Measuring states and traits in motivation and emotion.
A new model illustrated for the case of work engagement

Robert A. Roe 1, Ilke Inceoglu 2

1 Maastricht University, School of Business and Economics

2 Surrey Business School, University of Surrey

To be published in:

Leong et al. (Eds.), ITC International Handbook of Testing and Assessment.

New York: Oxford University Press. 2016.

Do not quote without permission of the authors.

Author Note

Correspondence address: Ilke Inceoglu, Surrey Business School, Faculty of Business, Economics and Law, University of Surrey, Guildford, GU2 7XH, Email: i.inceoglu@surrey.ac.uk.

Acknowledgements

The authors of this chapter gratefully acknowledge the willingness of Kimberley Breevaart and Arnold Bakker to make the data from their study ‘Daily transactional and transformational leadership and daily employee engagement’ published in the Journal of Occupational and Organizational Psychology (2013) available for secondary analysis in the research presented here. The analysis and interpretation of the data presented in this chapter are the sole responsibility of the authors.

The contribution of the first author was written while receiving medical treatment at the Department of Hematology and Oncology of the University Hospital Leipzig. The excellent treatment and care by Prof. Dietger Niederwieser, Dr. Simone Heyn and nursing staff of the Department, which was most helpful in expediently completing the text, is gratefully acknowledged.

Abstract

In this chapter we argue that psychological measurement in the field of motivation and emotion is marked by a considerable degree of ambiguity, partly because these phenomena are poorly defined, but mainly because they are dynamic – motivation and emotion are about changes in behavior – while measurement designs and techniques are predominantly addressing individual differences, which are typically assumed to be stable. Building on recent work, which has distinguished between differential and temporal approaches to measurement and prediction (see Roe, 2014), we discuss the merits and limitations of prevailing differential methods.

Next, we consider how researchers have tried to overcome the challenge of dynamic measurement with the help of state-trait models, and note that there are conceptual and logical problems, limiting the use of these models. To overcome these problems we propose a new measurement model, which focuses on individuals' dynamic trajectories, defined with reference to a time frame of length L , starting at moment M , and comprising N observations. We show how this model can be used to describe subjects' motivational states and to redefine traits in a dynamic way. The logic and utility of this approach is illustrated for work engagement – a well-investigated phenomenon in the current literature on work motivation.

Keywords: Motivation, Measurement, Stability, Change, Time, State-Trait, Work Engagement, Employee Engagement, Motives, Emotion, Affect

Measuring states and traits in motivation and emotion.

A new model illustrated for the case of work engagement

Introduction

Considering the scope and volume of the body of theory and research dedicated to motivation and emotion it would be excessively pretentious to cover the measurement of both phenomena in a book chapter like the present one. Even if we confined ourselves to motivation and emotion in relation to work, it would be beyond the limits of our competence and ambition to give a comprehensive and informative treatment of measurement issues. Apart from lack of agreement on definitions, the psychological processes involved in the *arousal, direction, intensity, and persistence of behavior* (motivation; cf. Cofer & Apply, 1964), and in the *affective experiences associated with internally or externally triggered bodily states* (emotion; Kleinginna & Kleinginna, 1981a, 1981b) are simply too wide-ranging to all be captured in a single text. For general publications on this matter – mostly focusing on motivation and emotion at work – we gladly refer to other sources (Ruth Kanfer, Chen, & Pritchard, 2008; S. Kaplan, Dalal, & Luchman, 2013; Lane, 2004; Latham & Pinder, 2005; Ployhart, 2008; N. L. Stein & Oatley, 1992).

Our focus will be on one particular issue that poses a formidable challenge to scholars as well as practitioners, namely the fact that motivational and emotional phenomena - as the very words express - are essentially dynamic. They are very much related to what Kanfer (1990) refers to as “the continuing stream of experiences”, or what Barker calls the “stream of behavior” (Barker, 1963), and James the “stream of consciousness” (James, 1890). Surprisingly, there are hardly any works on the methodology of measuring motivation and emotions from the view perspective of how they influence the evolving behavior stream.

Measurement methods have sometimes been dynamic, but rarely comprehend the inner processes and the observable behaviors.

What methods would be needed to measure motivation? Obviously, they would depend on the nature of the motivational process. Homeostatic processes, which involve equilibrium-seeking and the management of deviations from a naturally given or deliberately chosen set-point, require methods that show the wavelength, amplitude and frequency of the change trajectories. This approach has, in fact, been applied to numerous motivational processes with a physiological substrate. Motivation as processes of (adaptive) responses to ongoing environmental changes needs measurement methods that capture all such changes – as a sequence of discrete and qualitative different episodes. These methods should assess sequences, speed / duration of transitions, duration of each part, and so on. Of course, the intensity (cf. amplitude) of responses might also be included. Preference or goal-based forms of motivation would require a similar approach to measurement, capturing ongoing sequences in behavior and the changes in underlying states and goals.

The literature provides only few examples acknowledging motivational processes related to change of activities (e.g. Atkinson & Birch, 1970; Dalal & Hulin, 2008; Kirchberg, 2014; Mitchell, Harman, Lee, & Lee, 2008). Most of them chose a limited time frame and a limited number responses or actions. The large majority of studies do not cover multiple actions, but just one! This is remarkable, because motivation is essentially about the change of behavior from A to B, B to C etc.

As for emotions, which are often seen as being evoked by perceived events and occurrences, but which may also accompany or elicit (subsequent) responses and actions, similar types of measurement methods would be needed. They should either fit the notion of a dynamic-equilibrium – with deviations seeking a neutral or central mood point – or capture the sequence of dominant emotional states (i.e. states of elevated activation) or the way in

which multiple emotional states evolves over time (e.g., Ruef & Levenson, 2007; Schubert, 1999)

We should realize that the multiplicity of actions and emotion triggering events, even in a time window as short as a day is – for many people – very large indeed, and that their proper measurement would result in something comparable to the output of a range of multi-channel polygraphs. This even applies to the subset of work-related activities, which is embedded in numerous other activities. Information on the number of activities performed per day is virtually lacking in the literature. We only know from diary studies of managers that they engage in many hundreds of different activities in a single day (Mintzberg, 1973; Tengblad, 2002). Although research has shown that people can experience multiple emotions at the same time (e.g., positive and negative affect; Warr, Bindl, Parker, & Inceoglu, 2014), we know very little about how many emotions people experience simultaneously and sequentially within a single day, or across multiple days. What comes closest are studies about multiple emotions in moral judgment (U. Kaplan & Tivnan, 2014) or while listening to music (Schubert, 1999). Yet the majority of studies deal with one or two emotional states at the time, singled out from all other states.

In contrast to what would be needed, and hard to reconcile with the dynamic nature of the phenomena, the models and methods for measurement are predominantly based on the assumption of stable individual differences (e.g., Ployhart, 2008). They pertain to the intensity or amplitude of a single attribute of a single motivational or emotional phenomenon that is supposed to characterize a person in general. The measurements they produce indicate the typical strength of a person's emotion or motivation such as a person's typical level of achievement motivation or anxiety. Diary researchers no longer subscribe to this idea, as they assume that the measure characterizes the person in a limited time-window, such as a day.

In the following sections we will take a closer look at methods and models of measurement, assess the current state of the art, and propose a new approach. After that, we will consider a specific field of research, i.e., that of work engagement, briefly assess the current state of measurement and apply the newly developed method.

Current approaches to research and measurement: differential and temporal

During the past decade researchers have increasingly realized the need to differentiate between research designs that analyze individual differences (between-subject designs) and those that analyze changes over time (within-subject) (Molenaar, 2004; Navarro, Roe, & Artiles, 2014; Roe, Gockel, & Meyer, 2012; Van de ven & Poole, 2005). Yet, many researchers still seem to believe that evidence on between-subject and within-subject variation and covariation is exchangeable, which is generally not the case (e.g., Roe, 2014a). There is no logical ground for this belief and the conditions under which a relationship is to be expected on theoretical grounds (“ergodicity”, which means that each subject shows the exactly same dynamical pattern, and that single pattern is time-invariant; Molenaar, 2007) are extremely rare (Molenaar & Campbell, 2009). The vast majority of studies in psychology, certainly in such domains as work or education, are of differential nature and have used a single moment in time, even though the number of studies using more time moments has begun to increase. For both reasons these designs are inherently unsuited to assess change, even though results are sometimes interpreted as showing that “a change in one variable is associated with a change in another variable”. One typical example comes from the leadership domain, where between-leader style–outcome relationships have often been interpreted as within-leader style–outcome relationships and, implicitly assuming causality, been used as the basis for training leaders with the expectation of getting better outcomes.

Differential designs have also been used in longitudinal studies, taking measurements at multiple time-moments. Many studies have taken only 2 to 5 time moments, recent diary-studies go beyond this and use daily measures during one or two weeks, sometimes a month. Although the data would allow analyzing changes within subjects, researchers have typically used differential analyses (e.g., cross-lagged panel analysis, growth modeling), which means that the focus is on how differences between subjects established at various time moments co-vary with each other. The hallmark of differential research is “prediction”, which is conceived as explaining variance in a dependent variable from an independent variable, measured simultaneously (or earlier). As the measurement of the variables has been completed at the moment the prediction is made, there is no direct implication for what will happen in the future. That is, differential prediction is postdiction, and one has no bases for extrapolating to the future, unless one assumes that a relationship that was found in the past will also hold in the future. For instance, the validity of a selection test can only be assessed after predictor and criterion data have been collected, and shows how well the criterion *was* predicted over a time interval (L) starting at moment (M1). One needs to assume that the relationship between predictor and criterion measures remains the same for intervals of similar length starting at any later moment (M2, M3 etc.) – in other words that the validity generalizes over time.

For studying change one needs temporal designs, which allow analyzing variation and covariation within subjects over a number of time-moments. Just like one time-moment suffices to apply a differential design, a single person suffices to apply a temporal design. Of course, combined differential-temporal designs, with multiple subjects and multiple time-moments are more informative. For the study of change, the preferred analysis is one that starts with assessing the change trajectory of every single subject, and proceeds with a comparison of individual change trajectories (Molenaar & Campbell, 2009). This leads to the

between-subject variability of within-subject variability analyses that are well known from research in developmental psychology (Nesselroade, 1991; Nesselroade & Ram, 2004). Time series based on a single subject or multiple subjects can be used to make “forecasts” of what will happen in the future, based on the dynamic pattern inherent in it – which is an important capability that is lacking from purely differential prediction.

We should briefly mention that the two approaches are also different in their ontological assumptions. The differential approach assumes that psychological phenomena are universal, that is, manifest themselves in different degrees in different subjects. It is based on the postulate of Uniformity of Nature (Hume, 1748; see Salmon, 1953) which states that any sample of matter is suitable for scientific study of its properties – in this case of the human species. The assumption is that individuals are exchangeable and will reveal the same stable characteristics, apart from errors. One could think of the stability of intelligence, self-efficacy, or anxiety. To the degree that change occurs, it is conceived as a transition from one stable level to another one. The temporal approach, in contrast, assumes that everything in human life is subject to change (cf. Heraclitus’ “*Panta Rei*”) and that stability is a special form of change only occurring in episodes (see radical temporalism below; Roe, 2005; Roe, 2008b).

