of mood using a paper and pencil daily diary technique, with five women in the months preceding and following childbirth. Using the same P-factor technique, Zevon and Tellegen (1982) found two reliable dimensions of affect for 25 undergraduate and graduate students. Some of the more recent work examining mood variation takes advantage of an exciting change in technology over the past few years which has made it far easier to collect within subject mood measures across time for many subjects. Although previously possible using paper and pencil diary techniques, the introduction first of “Personal Digital Assistants” (PDAs) and subsequently of “smart phones” and “apps” has led to an explosion of within subject mood measurement with multiple mood measurements per day repeated over multiple days even when the subjects are far from the lab (Wilt, Condon, & Revelle, 2011). This, in combination with easy access to multilevel modeling software in the open source statistical system R (R Core Team, 2015) [e.g., lme4 Bates, Maechler, Bolker, & Walker (2015), and nlme Pinheiro, Bates, DebRoy, Sarkar, & the R Core team (2015)] for examining the within and between subject structure of mood has truly revolutionized the way we can think about data. Analyzing this kind of data is what makes the Nesselroade and Molenaar (2015:in press) article so timely.

Within subject correlations are stable and predictable

We have used daily diaries, PDAs, and smart phones to examine individual differences in the structure of mood (Rafaeli, Rogers, & Revelle, 2007; Wilt, Funkhouser, & Revelle, 2011). In doing so, we have found consistent patterns across subjects of within subject factor loadings but reliable individual differences in the correlation of these factors. We have found this when examining the factor structure of positive and negative affect (PA/NA: Watson & Tellegen, 1985) as well as energetic and tense arousal (EA/TA: Thayer, 1970, 1978)). At the aggregate (between subject) level, the latent factors of PA and NA are nearly orthogonal, and those of EA and TA are orthogonal (Rafaeli & Revelle, 2006; Rafaeli, Rogers, & Revelle, 2007).³ However, at the individual level (within subjects) the between factor correlations of EA and TA range from synchrony (positive correlations exceeding .3) to asynchrony (zero correlations) to desynchrony (negative correlations less than -.3). Using diary studies and PDAs we have shown that these differences in correlational structure are temporally stable across at least a two week interval ($r=.65$) (Rafaeli et al., 2007).

Not only are the within subject factor correlations stably different, they are also predictable (Rafaeli et al., 2007). Although standard trait measures of temperament such as Extraversion or Neuroticism did not reliably predict the within subject correlation, measures of individual differences in valence focus (Feldman, 1995) did. Valence focus refers to the tendency for some individuals to emphasize the valence of mood items, whereas others focus on the arousal aspects. Rafaeli et al. (2007) measured valence focus by finding the individual weights in the perception of affective similarities using an INDSCAL analysis (Carroll & Chang, 1970). Each subject was asked to rate the conditional probability between each of 120 pairs of 16 circumplex affect markers (e.g., how likely would one experience feeling

³These data consist of 3,890 cases on 75 items collected in multiple different studies over 10 years and are available as the msq data set in the psych package (Revelle, 2015) in R.

surprise if feeling happy?). These probabilities were converted into distances and then the nomothetic solution for all subjects was found using multidimensional scaling. After rotating the MDS solution to match that reported by Feldman (1995) (representing a valence and an arousal dimension) we used OLS to solve for the rated distances between stimuli (d_{ij}) for each individual (I) in terms of an individual weighting (w_{lk}) of the two dimensions ($k=1,2$) and the group MDS solution:

$$d_{ijl} = \sqrt{w_{lk}(s_{ik} - s_{jk})^2} \quad (1)$$

where the d_{ijl} represents the distance between the i th and j th affective marker for the l th individual, and the w_{lk} are the weights for dimension k (1 to 2) for the l th individual, and the s_{ik} and s_{jk} represent the MDS based locations of the i th and j th stimuli on the k th dimension. If everyone weights the dimensions equally, the distances are just those found in the group space. But, if the weight is greater for valence than for the arousal dimension for some individuals and less for other individuals, we can capture the degree of valence focus of the individuals. Thus, the individual subject weights (w_{lk}) on the valence dimension were used to predict the factor correlation between PA and NA from the within subject analysis. Those participants who emphasized valence more than arousal were more likely to be affectively desynchronous (to show a negative correlation between PA and NA) than were those who demphasized the valence dimension but emphasized the arousal dimension.

Two subsequent studies (Wilt, Funkhouser, & Revelle, 2011), using cell phone text messaging to elicit responses six times a day for two weeks, found that the within subject correlation of EA and TA related to the within subject correlation of the perception of Threats and Incentives: those subjects who perceived threats as incentives had a higher correlation between EA and TA than did those who saw threats negatively. In particular, those subjects who viewed threatening situations as challenging tended to be energized and tense, whereas those who viewed those situations as threatening tended to be deenergized (tired) and tense. That is, variations in the factor correlations were the important signals, not the individual within subject factor loadings.

