Jana Uher , Elsa Addessi, Elisabetta Visalberghi

Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (Cebus apella)

Article (Accepted version) (Refereed)

Original citation:
Uher, Jana, Addessi, Elsa and Visalberghi, Elisabetta (2013) *Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (Cebus apella).* Journal of Research in Personality, 47 (4). pp. 427-444. ISSN 0092-6566

DOI: 10.1016/j.jrp.2013.01.013

REPRINT

Original research article

Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (*Cebus apella*)

Jana Uher [*1], Elsa Addessi[2], Elisabetta Visalberghi[2]

1 Comparative Differential and Personality Psychology, Department of Educational Science and Psychology, Freie Universität Berlin, Germany
2 Unit of Cognitive Primatology and Primate Centre, Institute of Cognitive Sciences and Technologies, National Research Council of Italy (ISTC-CNR), Rome, Italy

*Corresponding author:
Jana Uher
Comparative Differential and Personality Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin, Germany
mail: uher@primate-personality.net
fon: +49-(0)30-838 55 600
fax: +49-(0)30-838 55 789

Highlights

- Personality differences studied with a new philosophy-of-science paradigm featuring a behavioural research framework.
- A non-lexical taxonomic approach applied for comprehensive emic/bottom-up construct generation.
- In 15 test and 2 group situations, 146 contextualised behaviours measured repeatedly.
- This is the first comprehensive study on personality differences in capuchin monkeys.
- Overall no age and sex differences found, but long-term effects of early life experiences.

Abstract

We applied a new framework for behavioural research on personality differences in 26 adult tufted capuchin monkeys. Using the Behavioural Repertoire x Environmental Situations Approach, we generated systematically 20 non-lexical emic personality constructs that have high ecological validity for this species. For construct operationalisation, we obtained 146 contextualised behavioural measures repeatedly in 15 experimental situations and 2 group situations using computerised and video-assisted methods. A complete repetition after a 2-3-week break within a 60-day period yielded significant test-retest reliability from individual-oriented and variable-oriented viewpoints at different levels of aggregation. In accordance with well-established findings on cross-situational consistency, internal consistency was only moderate. This new and important finding highlights fundamental differences between behavioural approaches and judgment-based approaches to personality differences.

Key words: age differences; Behavioural Repertoire x Environmental Situations Approach; capuchin monkeys; cross-situational consistency; early life experiences; individual behaviour; methodology; personality differences; sex differences; situation-behaviour-profiles

1. Theoretical Background

Psychologists increasingly advocate for taking behavioural measurements of personality differences in greater consideration (Baumeister, Vohs, & Funder, 2007; Furr, 2009). Methods for studying the individuals' *ongoing* behaviour, physiology, and experience, as well as the environmental contexts in which they occur are currently being revived and further developed. Contextualised behavioural measurements are obtained, for example, in experimental "miniature situations" in objective personality tests sensu Cattell (Cattell & Warburton, 1967; Kubinger, 2009) or in daily life settings in the field of ambulatory monitoring (Gunthert et al. 2007; Mehl & Conner, 2012; Mehl & Pennebaker, 2003) to which Cattell (1957) referred to as Test-data and Life-data respectively.

The behavioural study of individual differences involves more than just behavioural measurements, however. It requires meta-theoretical concepts and methodological approaches that differ in parts from those that have been established for judgment-based investigations of individual differences. This is because behavioural measurements and judgments capture different types of phenomena (see below; Uher, 2013). For comprehensive research on individual *behaviour*, an integrative meta-theoretical and methodological framework that provides the necessary conceptual and analytical foundations was therefore elaborated as part of a novel philosophy-of-science paradigm for personality psychology. This framework was derived from established concepts of various disciplines, among them personality psychology, differential psychology, cross-cultural psychology, comparative psychology, and behavioural biology. Because their separate conceptualisation and application hindered comprehensive investigations, these concepts were coherently integrated and expanded by new concepts and approaches that relate individual behaviour to different contexts on various population levels including the species level (Uher, 2013, 2008a, b, 2011a, under review a).

