

REPRINT

Original Article

Individual behavioral phenotypes: An integrative meta-theoretical framework. Why 'behavioral syndromes' are not analogues of 'personality'

Jana Uher *

*Comparative Differential and Personality Psychology,
Free University Berlin*

*Correspondence:

Comparative Differential and Personality Psychology
Department of Educational Science and Psychology
Free University Berlin
Habelschwerdter Allee 45
D-14195 Berlin, Germany
email: uher@primate-personality.net

Abstract

Animal researchers are increasingly interested in individual differences in behavior. Their interpretation as meaningful differences in behavioral strategies stable over time and across contexts, adaptive, heritable, and acted upon by natural selection has triggered new theoretical developments. However, the analytical approaches used to explore behavioral data still address population-level phenomena, and statistical methods suitable to analyze *individual* behavior are rarely applied. I discuss fundamental investigative principles and analytical approaches to explore whether, in what ways, and under which conditions individual behavioral differences are actually meaningful. I elaborate the meta-theoretical ideas underlying common theoretical concepts and integrate them into an overarching meta-theoretical and methodological framework. This unravels commonalities and differences, and shows that assumptions of analogy to concepts of human personality are not always warranted and that some theoretical developments may be based on methodological artifacts. Yet, my results also highlight possible directions for new theoretical developments in animal behavior research.

Keywords:

animal personalities; behavioral profiles; behavioral style; behavioral syndromes; behavioral types; characters; correlated traits; individual differences; individuality; personality; personality traits; temperament

Table of contents

Introduction	3
Central characteristics of meaningful individual differences in behavior.....	4
Meta-theoretical foundations and basic methodological implications	5
Repeated observations and the Principle of Aggregation	5
Differential standardization	5
Test-retest reliability	6
Two complementary research approaches: Variable- and individual-oriented analyses.....	7
Variable-oriented approaches	7
Individual-oriented approaches	8
The nature of individual behavioral phenotypes and issues of terminology.....	9
Jingle-Jangle-Fallacies.....	10
Different kinds of terms	10
Various meta-theoretical concepts of consistency	11
Internal consistency of empirical estimates	11
Individual response specificity	11
Classification of environmental situations	12
Cross-situational consistency	13
Individual situation-behavior profiles.....	14
Individual-situation transactions.....	15
Hierarchical taxonomic models.....	15
Systematic methodological approaches to identify multi-dimensional structures in individual differences.....	17
Basic research principles and classes of methodological approaches	17
The Behavioral Repertoire x Environmental Situations Approach.....	18
Types of individuals.....	19
Longitudinal stability	20
Differential stability and stability across age groups.....	20
Variable-oriented and individual-oriented concepts of longitudinal stability.....	20
Discussion.....	21
Terminological misunderstandings about 'trait', 'character', and 'personality structure'.....	21
Conceptual misunderstandings about 'correlated traits' and 'independent traits'	22
Conceptual misunderstandings about 'personality' and 'animal personalities'	22
Which species do actually exhibit multi-dimensional structures of individual differences?.....	22
Meta-theoretical and methodological concepts underlying 'behavioral syndromes'	23
General limitations of syndromal concepts	24
'Reaction norms' versus individual situation-behavior profiles	25
Ambiguities in concepts of 'consistency across contexts and situations'.....	25
Conclusions.....	26
Acknowledgements	27
References.....	27
Glossary of meta-theoretical concepts, and basic principles and concepts of measurement for the empirical study of individual behavioral phenotypes.....	33

Introduction

Individual differences in animal behavior have become a new focus of interest. The four key questions on the causation, ontogeny, adaptivity, and phylogeny that Tinbergen (1963) originally formulated to study behavior at the species level are increasingly refined to the individual level. Animal researchers study genetic variation (Fidler, van Oers, Drent, Kuhn, Mueller, & Kempenaers, 2007), epigenetic influences (Groothuis & Carere, 2005), psychobiological parameters (Kinnally, Lyons, Abel, Mendoza, & Capitanio, 2008; Schneider & Suomi, 1992), and developmental conditions (Fairbanks & McGuire, 1993) associated with behavioral differences among individuals. Assuming that “nothing in biology makes sense except in the light of evolution” (Dobzhansky, 1973), sophisticated models are developed to explain their emergence from evolutionary viewpoints (Dall, Houston, & McNamara, 2004; Wolf, van Doorn, Leimar, & Weissing, 2007). Accordingly, individual behavioral differences are currently understood as adaptive strategy differences based on ecological trade-offs with different costs and benefits. Their investigation seems therefore illuminative for evolutionary theory (Bell, 2007a).

Until recently, however, animal individual differences were not considered significant variation (Bell, 2007a; Sih, Bell, & Johnson, 2004; Sih, Bell, Johnson, & Ziemba, 2004). This dramatic change in the theoretical interpretation of between-individual behavioral variation is surprising because random variation is apparent in any empirical investigation, and the ways that animal researchers collect and analyze behavioral data have remained more or less the same. Many studies continue to compare averages among groups of individuals, or to analyze sample correlations, whereas suitable methods to analyze the behavior of the *individual* are rarely applied (Uher, 2011). Hence, what behavioral phenomena do animal researchers actually study when they refer to “individual differences in behavior”? Is it justifiable to interpret any individual behavioral variation as meaningful? Do some species actually exhibit individual behavioral strategies, or are these theoretical ideas developed on the basis of methodological artifacts?

The empirical database and the body of theories explaining origins and mechanisms of individual behavioral differences are rapidly growing. But the meta-theoretical and methodological foundations needed to describe the empirical phenomena to be explained in the first place, and to test the proposed causal hypotheses, concepts, and models empirically are still largely lacking (Uher, 2008a, b). In this article, I introduce fundamental investigative principles needed to explore systematically and empirically whether, in what ways, and under which conditions individual differences in behavior are actually meaningful. Such explorations are essential to unravel the mechanisms and processes of their causation, ontogeny, adaptivity and phylogeny empirically, and to avoid basing theoretical developments on mere methodological artifacts.

Rather than focusing on particular concepts, I explore the basic theoretical ideas underlying various theoretical concepts to reveal commonalities and differences among them. Unlike most animal researchers, my focus is therefore on meta-theoretical concepts that describe the behavioral phenomena of interest on the most abstract level. I thereby refer primarily to *structural concepts* describing empirical structure; not to *causal concepts* explaining biological causation on which most researchers focus (Eysenck & Eysenck, 1985; Westmeyer, 1997). I integrate these concepts within an overarching meta-theoretical and methodological framework that can guide research on more specific concepts and empirical phenomena. For a better orientation, all meta-theoretical concepts and basic methodological principles introduced here are summarized in a glossary in the appendix.

This framework draws heavily on concepts and principles established for research on humans primarily developed in ‘differential psychology’ and ‘personality psychology’. The advantages of this cross-disciplinary approach are twofold. First, rather than proposing entirely new concepts of unknown empirical value, it relies on the comprehensive empirical experiences, methodological, statistical, and theoretical developments made in those disciplines over the last century since Francis Galton (1869) and William Stern (1911) laid their first statistical and methodological foundations. I show that many meta-theoretical

concepts and methodologies developed for humans are equally applicable to animal species, and that some concepts developed for animals are almost identical.

Second, this explicitly cross-disciplinary perspective allows me to evaluate and integrate new concepts, such as those of 'animal personalities' or 'behavioral syndromes' (Groothuis & Carere, 2005; Sih et al., 2004a, b). I differentiate these concepts from meta-theoretical viewpoints and argue that their equation with concepts of human 'personality' is not always warranted. I discuss that, at least in some cases, hypotheses of analogy reflect fundamental conceptual misunderstandings that derive from discipline-specific usages of basic terms. This needs to be resolved to enable future exchange and comparisons of theoretical and empirical developments across disciplines and species.

Finally, I highlight that theoretical models explaining how and why multidimensional patterns of meaningful individual differences in behavior emerge in some species are still missing in behavioral ecology and evolutionary theory. This shows that meta-theoretical and methodological considerations can also trigger new theoretical developments in biology.

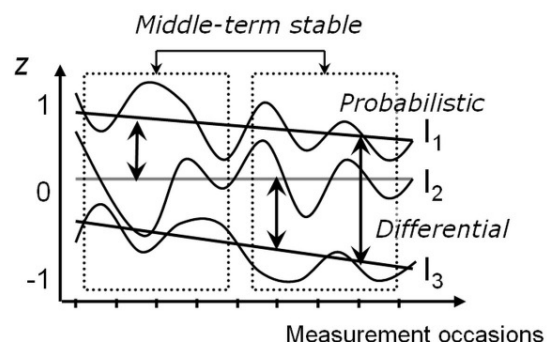
Central characteristics of meaningful individual differences in behavior

Researchers of animal behavior are confronted with tremendous methodological challenges when they attempt to determine whether individual differences are meaningful in a given species because individual behavior changes flexibly at any moment and often shows considerable plasticity over the lifetime. Keeping this in mind, I now consider three central characteristics of meaningful individual differences in behavior.

First, as members of the same species, individuals are necessarily similar in their behavior to a considerable extent. Yet they differ from one another in the probabilities of displaying particular species-typical behaviors. Individual differences are thus *probabilistic* patterns of behavior. Second, individual differences can only be regarded as meaningful and hypothesized to be adaptive, heritable, and acted upon by natural selection (Bell, 2007a,b; Sih et al., 2004a,b) if they reflect *individual-specific* patterns. This means that the individuals' behavioral probabilities must be relatively *stable* at least for some time. Third, relatively stable probabilistic behavioral patterns are informative about the individual only if they deviate from those of other individuals. This means they must be set in relation to those of other individuals. Hence, concepts of meaningful individual differences are necessarily *differential* concepts. These characteristics of meaningful differences in behavior are always implied when I refer to *individual behavioral phenotypes* in this article (Figure 1).

Figure 1
Three core characteristics of individual behavioral phenotypes and the Principle of Aggregation

Individual behavioral phenotypes denote individual-specific patterns of behavior that are a) *probabilistic* b) *differential*, i.e. differ among individuals (I), and c) relatively *stable* at least over some time. Such patterns can be extracted from the fluctuations of spontaneous behavior by sufficient aggregation over repeated measurements (*Principle of Aggregation*), and by differential standardization of scores, such as z-standardization across individuals (z), in which the sample mean is set to zero and its standard deviation to one.



Given these characteristics it becomes apparent that studying individual behavioral phenotypes requires fundamentally different concepts and methodologies than studying the average behavior patterns of particular groups of individuals, such as species, sex or age groups. This seems still not well recognized in animal research. Yet without proper measurement of the phenomena to be explained, all analyses are necessarily compromised

and theoretical models cannot be tested empirically. If research methodology is insufficient, the relevance and heritability of individual behavioral phenotypes may be underestimated or even completely overlooked. Given the complexity and multi-factorial genesis of natural systems, relations to genetic, physiological, or environmental parameters may be only moderate in strength; these may be hard to detect unless reliable empirical behavioral measures are used.

Meta-theoretical foundations and basic methodological implications

Above I specified that the nature of individual behavioral phenotypes is necessarily probabilistic and individual-specific, and thus also differential and relatively stable within the individual. Next I introduce the meta-theoretical and methodological foundations needed to establish empirically that some variations in behavioral data meet these requirements.

Repeated observations and the Principle of Aggregation

Identifying probabilistic individual patterns presupposes, for mathematical reasons, *repeated observations* of the same individuals. Moreover, individual behavioral phenotypes are always partially masked by considerable transient variations derived from behavioral flexibility, ephemeral environmental conditions, and, like any empirical measurement, from measurement error (Bell, 2007a, b; Groothuis & Carere, 2005; Mischel, Shoda, & Ayduk, 2007; Sih et al., 2004a, b). According to classical theories of error in measurement, any measure for an individual at any occasion differs from a hypothetical true score on account of random error. If measurements for an individual are obtained repeatedly on many occasions, the scores will vary around the individual's true score. The mean of this distribution approximates the true score (Figure 1; Kline, 2000).

These considerations also underlie the *Principle of Aggregation* stating that measurements should be sufficiently aggregated to obtain reliable empirical estimates (Epstein, 1979, 1980; Mischel, Shoda, & Mendoza-Denton, 2002; Rushton, Brainerd, & Pressley, 1983). Aggregation is particularly important for measurements of behavioral phenomena, in which error variation is higher than in measurements of less dynamic phenomena, such as morphological phenotypes. Consequently, individual behavioral phenotypes can only be estimated empirically through *aggregated measures*. This is also implied by their labeling as individual behavioral strategies or tendencies that, by definition, reflect *repeated* patterns of behavior.

Differential standardization

Probabilistic behavioral patterns can only be individual-specific if they differ among individuals; that is, if they are differential. Absolute behavior scores confound the sample's average behavior patterns with those that may be individual-specific. Therefore, empirical estimates must be set in relation to those of other individuals, such as through *statistical standardization* across all individuals. In z-standardization, for example, the individual scores are centralized and standardized, such that the sample mean is 0, and its standard deviation is 1. It does not change the skewness or kurtosis, that is, it does not presuppose norm-distribution. Hence, a z-standardized score is a measure of deviation from the sample's mean that informs about a particular individual (Figure 1).

When behavioral variables show different score distributions (e.g. skewed or norm distributions) or when they are of different measurement types (e.g. frequencies or durations), scores cannot be directly compared. For example, findings that all individuals visually inspect the environment more frequently than they show agonistic behaviors (Suomi, Novak, & Well, 1996) can obscure that some individuals are less vigilant than others, and that some other individuals are more agonistic than others. Standardization is a statistical means to make such variables comparable to one another in regard to the between-individual variation they reflect. It allows three basic kinds of comparisons. 1) An individual's

score can be placed within the context of the score distribution of the population; for example, an individual may score higher than 85% of the population. 2) Specific individuals can be compared with one another; for example, this particular individual may still score lower than another one. 3) An individual's scores on different variables can be rank-ordered from the most to the least pronounced of its profile.