For this chapter it is important that the differential-temporal distinction does not only affect the logic and design of research studies, but also measurement. We should emphasize that the psychometric theories and techniques commonly used in psychological research are rooting in the differential paradigm and that for reasons we will explain they are ill-suited to measure change (also Molenaar, 2008).

The best known and most used psychometric theory is Classical Test Theory (CTT; Gulliksen, 1950; Lord, Novick, & Birnbaum, 1968); it starts from a model in which a person’s score X on a test is the sum of a “true score” T and an uncorrelated error term: $X = T$

+ e . The theory provides a number of methods to estimate and reduce e , based on the notion of maximizing reliability. It also covers the relationship between scores on two tests, which is known as validity. Without going into detail, three things are worth noting:

1. all estimations are based on test scores (or item scores) of different individuals;
2. the model has a single true score for each individual on each tested attribute
3. e can be reduced by minimizing change over time.

In practical terms: to determine a person's true score one needs information of other people, and the test needs to be insensitive to change. These characteristics place CTT firmly into the differential research approach, and outside of the temporal approach.

In the last few decades researchers have increasingly embraced Item Response Theory (IRT), which “is a rubric for a family of measurement models that describe the relationship between an individual's performance on a test item and his or her standing on a continuous latent trait”, indicated by a “theta score” (Reise & Waller, 2002, p. 88) The basic tenet of IRT is that individuals with a higher standing on a latent trait (e.g. higher levels of ability, more favorable attitudes towards something) are more likely to pass an item (or endorse a response) that reflects the underlying latent trait (Guion, 2011). IRT models are similar to the CCT model with regards to the differential approach, but they are explicitly made for multi-item tests and have one or more additional parameters that refer to properties of the test items and guessing. There are no time parameters in IRT-models and items susceptible to change are likely to be removed in test construction, since items that behave differently in several trials while other items keep behaving the same are more likely to be dropped. Thus, like CCT models, they are strongly rooted in the differential paradigm, they require measured qualities to be stable, and they have no capability to capture change.

All this does not mean that researchers have not used CCT and IRT to develop measures to study change. On the contrary, there are countless “longitudinal” studies in

which tests and surveys constructed on the basis of these psychometric theories have typically been administered at two, three or more moments in time. Obviously, researchers have failed to see the contradiction between the notion of a single true score or theta score that – without constraining conditions – hold forever and the notion of change. This is not just a matter that can be argued away by introducing a post hoc assumption that a true score or theta holds for a certain moment and that one can define as many true scores or thetas as one wants. A fundamental issue is that the models and the measurement scales derived from them are built to be maximally sensitive to differences between subjects and to be minimally susceptible to change, whereas what one would need to measure change is sensitivity to change and robustness against re-use (Roe, 2008a). This means that repeated measurements as such do not change the measurement values, which is an essential requirement since one cannot measure unless the measurement standard remains the same. It may be argued that tests and surveys based on CTT or IRT may still show differences when applied multiple times in a longitudinal study. In fact, they often do; but they may fail to pick up significant changes because test construction and calibration favor items that consider change as noise. Moreover, they often show learning or practice effects, which limits their suitability for measuring change. More generally, it is difficult for multi-item instruments to show full measurement equivalence over time. An additional problem is that CTT and IRT produce measures with interval properties, which limit their capability to measure change.

An ideal way to measure change is with instruments that produce ratio scales (Stevens, 1958) and that allow continuous measurement over time (not to be confused with measurement using a continuous scale) – as in ECG, EEC and other physiological measurements. Ratio scales are preferable because there is a zero point, which allows for the case that the phenomenon is absent (e.g. no commitment to an employer because one is not employed), a property dearly missing from interval scales. “The variable perspective restricts

our view of people's actions and interactions in organizations. It produces the illusion that the behaviors under study are always present, and prevents us from seeing how they emerge and disappear during phases of the individual's or organization's life time" (Roe, 2005, p. 17).

Ratio scales can also show a rate of growth or decline. Continuous measurement means that measurements can go on for extended periods of time and with arbitrary grids, without affecting the measured values. We are not aware of general psychometric theories for such measurement, but there are several sources discussing such issues as setting a base-line and assessing periodicity (e.g., Fishel, Muth, & Hoover, 2007; Wieland & Mefferd, 1969).

However, there are also other techniques, which are based on gains in performance or other learning effects (e.g., Guthke, 1993)

The research literature on motivation and emotion shows that researchers have used both differential and temporal approaches to measurement, with a clear prevalence of the first ones. Differential approaches clearly prevail. Numerous researchers have used notions and methods based on psychometric theories that were originally developed to investigate differences between people in ability and personality. See for example Erez and Judge (2001; 2010); Fernet, Gagné and Austin (2010), Hopp, Rohrmann and Hodapp (2012), Kooij, Bal, and Kanfer (2014). Thus, with some significant exceptions, which deserve attention and credit, the field has largely developed trait-like notions and test-like methods. Many of these measure subjects' *level* of motivation (ranging from low to high) using a single overall indicator of motivation, two or more types of motivation (like intrinsic and extrinsic), or presumed sources of motivation which might play a role in *arousing* motivation (like the Achievement Motivation Inventory; Byrne et al., 2004; or in the SHL Motivation Questionnaire; SHL, 1992) . They typically consider the motivation for a particular type of behavior or performance, which is valued in an organization or educational context. This means that a qualitative change in behavior (for instance: stopping computer work, attending

another client, working on another paper, writing an email, attending a meeting, making a phone call with home, talking to colleagues) – which is the hallmark of motivation – is not being considered. A noteworthy exception is Value-Expectancy Theory (Vroom, 1964), which predicts which of several alternative behaviors a person will chose, based on valences, expectancies, and instrumentalities. Noteworthy is that the behavioral options are chosen by the researcher, not by the acting person; which clearly shows the differential basis of Vrooms model: a comparison of subjects by the researcher. The persistence of behavior amidst pressing or tempting alternatives is rarely considered, which is also not surprising in the light of the differential nature of the measures.

Many researchers have used also methods from psychophysiology and biochemistry to assess motivational and emotional states. These comprise electro-physiological measures such as skin conductivity, muscle tension, heart rate, evoked potentials etc. and endocrine measures, such as levels of various hormones (Balthazart, de Meaultsart, Ball, & Cornil, 2013; Capa, Audiffren, & Ragot, 2008; Frank & Fossella, 2011; Kreibig & Gendolla, 2014; Kukolja, Popović, Horvat, Kovač, & Ćosić, 2014; L. Liu, Zhang, Zhou, & Wang, 2014; Schmidt, Lebreton, Cléry-Melin, Daunizeau, & Pessiglione, 2012; Shimomitsu & Theorell, 1996; M. Stein, Egenolf, Dierks, Caspar, & Koenig, 2013; Vecchiato et al., 2014; Wise, 2004). As said, these methods have the advantage of producing measurements on a ratio scale, which allow measuring changes within persons as well as comparing levels and changes between persons. However, there are issues related to differences in baseline levels, which should be considered in comparing results from difficult subjects. Psychophysiological measures may also be combined with behavioral assessment techniques, subjective rating scales, and questionnaires (e.g., Magnusson & Endler, 1977). The ultimate choice of measures (e.g. focus on psychological, physiological or social well-being) should of course be guided by the research objectives (Warr, 2013).

The measurement of states and traits: current practice and a new model

An interesting development regarding the measurement of the dynamics of motivation and emotion is the adoption of state-traits models, based on the distinction between variable states and stable traits. In this section we will take a look at the state-trait distinction and the way in which it has been applied in theory and research. We note that state-trait models do not resolve the problems signaled above, since they are rooted in differential thinking. Therefore, we will propose a new method that starts from temporal thinking and gives a better view of temporal dynamics and individual differences, and is free of inconsistencies. Since we are dealing with a vast research domain we will from here on focus on motivation, under the assumption that much of what we have to say regarding its measurement will *mutatis mutandis* also – perhaps even better - apply to emotions.

State-trait models

State-traits models are based on the idea that a single phenomenon can both show change and stability. The change is conceived in terms of variability around a certain average level that does not change. Thus, the phenomenon manifests itself as a state (indicated by variability in level) and as a stable trait (indicated by the average level). The best known example from the general psychological literature is state and trait anxiety (Spielberger, 1975). A recent example from the field of work psychology is state and trait engagement (Xanthopoulou & Bakker, 2013). Early research on anxiety used physiological indicators such as heart rate and systolic blood pressure to measure people's current level of anxiety, and a questionnaire to assess trait anxiety (Endler & Magnusson, 1977; Johnson & Spielberger, 1968). Although this seems to make sense, because the emotional state of anxiety has clear bodily components, later research discontinued this practice. In fact,

Spielberger (1968) introduced an significant simplification in the method of measurement by using the same multi-item scale with different instructions. Respondents were required to indicate either “how they feel now” on particular occasions or “how they generally feel”. This simplification was welcomed by many and has likely contributed to the popularity of Spielberger’s scale. However, it has introduced problems of measurement that have inadvertently corroded the work of many researchers until the current date. First, from a temporal point of view neither “now” nor “generally” have clear temporal referents. (For instance, “now” could be understood as today, this morning, between 10 and 11, the past 15 minutes, this instant in milliseconds). As a result they have an imprecise meaning, which hinders an assessment of changes in the current state as well as the stability of the trait. Second, it is uncertain whether the multi-item instruments used to measure states have measurement equivalence over time, that is, whether they can be used for repeated measurements. Third, and more importantly, the instruments are based on classical test theory, which as we have noticed is ill-suited to measure change, and implies an inconsistency between the measure and the state construct.

Building on this inconsistent and flawed understanding of states and traits, a number of researchers have made suggestions to establish a bridge between state and trait measurements, by integrating them into the same latent trait model. (e.g., Hamaker, Nesselroade, & Molenaar, 2007; Schmukle, Egloff, & Burns, 2002). Another approach has been proposed by Inceoglu and Fleck (2010) who see state and trait engagement are poles of a single continuum, and posit that the degree of dynamism will depend on the length of the time interval. *Ceteris paribus*, long intervals will show trait engagement, short intervals state engagement. Although the proposed models represent a step forward in a certain sense, they leave the main issues unresolved, namely the reliance on the assumption of a stable trait

represented by a single true score for every individual, a limited sensitivity to time, and a lack of explicit references to time - symptoms a time-impooverished view of reality (Albert, 2013).

To overcome these limitations we will propose an alternative model that abandons the notion of stability and starts from the idea that states show within-subject variation without restrictions regarding the degree and form of dynamics, and traits (plural) differentiate the dynamic features of the states between subjects. Thus, it fits the general notion of “between-subject variability in within-subject variability” mentioned before. Our model is temporally referenced, in the sense that it specifies the length of the interval during which the state and the trait are being measured, the moment at which the episode starts, and the number of observations in the interval.