1.1 What is behaviour?

Behaviour is the central object of research in many psychological and biological disciplines. But definitions of behaviour are rarely discussed and the few proposed are only operational or nominal (e.g., Furr, 2009). Uher's new research paradigm relies on an elaboration of a meta-theoretical definition from philosophy that defines behaviour as "external activities or externalisations of living organisms that are functionally mediated by the environment (Millikan, 1993) in the present" (Uher, 2013). This definition highlights that behaviour is intrinsically related to properties of its immediate external environment that are defined as environmental situations. It also emphasises that behaviour is inherently bound to the present, which necessitates realtime measurement. Externality differentiates behavioural phenomena from psychological phenomena, which are also bound to the present, but are entirely internal phenomena. The meta-theoretical definition generally refers to all behavioural phenomena, yet without specifying particular ones, such as goal-directed action, because these concepts include a priori assumptions about underlying psychological processes that are considered separate phenomena in this paradigm.

Personality judgments, in contrast, typically involve assumptions of underlying psychological processes—be they accurate or not. Judgments capture constructs and representations that people develop of individual behaviour and its possible causes and consequences. Judgments are based on abstractions of behaviours and other events that have been perceived in the past and that are therefore being (re)constructed in often decontextualised ways.

1.2 What is individual behaviour?

Given the steady fluctuations in behaviour, individuals can be characterised only by probabilities of behaviour. But behavioural probabilities that characterise all individuals in a given population or species are not individual-specific; they must differ among individuals in relatively stable ways. To reflect individual-specificity, differential probabilities of behaviour must be stable across time periods longer than those in which the probabilities were first ascertained and in ways that are considered to be meaningful as defined, for example, by the strength of correlations over specified time periods (Uher, 2013; see Section 4). The three defining criteria of individual-specificity in behaviour—probability, differentiality, and temporal stability—have important implications for research designs. They require *1)* aggregations of realtime measurements over repeated occasions to obtain measures reflecting individual probabilities of behaviour; *2)* standardisation of these aggregate scores to obtain measures reflecting differential patterns in these probabilities; and *3)* evidence of temporal reliability in the standardised aggregate scores over some time to show that the obtained behavioural measures do reflect individual-specific patterns and can be constructed as personality differences (Uher, 2011a).

1.3 Objective quantifications and comparisons of individual behaviour

To meet the peculiarities of the momentary fluctuations of behaviour and its ontogenetic changeability, Uher's philosophy-of-science paradigm builds on a new type of probability—time-relative probability—that relates the empirical occurrences (e.g., frequencies and durations) of specified behavioural events to specified time periods. Time-relative probabilities enable ratio-scaled quantifications, which are essential for comparisons across different situations, groups, populations, and species (Uher, 2013). They allow for quantifications and comparisons of personality differences that are *objective* in the sense that they are based on statistical computations of a set *B* of explicitly known behaviours studied in a set *S* of known environmental situations, across a set *T* of known occasions and spans of time in a set *I* of explicitly known individuals. When people judge individuals, in contrast, it remains unknown which particular elements of these four sets *B, S, T,* and *I* they implicitly consider and how they mentally compute interrelations within and among these sets. It follows that quantifications and comparisons of individual-specific patterns using judgment-based methods are necessarily *subjective* (Uher, in preparation).

1.4 Contextualised concepts of individual behaviour

The new meta-theoretical framework also allows for the systematic study of the situation-to-situation variability of individual behaviour. It is well established in psychology that cross-situational consistency is often only moderate (Hartshorne & May, 1928; Mischel, 1968, 1977). Interestingly, cross-situational consistency was also only moderate in nonhuman primate species, such as rhesus macaques (Stevenson-Hinde, Stillwell-Barnes, & Zunz, 1980) and great apes (Uher, Asendorpf, & Call, 2008), and rarely exceeded the limit of r = .30 that Mischel (1968) identified for humans. Individual-specificity emerges at the level of distinct *individual-specific situation-behaviour profiles* (Mischel, Shoda, & Mendoza-Denton, 2002) or *if...then...profiles* (Mischel & Shoda, 1995) that are relatively stable over time in both humans and nonhuman primates. These findings show that the core issues of the long-lasting person-situation debate (Fleeson, 2004; Funder, 2006) are not uniquely human, but apply to other species as well (Uher, 2011a, 2011b).