Test-retest reliability

Individual-specificity furthermore implies stability in how individuals behave. Measurement of individual phenotypes requires evidence that an individual's probabilistic, differential patterns are relatively stable across repeated occasions (Figure 1). A basic measurement criterion is therefore considerable *test-retest reliability* (also called temporal reliability or repeatability; Bell, 2007a; John & Soto, 2007; Réale, Reader, Sol, McDougall, & Dingemanse, 2007; Sih et al., 2004b). Reliability can be estimated through test-retest correlations, such as intra-class correlations (for statistical coefficients see Shrout & Fleiss, 1979). A reliability score of $r = 0.50$ indicates that about half of the variance of the empirical measure is attributable to variation of the hypothetical true scores, and half is attributable to transient or error variation. For example, measurements of exploratory behavior of individual great tits (*Parus major*) showed low to moderate test-retest reliabilities of $r = 0.48$ for 111 males, but fairly low test-retest reliabilities of $r = 0.27$ for 74 females (Dingemanse, Both, Drent, van Oers, & van Noordwijk, 2002).

A critical point in reliability estimation is the time span between measurement repetitions. Two opposite effects must be considered. First, the greater random and error variation, the more measurement occasions are needed to obtain reliable estimates. More data must be aggregated. Aggregation is particularly important in naturalistic studies in which many environmental conditions may influence individuals in uncontrolled ways. Therefore, most measures are obtained in laboratory settings (e.g. Janczak, Andersen, Færevik, Bøe, & Bakken, 2002; van Oers, Klunder, & Drent, 2005), in which ecological validity is compromised. Second, although individual behavioral phenotypes are inherently stable over some time, they also show gradual change and long-term development. Naturally, changes in the individuals' true scores are more likely with increasing time spans; this limits the time periods that are useful to estimate true scores at any one time. Such baseline estimations are, however, needed to study the longitudinal changeability of true scores (see below).

The proportions of true and random variation and the changeability of true scores cannot be estimated in advance. The number of repetitions and the optimal time spans in which they can be obtained in a given species is always an empirical question. For example, a great ape study measured 76 behavior variables repeatedly over 2-3 weeks, and after a fortnight's break again repeatedly in a second 2-3 weeks period. In total, each individual was observed for 67 hours within the overall 50-day period. The behavioral raw data were aggregated within each period; their average test-retest correlation of $r = .78$ showed that this design of intense and repeated observations within the studied time span allowed obtaining reliable estimates of individual behavioral phenotypes (Uher et al., 2008).

A frequently raised concern about repetitions is their potential to induce habituation and learning effects (Carere & van Oers, 2004; Kralj-Fišer, Scheiber, Blejec, Moestl, & Kotrschal, 2007). Yet this matters only if the experimental conditions lose their relevant properties with repetition, such as their degree of novelty or risk. Changes in the individuals' absolute behavior scores need not affect the relative differences among them as studied with standardized measures. Therefore, rather than repeating measurements in exactly the same experimental conditions, variations of standard procedures can and should be studied more often to reduce the potential of habituation and learning effects and to measure individual behavioral phenotypes reliably (Réale et al., 2007; Uher, Asendopf, & Call, 2008; Verbeek, Drent, & Wiepkema, 1994).

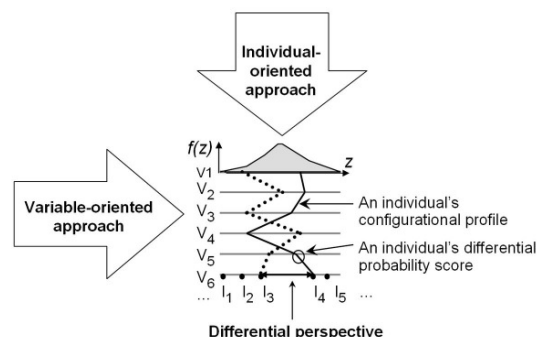
Two complementary research approaches: Variable- and individual-oriented analyses

Studying many individuals in multiple behaviors results in a data matrix of X_i individuals by Y_j variables with i as the number of individuals and j as the number of variables. In this matrix, behavioral probabilities can be explored for individual-specific patterns from two orthogonal perspectives (Figure 2; Stern, 1911; cf. Lamiell, 2003).

Figure 2
Two complementary approaches to study individuals

Variable-oriented approaches focus on the differences in behavioral probabilities among the individuals (I) of a population that can be described with (dimensional) latent theoretical variables (V). They explore properties of the population. Individual-oriented approaches, in contrast, focus on the single individuals (I) and their individual-specific configuration of behavioral patterns. If individual profiles are based on

standardized measurement scores, such as z-scores (z), they illustrate the configuration of individuals' differential scores and explore properties of single individuals. These two complementary approaches allow quantifying an individual's behavioral uniqueness (using individual-oriented approaches) based on empirical comparability with other individuals (using variable-oriented approaches).



Variable-oriented approaches

The *variable-oriented perspective* focuses on the j measurement variables and studies their score distributions across all i individuals. It studies the population, but not the single individual. For example, piglets (*Sus scrofa domestica*) differ in their frequencies of struggling when being held on the back (D'Eath & Burn, 2002). These frequencies were unimodally distributed, whereas exploratory behavior in bi-directionally selected great tits was bimodally distributed (Drent, van Oers, & van Noordwijk, 2003). Test-retest reliabilities are analyzed as correlations of between-individual variation over time (Figure 3). The reliabilities for great tits (Dingemanse et al., 2002) and great apes mentioned above were analyzed this way.

The distributions of the individuals' true scores across the composite of the population, as estimated by test-retest reliable measurements, can be modeled by latent variables. These are hypothetical variables describing the structures that underlie the empirical data. These may be dimensional, but also non-dimensional types of variables, such as those describing distinct clusters of individuals. The piglets' struggling estimates, for example, can be described by a dimensional latent variable that differentiates individuals with high probabilities of struggling from those with moderate and low probabilities.

The individuals' relative positions on such latent variables naturally depend on the variation among those individuals to which they are compared. Piglets scoring high within their own litters may obtain much lower scores when compared to large collections of litters. Quantifying individual behavioral phenotypes therefore always requires specification of the *reference population*; this is also basic to comparisons among populations (Uher, 2008a, b).

Variable-oriented analyses exploring single variables are called '*variation research*'; those studying covariations of individual differences among different variables are called '*correlation research*' (Stern, 1911). Intercorrelational patterns among multiple variables are identified statistically, such as with *factor analysis* (Basilevsky, 1994; Floyd & Widaman, 1995; Goldberg & Digman, 1994; Gorsuch, 1983; Lee & Ashton, 2007). These methods detect latent structures in the empirical relations among many variables and gather together those that are most highly correlated. They indicate to what degree empirical measurement variables can be explained and summarized by a much smaller number of synthesized variables, so-called *factors*. For statistical factor extraction, the sample size should exceed the number of variables. This is not always considered in animal research. Ideally, the sample size should be ten times the numbers of variables, yet this requirement is often

difficult to meet in animal studies. Other recommendations, suggesting individual-to-variable ratios of 4-5:1 (Floyd & Widaman, 1995) can be more easily fulfilled. Variable-oriented types of factor analysis that extract a few broad latent dimensions explaining between-individual variations are called *R-factor analysis*.

The meaning of a factor, and hence its label, depends on the meaning of those measurement variables from which it is synthesized. Factor labels are therefore necessarily more abstract. As synthesized variables, factors can be estimated empirically by calculating composite estimates of those measurement variables that they statistically explain. The summation of the composite estimates is called a *scale*. For example, to study shyness-boldness differences in dumpling squids (*Euprymna tasmanica*), Sinn and colleagues (2008) developed measurement scales by summing several standardized behavioral estimates that showed robust interrelations across years and in independent populations. Pools of measurement variables covering several scales, and thus different, independent factors, are called *inventories* (Costa & McCrae, 1992) or *test batteries* (Gorlyn, Keilp, Tryon, & Mann, 2005).

Animal researchers often use principal component analyses that are conceptually different from factor analyses. Principal component analyses are based on the assumption that the measurement variables are the causes of the composite variables; their aim is data reduction. Factor analyses, in contrast, are based on the assumption that latent variables, the factors, cause the empirical variation in the measurement variables. Their aim is to explain the variation in the measurement variables by construing latent variables, which also enables comparisons among and generalizations across studies. The meta-theoretical concept of individual behavioral phenotypes suggests their description as latent variables contributing to an individual's observable behavior. Factor analyses are therefore conceptually better suited for their empirical investigation than principal component analyses (Capitanio, 1999),

To analyze whether individual differences are meaningful, animal researchers sometimes use analyses of variance (ANOVA), such as general linear models (GLM), and include individual identity as a random factor. However, unless the behavioral data entered in the analyses reflect test-retest reliable differential scores, such analyses can only test whether there is between-individual variation in the particular behaviors examined; yet this could also be due to ephemeral variation and measurement error. Moreover, analyses of variance cannot test whether the individuals differ in consistent ways in all the behaviors examined (see below), as this is possible for factor analysis.

Further suitable statistical methods are, for example, multilevel analysis (Hox, 2002; Raudenbush & Bryk, 2002; Reise, Ventura, Nuechterlein, & Kim, 2005; Snijders & Bosker, 1999), cluster analysis (Aldenderfer & Blashfield, 1984), configurational frequency analysis (von Eye, 2002), structural equation modeling (Hoyle, 1995, 2007; Kline, 2005), multi-dimensional scaling (Borg & Groenen, 2005; Cox & Cox, 2001), or item response models (Morizot, Ainsworth, & Reise, 2007). Some of these methods permit probabilistic approaches that are most appropriate for analyses of probabilistic concepts like individual behavioral phenotypes.

Individual-oriented approaches

The *individual-oriented perspective* explores the matrix of X_i individuals by Y_j variables from an orthogonal view and studies each individual's scores on all j variables. This score configuration can be illustrated as a profile that plots the individual's differential probabilities of behavior against the variables on which they are obtained. It depicts the individual's deviations from the average behavioral patterns of its reference population, and thus reflects a property of the individual, not of the population (Figure 2; Bergman & Trost, 2006; Stern, 1911; Uher et al., 2008).

These two orthogonal perspectives on variables and individuals are complementary. Variable-oriented analyses are needed to identify latent variables along which individuals of a population differ. Individual-oriented analyses, in turn, are needed to estimate each

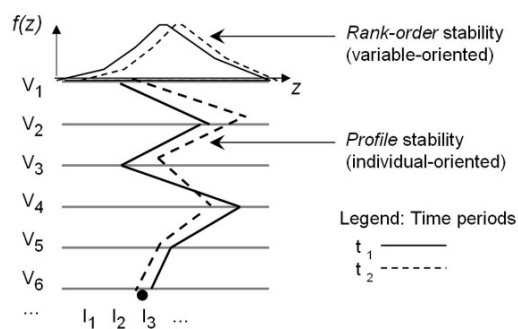
individual's unique configuration of differential scores on multiples of such variables. This allows quantifying the individuals' behavioral uniquenesses—their individual behavioral phenotypes—based on empirical comparability with other individuals (Mischel et al., 2002; Stern, 1911).

Individual-oriented analyses can also focus on multiple individuals to identify groups of similar individuals based on their *distinct profile shapes* (Figures 2 and 3). This is called 'comparison research' (Stern, 1911). Individual-oriented types of factor analyses that extract prototypes of individuals are called inverse or *Q-factor analysis*. Their covariance matrices are transposed from the covariance matrices used for variable-oriented R-factor analysis such that the j variables are the cases and the i individuals are the variables (Basilevsky, 1994; Floyd & Widaman, 1995; Goldberg & Digman, 1994).

Figure 3

Variable- and individual-oriented stability

Variable-oriented stability refers to the stability of the differences among individuals (I) between two periods of measurement (t_1 and t_2). This rank-order stability characterizes properties of the population. *Individual-oriented stability* refers to the stability of the single individual's configuration of differential scores on various latent theoretical variables (V). This profile stability characterizes properties of the individual, not of the population. Empirical stability over intermediate periods is needed to demonstrate



that the empirical measures are reliable estimates of individual behavioral phenotypes. Empirical stability of test-retest reliable measures over longer-periods of time refers to the stability and changeability of their hypothetical true scores.

In the great ape study mentioned above, individual profiles were construed from the individuals' standardized probability scores on various behavioral variables. The profile shapes were individually unique because they were based on relative, not on absolute, scores that reflected each individual's deviations from the sample mean on multiple variables (Figures 2 and 3). This was done separately for two non-overlapping time periods. The average test-retest reliability of $r = 0.76$ to 0.80 showed that these differential profiles were individual-specific, and could be considered reliable estimates of individual behavioral phenotypes (Uher, 2011; Uher et al., 2008).

The nature of individual behavioral phenotypes and issues of terminology

Before I elaborate specific concepts, I wish to emphasize that individual behavioral phenotypes are merely theoretical in nature because the empirical phenomena they denote are temporal and differential patterns in behavior that, in contrast to single behavioral acts, cannot be directly observed at any one time. Individual behavioral phenotypes can therefore only be *inferred* from repeated observations of multiple individuals over time. They are mere theoretical conceptions, so-called *constructs*, that laypeople and researchers construe to describe individual-specific patterns that they either notice intuitively from mental aggregations across memorized behavior observations—with all the biases this entails—or that they extract statistically from behavioral data (Kelly, 1955; Mischel et al., 2007; Uher & Asendorpf, 2008).

The difference between concepts of directly and indirectly observable phenomena of behavior is also often reflected in terminology; single behavioral acts are labeled *aggression*, whereas corresponding individual behavioral phenotypes are labeled *aggressiveness*. Terminology is *per se* arbitrary. Laypeople and researchers are free to conceive theoretical ideas to describe and explain empirical phenomena, and to label them with particular terms (Kelly, 1955; Tellegen, 1993). Plurality in concepts and terms is therefore inevitable. Those concepts and terms denoting indirectly observable phenomena are particularly diverse,

whereas those denoting directly observable phenomena, at which we can point, are much less diverse. People agree substantially on what they consider tables or aggressive acts, although the latter need more clarification than the former because they are not static by nature.

Establishing a coherent terminology for individual behavioral phenotypes is additionally complicated because the terms denote phenomena of human everyday life. Researchers investigate these phenomena only after they have acquired, in their pre-scientific life, particular ideas, so-called *personal* and (*locally*) *socially shared constructs* (Kelly, 1955; Tellegen, 1993), and their labels. Despite all scholarly efforts, they cannot completely prevent this from influencing their concepts and terms as researchers. Hence, unlike researchers of quantum physics, researchers of individual behavioral phenotypes are not entirely free to create new terms. Some even refer explicitly to lay definitions in general dictionaries to differentiate *scientific constructs* (Réale et al., 2007); this is impossible for quantum physicists. The inevitable interference of everyday constructs and scientific constructs, and the resulting resemblance of folk terms and scientific terms (Tellegen, 1993) lead to various misunderstandings between and among scientists and laypeople.