The conceptual basis for this approach is the paradigm of ‘radical temporalism’ (Roe, 2005, 2008b), which starts from the assumption that all psychological phenomena, including motivation, are subject to change, and proposes a research strategy that is alternative to that of differential psychology. It proposes that the subject matter of research should be conceptualized in terms of ‘phenomena’ rather than ‘variables’, and that one should use verbs rather than nouns to designate these. The main argument is that variables are generally understood as being able to capture intra- en inter-subject variation, which promotes the likelihood of confusion the two. Phenomena remind one of the fact that psychological studies deal with “things that happen” during people’s (work) life. The research strategy encompasses three stages, i.e. establishing temporal features of phenomena, temporal relationships between multiple phenomena, and long-term stability and change, of which here we consider only the first. A crucial idea in radical temporalism is that the temporal features of a phenomenon depend on the time scope of a study. This implies that one cannot meaningfully speak about change unless one defines a time window during with the phenomenon of interest is being observed and measured. Although a rigorous anchoring of a

study will involve more characteristics (Albert, 2013; Roe, 2008b), we define three temporal referents, namely a time frame of length L , starting at moment M , and comprising N observations (see figure 1). The rationale for choosing these referents is that a longer or shorter length (e.g., a week, month or year), a later or earlier starting point (e.g., Monday or Friday, Spring or Fall, 1990 or 2020), and a larger or smaller number of observations (e.g., 2 or 4 or 30 or 260) will produce different observations and measurements, which will generally show very different images of reality. These three parameters do not directly define the measurement outcomes but rather moments of observation. They do have an influence on the pattern of measurements (measurement trajectory), though. For an illustration of how L and N matter, we refer to a recent review of temporal studies on performance and motivation (Roe, 2014a). The importance of M derives from natural cycles (e.g. circadian, weekly, and seasonal) as well as events on the historical calendar. Its significance has been underlined by Spain et al. (2010, p. 621), who state: “Often researchers wade into the stream of events with no real care as to when they do so. Put simply, Time 1 often is not really Time 1 but an arbitrary starting point for the study. Likewise, the studies often end at an equally arbitrary point in time”.

See Figure 1

The measurements taken at the moments defined by the three parameters build a time-series or temporal trajectory that can be graphed. Figure 2 gives two examples that illustrate the importance of the time window. The first time window, which starts earlier and has fewer observation points (parameters L_1 , M_1 and N_1) than the second one (parameters L_2 , M_2 and N_2) shows a declining trajectory, whereas the second one shows a cyclical trajectory. Such

differences can easily emerge when one moves from a single measurement per day to two or more per day.

Here Figure 2.

In a temporal approach, following the recommendations by Molenaar and others (Molenaar, 2007; Nesselroade & Molenaar, 2010) the trajectory of a single subject shall be treated as self-standing and not a priori be assumed to be similar to that of other subjects. This is an important point that clearly deviates from differential methods. Each person's trajectory can be characterized in terms of measurement parameters of the raw trajectory or a fitted function, such as: mean level, variability (within person SD, for example), ruggedness etc. (Solinger, van Olffen, Roe, & Hofmans, 2013) or in terms of the parameters of a mathematical function fitted to it: linear, quadratic, cubic etc. These parameters can be investigated post-hoc for similarity, without making the assumption of random variation (between-subject normality) that follows from the Uniformity of Nature postulate underlying differential measurement.

Our model has three distinguishing features. First, the dynamic trajectories may differ in various respects, not only in (within-person) mean level. Examples of measurement parameters are: initial level, slope, SD, magnitude of change, change frequency, intervals between inflections, pattern ruggedness and so on. Differences between such parameters can provide multiple trait measures, not just one. For example: some subjects may show frequent changes in their trajectory, while others may show only few changes. Some may show an increase of motivation, others a decline. The duration until the onset of decline – corresponding with the persistence aspect of motivation - may be a third characteristic in which people differ. Of course, there can also be a difference in overall level, as is assumed

in classical state-trait levels – but this is just one of many possibilities. For us, a trait would be any individual difference in within-subject pattern.

Second, there can be qualitative (in addition to quantitative) differences in subjects' states and traits. Differential models typically assume similarity between subjects and differences to accord to normal distributions (as in Random Coefficient Modeling or Latent Growth Modeling, applied in longitudinal research). They follow a “top-down logic”, assuming that a pattern found at the level of the sample will be found in all subjects, except from random deviations. Temporal models, in contrast, follow a “bottom-up logic” (Molenaar, 2004; Molenaar & Campbell, 2009; Roe, 2014a) and allow for the possibility that temporal trajectories are heterogeneous. The analysis of intra-individual variation should precede that of inter-individual variation. See for an illustration the “Spaghetti-plots” in job satisfaction observed by Liu, Rovine and Molenaar (2012) and in team conflict by Li and Roe (2012). Clustering methods can be used to identify similarities in their temporal trajectories. An example of trajectory clustering, based on a mixed modeling approach, can be found in research on commitment over time (Solinger et al., 2013).

Third, state and trait characteristics are conditional upon the three time parameters L (time frame length), M (starting moment) and N (number of observations). With other values of these parameters, researchers will normally find different state and trait parameters. As noted before, this has been observed in empirical studies that used different time frames and grids. There is a logical reason to expect this: except for trajectories showing stationary changes, like a stable or sinusoid trajectory, it is impossible to find the same parameters. For instance, one could not find the same parabolic trajectory within a day and a week. Each psychological phenomenon will have its own array of temporal trajectories within a given time window – this also applies to motivational and emotional trajectories. With minimal a priori constraints, the motivational trajectories defined by our model will allow identifying

people with highly volatile levels of motivation, extreme ups and downs, rather constant level of motivation and so on. This is a significant extension compared to current state-trait models (and trait models!).

In presenting our model we will restrict it to a single dimension of motivation, just as is the case in current state-trait models. That is, we consider a person's motivation to opt for and persistently pursue a particular goal or perform a particular task or role, which we designate as A. However, we should keep reminding ourselves that motivation is much broader and that crucial issues reside in the person oscillating between A and B, struggling with multiple goals, abandoning A for the benefit of B, etc. – topics that have been addressed in, e.g., Lewin's field theory (Lewin, 1951) and Atkinson's and Birch (1970) dynamic theory of action. The real merits of our model can therefore become particularly clear in a multidimensional version.

Operationalization procedures

What we discussed in the previous section concerns two important topics that are normally overlooked in motivation measurement, i.e., the demarcation of moments of observation and measurement proper, that is, the temporal assessment of values measured at these moments. With "measurement proper" we refer to the process of operationalizing the phenomenon under study, which includes making, recording and coding observations, and turning them into quantitative values. It must be noted that this is a topic that got some attention in earlier days (e.g., Lorge, 1951; Sanford, 1961) but virtually disappeared from the measurement theory (see however: Hartmann, Barrios, & Wood, 2004). Most texts start from the assumption that responses from subjects to some set of stimuli or items are "simply there", and that they just need to be modeled in an appropriate way. Currently used state-traits models use a simple questionnaire format: respondents typically use a (Likert type) rating

scale to report on “how they feel now” or “how they generally feel”, and the sum or average of the item scores represents the scale score. Such procedures are not satisfactory in a temporal approach, though. Temporal measurement comes with special requirements regarding the way in which the phenomena are observed and measured, i.e. sensitivity to change and robustness against re-use, which make it desirable to take a step back and consider ways in observations are gathered and quantified.

There is a fundamental difference between the operationalization of constructs in the differential paradigm and the operationalization of phenomena in the temporal paradigm. The first is based on the principle of homogeneity, that is, items should be indicators of the same latent construct, in order to measure this construct reliably¹. The notion of homogeneity is conceived differentially, that is, subjects’ scores on different indicators should inter-correlate highly with each other. Such evidence has no particular value within a temporal perspective, where homogeneity – as assessed by dynamic within-subject factor-analysis – would rather mean that indicators should synchronous change (cf. Molenaar, 2008).

While multiple items that are similar but slightly different in content and measurement qualities can perform well in distinguishing subjects at a single point in time, they may not be suited for measuring change. A main reason would be that the content domain covered by the items lacks an a priori definition and delineation that retains the same meaning over time. This is particularly troublesome in a research area like motivation, where quite different instruments are used to operationalize a construct with the same name (e.g., goal orientation, extrinsic motivation). Since change cannot be ascertained unless the content of what is supposed to change is sharply defined, multi-item instruments that are considered adequate for differential research are generally not suited for temporal measurement. A

¹ Some instruments cover multiple dimensions that are supposed to be part of the same overarching construct (e.g. cognitive ability, engagement).

related issue is that the measures must have the same meaning to participants at each time period (Ployhart, 2008), or more specifically that they show temporal measurement equivalence, to be understood as (between-subject) measurement equivalence not just over time moments, but across different time scopes, defined by L, M, and N. We cannot rule out that some well-developed tests may show such measurement equivalence across a wide range of conditions, but with the needed evidence lacking this cannot generally be assumed to be the case. From a temporal perspective one would rather prefer a single well-defined indicator (or a few synchronously changing indicators) for each phenomenon – even though that contradicts the canon of differential measurement (Ployhart, 2008). This would give the indicator the needed precision and avoid issues of measurement inequivalence.

Once the content issue has been settled, a suitable method of observation, recording and quantification (measurement proper) must be chosen. Of immediate relevance is the distinction between direct (or objective) methods that do not involve the subject in any way, and indirect (or subjective) methods in which subjects are involved. Direct methods, such as recording by some technical device (e.g., video) and observation/recording by a researcher, can be made unobtrusive, which enhances robustness against re-use because the subject is not aware of and has no way to directly influence the outcomes. They may also be constructed in ways that provide information about smaller changes and avoid stability bias. Indirect methods include tests that gauge subjects' responses to item content, and questionnaires and diaries that call for self-observation and reporting by the subject. Such methods give subjects a certain degree of control over the responses, which can make them less robust to re-use and limit their sensitivity to change.

Direct measures with unobtrusive qualities are rare and more used by researchers outside of psychology (e.g., Fulmer & Frijters, 2009; Lopatovska & Arapakis, 2011; Truong, van Leeuwen, & Neerincx, 2007; Westerink, van den Broek, Schut, van Herk, &

Tuinenbreijer, 2008). They have the advantage of picking up smaller changes that are not filtered out by the generalizing instructions of differential measures or subjects impression-management. When variations fail to make sense and appear to be “noise” (Ployhart, 2008, p. 33) one can adjust the M or N parameter of the time scope or fit a smoothing function. In other words, changing the resolution of the time window (e.g., from days to hours or less) or the moment in which observations start (e.g., early morning, rather than mid-day) in order to pick up a temporal signal may give a more meaningful picture and take away the impression of noise. This approach is what Roe (2013) has called ‘temporal zooming’.

An example of an indirect method can be found in the work of (Solinger, 2010; Solinger et al., 2013) on organizational commitment. He used a single indicator for a subject’s affective commitment towards an organization. The indicator was quantified by the subject with the help of a 0-100% graphic analogue scale, which – due to the instruction - can be supposed to have ratio, rather than interval qualities. Thus, a person can have no commitment at all (zero), or a very strong commitment (hundred). In fact, Solinger conceived of commitment as an attitude and postulated a three-dimensional model of the commitment phenomenon, with cognition, affect and action readiness) as dimensions. These dimensions were each measured by a single indicator and since analyses over a certain time window showed considerable synchrony, the three indicator scores were averaged to produce an overall commitment score.

Summary of our model

In this section we have presented a model for the measurement of motivation that offers a wide range of options to capture the dynamics of states and allows identifying multiple individual differences or traits. The model is based on the temporal paradigm and starts from time-series or trajectories that are defined in terms of length of time frame (L), starting at a particular moment on a historic time line (M), and the number of observations

(N). A distinguishing feature of the model is that the measurement data are subject to within-subject analysis to assess motivational states, and that one or more dynamic motivational traits are determined in a subsequent between-subject analysis. The dynamic traits – of which stable traits can be considered to be a special case - are identified post hoc and not assumed to exist ex ante, as in conventional psychometrics.