The studies just described (Rafaeli et al. (2007); Wilt, et al. (2011)) demonstrated individual differences in within-person correlations between constructs, yet they are both limited by the reliance on traditional multilevel modeling (MLM) procedures. This is because traditional MLM works at the level of observed variables instead of latent constructs and thus by default constrains the loadings of indicators of composite manifest variables to be equivalent across people without explicitly modeling that assumption (Bliese, Chan, & Ployhart, 2007). Therefore, our findings may have been artifacts of the very problem that Nesselroade and Molenaar highlighted, that is, manifest variables might be indicators of different latent variables for different people. Multilevel factor analytic techniques would provide a stronger test of the general idea that the latent within-person correlation of affects is reliably different and predictable across people, as these techniques allow for estimation of the factor loadings of manifest variables on latent factors and tests of competing models of the latent structure at the within-person level (Muthén & Asparouhov, 2011; West & Ryu, 2007). Brose,

Voelkle, Lövdén, Lindenberg, & Schmiedek (2015) used multilevel factor analyses to determine the average within-person associations between positive and negative affect, and also to reveal that this association did indeed vary between people as a function of trait positive affect (positively), personal growth (positively), and major life stressors (negatively).

Between subject differences in the within subject diurnal rhythm

In a further analysis of the Wilt et al. (2011) data set, we have examined how within subject measures of EA and TA relate to the time of day at which the measures were taken. Using item level (e.g., responses to the words: energetic, alert, sluggish, irritable, relaxed) or scale level (EA, TA, PA, and NA composites) we can find the circular correlation between the items/scales and time of day⁴. From this, using basic circular statistics (Jammalamadaka & Lund, 2006; Pewsey, Neuhaus, & Ruxton, 2013; Winfree, 2001) or cyclic analysis (Chow, Hamaker, Fujita, & Boker, 2009), it is possible to find individual differences in the acrophase or phase delay for our participants⁵. For variables that show a circadian rhythm, the acrophase is that time of day at which the relevant state achieves a maximum. It may be estimated by finding the regression of the state variable S_i for the i th subject, S_i , with the sine and cosine of the time of day (expressed in radians) for each subject (Chow et al., 2009; Jammalamadaka & Lund, 2006):

$$S_i = a_i \cos(t) + b_i \sin(t). \quad (2)$$

The acrophase, ϕ , is then the arc tangent of these two slopes. That is:

$$(\phi_i = \tan^{-1}(-b_i/a_i)) \quad (3)$$

This phase angle for each subject, ϕ_i , may then be correlated⁶ with some trait level measure, e.g., extraversion. When doing so, we are finding an index, based upon the within subject correlation with time of day and the between subject values of the trait. For the first data set from Wilt et al. (2011), ($N=40$) the correlation of trait level extraversion with the phase of Energetic Arousal was .37 and -.34 with Negative Affect. The phase of Energetic Arousal correlated also with Agreeableness (.50) and Openness (.38), suggesting that people with later diurnal rhythms of alertness are outgoing, agreeable, and open. Once again, it was the correlation of the latent scores (the EA and TA factors) with an external measure (time of day) that carried the important signal.

⁴In the following discussion, we use terms taken from the biological study of circadian rhythms that have a phase of roughly 24 hours. An example of such a rhythm is body temperature which shows striking individual differences in the phase delay of when one achieves a minimum temperature (Baehr, Revelle, & Eastman, 2000). For such data, basic descriptive statistics such as the mean, variance, and correlations need to take into account the cyclical nature of the data. As Winfree (2001) discusses, these are projections from linear time on to a circle. "... the practical mathematics of things periodic and rhythmic is mostly about points on the circle" (Winfree, 2001, p 6). Thus, the mean time of observations at 22:00 and 2:00 is 24:00, not 12:00. Similar issues arise when examining directions in migratory birds, or when averaging wind directions. More advanced cyclic statistics can examine other periods (e.g., weekly) in the data (e.g., Chow et al., 2009; Liu & West, 2015).

⁵We used the cosinor function in the psych package in R to find ϕ_i for every subject.

⁶The correlation is a linear by circular correlation rather than the conventional Pearson correlation.

Conclusion

Nesselroade and Molenaar's emphasis upon decomposing the data box to analyze patterns at the single subject level is to be commended. However, in contrast to their recommendation to look for invariance in the structural relationships between constructs, we have given three examples of when the within-subject correlational pattern of observed or latent scores is itself an important signal.

Acknowledgments

Preparation of this manuscript was funded in part by grant SMA-1419324 from the National Science Foundation to WR and by a National Institute for Mental Health, National Research Service Award Grant F31-MH093041 to JW. We wish to thank Dan Ozer and Scott Hofer for helpful discussions of the issues raised in this article and Stephen West for his extensive comments and suggestions.