Given this, a bewildering diversity of terms labeling individual behavioral differences in humans and animals emerged over the last century (Gosling, 2008; Réale et al., 2007; Sih et al., 2004b; Uher, 2008a). The most frequent terms, in alphabetical order, are 'animal personalities', 'behavioral profiles', 'behavioral strategies', 'behavioral styles', 'behavioral syndromes', 'behavioral types', 'characters', 'coping strategies', 'coping styles', 'dispositions', 'individual (behavioral) characteristics', 'individual distinctiveness', 'individuality', 'personality', 'personality dimensions', 'personality factors', 'personality types', 'personality traits', 'predispositions', 'reactivity', 'responsiveness', 'temperament', 'traits', and 'trait dimensions'. This variety substantially impedes literature searches (Gosling, 2008) and cross-disciplinary exchange (Uher 2008b).

Jingle-Jangle-Fallacies

Many researchers currently assume these terms denote the same or similar theoretical concepts (Bell, 2007b; Fidler et al., 2007; Gosling, 2008; Groothuis & Carere, 2005; Réale et al., 2007; Sih et al., 2004b; Weinstein, Capitano, & Gosling, 2008; Wolf et al., 2007). Some even advocate a more unified usage of particular terms to facilitate exchange across disciplines (Gosling, 2008). Others include terms of neighboring disciplines in their publication titles and keywords (Bell & Sih, 2007; Weiss, King, & Murray, 2011). But is equating these terms really justified? The relative arbitrariness in labeling concepts masks differences and similarities among them, which hampers direct comparisons effectively. Two fallacies are known. *Jingle-fallacies* occur when the same term refers to different concepts (Thorndike, 1903); *jangle-fallacies* when different terms refer to the same concept (Kelley, 1927).

Whether different terms are actually exchangeable and denote the same theoretical concepts requires comparison of their underlying meta-theoretical concepts. There are difficulties in doing this because terms are sometimes insufficiently defined or used inconsistently. To reveal basic conceptual principles, I therefore focus on the most frequent terms that are either well established, or for which extensive definitions are published. I wish to emphasize that, because researchers differ in their definitions and usages of terms, I can only—and only want to—outline overall meta-theoretical structures that I hope provide some orientation in basic terminological issues on which researchers of various disciplines can build.

Different kinds of terms

A first differentiation among terms can be made based on the *degree of abstraction* with which they describe the phenomena of interest. Those that do not specify any particular kinds of behavioral variation are terms of higher order than those that specify, though to

varied degrees of abstraction, particular kinds of behavioral variation. Most terms listed above denote individual behavioral patterns on very abstract levels. Others, such as 'coping strategies', 'coping styles', 'reactivity', and 'responsiveness' refer to particular subsets of behaviors that reflect how individuals cope with, react or respond to particular, mostly challenging environmental conditions. 'Temperament' also refers to particular subsets of behaviors that mostly indicate some kind of affect, activity, and attention (Rothbart & Bates, 1998), whereas it is typically not used, for example, for feeding or problem-solving behaviors. Although researchers disagree considerably on which particular kinds of behaviors they refer to as 'temperament', they agree that 'temperament' does not cover any kind of behaviors as some other terms do. Further terms not listed above refer to even more specific kinds of behavior, such as 'aggressiveness', 'emotionality', 'neophobia-neophilia', 'risk-taking', or 'shyness-boldness'.

A second differentiation of terms can be made based on their reflecting *variable-oriented or individual-oriented* concepts, which is often, though not always, indicated by their labels. This distinction is most prominently reflected in the name of the discipline studying individual behavioral phenotypes in humans, 'differential and personality psychology'. 'Differential psychology' is concerned with the description and explanation of the ways in which individuals differ from one another, for which variable-oriented concepts were developed. 'Personality psychology', in contrast, is concerned with the description and explanation of the coherent and integrated set of the behavioral patterns of the single person, for which (person-) individual-oriented concepts were developed.

This distinction is also reflected in terminology. The theoretical (often dimensional) variables inferred from stable rank-order distributions are typically called '*personality dimensions*'. The individual's relatively stable configuration of scores across different 'personality dimensions' is typically referred to as its '*personality*' or '*personality profile*' (Figure 3). For reasons of brevity, psychologists often briefly refer to both as 'personality'. This obscures the conceptual differentiation between the *population-level* concept of 'personality dimensions' and the *individual-level* concept of an individual's 'personality'. This terminological imprecision caused various cross-disciplinary misunderstandings discussed below.

Various meta-theoretical concepts of consistency

The distinction of variable-oriented (i.e. population-level) and individual-oriented (i.e. individual-level) terms is also basic to more fine-grained concepts of individual behavioral phenotypes that focus on different kinds of *consistency*.

Internal consistency of empirical estimates

More complex and abstract latent variables require more diverse empirical estimates for reliable measurements (Costa & McCrae, 1995; Eysenck, 1990; Uher, 2008b). They can be analyzed for how well each measure relates independently to the other measures, and how they are related overall. This variable-oriented reliability of composite empirical estimates is called *internal consistency* (for statistical coefficients see Shrout & Fleiss, 1979). In squids studied in feeding tests, for example, number of feeding strikes, latency to first feeding strike, and feeding rate (both reverse scored) showed high internal consistency as empirical estimates of a latent 'shyness-boldness' dimension (Sinn, Gosling, & Moltschanivskyj, 2008). Conversely this means that estimating complex concepts like 'risk taking' versus 'risk avoiding' or 'shyness-boldness' with single measures (as in Bell & Sih, 2007; van Oers et al., 2005a) compromises the reliability and validity of their measurement.

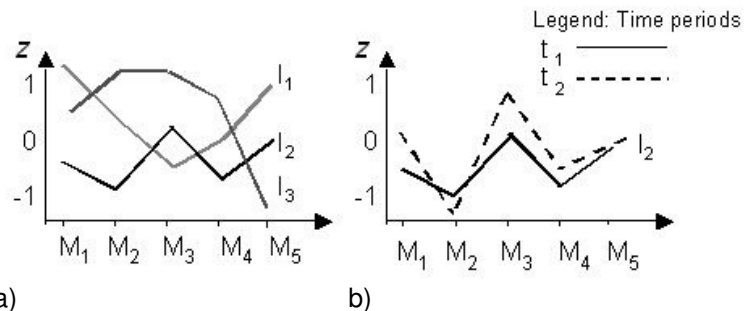
Individual response specificity

Sometimes, internal consistency is low although the studied behaviors have similar functions and meanings. However, this need not indicate lack of relatedness or validity. It can

derive from more fine-grained differences among individuals. For example, prior to their feeding, chimpanzees (*Pan troglodytes*) typically show various behaviors of arousal, such as rocking, grinning, vocalizing, or pacing. Yet which and how many of these behaviors are typically displayed varies substantially among individuals (Figure 4a). Some chimpanzees primarily rock and vocalize, others primarily vocalize and pace, and some others show many pleasure grins, but hardly ever rock, vocalize, or pace (Uher, 2011). The various behaviors indicating arousal are not displayed by all individuals. If data collection is restricted to only a few behavioral measures, those individuals who primarily show other behaviors representing the same general construct will be misclassified, with the effect that all analyses based on these measures will be compromised (Asendorpf, 1988; Lacey, 1950; Shoda, 1999; Uher et al., 2008).

Figure 4
Individual response specificity and individual response profiles

a) Individuals (I) can differ in their probabilities with which they show various meaningfully or functionally related behaviors (M) in a particular situation. Variable-oriented intercorrelations are then only low to moderate. b) Individual-oriented



approaches can reveal whether this is due to response specificity. Then individual response profiles that plot each individual's configuration of relative probabilities (z) against various such behaviors (M) are stable over some time periods (t_1 and t_2). Individuals can be compared in their distinct profile shapes and mean profile levels.

This phenomenon, called *individual response specificity*, emerges in behavior (Asendorpf, 1988; Uher, 2011), and especially in physiological responses (Fahrenberg, 1986; Foerster, Schneider, & Walschburger, 1983; Lacey, 1950). It is analyzed from individual-oriented viewpoints and illustrated with *individual response profiles* that plot the relative probabilities against various behaviors of similar functionality and meaning (Figure 4a). As ascertained above, the “particular configuration of behaviors that an individual expresses” can be considered a property of the individual (Bell 2007a, p. 755), *only* if it reflects relatively stable individual-specific deviations from average species-specific patterns. The profiles must therefore be based on standardized scores, and the particular shape of an individual's response profile must be shown to be test-retest reliable (Figure 4b).

Clearly, individual-specific response patterns can emerge only when the behavioral measures do not covary much on the sample level (Asendorpf, 1988). Thus, if individual-oriented analyses reveal reliable individual response profiles, variable-oriented correlations can only be low to moderate. Distinct individual response profiles can be empirically compared in both their profile shapes and their mean profile levels, provided they are based on standardized scores.

Classification of environmental situations

Further fine-grained variations of individual behavioral phenotypes can be identified by differentiating the situations in which they occur. Situations are complex constellations of stimuli or conditions in the abiotic, biotic, and conspecific environments (Capitanio, 2004) that can affect individuals by both their presence and their absence, such as presence of same-sex conspecifics, and absence of predator cues. Situations can be classified by the specific properties they represent for particular individuals, groups of individuals, or populations (Sih et al., 2004b; Ten Berge & De Raad, 1999; Uher, 2008b), such as degrees of competition or predation risk. These may differ, for example, among age groups, such that, for snake-naive infant monkeys, snakes may represent less strong threats than for snake-experienced adult monkeys (Cook & Mineka, 1989).

Not every situation is suited for the empirical study of individual behavioral phenotypes. Whether individual differences can emerge at all depends on two aspects, situational relevance and situational strength. *Situational relevance* refers to the *qualitative* properties of situations, that is, to the type of information to which the individuals respond (Funder, 1995, 1999; Tett & Guterman, 2000). That particular behaviors are not shown does not necessarily mean that the individual is scoring low, or that the latent dimension they are supposed to measure is an invalid construct (Capitanio, 2004). Situations have to feature cues indicating for the individuals that particular behaviors may be functional so that they will display these behaviors with some probability.

Situational strength, in contrast, refers to the *quantitative* properties of situations. It denotes how compelling situations are for the individuals, and how much they restrict their behavior. Strong situations force the individuals' behavior into specific channels, thus preventing the emergence of differences among them (*cf.* Réale et al., 2007); whereas weak situations permit individual-specificity to emerge (Mischel, 1977; Tett & Guterman, 2000). Note that the labeling of situations as strong or weak has nothing to do with their meaning for the individuals and for human observers, such as the risk of falling prey to predators.

If particular responses to predators are used to estimate 'shyness-boldness', the proportion of individuals classified as 'shy' or 'bold' depends on the strength of the threat induced by the particular predators or predator models in the particular study situation. If the threat is too strong, it will make most individuals appear 'shy', including those that would be classified as 'bold' when confronted with a more moderate threat. For example, trapped wild Grey mouse lemurs (*Microcebus murinus*) showed considerable individual differences in an open field test that did not emerge in highly risky feeding situations on the ground (Dammhahn, 2009).

In summary, study situations have to differentiate well among the individuals; they must have high *discriminatory power* (Kline, 2000). Adjusting situational strength in empirical studies is important to avoid floor or ceiling effects, and to obtain distributions that are comparable across studies or even species (Uher, 2008a,b).

Cross-situational consistency

Since individual-specific patterns of behavior are inherently stable over some time, they are often assumed also to be consistent across situations (Hartshorne & May, 1928; Mischel, 1968). For example, individuals high scoring on 'aggressiveness' in one situation are expected also to be high scoring on 'aggressiveness' in another situation (Bell, 2007a,b; Groothuis & Carere, 2005; Sih et al., 2004a,b). However, this assumption may not be accurate because all individuals of a population may shift their behavior across situations in similar ways. For example, in salamander larvae (*Ambystoma barbouri* and *A. texanum*) exposure to predators generally decreases with increasing amounts of chemical predator cues (Sih, Kats, & Maurer, 2003). This behavioral variation reflects the species' general adaptation to specific environmental conditions. Individual-specific behavioral patterns, if they occur in a species at all, emerge as between-individual deviations around these shifts in the populations' mean scores.

Since population-level patterns may confound individual-level patterns in empirical data, studies on the cross-situational consistency of individual behavior have to apply differential methodologies. In variable-oriented analyses this means that the individuals' rank-orders on a particular behavioral variable measured in different situations, such as foraging probabilities measured in different habitats, are correlated. This is also called *carryover or spillover effects* across situations (Bell, 2007b; Sih et al., 2004b).

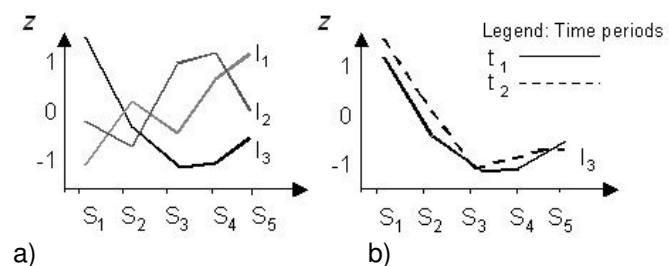
The empirical results on cross-situational consistency are varied. In greylag geese (*Anser anser*), 'aggressiveness' was highly correlated across different social situations ($r = 0.60-0.93$); individuals more 'aggressive' in one situation were also more 'aggressive' in other situations (Kralj-Fišer et al., 2007). Similarly, in salamander larvae, predator exposure measured in absence and presence of predator cues was correlated moderately to highly ($r = 0.43-0.80$; Sih et al., 2003).

Empirical findings in other species draw a more complex picture, however. In nonhuman primates, such as rhesus monkeys (Stevenson-Hinde, Stillwell-Barnes, & Zunz, 1980) or great apes (Uher, 2011; Uher et al., 2008), variable-oriented correlations across situations were only low to moderate. This means that, although the empirical estimates were test-retest reliable within each situation, the rank-orders of the same individuals on the same dimension varied considerably across different situations (Figure 5a). These findings parallel well-established findings in humans, in which variable-oriented cross-situational consistency rarely exceeds $r = .30$ (Borkenau et al., 2006; Fleeson, 2004; Hartshorne & May, 1928; Kenrick & Funder, 1988; Mischel, 1968, 1977; Mischel & Peake, 1982).