At the present stage the research can only be empirically-driven since, to our knowledge, there is no specific theory to guide the search for state trajectories and dynamic traits. Current theories based on relations between differential constructs are of no help because of their “time-impoverished” character (Albert, 2013); they lack the needed reference to time. Thus, strictly theory-driven assessment will have to wait until sufficient temporal data have been gathered and analyzed, and suitable theories have been developed. On the other hand, there are considerations from general motivation theory that can provide some guidance about what to expect in state measurements. The most important undoubtedly is that motivational states are likely to show a cyclic character, with upward and downward moves between the ends of the scale. Thus, motivation will never indefinitely continue to increase, but at best reach an upper asymptote that is maintained during a certain period of time, until the outside world or the subject initiate a change of activity (task, mission, role, job). It can also reach a lower asymptote, which implies a loss of motivation and will almost inevitable trigger a change in activity. Otherwise, changes may be more or less smooth, displaying regular waves or more rugged patterns, or show evidence of stabilization or destabilization (Roe, 2014a, 2014b).

In the next section we will illustrate how the new measurement principles and analytical procedure can be put to use by applying them to a key notion from the recent literature on work motivation, i.e. work engagement. We have chosen this topic because it has enjoyed a rapid increase in popularity among researchers in the field of work and

organization and because researchers have begun to distinguish between state and trait engagement. Like other state-trait studies this research suffers from ambiguity (Inceoglu & Fleck, 2010), which our approach may help to resolve.

Static and dynamic aspects of engagement

We begin with a brief introduction of research on work engagement and an assessment of current state-trait research. Next, we describe how engagement may be measured and recorded with novel tools (e.g., based on interactive gaming and wireless diaries). Then, we discuss how recorded data can be analyzed so as to produce state and trait measures. And finally we explore how these measures can be used to make within- and between-person assessments.

The notion of engagement has a remarkable history. While the term engagement had been used by researchers in earlier times (e.g., Kahn, 1990; Klinger, 1975), its ascent began in the late 1990's when Dutch researchers (Schaufeli, Bakker, and others) proposed to change the focus in work and health research from stress and burnout, which are generally seen as negative phenomena, to something positive, fitting into the emerging the spirit of positive psychology. The term that was chosen for this after some period of exploration (e.g., in Dutch language the term “bevlogenheid” was initially used; Schaufeli et al., 2001) was ‘engagement’. It was originally conceived as the counter pole of burnout and defined in a contrasting manner. Just like burnout was seen as a syndrome comprising exhaustion, cynicism and inefficacy, engagement was thought as comprising vigor, dedication, and absorption.

Engagement was initially measured with instruments designed for measuring burnout, scored in the opposite direction: the Maslach Burnout Inventory (MBI, Maslach et al., 1996) and the Oldenburg Burnout Inventory (OLBI; Demerouti, Bakker, Nachreiner, & Schaufeli,

2001). Later a new instrument for measuring engagement was developed, the Utrecht Work Engagement Scale (UWES; Schaufeli & Bakker, 2003). As the research with different (also translated) scales began to augment, it became clear that engagement cannot strictly be equated to the absence of burnout, and that perfect negative correlations between measures of burnout and engagement are hard to find. Thus, the construct began to “wander” (Bakker, Schaufeli, Leiter, & Taris, 2008 speak of an "emerging" concept) gravitating to its own definition and operationalization. Yet, the concept’s origin and meaning – as the positive counter-pole of burnout – was retained.

The way of naming and defining the concept had some interesting implications from a temporal point of view. Burnout in its original meaning is a relatively rare clinical phenomenon, with an estimated life-time prevalence of 4.2% and 12 month prevalence of 1.5% (Maske, Riedel-Heller, Seiffert, Jacobi, & Hapke, 2014). It is a temporary bounded phenomenon that spans a period of 4-11 months, depending on the treatment (Sonnenschein et al., 2008). Engagement, on the other hand, was initially thought of as a long-lasting phenomenon, with a trait-like character. In the words of Bakker (2014, p. 1) “an enduring, affective-motivational state of employees regarding their job”. This temporal discrepancy was not immediately obvious to psychological researchers, probably since they measured both burnout and engagement with multi-item questionnaires based on psychometric models designed for measuring stable traits (see above).

Recent research with diary methods, measuring engagement at the level of the day (Bakker, Albrecht, & Leiter, 2011; Breevaart, Bakker, Demerouti, & Hetland, 2012; Sonnentag, 2003; Sonnentag, Dormann, & Demerouti, 2010; Xanthopoulou & Bakker, 2013) changed this idea and made researchers recognize that engagement is not inherently stable but varies over time. In fact, they were quick to adopt the state-trait distinction from other domains in motivation and emotion research. In studies aiming to understand the dynamics of

engagement, two important issues seemed to have been overlooked, though. First, it remained unclear what the temporal profile of engagement should be like, given that its counterpart burnout happens once in the life of most persons and disappears after say 6 months. Is it present (and stable) during the whole lifetime except for these 6 months? Or is it present during these 6 months in the form of a lowered engagement level? Or does it have its own dynamics during the work career? Second, and more importantly, researchers did not examine whether the components of engagement share the same temporal profile; that is, whether vigor, dedication and absorption are fully synchronous. Meanwhile, there are several indications that this may not be the case (e.g., Sonnentag, Dormann, et al., 2010, p. 26), which raises doubts regarding the use a single overall engagement measure for assessing state engagement.

Apart from these issues, research on state and trait engagement suffers from the same measurement shortcomings as we discussed for state-trait research on other phenomena. State engagement was generally measured with the same scale as was used for measuring trait engagement, following Zuckerman's (1983) argument that one can measure enduring and state facets of the same construct by using instructions specifying a particular time frame, such as "in general" or "today". As explained, this practice rests on a differential psychometric model that is biased against change. Also it carries the risk that state measurements are contaminated with trait engagement, and that some items cannot be appropriately answered for short time spans (Sonnentag, Dormann, et al., 2010).

About a decade before Schaufeli, Bakker and others introduced their notion of engagement, a very different and very dynamic view of engagement was published by Kahn (1990). The aim of the study was to find out how people engage and disengage themselves during the course of their working days, that is, to describe the dynamics and develop a theoretical framework for understanding "self-in-role" processes. "My guiding assumption

was that people are constantly bringing in and leaving out various depths of their selves during the course of their work days. They do so to respond to the momentary ebbs and flows of those days and to express their selves at some times and defend them at others” (Kahn, 1990, pp. 692-693). Kahn identifies psychological conditions under which engagement arises and suggests that “a primary aim of future research might be to develop a dynamic process model explaining how the variables documented above combine to produce moments of personal engagement and disengagement” (Kahn, 1990; p. 717).

Due to the fact that this research (in two quite different work settings) used field observations, interviews etc. rather than surveys with pre-established scales, the picture of engagement and disengagement (and the factors involved in it) is highly dynamic and allows for many ways of interpretation. It seems to us that it offers a better basis for temporal measurement than the state-trait notions associated with the later work on engagement. This raises the question of why Kahn’s work was not integrated in the research by Bakker, Schaufeli and colleagues when they turned to the study of changes in engagement. One can find the first references to Kahn’s article in a publication by Bakker in 2008, but without recognition of the potential implications for enhanced dynamic theorizing and measurement. After 2009 Kahn’s work has been rarely mentioned by Bakker c.s. but it has inspired further research in the United States (Rich, LePine, & Crawford, 2010). Kahn’s work is also absent from research on detachment and disengagement by Sonnentag and others (Sonnentag, 2011; Sonnentag, Binnewies, & Mojza, 2010; Sonnentag, Mojza, Binnewies, & Scholl, 2008). However, it is compatible with our state-trait model, which allows for alterations between engagement and disengagement.

Overseeing the literature on work engagement, it seems that further research into its dynamics might profit from our extended state-trait model. We believe that following a temporal rather than a conceptualization and measurement approach, could deepen the

understanding of engagement and related phenomena. For instance, it could give a better view of the alternations of engagement and disengagement during shorter periods as proposed by Kahn (1990) of the variations in engagement (or in absorption) during flow (Csikszentmihalyi, 2000). In a wider time frame temporal research might elucidate the loss of engagement in the context of psychological contract violation (Schalk & Roe, 2007). In addition, it might clarify how engagement relates to burnout, temporally. As suggested in the clinical literature engagement may appear to be an important precursor of burnout, in the sense that high engagement combined with a lack of reward or accomplishment may be a trigger of burnout (Freudenberger, 1974; Längle, 2003). It would be equally interesting to see which engagement levels are typical after recovery from burnout. A final topic would be the upward spirals referred to in the work by Salanova et al. (Salanova, Llorens, & Schaufeli, 2011; Salanova, Schaufeli, Xanthopoulou, & Bakker, 2010) and Xanthopoulou et al. (2008). Spirals are complex temporal structures that cannot be ascertained with two or three measurement moments for each of the variables involved. One would need time-series generated with dense-measurement designs to be able to ascertain whether spirals do indeed occur, what their parameters are, and whether they ascend or descend. Considering the general constraint, mentioned earlier, that motivation couldn't continue to move upward and pass its higher asymptote spiral effects do not seem very likely. Also, they would be rather difficult to measure given the limited range of usual measurement scales. Besides, one should keep in mind that differential evidence of spiral effects may also mean that there are upward changes in some subjects and downward effects in other subjects: correlations between measurement points may become stronger without an upward trend in the scores.

All this makes us believe that engagement is an excellent domain for applying the model that we have proposed. They allow us (1) to show the benefits of an alternative approach to define and measure state engagement namely as temporal trajectories, (2) to

suggest multiple ways to model individual differences in engagement trajectories, (3) to include multiple time referents, that is, alternative sets of time windows and grids, which leads to a much richer view of engagement dynamics than a single state-trait dimension. In fact, this results in the distinction of multiple types of states and multiple types of traits – of which stable traits are just one specific case.

An empirical illustration

We will now illustrate how our model might be used with time-series data on engagement, using data from a study in which engagement was measured with a daily diary over a period of 34 days, involving 61 naval cadets (Breevaart et al., 2013). We should acknowledge that this data set has some limitations. First, we are dealing with a measurement instrument, the Utrecht Work Engagement Scale (Schaufeli & Bakker, 2003; see also Schaufeli, Bakker, & Salanova, 2006), which asks subjects to describe how they “experience” the state of engagement (not the state itself!). Second, the method is indirect since subjects have full control over their responses, which can lead to distortions. Third, the measurements are based on differential psychometrics, which we have criticized above for their bias towards stability and typically using interval scales.

Despite these limitations the data are interesting and useful to illustrate our model. They come from a study in which 61 naval cadets from a Norwegian Military University College completed a 9-item Utrecht Work Engagement Scale, once every day (Breevaart et al., 2013). The questions were adapted to capture engagement at the level of the day, e.g., “Today, my job inspired me” and “Today, I was very enthusiastic about my job.” The answers were scored on a 5-point Likert-scale (1 = totally disagree, 5 = totally agree). We do not know the exact starting point (M), but it is important to know that the time frame spanned a forty day sailing trip of naval cadets on a training ship. Recording took place at 34 of the 40

days, because of a break of 6 days. Thus, $N = 34$ and $L = 40$ (with a gap of 6 observation moments).