References

- Baehr EK, Revelle W, Eastman CI. Individual differences in the phase and amplitude of the human circadian temperature rhythm: with an emphasis on morningness-eveningness. *Journal of Sleep Research*. 2000; 9(2):117–127. [PubMed: 10849238]
- Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. 2015 Retrieved from <http://CRAN.R-project.org/package=lme4> (R package version 1.1–8).
- Bliese PD, Chan D, Ployhart RE. Multilevel methods: Future directions in measurement, longitudinal analyses, and nonnormal outcomes. *Organizational Research Methods*. 2007; 10(4):551–563.
- Brose A, Voelkle MC, Lövdén M, Lindenberg U, Schmiedek F. Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*. 2015; 29:55–71.
- Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*. 1970; 35:283–319.
- Cattell RB. Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology*. 1946; 36:88–102. [PubMed: 21015326]
- Cattell, RB. The data box: Its ordering of total resources in terms of possible relational systems. In: Cattell, RB., editor. *Handbook of multivariate experimental psychology*. Chicago: Rand-McNally; 1966. p. 67-128.
- Cattell, RB. The scientific use of factor analysis. New York, NY: Plenum Press; 1978.
- Chow S-M, Hamaker EL, Fujita F, Boker SM. Representing time-varying cyclic dynamics using multiple-subject state-space models. *British Journal of Mathematical and Statistical Psychology*. 2009; 62(3):683–716. [PubMed: 19200409]
- Embretson SE. A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*. 1998; 3(3):380–396.
- Feldman LA. Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*. 1995; 21:806–817.
- Jammalamadaka S, Lund U. The effect of wind direction on ozone levels: a case study. *Environmental and Ecological Statistics*. 2006; 13(3):287–298.
- Kelly, G. The psychology of personal constructs. New York, NY: Norton; 1955.
- Lebo MA, Nesselroade JR. Intraindividual differences dimensions of mood change during pregnancy identified in five p-technique factor analyses. *Journal of Research in Personality*. 1978; 12(2):205–224.
- Liu Y, West SG. Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*. 2015 n/a–n/a.
- Lord FM. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*. 1955; 20(1): 1–22.

- Lord FM. Some item analysis and test theory for a system of computer-assisted test construction for individualized instruction. *Applied Psychological Measurement*. 1977; 1(3):447–455.
- Muthén, B.; Asparouhov, T. Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In: Hox, JJ.; Roberts, JK., editors. *Handbook of advanced multilevel analysis*. New York, NY: Routledge/Taylor Francis; 2011. p. 15–40.
- Nesselroade JR, Gerstorf D, Hardy SA, Ram N. Focus article: Idiographic filters for psychological constructs. *Measurement: Interdisciplinary Research and Perspectives*. 2007; 5(4):217–235.
- Nesselroade JR, Molenaar PCM. Some behavioral science measurement concerns and proposals. *Multivariate Behavioral Research*. 2015 in press.
- Oravecz Z, Tuerlinckx F, Vandekerckhove J. A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*. 2011; 16(4):468–490. (Supplemental). [PubMed: 21823796]
- Ortony, A.; Norman, DA.; Revelle, W. Affect and proto-affect in effective functioning. In: Fellous, J.; Arbib, M., editors. *Who needs emotions? The brain meets the machine*. New York, NY: Oxford University Press; 2005. p. 173–202.
- Pewsey, A.; Neuhäuser, M.; Ruxton, GD. *Circular statistics in R*. Oxford, UK: Oxford University Press; 2013.
- Pinheiro J, Bates D, DebRoy S, Sarkar D. the R Core team. nlme: Linear and nonlinear mixed effects models [Computer software manual]. 2015 Retrieved from <http://CRAN.R-project.org/package=nlme> (R package version 3.1–121).
- R Core Team. Vienna, Austria: 2015. R: A language and environment for statistical computing [Computer software manual]. Retrieved from <http://www.R-project.org/>
- Rafaeli E, Revelle W. A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*. 2006; 30(1):1–12.
- Rafaeli E, Rogers GM, Revelle W. Affective synchrony: Individual differences in mixed emotions. *Personality and Social Psychology Bulletin*. 2007; 33(7):915–932. [PubMed: 17551163]
- Revelle W. psych: Procedures for personality and psychological research [Computer software manual]. 2015 <http://cran.r-project.org/web/packages/psych/>. (R package version 1.5.6).
- Revelle W, Condon DM. A model for personality at three levels. *Journal of Research in Personality*. 2015; 56:70–81.
- Thayer RE. Activation states as assessed by verbal report and four psychophysiological variables. *Psychophysiology*. 1970; 7(1):86–94. [PubMed: 5492738]
- Thayer RE. Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*. 1978; 2(1):1–34.
- Watson D, Tellegen A. Toward a consensual structure of mood. *Psychological Bulletin*. 1985; 98(2): 219–235. [PubMed: 3901060]
- West SG, Ryu E. Commentary: Assumptions and challenges of an idiographic-nomothetic approach to measurement: A comment on Nesselroade, Gerstorf, Hardy and Ram. *Measurement*. 2007; 5(4): 259–263.
- Wilt, J.; Condon, D.; Revelle, W. Telemetrics and online data collection: Collecting data at a distance. In: Laursen, B.; Little, TD.; Card, N., editors. *Handbook of developmental research methods*. New York, NY: Guilford Press; 2011. p. 163–180.
- Wilt J, Funkhouser K, Revelle W. The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*. 2011; 45:309–321.
- Winfree, AT. *The geometry of biological time*. 2nd. New York, NY: Springer Verlag; 2001.
- Zevon MA, Tellegen A. The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*. 1982 Jul; 43(1):111–122.