At first sight, these findings seem to challenge the idea of individual behavioral phenotypes fundamentally because individual behavior appears to be highly situation-specific rather than individual-specific. However, low to moderate cross-situational consistency may result from more fine-grained differences among individuals that are studied with individual-oriented analyses.

Figure 5
Cross-situational consistency and individual situation-behavior profiles

a) Individuals (I) can differ in the probabilities with which they show a particular behavior (M_1) in various situations (S). Then variable-oriented cross-situational consistency is only low to moderate. b) This phenomenon is



analyzed from individual-oriented viewpoints and illustrated in individual situation-behavior profiles that plot the individuals' standardized probabilities (z) in a particular behavior (M_1) against the environmental situations (S) in which they are measured. Temporal stability between time periods (t_1 and t_2) in distinct individual situation-behavior profiles provides evidence that they are individual-specific and reflect fine-grained variations among individual behavioral phenotypes. Individuals can be compared to one another in their profile shapes and mean profile levels.

Individual situation-behavior profiles

Individuals of some species, such as primates, often show the *same* behavior to different probabilities in *different* situations, and they differ from one another therein. This phenomenon is analyzed empirically in *individual situation-behavior profiles* that plot the individual's behavioral probabilities against the different situations in which they are measured (Figure 5a). This is also called 'personality signatures', 'if...then...profiles', or 'individual x situation profiles' (Borkenau, Riemann, Spinath, & Angleitner, 2006; Mischel & Shoda, 1995; Mischel et al., 2002).

Individual-specific patterns of cross-situational (in-)consistency must be differentiated from average population-level shifts across situations (Sih et al., 2004b). Each individual's situation-behavior profile must therefore be based on differential scores that are standardized within each situation across all individuals (see above; Figure 5a). Standardization enables direct comparisons of individual scores across situations independent of the population's general shifts in mean scores. For example, while all individuals may generally respond more strongly to predators than to strange objects, some individuals may respond less strongly than others to predators, or more strongly than others to strange objects (Mischel et al., 2007; Uher, 2011).

The great ape study mentioned above (Uher et al., 2008, 2011) obtained, amongst others, 'aggressiveness' estimates in four different laboratory situations. For each individual, two standardized situation-'aggressiveness' profiles were obtained independently in two nonoverlapping periods. The average test-retest profile correlation of $r = 0.77$ contrasted with the cross-situational correlation of 'aggressiveness' on the sample level of $r = 0.25$. However, if individuals show distinct situation-behavior profiles, necessarily, cross-situational consistency cannot be high; otherwise such individual variation could not occur. This shows

again that individual-oriented approaches can unravel fine-grained variations of individual behavioral phenotypes that variable-oriented approaches fail to identify (Bergman & Trost, 2006).

Analogous to individual response profiles, individual situation-behavior profiles can be empirically compared in their profile shapes and in their mean profile levels (Figure 5b). The latter are derived by averaging the individuals' standardized scores across situations. The composite estimates of those individuals that are more 'aggressive' than others in several situations will be higher than composite estimates of individuals that are more 'aggressive' than others in only one situation. However, this ignores individual particularities in profile shape (Mischel & Shoda, 1995; Mischel et al., 2002; Uher, 2011) that may reveal important information about transactional patterns in individual development (Magnusson, 1988).

Individual-situation transactions

Individual situation-behavior profiles are interpreted as reflecting relatively stable interactional and transactional patterns between environmental features and individual behavioral phenotypes over time (Magnusson, 1988; Mischel, 1977). For example, aggressions are responses to stimuli indicating that aggressive behavior might be functional. Some individuals may be more sensitive to such stimuli, and may respond more quickly or more intensely with aggression than other individuals that may be less sensitive to these stimuli and that may generally show lower probabilities of aggression (Capitanio, 2004; Tett & Guterman, 2000). Furthermore, there may be individuals who tend to seek out environmental situations that, in their population, are generally associated with higher probabilities of aggression, whereas other individuals tend to avoid such situations. That is, individuals may actively choose particular environments that are suited to their individual behavioral phenotypes.

Individuals may also evoke particular responses from their environments. For example, they may differ in their probabilities to evoke aggressive responses from conspecifics. Some are frequently involved in aggressive conflicts, whereas others are not. Individuals may also actively shape their environments, and they may differ in how they do that and in what environments they create. Environmental conditions, in turn, can also affect the individuals' behavior. Reciprocal effects between transactions over time are thought to be the mechanisms behind increasing matches between certain individual behavioral phenotypes and certain environments. They may facilitate the individuals' adaptations to their specific environmental conditions. This way, individual-situation transactions may influence the individuals' courses of development and account for certain continuity in development (Bell, 2007b; Borkenau et al., 2006; Capitanio, 2004; Magnusson, 1988; van Aken, 2008).

Hierarchical taxonomic models

The rank-orders of individuals' scores are often consistent across a broad range of measurement variables, which can be statistically summarized and described by composite latent variables, such as factors (see above). Strong correlations indicate that one latent variable can be construed that accounts for much of their shared variation. If some measurement variables are highly correlated among themselves, and others are also highly correlated among themselves, yet only poorly or moderately correlated to the first group of variables, then several latent variables can be construed that summarize the shared variation of different sets of intercorrelated measurement variables.

Moderate correlations among all measurement variables indicate, however, that only one latent variable can be construed; yet it accounts for less variation in the summarized variables than latent variables summarizing strongly intercorrelated variables. Latent variables describing moderately intercorrelated variables are thus more abstract and more general, whereas latent variables summarizing strongly intercorrelated variables are more specific and cover narrow domains of behavior.

Hence, depending on the strength of the correlation among measured variables, latent variables, mostly dimensions, can be construed at different levels of abstraction. For example, in great tits, dimensions of 'explorativeness', 'startle responsiveness', and intra-specific 'aggressiveness' correlate moderately to highly, which can thus be summarized and described by one more abstract latent dimension maybe labeled 'proactivity-reactivity' (Carere, Drent, Privitera, Koolhaas, & Groothuis, 2005; Groothuis & Carere, 2005). Such broad dimensions constitute parsimonious and comprehensive statistical descriptions of between-individual variation.

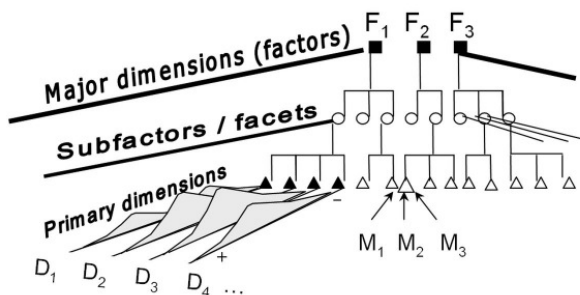
In many species, such as primates, various broad dimensions can be identified that show only low intercorrelations so that they cannot be further summarized into more abstract dimensions. These complex dimensions are relatively independent from one another and cover different sets of intercorrelating measurement variables. For example, differential research in humans yielded several broad dimensions that show only low intercorrelations, such as the so-called 'Big Five' factors (Goldberg, 1990). They describe fundamental dimensions of human individual differences on very abstract levels. Each of these factorial dimensions summarizes the shared variation of diverse inter-correlating lower-order dimensions. The factor 'extraversion' of the 'Five Factor Model', for example, is composed of the facets 'warmth', 'gregariousness', 'assertiveness', 'activity', 'excitement seeking', and 'positive emotions', each of which in turn is measured with several empirical variables of high internal consistency (see above; McCrae & Costa, 1997).

The intercorrelational structure of multiple variables can be organized in *hierarchical taxonomies*. These taxonomic models can be built by continuing the process of evaluating the correlations among the latent (mostly dimensional) variables extracted first, in order to identify higher orders of summarization. At the bottom of such hierarchies are the more specific dimensions that are subsumed within broader and more abstract dimensions, making the less complex dimensions sub-dimensions of the emergent, more complex major dimensions (Figure 6; Costa & McCrae, 1995; Eysenck, 1990). Each of the more narrow dimensions at the bottom reflects highly intercorrelated variables, whereas the broader dimensions at its top reflect broad ranges of co-varying variables that are relatively independent from one another.

Figure 6
Hierarchical taxonomic models describing complex intercorrelational structures of individual differences

Latent theoretical dimensions (D) describing relatively stable individual differences, each of which are inferred from multiple empirical measures (M), can be summarized at different levels of complexity and abstraction. Hierarchical taxonomic models can be built by studying the

intercorrelations among the dimensions (D) identified first and by identifying higher orders of summarization using statistical methods, such as factor analysis. At the bottom of such correlational hierarchies are more specific and often highly intercorrelated dimensions (D). They are statistically subsumed within broader and more abstract dimensions, making the less complex dimensions sub-dimensions of the emergent, more complex major dimensions often called factors (F). If taxonomic models with multiple factors (F) describe the relatively stable individual differences in a population the best, then their structure is multidimensional.



Such *descriptive* hierarchical taxonomies map the *between-individual variability* of individual behavioral phenotypes on the population-level. They should not be confused with *causal* hierarchies modeling hypothetical *intra-individual mechanisms*, such as interrelations among genes, neurobiological mechanisms, and behavior that may contribute to an individual's behavioral phenotype (*cf.* Réale et al., 2007, p. 3). Various methods of variable-oriented statistical analysis (such as factor analysis, see above) can be used to construe latent (often dimensional) variables at varying degrees of complexity and abstraction that

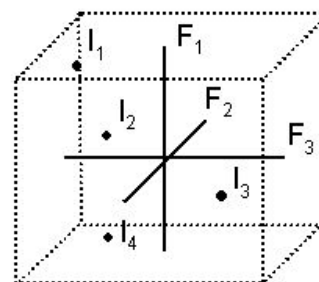
represent *different levels in the hierarchical taxonomy* (Figure 6; for empirical examples see De Raad & Barelds, 2008, p. 355; Zuckerman et al., 1991, p. 938).

Systematic methodological approaches to identify multi-dimensional structures in individual differences

Multiple, linearly relatively independent dimensions span a multi-dimensional descriptive space (Figure 7) in which the possibilities to describe complex diversity among individuals and unique individual profile configurations increase with the number of dimensions studied (Capitanio, 2004; Uher, 2008b). Since all studies ultimately rely on the selection and definition of the theoretical dimensions they investigate, it is of foremost importance to empirically establish comprehensive taxonomic models for all major study species (Uher, 2008a,b). This is a voluminous task; it may be as voluminous as mapping the entire genome sequence of an organism. The human 'Big Five' taxonomy, for example, was derived from scanning half a million entries in the lexicon for personality descriptors (Allport & Odberg, 1936). The results of this enormous effort laid the foundations for much of today's human differential and personality psychology. Once reference models are established empirically, they can put on a firm footing other studies that focus on single dimensions, or that can obtain only single measures for practical reasons. This still has to be done for most nonhuman study species.

Figure 7
Multidimensionality in the intercorrelational structure of relatively stable individual differences

Linearly relatively uncorrelated (orthogonal) major dimensions or factors (F) describe multidimensional structures of relatively stable individual differences. They span a multidimensional descriptive space in which the possibilities to describe complex diversity among individuals (I) and unique individual profile configurations increase with the number of dimensions (F) studied.



Establishing comprehensive taxonomic models requires not only suitable statistical methods (see above), but also comprehensive and systematic selections of those behaviors that shall be explored empirically for individual behavioral phenotypes. This is still not well recognized in animal research (Uher, 2008a,b). So far, animal researchers mostly restrict their empirical investigations to a few behavioral variables that are assumed to “best” capture a population's individual differences (Réale et al., 2007). However, such approaches may be incomprehensive, and may even miss important kinds of variation completely.

Basic research principles and classes of methodological approaches

Establishing comprehensive taxonomic models encounters two crucial bottlenecks. First, variable selection must represent the existing stable individual differences appropriately and comprehensively. Second, reduction to a few broad dimensions should be representative for the original pool of variables. Bias in either can cause the entire process of establishing comprehensive taxonomic models to succeed or fail (Uher, 2008a, b). Methods of reduction mostly rely on statistical analysis and are widely discussed (see above), but the explicit or implicit strategies used to gather pools of variables to be studied empirically receive surprisingly little attention. But without clear concepts of what to study, research on individual behavioral phenotypes can become directionless, and split up into different traditions each focusing on particular concepts (Bell, 2007a; Buss & Craik, 1985). To ensure that a comprehensive pool of variables is entered into the identification process, selection strategies require stringent rationales.

Various methodological approaches are used to guide variable selections in human and animal research. Based on their underlying selection strategy, they can be grouped into five major *classes of approaches* that differ in starting points and theoretical rationales. 1)

Nomination approaches rely on human observers, such as researchers, keepers, or laypeople, who nominate theoretical concepts or measurement variables that they deem important for the respective population. 2) Adaptive approaches derive theoretical concepts and measurement variables from ecological and evolutionary theories about interactions between populations and their environments to identify kinds of individual behavioral phenotypes that evolved in response to present and/or past adaptive problems. 3) Bottom-up or emic approaches start from naturally evolved systems inherent to the population, such as neurobiological, behavioral, or language systems (in humans) that are directly measurable or observable in order to derive theoretical concepts and measurement variables. 4) Top-down or etic approaches import theoretical concepts and measures from findings in other populations. 5) Eclectic approaches capitalize on empirical results and/or methods of multiple other approaches (Uher, 2008a).

Since these approaches were developed for various aims and purposes, not all of them formulate selection strategies that are suitable for the purposes of comprehensive taxonomizations. Some approaches may be prone to biasing or even ignoring important kinds of stable individual differences. This may happen, for example, because they are not salient to human observers or not exhibited by other species, or because theoretical concepts and measures imported from other species may be not applicable to the particular study species and therefore lacking ecological validity. But some approaches, in particular bottom-up/emic and adaptive approaches rely on rationales formulating selection strategies that enable systematic and comprehensive selections (for detailed discussion on the advantages and disadvantages of the different approaches see Uher, 2008a, b).