To circumvent stability bias resulting from classical (CTT-based) scale construction, we selected a single item as indicator of work engagement, the one with the greatest within-subject standard deviation. This is UWES item #4: “Today, my job inspired me”. Next, we examined the data and noted that there were 6 only among the 61 subjects with a complete record of 34 observations. The records of 28 subjects appeared to be censored, that is, subjects had stopped responding from a certain moment; in fact, 1 to 16 (average 7) of the last observations were missing. In 55 subjects there were also missing observations from earlier moments, on average 4 per subject. We decided to discard 7 subjects with 10 or more missing values before the end of their last observation, which leaves 54 subjects with potentially useful records.

Next, we plotted the trajectories for each of the subjects and compared these with each other, concentrating on patterns of variation they displayed. As expected in a temporal analysis, there was a great variety of unlike patterns between subjects. It demonstrated that daily variation in engagement comes in many forms and that differences between subjects are not just a matter of mean level as one would expect on the basis of a general state-trait model. Figure 3 gives two examples of contrasting trajectories (408 and 201). Figure 4 presents the state engagement trajectories for all 54 subjects with a valid record. The “spaghetti-plot” data (S. Liu et al., 2012) show an impressive variety of patterns, with some people moving from high to low engagement, others starting low and remaining low and so on.

Here Figure 3

Here Figure 4

Somewhat unexpectedly, most subjects revealed strong fluctuations across the days². It is worth noting that many subjects appeared to be low-engaged (scoring below “3”) for most of the time and that the overall trend is a declining one. This may reflect the circumstances under which the data were collected, namely during a long and challenging sailing trip.

From a differential perspective, differences would be readily categorized as errors and overall variation would be considered as noise (Ployhart, 2008), but within a temporal perspective this is not so. Under the assumption of a proper within-subject measurement, each trajectory is seen as expressing a subject’s genuine and unique development of *state engagement* during the particular time window chosen for the study. It should be considered that the changes in a subject’s trajectory may have many different origins; such as, variations in perceived job resources, experiences of demand and reward, leader support and hindrances, social comparisons, etc. but also variations in and personal resources, fatigue, emotions, vitality etc. To properly interpret the trajectory one would need information about such endogenous and indigenous factors also measured along the time line. The rise of engagement after the 6-day break, observed in about half of the subjects, is illustrative in this case.

As expected on the basis of our model, there are multiple ways to define *engagement traits*. That is, differences in subjects’ state trajectories are not limited to levels. In fact, differences in average level of engagement – the traditional meaning of trait engagement – are least informative because of the great variety in dynamic patterns. Following a heuristic approach, that is, not claiming generalizability to other time windows and/or other persons,

² This may seem at odds with our statement that CTT-based questionnaires are biased towards stability. However, we did not use the full UWES but selected the item with the greatest within-subject SD. The underlying work setting (ocean sailing trip) may also have contributed to this result.

and building on an exploratory analysis we propose that there are six types of engagement traits fitting to the present dataset. Table 1 list and defines these traits and gives three examples of the difference. An explanation follows in the next section.

Here Table 1

Before discussing the differences, we should explain that the proposed traits are all conditioned on a time window (with parameters L, M and N) and that some of them lack the connotation of lifetime stability that prevails in personality and intelligence research. Although this connotation is commonly present in differential psychology, the empirical evidence to support it is remarkably scarce. There are many studies showing reasonable test-retest correlations over longer periods, but these show between-subject covariance of relative scores and fail to capture changes in subjects' individual scores or sample mean scores. The rare studies that have examined score levels typically show change rather than stability (see for example Deary, Pattie, & Starr, 2013). Leaving the issue of temporal measurement adequacy aside, all studies are conditioned by the time window in which they were conducted – just like in our secondary analysis of engagement. As for content, one should realize that engagement is an inherently dynamic phenomenon: one cannot expect a person's engagement to show the same features during years or decades, if it were only for the reason that the particular work role or job typically changes in the course of years.

We designate trait 1 as '*engagement level*' and define it as a subject's average level of engagement during the observed time window (operationalized as Mean_{LMN}). Although the level of engagement seems the most meaningful trait in the differential state-trait model, it makes limited sense in a temporal model because of the large degree of variation between state engagement trajectories. Its descriptive value is limited to trajectories that show limited variability (see trait 2) and no downward or upward trend (see trait 4) – as is the case with the

three examples given in the table. In total there are only 9 such cases, 2 at a high level, 5 at a middle and 2 at a low level.

We call trait 2 ‘*engagement variability*’ and define it as overall degree of variation in engagement during the observed time window (operationalized as SD_{LMN}). The table gives three examples showing differences in the variability in state engagement trajectories.

Trait 3, called ‘*engagement polarity*’ (operationalized as $Range_{LMN}$) helps to make clear that not all forms of variability in state engagement are the same. In the first example scores span 4 scale points, ranging from the highest scale-end to the lowest scale-end; in the second and third example it spans 3 and 2 levels. Of course, the five-point interval scale can show limited differentiation only; more informative differences can be observed with ratio scales with more scale points.

Trait 4 refers to the trend in engagement during the time window. Since the overall trend is negative, we decide to name this trait ‘*engagement decline*’ and defined as the tendency to disengage during the time window (operationalized as Slope of linear decline $_{LMN}$). We do not know whether this pattern is unique to this study or occurs in other studies as well. It bears similarity to what has been called the “honeymoon-hangover” effect in satisfaction and commitment research (Boswell, Tichy, & Boudreau, 2005; Solinger et al., 2013). The first example shows a large decrease, the second no change (stability), and the third one an increase (this is the only case of increase in the data-set).

We labeled trait 5 ‘*engagement irregularity*’ and defined it as the unevenness of change in state engagement, which is a specific form of variability not captured by the standard deviation (trait 2). Irregularity can be compared with the ruggedness of a mountainous landscape, as studied by Sappington and colleagues (Sappington, Longshore, & Thompson). Like Solinger (2010) we applied their 3-D to our 2-D trajectories, and measured it as the standard deviation of the slopes of all adjacent trajectory segments. We also explored

other another measure, namely the count of the number of upward and downward changes following a horizontal trajectory segment. Our three examples are based on cases where ruggedness-indices coincided; they show high, medium and low ruggedness.

Finally, trait 6 was labeled '*engagement persistence*' and defined as the maximum duration of the interval during which state engagement is positive. Persistence may be measured in many ways: we counted the number of days on which subject scored 3, 4 or 5 (on the 5-pointscale). We noted that few subjects remain engaged for almost the whole period, most for a much shorter spell, while some do not reach the level 3 at all! This is shown in the three examples. We should keep in mind that these duration figures are inflated since missing values were replaced by interpolated values; thus, some were estimated to be 3 or higher while the actual state engagement might have been lower.

Since the purpose of our analyses was to just show what kinds of traits might emerge from our approach we did not conduct any further analysis. We neither searched for additional traits nor conducted any grouping of subjects by formal clustering algorithms. We should however point out that with other data sets or other time windows, additional types of traits might emerge. Quantitative methods may be helpful to group subjects on the basis of similarity in their state trajectories. Researchers might for instance use mixed modeling methods for this purpose (e.g., S. Liu et al., 2012; Solinger et al., 2013).

Our exploration of engagement trajectories, allows us to make a final note, namely that motivation generally fluctuates – regularly or irregularly, within a certain range of values (ultimately the ends of the measurement scale). A person may remain highly engaged in a certain type of work, but the level will reach an asymptote (ultimately defined by the upper end of the measurement scale) and sooner or later decline. This argues against the idea of an upward engagement spiral, which we have encountered before.

In concluding this section, we should remind our readers that the material used to illustrate our methods has several limitations, most obviously in the measurement methods. The instrument used was not created to maximize within-person change and used an interval scale with very few scale-points. Although this has limited the possibility to display engagement state and to find traits, we hope that the rationale of our approach has yet been sufficiently demonstrated. It should at least have become clear that within a temporal approach there is no a priori reason to expect flat state trajectories, that individual differences between subjects in state trajectories can be many, and that the form of both state and trait trajectories are conditioned by the time window of the study – characterized by its length, the number of observation points, and the starting moment (which defines the location on an historic time axis).

Discussion

In this chapter we have pointed out that the prevailing differential approach to measurement, largely based on CTT, is of limited value when it comes to the study of motivation and emotion. We argued that measurement of such inherently dynamic phenomena calls for other methods based on a temporal paradigm. The core of our position is that state-traits models, which were introduced some fifty years ago and which have become increasingly popular in recent years, are based on differential premises that limit their capability to grasp the dynamics of emotion and motivation. The alternative state-trait model that we presented starts from a temporalist position and posits that states shall be measured within each individual subject and that individual differences in state trajectories shall be assessed afterwards. This allows for multiple traits to emerge as we have illustrated for the case of work engagement. We are not the first to note that individuals do not only differ in their average score but also in the variation of their scores over time. Cattell (1978) referred

to changes due to learning, maturation etc. and used the term ‘trait change’, indicating that traits are not always stable – a topic that has attracted much attention in the recent literature (e.g., Allemand, Zimprich, & Hertzog, 2007; Klimstra, Bleidorn, Asendorpf, van Aken, & Denissen, 2013; Woods & Sofat, 2013). Nesselroade (1991) pointed at the degree of variability and spoke of ‘state variation’. Fleeson et al. (2001) conducted a number of experiences sampling studies using state-trait measures of personality spanning periods of 2-3 weeks. They noted a substantial variability, such that a person with a certain trait level of extraversion would show all state levels from high to low. Rather than analyzing trajectories Fleeson concentrated on the ‘density distribution’ of state scores, which, in his view, is best represented by the mean score. This mean score was found to be stable over time, and indicates the person’s trait level.

In spite of some superficial resemblance (e.g., trait change seems similar to our ‘*engagement decline*’, and state variation to our *engagement variability*), this works differ fundamentally from our state-trait model. First, they are firmly based in differential reasoning and psychometrics and take the “existence” of traits for granted. Within-person changes are therefore interpreted as deviations³ from the stable (or changing) trait level (Hamaker et al., 2007). Our model does not make any a priori assumptions regarding the existence of stable (or changing) traits. It starts from within-person changes and lets the findings determine which and how many traits there are. It does not accept the idea that the various trajectory types should be interpreted as deviations from a mean trajectory (our ‘*engagement level*’). This would make sense in cases where some degree of stability is present, but not in cases of declining, inclining or rugged trajectories⁴.

³ In differential reasoning one could also speak of them as ‘residuals’ (e.g., Borsboom, Cramer, Kievit, Scholten, & Franić, 2009)

⁴ It would be similar to stating that the altitude profile of Switzerland is characterized by deviations from its average elevation of 1,350 meters.

Second, our approach requires that any investigation of states and traits shall specify the length (L) of the time interval, the moment at which this interval starts (M), and the number of observations (N). This implies that all findings regarding traits and states become contingent upon a specific temporal windows and that claims of stability should be tempered, unless replications over longer time periods with similar starting moments (M) have confirmed that stability is indeed present. Thus, for example, if Fleeson (2001) states that “individual differences in central tendencies of behavioral distributions were almost perfectly stable” in three studies of personality states across 2 to 3 weeks of everyday life, that should be understood as stable *within those 2-3 week intervals*. As there were no replications over similar intervals spread over a larger period of time, we do not know how long stability lasts.