The Behavioral Repertoire x Environmental Situations Approach

The *Behavioral Repertoire x Environmental Situations approach* (Uher, 2008a,b), for example, is a systematic behavioral bottom-up/emic approach that was specifically developed to establish comprehensive taxonomic models for humans and animals. Its rationale is grounded in the notion of individual behavioral phenotypes as theoretical concepts describing variations of individual-specific patterns in the conditional probabilities to display behaviors of particular categories in environmental situations of particular categories. The approach thus explicitly considers their differential, probabilistic, and individual-specific nature. It also builds on established findings of individual-situation transactions, and the crucial distinction of individual-specific versus species-specific situation-behavior profiles discussed above.

Consequently, the approach bases its selection strategy on the behavioral repertoire and important environmental situations that behavioral scientists already identified as meaningful for a particular study population (usually a species). In a broad-based review of the pertinent literature, all major behavioral categories describing its known behavioral repertoire are compiled and plotted systematically against all categories of environmental situations in which they are reported to be typically displayed. These behavior-situation units are used to derive theoretical constructs describing *potential* kinds of variation in individual behavioral phenotypes. Only thereafter are these theoretical construct each translated into multiple empirical measurement variables and suitable experimental or observational situations for empirical investigations.

Subsequently, the obtained data are analyzed for individual-specific patterns as described above. These empirical analyses are essential because the theoretically generated constructs may not reflect actual domains of high between-individual variability in the considered population. If individual differences or temporal stability therein cannot be shown empirically, then the particular theoretical concept has to be discarded. Final analyses concern the intercorrelational structure of the empirically secured between-individual variations, and their statistical reduction to a few latent theoretical dimensions.

To realize such a systematic and comprehensive selection strategy, the Behavioral Repertoire x Environmental Situations Approach features two particularities. First, unlike most other approaches, it generates specific *theoretical concepts*, such as 'aggressiveness

in social situations', but not measurement variables to study these constructs empirically. Some animal researchers seem to have misunderstood this point (e.g., Carere & Maestriperi, 2008; Fox & Millam, 2010), probably due to discipline-specific meanings of the term 'trait' (see below). This theoretical step is vital because the targeted use of empirical variables to measure the generated theoretical concepts helps to keep their number manageable for comprehensive empirical studies.

Second, the approach generates specific theoretical concepts *from behavioral and situational categories of known meaning and function* instead of studying behavior from scratch. That is, the approach capitalizes on the existing expert knowledge that the behavioral sciences have gained on the behavior of the *average* individual of the study population. It therefore considers behavioral categories of various type and function that may also overlap in parts, such as behaviors related to the social and ecological system that can also be classified within categories of locomotion. This is fundamentally different from ethogrammes that are confined to homogeneous and disjunctive categories, and that are therefore insufficient as starting points for attempts to taxonomize individual behavioral phenotypes comprehensively. For discussions of potentials, limitations, and empirical examples see Open peer commentaries (2008); Uher (2008a,b, 2011); Uher et al., (2008), Uher and Asendorpf (2008).

Types of individuals

Comprehensive taxonomic models, which are established with variable-oriented analyses, are also a prerequisite for the holistic study of single individuals, which relies on individual-oriented analyses. When multi-dimensional models can be construed, an individual can have different scores on each major dimension (Figure 7); for example, it can score high on one dimension, but moderately or low on other dimensions. This individual configuration is illustrated as a profile across dimensions (see above, Figure 2).

The individual-oriented perspective can also be applied to analyses of multiple individuals such that individuals sharing a similar profile can be grouped statistically into *configurational types* using Q-factor analysis (see above). Such groups represent distinct and discontinuous categories of prototypes. Empirically, the individual profiles cluster around those prototypical profiles that are best thought of as approximations of the underlying continuous reality (Stern, 1911). For example, a type repeatedly found in humans (labeled 'resilient') refers to prototypical individuals that are described as being more 'emotionally stable', more 'extraverted', more 'agreeable', and much more 'conscientious' than the average individual, whereas another type (labeled 'overcontrolled') refers to individuals that are much less 'emotionally stable', particularly more 'shy' and 'inhibited', and slightly less 'agreeable' and 'conscientious' than the average (Asendorpf, Borkenau, Ostendorf, & van Aken, 2001).

These discrete types based on *multi-dimensional* classifications are different from *behavioral types* that refer to extreme scores on *one-dimensional* concepts, such as more 'aggressive' versus less 'aggressive' types (Sih et al., 2004b). The latter is a special case of the former (Figure 2). Sometimes, behavioral types are described as *multi-variate* types, such that 'fast explorers' in great tits are, as compared to 'slow explorers', also more 'allo-aggressive', more 'neo-philic', and more likely to form routines (Carere et al., 2005; Groothuis & Carere, 2005). This is because the individuals' rank-orders in these behaviors correlate on the population-level, which can be described by one latent dimension often labeled 'proactivity-reactivity'.

It would be more precise to label these types as 'reactive' versus 'proactive' in order to clarify that 'explorativeness', 'aggressiveness', 'neophilia', and 'speed of routine-formation' generally correlate in this species. This avoids misinterpreting the so-called 'fast-explorers' to be discrete configurational types based on multi-dimensional taxonomic models. The latter would be the case if there were another type of individuals that is describable as 'fast explorer', but that is low in 'aggressiveness', intermediate in 'neo-philic', and perhaps more likely to form routines than the average individual. This is just a hypothetical example to

illustrate that labeling a type for a sub-dimension rather than for the overall summarizing concept may mislead us to infer that multi-*variate* descriptions also indicate underlying multi-*dimensional* concepts.

Since behavioral probability scores are often continuously distributed, types on one-dimensional concepts can be statistically identified only by arbitrarily demarking the distribution. For example, the outer quartiles of the distribution can be defined as representing types that are distinct from each other. But means or modal scores that animal researchers often use as cut-off scores fail to distinguish reliably among individuals in the middle of normally distributed dimensions, which can result in misclassifications. Defining types by such split-half methods is also inconsistent with theoretical considerations of generalist-specialist models, which argue that extreme groups acting as specialists are more stable than those in the middle acting more like generalists and being more flexible in their behavior than the specialists (Wilson, 1994).

Longitudinal stability

Regardless of their inherent intermediate stability, individual behavioral phenotypes can also be studied for their developmental plasticity. To investigate gradual change and long-term development, much longer time periods have to be studied than those needed to establish reliable measurements at any one time point. That is, cross-sectional designs (in the wider sense given that measurements always have to be based on repeated occasions) identify and classify individual behavioral phenotypes, and longitudinal designs reveal patterns of their ontogenetic differential development. For example, great tits were studied from juvenile phases to adulthood spanning 2 to 3 year intervals (Carere et al., 2005). Humans were studied from childhood to adulthood spanning ages 3 to 21 (Caspi, 2000); so far the longest studied time spans cover up to 45 to 50 years (Conley, 1984, 1985; Haan, Millsap, & Hartka, 1986; Soldz & Vaillant, 1999). Given the species differences in lifetimes, stabilities from 6 through 20 years of age in rhesus macaques (Suomi et al., 1996) cover similar extended periods of life.

Differential stability and stability across age groups

Two kinds of longitudinal stability are distinguished. *Differential stability* relates to stability in the individuals' probability scores relative to their particular reference population that can be defined by sex or age group. For example, individuals scoring high on a variable compared to other individuals of their age group may also be scoring high relative to their age group in later periods of life, though they may not be high relative to those who are younger or older. *Stability across age groups* signifies stability in the mean levels of the considered population on that behavioral variable over time. Some dimensions may show shifts in the mean level of individual scores across age groups. For example, in dumpling squids, average 'boldness' scores increase significantly through adulthood (Sinn et al., 2008); whereas average 'extraversion' scores in gorillas decline with age (see Figure 1 in Kuhar, Stoinski, Lukas, & Maple, 2006); and average 'agreeableness' scores in humans constantly increase with age (Roberts, Walton, & Viechtbauer, 2006).

Variable-oriented and individual-oriented concepts of longitudinal stability

Differential stability and stability across age groups are variable-oriented concepts that can be addressed by the individuals' rank-orders and the mean scores of all individuals within age groups over time respectively (Figure 3). Such population-level analyses may, however, mask changes that occur on the individual level (see above). For example, individual-oriented analyses in humans and some nonhuman species showed that the profiles of some individuals are much more stable than those of others. This suggests that some individuals are influenced more by environmental circumstances than others.

Consequently, behavioral stability itself may be a phenomenon that characterizes individuals to varying degrees (Carere et al., 2005; Caspi & Roberts, 1999; Uher 2011).

Many different methodological and statistical approaches are available for analyses of the changeability of individual behavioral phenotypes. Variable- and individual-oriented analyses allow for systematic study, for example, of rank-order stability, within-individual and population mean level stability, individual profile stability, and stability of profile type membership (Akse, Hale III, Engels, Raaijmakers, & Meeus, 2007; Conley, 1985; Biesanz, West, & Kwok, 2003). Such analyses will shed more light on fundamental issues in individual development across the life span in different species (Block, 1971; Magnusson, 1988; Mervielde & Asendorpf, 2000; van Aken, 2008).

Discussion

Above I elaborated meta-theoretical concepts underlying the idea of meaningful individual differences in behavior that I conceptualized here as individual behavioral phenotypes, and introduced suitable analytical methodologies for their empirical investigation. This now allows me to critically review and integrate terms and concepts that were launched comparatively recently in animal research in order to unravel commonalities and differences among them, to scrutinize frequent assumptions of analogy to concepts of human 'personality', and to highlight various cross-disciplinary misunderstandings.

Terminological misunderstandings about 'trait', 'character', and 'personality structure'

Animal researchers often misunderstand the psychological term '*trait*' because it has various meanings that are not obvious. First, 'trait' can signify the variable-oriented theoretical concept of a '*trait dimension*', thus describing population-level phenomena of relatively stable individual differences. 'Trait dimensions' can thereby be of different degrees of abstraction; all levels in taxonomic models can be labeled 'trait dimensions' (although some authors differentiate them into 'habitual acts', 'facets', 'sub-factors', 'factors', or 'super-factors' (Costa & McCrae, 1995; Eysenck, 1990). Accordingly, the entire empirical taxonomy is referred to as '*trait taxonomy*' or '*trait hierarchy*' (De Raad & Barelds, 2008; Matthews, Deary, & Whiteman, 2003). Second, 'trait' generally signifies the '*trait score*' of an individual on a given 'trait dimension', thus referring to individual-level phenomena. Individual configurations of 'trait' scores on multiple 'trait dimensions' are illustrated as '*trait profiles*', which are individual-oriented concepts.

In all these cases, however, the psychological term 'trait' refers to abstract theoretical constructs conceived to describe individual-specific patterns in measurement variables (Uher, 2008b). In biology and many other disciplines, in contrast, the term 'trait' is used to refer to any *empirical measure* obtained from an individual (Réale et al., 2007, p. 3; Sih et al., 2004a, b; van Oers, 2008), but not to the theoretical concepts that are inferred from such measures whether variable-oriented or individual-oriented. (Confusingly, this notion of 'trait' is sometimes also used in psychology, but rarely). The theoretical dimensions that psychologists label 'personality dimensions' or 'trait dimensions' are often called 'axes' or '*continua*' in biology.

The term '*character*' has also different meanings. In biology, it is often used synonymously with 'trait' and denotes an empirical measure. In psychology, in contrast, it refers to the theoretical concept of an individual's integrated set of behavioral tendencies that are inferred from multiple empirical measures. It is thus equivalent to the term 'personality'. In its narrower sense, 'character' denotes only those parts of an individual's 'personality' that concern morality and virtue. In any case, the psychological term 'character' refers to individual-oriented concepts, but not to variable-oriented concepts. This is also reflected by its meaning as the '*personality structure of an individual*'.

Confusingly, psychologists use the term 'personality structure' also to denote the variable-oriented concept of the (factor) structure of individual differences in a population. To avoid misunderstandings, this '*personality structure of a population*' may be labeled more precisely the 'structure of personality *differences* in a population'. These jingle fallacies are

major obstacles for cross-disciplinary understanding because they are associated with conceptual misunderstandings.

Conceptual misunderstandings about 'correlated traits' and 'independent traits'

For biologists, the central phenomena behind individual behavioral phenotypes are correlated behavioral variables that were previously not expected to be correlated and therefore studied independently. Now findings of so-called '*correlated traits*' or 'correlated characters' raise many questions from behavior ecological and evolutionary viewpoints (Bell, 2007b; Groothuis & Carere, 2005; Sih et al., 2004a,b). Given this, the psychologists' interest in statistically '*independent traits*' puzzles biologists (Dingemanse, 2007), whereas the biologists' focus on 'correlated traits' puzzles psychologists (Penke, Denissen, & Miller, 2007).

The point of misunderstanding that arises from the jingle-fallacies afflicted with the term 'trait' can be clarified by the concept of hierarchical taxonomies (see above, Figure 6). Biologists primarily focus on the correlated variables located at the bottom of the hierarchy; and thus study mostly one-dimensional concepts. Many psychologists also study homogeneous patterns of behavior, such as those of infantile 'aggressiveness' syndromes described below. But some other psychologists, so-called taxonomists, are concerned with the big picture of the intercorrelational structure of human individual differences. They focus on broad and abstract dimensions located at the top of hierarchical models that each summarize different sets of intercorrelated variables in parsimonious and comprehensive ways. These psychologists are concerned with multi-dimensional concepts.

Conceptual misunderstandings about 'personality' and 'animal personalities'

The conceptual focus on variable-oriented and mostly one-dimensional concepts in biology may have led to further terminological and conceptual misunderstandings among disciplines. As explained above, the psychological term 'personality' denotes the person's uniqueness within the diversity of individuals. This individual-oriented concept is based on multi-dimensional, variable-oriented concepts.

In contrast, the term '*animal personalities*' that was coined only relatively recently in biology denotes almost exclusively population-level phenomena of individual differences (Groothuis & Carere, 2005), but not individual-level phenomena. Its plural form and the typically variable-oriented data analyses also indicate this. In fact, individual-oriented analyses needed to study an individual's 'personality' are still rarely applied in animal research (Uher, 2011). Hence, '*animal personalities*' are equivalent to psychological concepts of 'personality dimensions' or 'trait dimensions', but not to an individual's 'personality'.

Moreover, research on '*animal personalities*' is currently focused on one-dimensional concepts, not on multi-dimensional concepts, which underlie the psychological term 'personality'. It is up to future research to negotiate whether the term 'personality' can also be applied to individuals of populations in which individual differences are just one-dimensional, and in which the diversity among individuals and the possibilities for individual uniqueness are fairly limited. Alternatively, the term 'personality' could be restricted to its meaning as the potentially unique individual configuration of behavioral tendencies along diverse dimensions as established for humans more than 100 years ago. In view of the considerable cross-disciplinary misunderstandings that already exist in the field, the latter appears to be preferable.

Which species do actually exhibit multi-dimensional structures of individual differences?

Whether multi-dimensional models are useful to describe individual differences in a species depends on the complexity of individual variation it exhibits. In some species, it may be best described with one-dimensional concepts, whereas other species may exhibit highly

complex structures of individual behavioral differences that are more accurately described with multi-dimensional concepts.

That some disciplines study one-dimensional concepts, whereas others study also multi-dimensional concepts, may derive from characteristics of their study species. In fact, in research on individual differences, the disciplines seem to have divided the study species up amongst themselves. Biologists focus on birds, fish, amphibians, insects, and spiders (Sih et al., 2004b), whereas psychologists focus on humans and nonhuman primates (Uher, 2008b). Domesticated animals, mice, and rats—i.e. mostly mammalian species—are among the few species studied by both disciplines.

The characteristics of the study species may have influenced not only the disciplines' theoretical concepts, but also their analytical methodologies. The psychological interests in multi-dimensional models led to the standard application of multivariate statistical methods enabling multi-dimensional analyses (see above). The biological focus on correlated behavioral variables and syndromal concepts, in contrast, may have led to the relative predominance of bivariate correlations and paired *t*-tests in pertinent empirical studies. With increasing complexity and numbers of variables to be analyzed, bivariate analyses meet their limits. Descriptive summaries and comparisons of results from series of bivariate analyses (e.g. in D'Eath & Burn, 2002) hinder the identification of empirical structures underlying the obtained data. To describe complex individual differences, suitable statistical methods such as factor analytic techniques are indispensable and should be used more frequently in differential research in biology (Uher, 2008b).

Meta-theoretical and methodological concepts underlying 'behavioral syndromes'

A further, relatively new concept is that of '*behavioral syndromes*' conceptualized as sets (suites) of correlated behaviors, and statistically defined by correlations significantly different from zero and stable over time (Bell & Stamps, 2004; Sih & Bell, 2008; Sih et al., 2004a, b). It is thus a variable-oriented concept describing population-level phenomena of individual differences, yet it is no individual-oriented concept. Accordingly, 'behavioral syndromes' could be analogous to 'personality differences', yet not to an individual's 'personality'.

However, correlations among behaviors on the population-level can be stable over time without necessarily being based on individual-level consistency as concepts of 'personality differences' are. Individual-specific patterns presuppose individual differences in the population; but the reverse need not be true. Not just any individual differences in behavioral data can be considered ultimate indicators of individual behavioral phenotypes (see above). For example, a study in brown rock fish (*Sebastes auriculatus*) found significantly positive correlations among time spent near predators, and feeding in absence and presence of predators. When measured 10 days later, the correlations were similar. These temporally stable correlations among behavior measures were therefore construed as 'behavioral syndromes' (Lee & Berejikian, 2008). But the crucial point is that the individuals' behavior was not consistent. Instead, the individuals' rank-order along the different behavioral variables changed considerably. No individual fish generally tended to feed and risk more than others. Rather, all individuals fed and took risks either at high rates or low rates, and alternated their behaviors flexibly from day to day. There was thus no indication of individual behavioral phenotypes.

Such findings may reflect the behaviors' functional similarity for ephemeral conditions like hunger. All individuals of a species, when they are hungry, may show high probabilities of feeding in both presence and absence of predators, and when they are not hungry, all individuals may show low feeding activities in both conditions. If, at the same points of measurement, the studied individuals differ in degree to which they are in these ephemeral conditions, such as in different degrees of hunger, then these behavioral variables are necessarily correlated, and these correlations are stable over time. But clearly, this has nothing to do with the phenomena observable at least in particular species that some individuals generally show particular behaviors with higher probabilities than others. Such

stable individual differences are correlated on the population-level for an entirely different reason than the stable correlations described by the concept of 'behavioral syndromes'. If 'behavioral syndromes' are, in fact, conceptualized as not being based on individual-specificity on both the meta-theoretical and methodological levels, then they are in no way equivalent to individual behavioral phenotypes, and thus also not to 'personality dimensions'.

Confusingly, however, 'behavioral syndromes' are regularly interpreted as reflecting individual behavioral phenotypes because the behavioral patterns they describe are often hypothesized to be adaptive, heritable, and acted on by natural selection (Bell, 2007b; Sih & Bell, 2008; Sih et al., 2004a, b). Yet this is only possible if they reflect properties of the individuals, not if all individuals show the same degrees of behavioral flexibility. Furthermore, according to classical theory of measurement (see above), not just *any* significant correlation of highly dynamic and fluctuating phenomena like behaviors may be meaningful as assumed by Sih and Bell (2008). Many studies construe 'behavioral syndromes' from relatively weak correlations ($r \sim 0.2-0.3$; Sih & Bell, 2008). Such syndromal measures largely reflect random variation, while the proportion of true score variation is comparably small.

Considering these methodological artifacts, researchers may currently overestimate the number of species that actually exhibit individual behavioral phenotypes. This may have misguided some theoretical developments on possible explanations of the proximate mechanisms, ontogeny, adaptivity, and phylogeny of meaningful individual differences. Future research should clarify the meta-theoretical and methodological conceptualization of the 'behavioral syndromes' concept, and critically review the theoretical developments that are based on its interpretation as reflecting individual-specific patterns of behavior.

General limitations of syndromal concepts

Syndromal concepts, including those that are actually based on individual-specific patterns, may be generally insufficient to unravel important kinds of meaningful individual differences. For example, a study in human children investigated whether 'aggressiveness' syndromes can be distinguished that are functionally different, and therefore require different educational treatment (Wright & Zakriski, 2003). Overall 'aggressiveness' scores as obtained through variable-oriented analyses failed to discriminate among children with similar aggression rates, but different responses to particular environmental situations. Some individuals responded 'aggressively' to threats, whereas others responded 'aggressively' to friendly overtures. This suggests that the proximate psychological and physiological mechanisms contributing to these infantile 'aggressiveness' syndromes may differ substantially.

Hence, syndromal measures studying variable-oriented co-occurrences of behaviors on the population-level can be insensitive to important functional differences among individuals who display similar overall probabilities of behavior. In other words, "syndromal similarity obscures functional dissimilarity" (Wright & Zakriski, 2003). This example also highlights that matching the method of analysis to the aim of scientific investigations is of paramount importance (Bergman & Trost, 2006). For conclusive interpretations of individual-level phenomena, such as individual plasticity or proximate mechanisms operating within the individual, variable-oriented analyses have to be complemented by individual-oriented analyses.

For example, ecological and evolutionary biologists currently interpret variable-oriented correlations across situations as reflecting constraints in the individuals' dynamic plasticity to discriminate and adapt to specific environmental conditions. These phenomena are therefore understood as reflecting individual trade-offs with different costs and benefits (Bell, 2007b; Sih et al., 2004b). However, if relatively stable individual situation-behavior profiles can be identified as the major source of moderate syndromal correlations, this would suggest much higher degrees of individual adaptive plasticity than hypothesized based on mere variable-oriented findings, while still conforming to the idea that infinite individual plasticity may be too costly (DeWitt, Sih, & Wilson, 1998; McDonald, Keeler, & McFarlane, 2007).

'Reaction norms' versus individual situation-behavior profiles

'Reaction norms' are individual-oriented concepts to some degree. They are empirical functions illustrated in graphs that plot the phenotypic response of particular genotypes against a continuous environmental parameter (van Oers, de Jong, Noordwijk, Kempenaers, & Drent, 2005b). The considered genotypes mostly refer to particular groups of individuals; yet the concept is also applicable to individual genotypes. In this regard, 'reaction norms' are similar to situation-behavior profiles; but the concepts also differ in essential ways.

First, graphs of behavioral 'reaction norms' typically rely on absolute behavior scores that do not readily inform about deviations from other individuals. To incorporate a differential perspective and to control for the effects that environmental conditions generally have on all individuals, 'reaction norm' graphs should be documented with standardized phenotypical scores. A second difference lies in the complexity of the considered environmental situations. Graphs of 'reaction norms' typically consider a *continuous* environmental parameter. But most situations are *complex and multivariate constellations* of environmental conditions. Situation-behavior profiles reflect this complexity by considering *nominal* situations that can be described in their full-blown complexity. If groups of individuals can be identified that behave similarly in the same group of situations, these situations can be analyzed for the particular environmental conditions they share. This information can be used to classify the situations according to the specific properties they represent for particular individuals or for particular populations (see above).

The complexity of environmental situations can also be documented in 'reaction norms' graphs by plotting phenotypic scores against *multiple* environmental variables. For example, if "feeding activity in the presence of predators" (Sih et al., 2004b, p. 244) is studied, and the 'reaction norms' are plotted only against variations of predator cues, important information about the individuals may be overlooked. For example, individuals may only show feeding activity when cues of food are present. Unless this is considered in the analysis, it cannot be detected if individuals perhaps differ in their sensitivity to perceive food cues or in their responses toward them. Food can vary in quality and quantity; all individuals may show higher feeding activity when high-quality food is available. Feeding activity may also generally increase with increasing population density or food scarcity respectively. Unless this is controlled in the experimental setup, empirical findings may be confounded with these average shifts in mean scores that are uninformative about the single individual. They may also be confounded with individual differences around these shifts that are related to intra-specific competition, but not necessarily to the presence or absence of predators.

Consequently, to identify in which ways environmental conditions actually affect individual behavior, all kinds of environmental variables represented by the studied settings should be considered. Only then can they be analyzed for the main, additive, or even interactive effects they may have on the studied behaviors. This is particularly important for explanatory approaches that seek to unravel functional differences underlying individual behavioral phenotypes.

'Ambiguities in concepts of 'consistency across contexts and situations'

Many animal researchers refer their concepts to behavioral 'consistency across contexts', for example, across feeding, anti-predator, mating, aggressive, and dispersal behaviors (Sih et al., 2004b, p. 242). In this connotation, 'context' is used to refer to *behaviors of different quality and function* that are displayed in the same or, necessarily, in different environmental situations. Hence, 'situation' and 'context' are not always used as synonyms for environmental variables. This is also indicated by concepts referring to 'consistency across contexts *and* situations' (Bell, 2007a; Réale et al., 2007; Sih et al., 2004a, b). The occurrence of particular behaviors may sometimes be bound to particular environmental situations, but many behaviors, such as aggression, can occur in environmental situations of very different kind.

Since these jingle-jangle fallacies may cause conceptual misunderstandings, researchers should differentiate clearly the behaviors they are studying from the environmental conditions in which they are studied, and specify both precisely. This also has implications for the meaning of the concept of 'context-specificity versus context-generality'. If 'context' is used synonymously with environmental situation, this concept would be analogous to the cross-situational consistency of a particular kind of behavioral variation (Figure 5a, b). But if 'context' refers to behavioral categories of different quality and function, then it would be analogous to a broad factorial dimension (Figure 6).

Conclusions

This article elaborated an overarching framework of investigative principles, meta-theoretical concepts, analytical and methodological approaches needed to disentangle meaningful individual differences in behavior from those based on behavioral flexibility, functional similarities among behaviors, and measurement error. First, I clarified that individual behavioral phenotypes are between-individual variations of individual-specific patterns of behavior that are inherently probabilistic, differential, and relatively stable. These characteristics determine the analytic methodologies required for their empirical exploration on both the population-level and the individual-level for which variable-oriented and individual-oriented approaches should be used complementarily.

Systematic comparisons and differentiations of the meta-theoretical ideas underlying frequently used terms and concepts revealed that some theoretical concepts denote equivalent phenomena in different species, whereas others do not. I hypothesized that conceptual differences among disciplines may also derive from characteristics of their typical study species, which may have led biologists to focus on one-dimensional concepts, and psychologists to focus on multi-dimensional concepts. I furthermore elaborated that unwarranted assumptions of conceptual analogy reflect fundamental misunderstandings originating from discipline-specific usages of terminology. This hinders cross-disciplinary exchange and can also lead theoretical interpretations and developments astray. To reduce these barriers, it is important that all researchers are more explicit in their terminological and conceptual definitions.

These new insights, uncovered from a meta-theoretical perspective, may dampen the enthusiasm of many researchers that phenomena analogous to human 'personality' can also be found in a broad range of species. It may well be possible that many species do not exhibit individual behavioral phenotypes at all. Perhaps, the phenomena currently labeled as 'animal personalities' or 'behavioral syndromes' reflect different phenomena than those studied in human and nonhuman primates as 'personality' and 'personality differences'.

Answering these questions requires systematic methodological efforts that involve suitable statistical techniques and broad methodological approaches, such as the Behavioral Repertoire x Environmental Situations Approach, in order to develop comprehensive taxonomic models of the structure of meaningful individual differences in the most important study species.

Ultimately, the meta-theoretical and methodological foundations elaborated here will also trigger new theoretical developments. For example, while evolutionary biologists currently focus on explaining the mechanisms and processes that govern the evolution of one-dimensional individual differences (e.g. Wolf et al., 2007; Wolf, van Doorn, & Weissing, 2008); future research has to develop models that explain how natural selection can act on populations in which individuals differ from one another along multiple, relatively independent dimensions.

Acknowledgements

I am very grateful to Wendy Johnson, John Capitanio, Hans Westmeyer, Hannelore Weber, Jochen Fahrenberg, and two anonymous reviewers for their valuable comments on previous drafts. I also thank Kees van Oers, Dik Heg, Michael Taborsky, Barbara Taborsky, and Claudio Carere for discussions on biological terminology. This research was partially supported by a research grant of Freie Universität Berlin, Germany, and a research grant of the Deutsche Forschungsgemeinschaft DFG (German Science Foundation, grant number UH249/1-1).