This raises an interesting point regarding our own study, namely that the dynamic traits that we found might show a certain degree of stability if the study were repeated at a later time. This would mean that the same cadets would make some additional transatlantic sailing trips, with intervals of e.g. quarters or years. Would the engagement traits be the same (or similar) at every trip, as the differential mode of thinking would suggest, or would one expect them to change due to habituation or other forms of learning, as the temporalism would suggest? As said, only empirical research can settle this issue.

Our model needs further elaboration, especially when it comes to the techniques of observing and measuring per se. We have outlined the basic requirements of sensitivity to change and robustness under repeated measurement, but a formal mathematical model that provides a suitable base for physiological measures, behavioral observations, and for instance diary-based self-observations, is still to be described.

However, even in its current form the new state-trait model can be of use in research on motivation and emotion, enhancing its capacity to observe changes in these phenomena and their antecedents, concomitants and consequences, and to avoid ambiguities and

contradictions present in current research. It can help to enhance the predictions based on differential knowledge, which sees subjects as randomly varying around a likely criterion value, by recognizing that subjects may in fact show different trajectories which leads to diverging forecasts. This may open the way to individualized prediction and treatment, similar to “individualized medicine” (Cortese, 2007). Finally, it can help to clarify causal relationships, which is something differential methods are particularly weak at.

Since the model produces time-based evidence, it may also help to improve practical applications at least those that aim to maintain subjects’ motivation or emotion or to change it, in the sense of preventing a downward turn or stimulating an upward turn.

A question inevitably rising when comparing the differential and the temporal approaches is where they meet, and how evidence obtained with both can best be applied and combined. It is good to see that these approaches – which are valuable in their own way – are not each other’s rival, but rather each other’s complement. They show distinct and complementary images of reality, which inform different interventions. The huge body of differential knowledge that has been acquired during three-quarters of a century gives good descriptions of between-subject variation in human attributes, and their power to explain variance in numerous outcomes. And this has allowed effective choice-based (selection, placement, allocation) interventions at work, in education and in clinical practice. The temporal approach has been very useful to understand idiosyncrasies, commonalities and differences in developmental and change processes, and served as a basis for effective change-based interventions.

To appreciate the differences it is helpful to refer to Cattell’s (1952) data-matrix of variables x subjects x time, since this clarifies that the differential approach concerns the (R) analysis of variables x subjects and the temporal approach the (P) analysis of variables x time subject. The differential approach aims at ‘prediction’ in the sense of explaining variance,

regardless of a specific timing of the criterion. The temporal approach aims at ‘forecasting’, i.e. stating what will likely happen at some specific later point in time. The matrix also shows the limitations of both: differential studies do not need to consider time; temporal studies do not need to consider subjects. As was said above: “Just like one time-moment suffices to apply a differential design, a single person suffices to apply a temporal design”. Of course, as the data matrix suggests there is a ‘common ground’ of studies with multiple subjects and multiple time-moments (Roe, 2014a; Roe et al., 2012).

This common ground can be accessed from both the differential and the temporal angle, but this does not give the same results. The differential analysis of multiple time moments has evolved from classical longitudinal studies with cross-lagged panel designs to contemporary studies using with multi-level designs (subjects x time moments) with latent traits. The older techniques analyze covariation between subjects across time moments and do not really capture intra-subject change. The more recent techniques model change by means of mathematical functions (linear, quadratic, polynomial), which are fitted to the data of all subjects simultaneously. They allow for random variation in function parameters between subjects, but not for qualitative differences. Therefore, they give a potentially valid, but at the same time limited view of variability and change within subjects, although it is possible to allow for systematic variation in change trajectories between subjects by using mixed modeling. Besides, these methods are handicapped by the absence of a good measurement model that is equipped to capture change and prevents that potentially meaningful variations are discarded as “noise”.

Temporal analyses do not need to remain limited to $N=1$ studies but can be extended to multiple subjects, as we have already indicated in this chapter. Our description of the state-trait model and the examples have been limited to the analysis of raw data trajectories, which we see as preferable because it makes the researcher aware of the real appearance of dynamic

phenomena. Yet, it is possible to make use of formal models that analyze growth functions that are fitted to raw trajectories, in such a way that both intra-subject variations and individual differences are captured. One should be aware of the trade-off between the advantage of having comprehensive and parsimonious outcomes and the fit of the functions to the raw trajectories. An example is integrated state-trait model by Hamaker et al. (2007). This has some resemblance with mixed modeling but offers room to model subjects' change trajectories individually.

As we have pointed out before, the temporal approach applied to multiple subjects does not give the same results as the differential approach applied to multiple time moments. The reason is that it works bottom-up, acknowledging differences in trajectories and taking together cases with common trajectories. The differential approach works top-down, assuming that the common pattern at the level of the sample applies to all subjects, unless there is evidence of a moderator effect. How large the differences can be, and which impacts they may have on theory-building is illustrated in a study on conflict dynamics in teams by Li and Roe (Li & Roe, 2012).

A general limitation of both differential and temporal studies that they lack temporal specification. We feel that future research shall be explicit about when research is done and within which time windows (Roe, 2014a, 2014c). Apart from the length of the time window (L) and the number of observations (N), it is particular important to be aware of the starting moment (M). In the study on engagement that we described above, all subjects began a new work experience at the same time. Many studies on motivation take an arbitrary starting point for the study, which means that the subjects may be at very different moments in the fulfillment of their job, team role, or project – which makes the observations difficult to compare and can lead to blurred results. It is also important that studies are properly

replicated at later time periods. Replications should preferably have the same time window (equal length, equal number of observations) but be far apart in starting moments.

We have argued that temporal research needs much more explicit attention to measurement issues than is usually paid to it. One cannot expect research to produce meaningful results with regard to variation and change if researchers keep using instruments based on CTT or even IRT. There are good reasons to scrutinize the whole measurement process starting with the way in which subjects are invited to participate and the conditions under which they are required to give responses, the cognitive processes involved in interpreting questions and generating answers to questions, the scoring of answers (i.e., assigning numbers to objects or events; Stevens, 1946), and so on. For the purpose of temporal measurement, instruments for measuring constructs in a differential way may not at all be adequate. This is not the place to elaborate on the problems of constructs (e.g., Borsboom et al., 2009) but the fact that they can be measured by instruments in different languages that comprise various sets of questions (which have sometimes been arbitrarily chosen and trimmed to optimize scale reliability, typically coefficient alpha) implies that they represent “semantic clouds” which lack the precision needed to measure change.

We favor methods that have a well-described content and that do not necessarily rely on the subjects’ self-evaluation. Our plea for direct, unobtrusive measures does not imply that this should be the only source of information. Given that emotion and motivation are phenomena involving biological, behavioral and experiential processes that cannot be captured by any specific type of measure, we think that research should try to find meaningful combinations of physiological recordings, behavioral observations, surveys and tests of which the results can be linked to each other. The methods should also cover changes of environmental conditions, which are supposed to trigger or be triggered by emotions and motivational states.

References

- Albert, S. (2013). *When. The perfect art of timing*. San Francisco: Jossey-Bass.
- Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-sectional age differences and longitudinal age changes of personality in middle adulthood and old age. *Journal of Personality*, 75(2), 323-358. doi:10.1111/j.1467-6494.2006.00441.x
- Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. New York: John Wiley & Sons.
- Bakker, A. B. (2014). Daily fluctuations in work engagement: An overview and current directions. *European Psychologist*, 19(4), 227-236. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-23163-001&site=ehost-live&scope=site>
bakker@fsw.eur.nl
- Bakker, A. B., Albrecht, S. L., & Leiter, M. P. (2011). Work engagement: Further reflections on the state of play. *European Journal of Work and Organizational Psychology*, 20(1), 74-88. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-03226-010&site=ehost-live&scope=site>
bakker@fsw.eur.nl
- Bakker, A. B., Schaufeli, W. B., Leiter, M. P., & Taris, T. W. (2008). Work engagement: An emerging concept in occupational health psychology. *Work & Stress*, 22(3), 187-200. doi:10.1080/02678370802393649
- Balthazart, J., de Meaultsart, C. C., Ball, G. F., & Cornil, C. A. (2013). Distinct neuroendocrine mechanisms control neural activity underlying sex differences in sexual motivation and performance. *The European Journal Of Neuroscience*, 37(5), 735-742. doi:10.1111/ejn.12102
- Barker, R. G. (1963). *The stream of behavior: Explorations of its structure & content*. East Norwalk, CT, US: Appleton-Century-Crofts.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz & R. W. Lissitz (Eds.), *The concept of validity: Revisions, new directions, and applications*. (pp. 135-170). Charlotte, NC, US: IAP Information Age Publishing.
- Boswell, W. R., Tichy, J., & Boudreau, J. W. (2005). The relationship between employee job change and job satisfaction: The honeymoon-hangover effect. *Journal of Applied Psychology*, 90(5), 882-892.
- Breevaart, K., Bakker, A., Hetland, J., Demerouti, E., Olsen, O. K., & Espevik, R. (2013). Daily transactional and transformational leadership and daily employee engagement. *Journal of Occupational and Organizational Psychology*, 87(1), 138-157. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-02554-008&site=ehost-live&scope=site>
breevaart@fsw.eur.nl
- Breevaart, K., Bakker, A. B., Demerouti, E., & Hetland, J. (2012). The measurement of state work engagement: A multilevel factor analytic study. *European Journal of Psychological Assessment*, 28(4), 305-312. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-27741-008&site=ehost-live&scope=site>
- Byrne, Z. S., Mueller-Hanson, R. A., Cardador, J. M., Thornton, G. C., Schuler, H., Frintup, A., & Fox, S. (2004). Measuring achievement motivation: Tests of equivalency for