References

- Akse, J., Hale III, W. W., Engels, R. C. M. E., Raaijmakers, Q. A. W., & Meeus, W. H. J. (2007). Stability and change in personality type membership and anxiety in adolescence. *Journal of Adolescence*, 30, 813-834.
- Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster Analysis*. Beverly Hills, CA: Sage Publications.
- Allport, G. W., & Odbert, H. G. (1936). Trait names: A psycholexial study. *Psychological Monographs*, 47, 1.
- Asendorpf, J. B. (1988). Individual response profiles in the behavioral assessment of personality. *European Journal of Personality*, 2, 155-167.
- Asendorpf, J. B., Borkenau, P., Ostendorf, F., & van Aken, M. A. G. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality*, 15, 169-198.
- Basilevsky, A. (1994). *Statistical factor analysis and related methods: Theory and applications*. New York, NY: John Wiley and Sons.
- Bell, A. M. (2007a). Future directions in behavioral syndromes research. *Proceedings of the Royal Society Biological Sciences Series B*, 274, 755-761.
- Bell, A. M. (2007b). Animal personalities. *Nature*, 447, 539-640.
- Bell, A. M., & Sih, A. (2007). Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecology Letters*, 10, 828–834.
- Bell, A. M., & Stamps, J. A. (2004). Development of behavioural differences between individuals and populations of sticklebacks, *Gasterosteus aculeatus*. *Animal Behaviour*, 68, 1339-1348.
- Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, 52, 377–389.
- Biesanz, J. C., West, S. G., & Kwok, O. M. (2003). Personality over time: Methodological approaches to the study of short-term and long-term development and change. *Journal of Personality*, 71, 905-941.
- Block, J. (1971). *Lives through time*. Berkeley, CA: Bancroft.
- Borg, I. & Groenen, P. (2005). *Modern Multidimensional Scaling: theory and applications* (2nd ed.). New York, NY: Springer-Verlag.
- Borkenau, P., Riemann, R., Spinath, F. M., & Angleitner, A. (2006). Genetic and environmental influences on person x situation profiles. *Journal of Personality*, 74, 1451–1479.
- Buss, D. M., & Craik, K. H. (1985). Why not measure that trait? Alternative criteria for identifying important dispositions. *Journal of Personality and Social Psychology*, 48, 934-946.
- Campbell, D. T., & Fiske D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 56, 81-105.
- Capitanio, J. P. (1999). Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. *American Journal of Primatology*, 47, 299-320.
- Capitanio, J. P. (2004). Personality factors between and within species. In B. Thierry, M. Singh & W. Kaumanns (Eds.) *Macaque societies* (pp. 13–33). Cambridge, UK: Cambridge University Press.

- Carere, C. & D. Maestriperi. (2008). The behavioral repertoire approach in comparative personality research: Inconsistencies between theory and practice. *European Journal of Personality*, 22, 457-459.
- Carere, C., & van Oers, K. (2004). Shy and bold great tits (*Parus major*): Body temperature and breath rate in response to handling stress. *Physiology and Behavior*, 82, 905-912.
- Carere, C., Drent, P. J., Privitera, L., Koolhaas, J. M., & Groothuis, T. G. G. (2005). Personalities in great tits (*Parus major*): Stability and consistency. *Animal Behaviour*, 70, 795-805.
- Caspi, A. (2000). The child is father of the man: Personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, 78, 158-172.
- Caspi A., & Roberts, B. W. (1999). Personality continuity and change across the life course. In L. A. Pervin, & O. P. John (Eds.) *Handbook of personality: Theory and research* (2nd Ed.) (pp. 300-326). New York, NY: Guilford Press.
- Conley, J. J. (1984). Longitudinal consistency of adult personality: Self-reported psychological characteristics across 45 years. *Journal of Personality and Social Psychology*, 47, 1325-1333.
- Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait-multimethod-multioccasion analysis. *Journal of Personality and Social Psychology*, 49, 1266-1282.
- Cook, M. & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology*, 98, 448–459.
- Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment.
- Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO Personality Inventory. *Journal of Personality Assessment*, 64, 21-50.
- Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional Scaling*. London, UK: Chapman and Hall.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- D'Eath, R. B., & Burn, C. C. (2002). Individual differences in behaviour: A test of 'coping style' does not predict resident-intruder aggressiveness in pigs. *Behaviour*, 139, 1175–1194.
- Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology letters*, 7, 734-739.
- Dammhahn, M. (2009). What Mediates Personality Variation in Grey Mouse Lemurs (*Microcebus murinus*)? *Folia Primatologia*, 80, 115.
- De Raad, B., & Barelds, D. P. H. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology*, 94, 347-364.
- DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology and Evolution*, 13, 77–81.
- Dingemanse, N. J. (2007). An evolutionary ecologist's view of how to study the persistence of genetic variation in personality. *European Journal of Personality*, 21, 593-596.
- Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, 64, 929-937.
- Dobzhansky, Th. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *The American Biology Teacher*, 35, 125-129.
- Drent, P. J., van Oers, K., & van Noordwijk, A. J. (2003). Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270, 45-51.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097-1126.
- Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist*, 35, 790-806.
- Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 244-276). New York, NY: Guilford.

- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum.
- Fahrenberg, J. (1986). Psychophysiological individuality: A pattern analytic approach to personality research and psychosomatic medicine. *Advances in Behaviour Research and Therapy*, 8, 43–100.
- Fairbanks, L. A., & McGuire, M. T., (1993). Maternal protectiveness and response to the unfamiliar in vervet monkeys. *American Journal of Primatology*, 30, 119-129.
- Fidler, A. E., van Oers, K., Drent, P. J., Kuhn, S., Mueller, J. C., & Kempenaers, B. 2007. Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings of the Royal Society of London B*, 274, 1685-1691.
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13, 83-87.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299.
- Foerster, F., Schneider, H. J., & Walschburger, P. (1983). The differentiation of individual-specific, stimulus-specific, and motivation-specific response patterns in activation processes: An inquiry investigating their stability and possible importance in psychophysiology. *Biological Psychology*, 17, 1-26.
- Fox, R. A., & Millam, J. R. (2010). The use of ratings and direct behavioral observation to measure temperament traits in cockatiels (*Nymphicus hollandicus*). *Ethology*, 116, 59-75.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652-670.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Galton, F. (1869). *Hereditary Genius: An Inquiry into its Laws and Consequences*. London, UK: Macmillan.
- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality & Social Psychology*, 59, 1216-1229.
- Goldberg, L. R., & Digman, J. M. (1994). Revealing structure in the data: principles of exploratory factor analysis. In S. Strack & M. Lorr (Eds.). *Differentiating normal and abnormal personality*. (pp. 216-242). New York, NY: Springer.
- Gorlyn, M., Keilp, J. G., Tryon, W. W., & Mann, J. J. (2005). Performance test correlates of component factors of impulsiveness. *Personality and Individual Differences*, 38, 1549-1559.
- Gorsuch, R. L. (1983) *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Gosling, S. D. (2008). Personality in nonhuman animals. *Social and Personality Psychology Compass*, 2, 985-1001.
- Groothuis, T. G. G., & Carere, C. (2005). Avian personalities: Characterization and epigenesis. *Neuroscience and Biobehavioral Reviews*, 29, 137-150.
- Haan, N., Millsap, R., & Hartka, E. (1986). As times goes by: Change and stability in personality over fifty years. *Psychology and Aging*, 1, 220-232.
- Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character*. Vol 1. *Studies in deceit*. New York, NY: Macmillan.
- Hox, J. (2002). *Multilevel Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications.
- Hoyle, R. H. (2007). Structural equation modeling in personality research. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 444-460). New York, NY: Guilford.
- Janczak, A. M., Andersen, I. L., Færevik, G., Bøe, K. E., & Bakken, M. (2002). Factor analysis of behaviour in the porcine and murine elevated plus-maze models of anxiety. *Applied Animal Behavior Science*, 77, 155-165.
- John, O. P., & Soto, C. J. (2007). The importance on being valid. Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.) *Handbook of*

- research methods in personality psychology. (pp. 190-206). New York, NY: Guilford Press.
- Kelley, T. L. (1927). Interpretation of educational measurements. Yonkers, NY: World Book.
- Kelly, G. A. (1955). The Psychology of Personal Constructs (Vol. 1 and 2). New York, NY: Norton.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person[^]situation debate. *American Psychologist*, 43, 23-34.
- Kinnally, E. L., Lyons, L. A., Abel, K., Mendoza, S. P., Capitanio, J. P. (2008). Effects of early experience and genotype on serotonin transporter regulation in infant rhesus macaques. *Genes, Brains, and Behavior*, 7, 481-486.
- Kline, P. (2000). The handbook of psychological testing (2nd ed). London, UK: Routledge.
- Kline, R. B. (2005). Principles and practice of structural equation modeling (2nd ed). New York, NY: Guilford Press.
- Kralj-Fišer, S., Scheiber, I. B. R., Blejec, A., Moestl, E., & Kotrschal, K. (2007). Individualities in a flock of free-roaming greylag geese: Behavioral and physiological consistency over time and across situations. *Hormones and Behavior*, 51, 239–248.
- Kuhar, C. W., Stoinski, T. W., Lukas, K. E., & Maple, T. L. (2006). Gorilla behavior index revisited: Age, housing and behavior. *Applied Animal Behaviour Science*, 96, 315-326.
- Lacey, J. J. (1950). Individual differences in somatic response patterns. *Journal of Comparative and Physiological Psychology*, 43, 338-350.
- Lamiell, J. T. (2003). Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism. Thousand Oaks, CA: Sage Publications.
- Lee, K., & Ashton, M. C. (2007). Factor analysis in personality research. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 424-443). New York: Guilford.
- Lee, J. S., & Berejikian, B. A. (2008). Stability of behavioral syndromes but plasticity in individual behavior: Consequences for rockfish stock enhancement. *Environmental Biology of Fishes*, 82, 179-186.
- Magnusson, D. (1988). Individual development from an interactional perspective: A longitudinal study. Hillsdale, NJ: Erlbaum.
- Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd Ed.). Cambridge, UK: Cambridge University Press.
- McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509-516.
- McDonald, D. G., Keeler, R. A., & McFarlane, W. J. (2007). The relationships among sprint performance, voluntary swimming activity, and social dominance in juvenile Rainbow trout. *Physiological and Biochemical Zoology*, 80, 619–634.
- Mervielde, I., & Asendorpf, J. B. (2000). Variable-centred and person-centred approaches to childhood personality. In S. E. Hampson (Ed.). *Advances in personality psychology* (Vol. 1, pp. 37-76). Hove, UK: Psychology Press.
- Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.
- Mischel, W. (1977). The interaction of person and situation. In D. Magnusson, & N. S. Endler (Eds.). *Personality at the crossroads: Current issues in interactional psychology* (pp. 333-352). Hillsdale, NJ: Erlbaum.
- Mischel, W., & Peake, P. K. (1982). Beyond déjà-vu in the search for cross-situational consistency. *Psychological Review*, 89, 730-755.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102, 246-268.
- Mischel, W., Shoda, Y., & Ayduk, O. (2007). *Personality*, (8th ed). New York, NY: Wiley.
- Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science*, 11, 50-54.
- Morizot, J. M., Ainsworth, A. T., & Reise, S. P. (2007). Towards modern psychometrics: Application of item response theory models in personality research. In R.W. Robins, R. C., Fraley, & R. F. Krueger (Eds.) *Handbook of Research Methods in Personality Psychology* (pp. 407-423). New York, NY: Guilford.

- Open peer commentaries (2008). Discussion on 'Comparative Personality Research: Methodological Approaches' by Jana Uher. *European Journal of Personality*, 22, 457–474.
- Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). Evolution, genes, and inter-disciplinary personality research. *European Journal of Personality*, 21, 639-665.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Newbury Park, CA: Sage.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating temperament within ecology and evolutionary biology. *Biological Review*, 82, 1-28.
- Reise, S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment*, 84, 126-136.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies. *Psychological Bulletin*, 132, 1-25.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology, Vol 3: Social, emotional and personality development* (5th ed) (pp. 105-176). New York, NY: John Wiley & Sons.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 94, 18-38.
- Schneider, M. L., & Suomi, S. J. (1992). Laboratory assessment of temperament and postrotary nystagmus responses in rhesus monkey infants (*Macaca mulatta*). *Physical and Occupational Therapy in Pediatrics*, 12, 37-52.
- Shoda, Y. (1999). A unified framework for the study of behavioral consistency: Bridging person \times situation interaction and the consistency paradox. *European Journal of Personality*, 13, 361-87.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 2, 420-428.
- Sih, A. & Bell, A. M. (2008). Insights for Behavioral Ecology from Behavioral Syndromes. *Advances in the Study of Behavior*, 38, 227–281.
- Sih, A., Bell, A. M., & Johnson, J. C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology and Evolution*, 19, 372-378.
- Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004b). Behavioural syndromes: an integrative overview. *The Quarterly Review of Biology*, 79, 241-277.
- Sih, A., Kats, L. B., & Maurer, E. F. (2003). Behavioral correlations across situations and the evolution of antipredator behaviour in a sunfish–salamander system. *Animal Behaviour*, 65, 29–44.
- Sinn, D. L., Gosling, S. D., & Moltschaniwskyj, N. A. (2008). Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Animal Behaviour*, 75, 433-442.
- Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.
- Soldz, S., & Vaillant, G. E. (1999). The big five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality*, 33, 208-32.
- Stern, W. (1911). *Die Differentielle Psychologie in ihren methodischen Grundlagen* (2. Auflage). [Differential Psychology in its methodological foundations (2nd ed.)]. Leipzig, Germany: Barth.
- Stevenson-Hinde, J., Stillwell-Barnes, R., & Zunz, M. (1980). Individual differences in young rhesus monkeys: Consistency and change. *Primates*, 21, 498-509.
- Suomi, S. J., Novak, M. A., & Well, A. (1996). Aging in rhesus monkeys: Different windows on behavioral continuity and change. *Developmental Psychology*, 32, 1116-1128.
- Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorders. *Psychological inquiry*, 4, 122-130.
- Ten Berge, M. A., & De Raad, B. (1999). Taxonomies of situations from a trait psychological perspective. A review. *European Journal of Personality*, 13, 337-360.
- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34, 397-423.