- English, German, and Israeli versions of the achievement motivation inventory. *Personality and Individual Differences*, 37, 203-217.
- Capa, R. L., Audiffren, M., & Ragot, S. (2008). The interactive effect of achievement motivation and task difficulty on mental effort. *International Journal Of Psychophysiology: Official Journal Of The International Organization Of Psychophysiology*, 70(2), 144-150. doi:10.1016/j.ijpsycho.2008.06.007
- Cattell, R. B. (1952). Three basic factor-analytic research designs: their interrelations and derivatives. *Psychological Bulletin*, 49, 499-520.
- Cofer, C. N., & Apply, M. H. (1964). *Motivation: Theory and Research*. New York: John Wiley.
- Cortese, D. A. (2007). A vision of individualized medicine in the context of global health. *Clinical Pharmacology And Therapeutics*, 82(5), 491-493. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=17952101&site=ehost-live&scope=site>
<http://onlinelibrary.wiley.com/doi/10.1038/sj.clpt.6100390/abstract>
- Csikszentmihalyi, M. (2000). Flow. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (Vol. 3, pp. 381-382). Washington, DC: American Psychological Association.
- Dalal, R. S., & Hulin, C. L. (2008). Motivation for what? A multivariate dynamic perspective of the criterion. In R. Kanfer, G. Chen, & R. D. Pritchard (Eds.), *Work motivation: Past, present, and future*. (Vol. 27, pp. 63-100). New York, NY US: Routledge/Taylor & Francis Group.
- Deary, I. J., Pattie, A., & Starr, J. M. (2013). The stability of intelligence from age 11 to age 90 years: The Lothian Birth Cohort of 1921. *Psychological Science*, 24(12), 2361-2368. doi:10.1177/0956797613486487
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, 86(3), 499-512. doi:10.1037/0021-9010.86.3.499
- Endler, N. S., & Magnusson, D. (1977). The interaction model of anxiety: An empirical test in an examination situation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 9(2), 101-107. doi:10.1037/h0081612
- Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology*, 86(6), 1270-1279.
- Fernet, C., Gagné, M., & Austin, S. (2010). When does quality of relationships with coworkers predict burnout over time? The moderating role of work motivation. *Journal of Organizational Behavior*, 31(8), 1163-1180. doi:10.1002/job.673
- Fishel, S. R., Muth, E. R., & Hoover, A. W. (2007). Establishing appropriate physiological baseline procedures for real-time physiological measurement. *Journal of Cognitive Engineering and Decision Making*, 1(3), 286-308. doi:10.1518/155534307X255636
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6), 1011-1027. doi:10.1037/0022-3514.80.6.1011
- Frank, M. J., & Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology: Official Publication Of The American College Of Neuropsychopharmacology*, 36(1), 133-152. doi:10.1038/npp.2010.96
- Freudenberger, H. J. (1974). Staff burnout. *Journal of Social Issues*, 30(1), 159-165.
- Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, 21(3), 219-246. doi:10.1007/s10648-009-9107-x

- Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions* (2nd ed.). New York, NY, US: Routledge/Taylor & Francis Group.
- Gulliksen, H. (1950). *Theory of mental tests*. New York :: Wiley.
- Guthke, J. (1993). Developments in learning potential assessment. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijsenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues*. (pp. 43-67). Lisse Netherlands: Swets & Zeitlinger Publishers.
- Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait--state model. *Journal of Research in Personality*, 41(2), 295-315. doi:10.1016/j.jrp.2006.04.003
- Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes, E. M. Heiby, S. N. Haynes, & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment, Vol. 3: Behavioral assessment*. (pp. 108-127). Hoboken, NJ, US: John Wiley & Sons Inc.
- Hopp, H., Rohrmann, S., & Hodapp, V. (2012). Suppression of negative and expression of positive emotions: Divergent effects of emotional display rules in a hostile service interaction. *European Journal of Work and Organizational Psychology*, 21(1), 84-105. doi:10.1080/1359432X.2010.539327
- Inceoglu, I., & Fleck, S. (2010). Engagement as a motivational construct. In S. L. Albrecht & S. L. Albrecht (Eds.), *Handbook of employee engagement: Perspectives, issues, research and practice*. (pp. 74-86). Northampton, MA, US: Edward Elgar Publishing.
- James, W. (1890). *The Principles of Psychology* (1 ed.). New York: Holt & Company.
- Johnson, D. T., & Spielberger, C. D. (1968). The effects of relaxation training and the passage of time on measures of state- and trait-anxiety. *Journal of Clinical Psychology*, 24(1), 20-23. doi:10.1002/1097-4679(196801)24:1<20::AID-JCLP2270240105>3.0.CO;2-9
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724. doi:10.2307/256287
- Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial & organizational psychology. (Vol. 1)* (pp. 75-170). Palo Alto: Consulting Psychologists Press.
- Kanfer, R., Chen, G., & Pritchard, R. D. (Eds.). (2008). *Work Motivation: Past, Present, and Future*. New York: Routledge Taylor Francis Group.
- Kaplan, S., Dalal, R. S., & Luchman, J. N. (2013). Measurement of emotions. In R. R. Sinclair, M. Wang, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology: Measurement, design, and data analysis*. (pp. 61-75). New York, NY, US: Routledge/Taylor & Francis Group.
- Kaplan, U., & Tivnan, T. (2014). Multiplicity of emotions in moral judgment and motivation. *Ethics & Behavior*, 24(6), 421-443. doi:10.1080/10508422.2014.888517
- Kirchberg, D. M. (2014). *The dynamics of multiple goal management: Diary studies at work*. Maastricht: Maastricht University Press.
- Kleinginna, P. R., & Kleinginna, A. M. (1981a). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4), 345-379. doi:10.1007/BF00992553
- Kleinginna, P. R., & Kleinginna, A. M. (1981b). A categorized list of motivation definitions, with a suggestion for a consensual definition. *Motivation and Emotion*, 5(3), 263-291. doi:10.1007/BF00993889
- Klimstra, T. A., Bleidorn, W., Asendorpf, J. B., van Aken, M. A. G., & Denissen, J. J. A. (2013). Correlated change of Big Five personality traits across the lifespan: A search

- for determinants. *Journal of Research in Personality*, 47(6), 768-777.
doi:10.1016/j.jrp.2013.08.004
- Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review*, 82, 1-25.
- Kooij, D. T. A. M., Bal, P. M., & Kanfer, R. (2014). Future time perspective and promotion focus as determinants of intraindividual change in work motivation. *Psychology and Aging*, 29(2), 319-328. Retrieved from
<http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-24959-013&site=ehost-live&scope=site>
t.a.m.kooij@uvt.nl
<http://psycnet.apa.org/journals/pag/29/2/319/>
- Kreibig, S. D., & Gendolla, G. H. E. (2014). Autonomic nervous system measuring of emotion in education and achievement settings. In R. Pekrun, L. Linnenbrink-Garcia, R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education*. (pp. 625-642). New York, NY, US: Routledge/Taylor & Francis Group.
- Kukolja, D., Popović, S., Horvat, M., Kovač, B., & Ćosić, K. (2014). Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications. *International Journal of Human-Computer Studies*, 72(10-11), 717-727. doi:10.1016/j.ijhcs.2014.05.006
- Lane, A. M. (2004). Emotion, Mood and Coping in Sport: Measurement Issues. In D. Lavallee, J. Thatcher, & M. V. Jones (Eds.), *Coping and emotion in sport*. (pp. 255-271). Hauppauge, NY, US: Nova Science Publishers.
- Länge, A. (2003). Burnout – Existential Meaning and Possibilities of Prevention. *European Psychotherapy*, 4(1), 107-121.
- Latham, G. P., & Pinder, C. C. (2005). Work motivation theory and research at the dawn of the twenty-first century *Annual Review of Psychology* (Vol. 56, pp. 485-516). Palo Alto: Annual Reviews.
- Lewin, K. (1951). *Field theory in social science: selected theoretical papers* (Edited by Dorwin Cartwright.). Oxford, England: Harpers.
- Li, J., & Roe, R. A. (2012). Introducing an intrateam longitudinal approach to the study of team process dynamics. *European Journal of Work and Organizational Psychology*, 21(5), 718-748. doi:10.1080/1359432x.2012.660749
- Liu, L., Zhang, G., Zhou, R., & Wang, Z. (2014). Motivational intensity modulates attentional scope: Evidence from behavioral and ERP studies. *Experimental Brain Research*, 232(10), 3291-3300. doi:10.1007/s00221-014-4014-x
- Liu, S., Rovine, M. J., & Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods*, 17(1), 15-30. doi:10.1037/a0026971
- Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction. *Information Processing & Management*, 47(4), 575-592. doi:10.1016/j.ipm.2010.09.001
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.
- Lorge, I. (1951). The fundamental nature of measurement. In E. F. Lindquist & E. F. Lindquist (Eds.), *Educational measurement*. (pp. 533-559): American Council on Education.

- Magnusson, D., & Endler, N. S. (1977). Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.), *Current issues in interactional psychology* (pp. 3-36). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Maske, U. E., Riedel-Heller, S. G., Seiffert, I., Jacobi, F., & Hapke, U. (2014). [Prevalence and Comorbidity of Self-Reported Diagnosis of Burnout Syndrome in the General Population.]. *Psychiatrische Praxis*. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=25158142&site=ehost-live&scope=site>
- Mintzberg, H. (1973). *The nature of managerial work*. Englewood Cliffs, NJ: Prentice-Hall.
- Mitchell, T. R., Harman, W. S., Lee, T. W., & Lee, D.-Y. (2008). Self-regulation and multiple deadline goals. In R. Kanfer, G. Chen, & R. D. Pritchard (Eds.), *Work motivation: Past, present, and future* (pp. 197-231). New York: Routledge.
- Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research and Perspectives*, 2(4), 201-218. doi:10.1207/s15366359mea0204_1
- Molenaar, P. C. M. (2007). Psychological Methodology will Change Profoundly Due to the Necessity to Focus on Intra-individual Variation. *Integrative Physiological and Behavioral Science*, 41(1), 35-40.
- Molenaar, P. C. M. (2008). Consequences of the ergodic theorems for classical test theory, factor analysis, and the analysis of developmental processes. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging: Interdisciplinary perspectives*. (pp. 90-104). Thousand Oaks, CA US: Sage Publications, Inc.
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112-117. doi:10.1111/j.1467-8721.2009.01619.x
- Navarro, J., Roe, R. A., & Artiles, M. I. (2014). *Taking time seriously. Changing practices and perspectives in Work/Organizational Psychology*.
- Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. (pp. 92-105). Washington, DC, US: American Psychological Association.
- Nesselroade, J. R., & Molenaar, P. C. M. (2010). Emphasizing intraindividual variability in the study of development over the life span: Concepts and issues. In W. F. Overton, R. M. Lerner, W. F. Overton, & R. M. Lerner (Eds.), *The handbook of life-span development, Vol 1: Cognition, biology, and methods*. (pp. 30-54). Hoboken, NJ, US: John Wiley & Sons Inc.
- Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, 1(1-2), 9-29. doi:10.1207/s15427617rhd0101&2_3
- Ployhart, R. E. (2008). The measurement and analysis of motivation. In R. Kanfer, G. Chen, & R. D. Pritchard (Eds.), *Work Motivation: Past, Present, and Future* (pp. 18-61). New York: Routledge Taylor Francis Group.
- Reise, S. P., & Waller, N. G. (2002). Item response theory for dichotomous assessment data. In F. Drasgow, N. Schmitt, F. Drasgow, & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis*. (pp. 88-122). San Francisco, CA, US: Jossey-Bass.
- Rich, B. L., LePine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, 53(3), 617-635. doi:10.5465/AMJ.2010.51468988

- Roe, R. A. (2005). No more variables, please! Giving time a place in Work and Organizational Psychology. . In F. Avallone, H. Kepir Sinangil, & A. Caetano (Eds.), *Convivence in organizations and society* (pp. 11-20). Milano: Guerini.
- Roe, R. A. (2008a). *Time and the study of behavior in organizations: Implications for assessment, selection, and development* Paper presented at the SHL Seminar, Thames Ditton, UK.
- Roe, R. A. (2008b). Time in applied psychology: The study of 'what happens' rather than 'what is.'. *European Psychologist*, 13(1), 37-52. doi:10.1027/1016-9040.13.1.37
- Roe, R. A. (2013). *Die Wiederentdeckung der Zeit und die Zukunft der Psychologie*. Paper presented at the Antrittsvorlesung, Leipzig.
- Roe, R. A. (2014a). Performance, motivation and time. In A. Shipp & Y. Fried (Eds.), *Time and Work. Vol. 1: How time impacts individuals* (pp. 63-110). London, UK: Routledge / Taylor & Francis.
- Roe, R. A. (2014b). The representation of time in psychological research on air pilots and air traffic controllers. In A. Droog (Ed.), *Aviation psychology: Facilitating change(s). Proceedings of the 31st EAAP Conference* (pp. [275]271-212). Valetta, Malta: EAAP.
- Roe, R. A. (2014c). Test validity from a temporal perspective: Incorporating time in validation research. *European Journal of Work and Organizational Psychology*, 23(5), 754-768. doi:10.1080/1359432X.2013.804177
- Roe, R. A., Gockel, C., & Meyer, B. (2012). Time and change in teams: Where we are and where we are moving. *European Journal of Work and Organizational Psychology*, 21(5), 629-656. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=2012-27739-001&site=ehost-live&scope=site>
bmeyer@sozpsy.uzh.ch
- Ruef, A. M., & Levenson, R. W. (2007). Continuous measurement of emotion: The affect rating dial. In J. A. Coan, J. J. B. Allen, J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment*. (pp. 286-297). New York, NY, US: Oxford University Press.
- Salanova, M., Llorens, S., & Schaufeli, W. B. (2011). “Yes, I can, I feel good, and I just do it!” On gain cycles and spirals of efficacy beliefs, affect, and engagement. *Applied Psychology: An International Review*, 60(2), 255-285. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=2011-04484-004&site=ehost-live&scope=site>
salanova@psi.uji.es
- Salanova, M., Schaufeli, W. B., Xanthopoulou, D., & Bakker, A. B. (2010). The gain spiral of resources and work engagement: Sustaining a positive worklife. In A. B. Bakker (Ed.), *Work engagement: A handbook of essential theory and research*. (pp. 118-131). New York, NY, US: Psychology Press.
- Salmon, W. C. (1953). The Uniformity of Nature. *Philosophy and Phenomenological Research*, 14(1), 39-48. doi:10.2307/2104014
- Sanford, F. H. (1961). Tests and Measurements in Psychology *Psychology: A scientific study of man*. (pp. 90-117). Belmont, CA, US: Wadsworth Publishing Company.
- Sappington, J. M., Longshore, K. M., & Thompson, D. B. (2007). Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. *Journal of Wildlife Management*, 71(5), 1419-1426. doi:10.2193/2005-723
- Schalk, R., & Roe, R. A. (2007). Towards a Dynamic Model of the Psychological Contract. *Journal for the Theory of Social Behaviour*, 37(2), 167-182. doi:10.1111/j.1468-5914.2007.00330.x

- Schaufeli, W., & Bakker, A. (2003). *The Utrecht Work Engagement Scale (UWES)*. *Test manual*. Retrieved from Utrecht, The Netherlands:
- Schaufeli, W., Bakker, A., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66, 701-716.
- Schaufeli, W., Taris, T., Le Blanc, P., Peeters, M., Bakker, A., & De Jonge, J. (2001). Maakt arbeid gezond? Op zoek naar de bevlogen medewerker. *De Psycholoog*, 36, 422-428.
- Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., & Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biology*, 10(2), e1001266-e1001266. doi:10.1371/journal.pbio.1001266
- Schmukle, S. C., Egloff, B., & Burns, L. R. (2002). The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality*, 36(5), 463-475. doi:[http://dx.doi.org/10.1016/S0092-6566\(02\)00007-7](http://dx.doi.org/10.1016/S0092-6566(02)00007-7)
- Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, 51(3), 154-165. doi:10.1080/00049539908255353
- Shimomitsu, T., & Theorell, T. (1996). Intraindividual relationships between blood pressure level and emotional state. *Psychotherapy and Psychosomatics*, 65(3), 137-144.
- SHL. (1992). *Motivation Questionnaire manual and user's guide*. Thames Ditton, UK.: CEB SHL.
- Solinger, O. N. (2010). *Capturing the dynamics of commitment: Temporal theory, methods, and an application to the organizational commitment attitude*. (PhD), Maastricht Univeristy, Maastricht.
- Solinger, O. N., van Olffen, W., Roe, R. A., & Hofmans, J. (2013). On becoming (Un)committed: A taxonomy and test of newcomer onboarding scenarios. *Organization Science*, 24(6), 1640-1661. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=2013-43675-003&login.asp&site=ehost-live&scope=site>
joeri.hofmans@vub.ac.be
r.roe@maastrichtuniversity.nl
w.vanolfen@tiasnimbass.edu
o.n.solinger@vu.nl
- Sonnenschein, M., Sorbi, M. J., Verbraak, M. J. P. M., Schaufeli, W. B., Maas, C. J. M., & Van Doornen, L. J. P. (2008). Influence of sleep on symptom improvement and return to work in clinical burnout. *Scandinavian Journal of Work and Environmental Health*, 34(1), 23-32.
- Sonnentag, S. (2003). Recovery, Work Engagement, and Proactive Behavior: A New Look at the Interface Between Nonwork and Work. *Journal of Applied Psychology*, 88(3), 518-528. doi:10.1037/0021-9010.88.3.518
- Sonnentag, S. (2011). Recovery from fatigue: The role of psychological detachment. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*. (pp. 253-272). Washington, DC US: American Psychological Association.
- Sonnentag, S., Binnewies, C., & Mojza, E. J. (2010). Staying Well and Engaged When Demands Are High: The Role of Psychological Detachment. *Journal of Applied Psychology*, 95(5), 965-976. doi:10.1037/a0020032
- Sonnentag, S., Dormann, C., & Demerouti, E. (2010). Not all days are created equal: The concept of state work engagement. In A. B. Bakker (Ed.), *Work engagement: A handbook of essential theory and research*. (pp. 25-38). New York, NY, US: Psychology Press.

- Sonnentag, S., Mojza, E. J., Binnewies, C., & Scholl, A. (2008). Being engaged at work and detached at home: A week-level study on work engagement, psychological detachment, and affect. *Work & Stress*, 22(3), 257-276. doi:10.1080/02678370802379440
- Spain, S. M., Miner, A. G., Kroonenberg, P. M., & Drasgow, F. (2010). Job performance as multivariate dynamic criteria: Experience sampling and multiway component analysis. *Multivariate Behavioral Research*, 45(4), 599-626. doi:10.1080/00273171.2010.498286
- Spielberger, C. D. (1968). State-Trait Anxiety Questionnaire for Adults. Redwood City CA: Mind Garden.
- Spielberger, C. D. (1975). Anxiety: State-trait process. In C. D. Spielberger & I. G. Sarason (Eds.), *Stress and anxiety (Vol. 1)* (pp. 115-143). New York: Hemisphere.
- Stein, M., Egenolf, Y., Dierks, T., Caspar, F., & Koenig, T. (2013). A neurophysiological signature of motivational incongruence: EEG changes related to insufficient goal satisfaction. *International Journal of Psychophysiology*, 89(1), 1-8. doi:10.1016/j.ijpsycho.2013.04.017
- Stein, N. L., & Oatley, K. (1992). Basic emotions: Theory and measurement. *Cognition and Emotion*, 6(3-4), 161-168. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psych&AN=1993-00400-001&site=ehost-live&scope=site>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680. doi:10.1126/science.103.2684.677
- Stevens, S. S. (1958). Measurement and man. *Science*, 127, 383-389. doi:10.1126/science.127.3295.383
- Tengblad, S. (2002). Time and space in managerial work. *Scandinavian Journal of Management*, 18, 543-565.
- Truong, K., van Leeuwen, D., & Neerincx, M. (2007). Unobtrusive Multimodal Emotion Detection in Adaptive Interfaces: Speech and Facial Expressions. In D. Schmorow & L. Reeves (Eds.), *Foundations of Augmented Cognition* (Vol. 4565, pp. 354-363): Springer Berlin Heidelberg.
- Van de ven, A. H., & Poole, M. S. (2005). Alternative approaches for studying organizational change. *Organization Studies* (01708406), 26(9), 1377-1404. doi:10.1177/0170840605056907
- Vecchiato, G., Cherubino, P., Maglione, A. G., Ezquierro, M. T. H., Marinozzi, F., Bini, F., . . . Babiloni, F. (2014). How to measure cerebral correlates of emotions in marketing relevant tasks. *Cognitive Computation*, 6(4), 856-871. doi:10.1007/s12559-014-9304-x
- Vroom, V. H. (1964). *Work and motivation*. New York: Jon Wiley.
- Warr, P. (2013). How to think about and measure psychological well-being. In R. R. Sinclair, M. Wang, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology: Measurement, design, and data analysis* (pp. 76-90). New York, NY, US: Routledge/Taylor & Francis Group.
- Warr, P., Bindl, U. K., Parker, S. K., & Inceoglu, I. (2014). Four-quadrant investigation of job-related affects and behaviours. *European Journal of Work and Organizational Psychology*, 23(3), 342-363. doi:10.1080/1359432X.2012.744449
- Westerink, J. D. M., van den Broek, E. L., Schut, M. H., van Herk, J., & Tuinenbreijer, K. (2008). Computing Emotion Awareness Through Galvanic Skin Response and Facial Electromyography. In J. D. M. Westerink, M. Ouwerkerk, T. M. Overbeek, W. F. Pasveer, & B. de Ruyter (Eds.), *Probing Experience* (Vol. 8, pp. 149-162): Springer Netherlands.

- Wieland, B. A., & Mefferd, R. B., Jr. (1969). Identification of periodic components in physiological measurements. *Psychophysiology*, 6(2), 160-165. doi:10.1111/j.1469-8986.1969.tb02895.x
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews. Neuroscience*, 5(6), 483-494. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=15152198&login.asp&site=ehost-live&scope=site>
- Woods, S. A., & Sofat, J. A. (2013). Personality and engagement at work: The mediating role of psychological meaningfulness. *Journal of Applied Social Psychology*, 43(11), 2203-2210. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-40168-004&site=ehost-live&scope=site>
s.a.woods@aston.ac.uk
- Xanthopoulou, D., Baker, A. B., Heuven, E., Demerouti, E., & Schaufeli, W. B. (2008). Working in the sky: A diary study on work engagement among flight attendants. *Journal of Occupational Health Psychology*, 13(4), 345-356. doi:10.1037/1076-8998.13.4.345
- Xanthopoulou, D., & Bakker, A. B. (2013). State work engagement: The significance of within-person fluctuations. In A. B. Bakker & K. Daniels (Eds.), *A day in the life of a happy worker*. (pp. 25-40). New York, NY, US: Psychology Press.

Figure 1: Time window defined by parameters L , M ,

N

Figure 2: Measurements in two overlapping time windows

Figure 3: Two contrasting examples of engagement state trajectories

Figure 4: Spaghetti-plot of engagement state trajectories (N=54)

Table 1: Proposed engagement traits

		High	Middle	Low
Trait1: Engagement level	The average level of engagement (mean) during the observed time window ($Mean_{LMN}$)	<p>213</p>	<p>412</p>	<p>212</p>
Traits 2: Engagement variability	The overall degree of variation in engagement during the observed time window (SD_{LMN})	<p>110</p>	<p>204</p>	<p>109</p>
Trait 3: Engagement polarity	The distance between the maximum and minimum level of engagement during the observed time window ($Range_{LMN}$)	<p>302</p>	<p>301</p>	<p>210</p>