- Thorndike, E. L. (1903). Notes on child study. (2nd Ed.). New York, NY: Macmillan.
- Tinbergen, N. (1963). On aims and methods in ethology. *Zeitschrift für Tierpsychologie*, 20, 410-433.
- Uher, J. (2011). Personality in nonhuman primates: What can we learn from human personality psychology? In A. Weiss, J. E. King, & L. Murray (Eds.). *Personality and behavioral syndromes in nonhuman primates* (pp. 41-76). New York, NY: Springer.
- Uher, J. (2008a). Comparative personality research: Methodological approaches (Target article). *European Journal of Personality*, 22, 427-455.
- Uher, J. (2008b). Three methodological core issues in comparative personality research (Author's reply). *European Journal of Personality*, 22, 475-496.
- Uher, J., & Asendorpf, J. B. (2008). Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality*, 42, 821-838.
- Uher, J., Asendorpf, J. B., & Call, J. (2008). Personality in the behaviour of Great Apes: Temporal stability, cross-situational consistency and coherence in response. *Animal Behaviour*, 75, 99-112.
- van Aken, M. A. G. (2008). The next step: towards personality development in animals? (Invited commentary). *European Journal of Personality*, 22, 468-470.
- van Oers, K. (2008). Animal Personality, Behaviours or Traits: What Are We Measuring? (Invited commentary). *European Journal of Personality*, 22, 470-472.
- van Oers, K., Klunder, M., & Drent, P. J. (2005a). Context dependence of avian personalities: risk-taking behavior is domain specific. *Behavioral Ecology*, 16, 716-723.
- van Oers, K., de Jong, G., van Noordwijk, A. J., Kempenaers, B., & Drent, P. J. (2005b). Contribution of genetics to the study of animal personalities: A review of case studies. *Behaviour*, 142, 1185-1206.
- Verbeek, M. E. M., Drent, P. J., & Wiepkema, P. R. (1994). Consistent individual differences in early exploratory behaviour of male great tits. *Animal Behaviour*, 48, 1113-1121.
- von Eye, A. (2002). *Configural Frequency Analysis - Methods, Models, and Applications*. Mahwah, NJ: Lawrence Erlbaum.
- Weinstein, T. A. R., Capitanio, J. P., & Gosling, S. D. (2008). Personality in animals. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3rd ed.) (pp. 328-348). New York, NY: Guilford Press.
- Weiss, A., King, J. & Murray, L. (Eds.). (2011). *Personality and behavioral syndromes in nonhuman primates*. New York, NY: Springer.
- Westmeyer, H. (1997). On the causal status of structural concepts in personality psychology. In J. Bermudez, B. de Raad, J. de Vries, A. M. Perez-Garcia, A. Sanchez-Elvira & G. L. van Heck (Eds.), *Personality Psychology in Europe*. Volume 6 (pp. 17-26). Tilburg: Tilburg University Press.
- Wilson, D. S. (1994). Adaptive genetic variation and human evolutionary psychology. *Ethology and Sociobiology*, 6, 219-236.
- Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, 447, 581–584.
- Wolf, M., van Doorn, G. S., & Weissing, F. J. (2008). Evolutionary emergence of responsive and unresponsive personalities. *Proceedings of the National Academy of Science, US*, 105, 15825-15830.
- Wright, J. C., & Zakriski, A. L. (2003). When syndromal similarity obscures functional dissimilarity: Distinctive evoked environments of externalizing and mixed syndrome children. *Journal of Consulting and Clinical Psychology*, 71, 516-527.
- Zuckerman, M., Kuhlman, D. M., Thornquist, M., & Kiers, H. (1991). Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences*, 12, 929-941.

Appendix

Glossary of meta-theoretical concepts, and basic principles and concepts of measurement for the empirical study of individual behavioral phenotypes

Causal concepts: Theoretical concepts (*constructs*) explaining possible mechanisms and processes of proximate and ultimate causation of the behavioral phenomena described by *structural concepts*. (What causes the phenomena?)

Comparison research: *Individual-oriented analyses* of multiple individuals for groups of similar individuals (*types*) based on statistical similarity of the distinct score configurations in their *individual profiles* (e.g. using *Q-factor analysis*).

Configurational profile: see *individual profile*.

Configurational type: see *types*.

Construct: Complex theoretical concept describing and/or explaining (*structural concepts, causal concepts*) complex phenomena that are not directly observable. Constructs and their empirical measurements must therefore be validated systematically to provide evidence that the theoretical assumptions and their empirical measures are both logically and empirically interrelated and coherent. Important criteria of measurement quality are *reliability* and *validity*.

Correlation research: *Variable-oriented analyses* of the intercorrelational patterns among multiple variables in a population (e.g. using *R-factor analysis*).

Cross situational consistency: A *variable-oriented concept* denoting the phenomenon that the individuals' probabilities of showing a particular kind of behavior are (rank-order) consistent across different environmental *situations*. In humans and nonhuman primates, cross-situational consistency is often only moderate because individuals differ from one another in how they behave in different situations. This phenomenon is analyzed from *individual-oriented* viewpoints and shown in *test-retest reliable* and distinct *individual situation-behavior profiles*.

Differential concepts, differential perspective, differential scores: Differential concepts describe and explain (*structural concepts, causal concepts*) relative differences among individuals rather than the average individual of a species or of particular groups of individuals (e.g. defined by age or sex). The differential perspective is inherent to all concepts of *individual behavioral phenotypes* and often requires *standardization* of measurement scores.

Factor, factor analyses: Factor analyses are statistical methods that detect *latent structures* in the empirical relations among many variables and gather together those that are most highly correlated. They indicate to what degree empirical measurement variables can be explained and summarized by a much smaller number of synthesized *latent variables* called factors. *Variable-oriented* types of factor analysis are called *R-factor analysis (correlation research)*; *individual-oriented* types of factor analyses are called inverse or *Q-factor analysis (comparison research)*.

Hierarchical taxonomy: see *taxonomic models*.

Individual behavioral phenotypes: *Individual-specific patterns of behavior* that occur as variations within species-specific patterns of behavior from which they must be disentangled.

Individual-oriented analyses, individual-oriented approaches: These explore the matrix of X_i individuals by Y_j variables from the viewpoint of the individuals and study their individual configurations of *differential scores* on various variables. These configurations can be illustrated as *individual profiles*. The individual-oriented comparison of multiple individuals is called *comparison research*; it often aims at identifying *types* of individuals (e.g. using *Q-factor analysis*).

Individual profile: Illustration of an individual's *test-retest reliable* configuration of *differential scores* on multiple variables based on *individual-oriented analyses*.

Individual response specificity, individual response profiles: Individual response specificity is an *individual-oriented concept* denoting the phenomenon that individuals differ in which and how many of several meaningfully or functionally related behaviors they tend to show in a particular *situation*, and that these differences are stable over some time. It is illustrated in individual response profiles that plot the different behaviors against the *differential* probabilities with which an individual shows them. These *configurational profiles* of individuals must be shown to be *test-retest reliable*.

Individual situation-behavior profiles: These denote and illustrate the phenomenon that individuals show a particular kind of behavior with different probabilities in different *situations* and that these patterns are relatively stable over some time, yet differ among individuals. Individual situation-behavior profiles plot the individual's *differential* probabilities against the different situations in which they are measured, and are analyzed from *individual-oriented* viewpoints. These *configurational profiles* of individuals must be shown to be *test-retest reliable*.

Individual-situation transactions: Continuous interactions of *individual behavioral phenotypes* and environmental features over time that may shape patterns of *longitudinal change and development* during ontogeny.

Individual-specific patterns of behavior: As members of the same species, individuals show similar behaviors. Since behaviors are highly fluctuating and dynamic phenomena, individuals can only differ from one another in the probabilities with which they show particular behaviors. Probabilistic behavior patterns can only characterize a particular individual if they are stable over some time (*test-retest reliability*) and if they differ among individuals, i.e. are *differential*. Consideration of individual-specific patterns therefore presupposes population-level patterns of individual differences in behavioral data; yet not all between-individual variation on the population level is due to individual-specific patterns of behavior (*individual behavioral phenotypes*).

Internal consistency: *Reliability* of composite empirical estimates (*scale*) of *latent variables* using *variable-oriented analyses* that describes how well each measure contributing to the composite relates independently to the other measures, and how they are related overall. It is a basic criterion of measurement quality in research on *individual behavioral phenotypes*.

Inventory: Pool of measurement variables covering several *scales*.

Jingle-Jangle-Fallacies: These describe systematic inconsistencies in terminology. Jingle-fallacies occur when the same term refers to different concepts; jangle-fallacies occur when different terms refer to the same concept.

Latent structure, latent variable: Latent variables are variables that cannot be directly observed and that are construed from statistical patterns of observable variables; the latter are therefore also called manifest variables. That is, latent variables are hypothetical variables that refer to abstract concepts, such as *individual behavioral phenotypes*. For example, *factor analysis* creates latent hypothetical variables that describe and summarize the latent data structure underlying many observable variables in parsimonious ways.

Longitudinal change and development: Refers to the *stability* and changeability of the true-score variation of *individual behavioral phenotypes* over longer periods of time during ontogeny. Various kinds of longitudinal stability can be analyzed from both *variable-oriented* and *individual-oriented* viewpoints.

Meta-theoretical concepts ; meta-theory: Theoretical concepts underlying other theoretical concepts; theory about theory.

Methodological approaches to identify multi-dimensional structures of individual differences: The rationales and strategies used (explicitly or implicitly) to decide which concepts (*constructs*) and variables to study in order to identify *individual behavioral phenotypes*. Different classes of methodological approaches can be distinguished, of which only some enable systematic and comprehensive selections

to establish comprehensive *taxonomic models*, such as bottom-up/emic approaches like the Behavioral Repertoire x Environmental Situations Approach.

Principle of Aggregation: Measurements of *individual behavioral phenotypes* should be sufficiently aggregated to obtain *reliable* empirical estimates because a) probabilistic *individual-specific patterns* can only be determined through measurement repetitions, b) behaviors are fluctuating and dynamic phenomena showing considerable ephemeral variation that superimposes individual-specific variation, the measurement of which is c) therefore always afflicted with considerable measurement error.

Profile shape: see *individual profile*.

Q-factor analysis: see *factor analysis*.

R-factor analysis: see *factor analysis*.

Reference population: The population of individuals to which an individual's scores are compared. Its specification is important because quantifications of an individual's *differential scores* depend on the variation among the scores of those individuals to which it is compared.

Reliability: A criterion of measurement quality that refers to the consistency with which a particular kind of measurement yields consistent results. It is a necessary, yet not sufficient prerequisite for measurement *validity*. Important kinds of reliability are *internal consistency* and *test-retest reliability*.

Scale: Composite empirical estimate of a synthesized theoretical variable (*latent variable*), such as a *factor* (*factor analysis*). It is computed from multiple empirical measures in *variable-oriented analyses* (*correlation research*).

Situation (environmental): Complex constellation of stimuli or conditions in the abiotic, biotic, and conspecific environments that can affect individuals by both their presence and their absence.

Situational relevance: Theoretical concept describing the qualitative properties of environmental *situations* to which the individuals respond. It refers to the suitability of the situation to study the particular kinds of behaviors of interest empirically.

Situational strength: Theoretical concept describing how compelling the quantitative properties of environmental *situations* are for the individuals, and how much they restrict behavior. Strong situations force the individuals' behavior into specific channels, thus preventing the emergence of differences among them. For empirical measurements of *individual behavioral phenotypes*, weak situations should be studied that permit individual differences to emerge and that therefore have high discriminatory power.

Standardization: Absolute behavior scores do not readily inform about the ways in which individuals differ from one another because they confound the sample's average behavior patterns with those that may be *individual-specific*. Empirical measures must therefore be set in relation to those of other individuals (*differential scores*, *reference population*) using statistical centralization and standardization of scores, such as with z-standardization.

Stability: Refers to *longitudinal change and development* in the hypothetical true scores (*latent variables*) of *individual behavioral phenotypes*. It should be differentiated from *test-retest reliability* that indicates the quality of the empirical measurement of individual behavioral phenotypes at any one time point, which inherently involves stability over intermediate periods of time. Since the word 'stable' can imply both, authors should indicate clearly whether they refer to measurement reliability or to true score stability.

Structural concepts: Theoretical concepts (*constructs*) describing the empirical structure of the behavioral phenomena of interest. (What are the phenomena?)

Taxonomic models, taxonomic hierarchies: *Variable-oriented* descriptive model (*structural concept*) mapping the empirical intercorrelational structure (*latent structure*, *correlational research*) of variations of *individual behavioral phenotypes* in a population. At the bottom of such hierarchies are more specific and often highly

intercorrelated dimensions. They are statistically subsumed (e.g. using *factor analysis*) within broader and more abstract dimensions, making the less complex dimensions sub-dimensions of the emergent, more complex major dimensions often called *factors*. The major dimensions at the hierarchical top each summarize different and often relatively independent sets of often less strongly intercorrelated variables. In such cases, taxonomic models are multidimensional. Establishing comprehensive taxonomic models requires systematic *methodological approaches* that enable representative selections and reductions of *constructs* and variables to be studied. This is of vital importance because all studies are ultimately based on the selection and definition of the theoretical concepts they investigate.

Test battery: Pool of measurement variables covering several *scales*.

Test-retest reliability (temporal reliability, repeatability): Criterion of measurement quality that estimates the proportion of true score variation attributable to *individual behavioral phenotypes* in relation to total variation that includes ephemeral and measurement error variation in empirical measurements. Requires sufficient consideration of the *Principle of Aggregation*. It should be differentiated from *stability* over longer periods of time, although test-retest reliability naturally also implies stability, but only over intermediate periods of time.

Types (configurational): These are statistically identified groups of individuals sharing similar distinct *profile shapes* (e.g. using *Q-factor analysis*) that represent discontinuous categories of prototypical individuals. Configurational types are analyzed from *individual-oriented* viewpoints based on multi-dimensional descriptive models (*structural concepts*) of individual differences on the population level (*taxonomic models*) that are developed using *variable-oriented approaches*.

Validity: Because *individual behavioral phenotypes* denote complex phenomena that cannot be observed directly, it must be shown that the theoretical *constructs* describing and/or explaining them (*structural concepts, causal concepts*), and their empirical measures are both logically and empirically interrelated and coherent. This criterion of measurement quality describes the extent to which the particular measures do actually measure what they are intended to measure, which is important for accurate interpretations of results. For different types of validity the interested reader is referred to the classics of Cronbach and Meehl (1955), for methods of construct validation to those of Campbell and Fiske (1959).

Variable-oriented analyses, variable-oriented approaches: These explore the data matrix of X_i individuals by Y_j variables from the viewpoint of the variables and study their score distributions across all individuals. They reveal population-level patterns of between-individual differences in behavioral data (rank-order differences) that may, but need not be based on individual-level phenomena of *individual-specific patterns*. *Variation research* and *correlation research* are distinguished.

Variation research: *Variable-oriented* analyses of single variables.