Cross-situational comparisons must also consider that all individuals in a population may shift their behavioural probabilities across situations in similar ways. These shifts reflect their general adaptation to specific environmental conditions. This can be illustrated with *population-specific situation-behaviour profiles* that depict the population's average probability for displaying a particular behaviour across different situations. Individual-specific patterns emerge as stable between-individual deviations around these shifts in the population's averages. Population-specific and individual-specific behavioural probabilities are disentangled with differential techniques, such as statistical standardisation within each situation (see above, Uher, 2011a, 2011b, submitted for publication a).

In the contextualised study of individual behaviour, two situational concepts must be considered. *Situational strength* denotes how compelling a situation is for a particular population of individuals. Weak situations permit easy emergence of individual differences and differentiate well among individuals. Strong situations, in contrast, mask individual differences because they either substantially inhibit particular behaviours or because they evoke heightened responses from all individuals (Mischel, 1977; Tett & Guterman, 2000). Individual differences emerge most clearly in situations that typically elicit moderate levels of the considered behaviours in a given population. *Situational relevance* refers to the functions that situational properties may have for particular behaviours in a particular population (see the Section 1.1). That particular behaviours cannot be observed in a given situation does not mean that the individuals' probabilities to exhibit them are low. Rather, it means that the situational properties that are functionally mediating these behaviours may be absent (Tett & Gutermann, 2000; Uher, 2011a, 2011b).

1.5 A new non-lexical taxonomic approach

These meta-theoretical definitions and concepts lay the foundations of a new methodological approach developed for taxonomic research on individual differences in behaviour—the Behavioural Repertoire x Environmental Situations Approach (BR_xES-Approach; Uher, 2008a, b, 2011a, 2011b). This approach links behaviours conditionally ("x") to the environmental situations in which they occur at the levels of both individuals and populations (see Section 1.4). The BR_xES-Approach is a non-lexical emic/bottom-up approach (also called a manifest system approach, see Uher, submitted for publication b) that allows researchers to generate constructs of individual-specificity systematically from within the behavioural-ecological system of a population. The approach capitalises on the existing knowledge that scientists already have about the behavioural repertoire of the *average* individual of a given population. It relies on scientifically described categories of behaviours and situations of known meaning and function that have been derived from scientific observation and analysis.

Lexical approaches, in contrast, on which many important models of human personality differences are based, built on the lexical repertoires of human language communities. These repertoires reflect the lexically encoded representations that people have developed from incidental everyday observations of individual differences and their socioculturally shared interpretations and appraisals. Lexical descriptors and lexically derived constructs often cannot be traced back to the particular behaviours and situations to which they refer, as this is the case for the BR_xES-Approach, however. For this reason, lexical approaches explore individual differences with primarily generic and decontextualised (i.e., nonconditional) descriptors and constructs (cf. McAdams, 1992). Moreover, constructs derived from everyday language are frequently attributed a causal status (Uher, 2013), whereas the constructs generated by the BR_xES-Approach are a priori merely descriptive. Given their origins in the ethnological and behaviour-scientific knowledge bases, BR_xES-Approach-generated constructs are labelled with terms that are much less colloquial than those derived from the human everyday languages (Uher, submitted for publication b). For example, individual differences in the tendency to pay attention to food and to invest efforts to obtain also non-preferred food are labelled Food orientation rather than Gluttonousness (see Section 2). This meets efforts to reduce the impact of implicit meanings that often vary socioculturally, thus introducing anthropomorphic biases to nonhuman studies. A further difference is that the BR_xES-Approach generates constructs and not measurement variables that are considered only in a second step. Their targeted selection for construct operationalisation helps to keep their number manageable for empirical studies. Lexical approaches, in contrast, select person-descriptors from the human lexica and use them as the measurement variables in empirical studies. Their number therefore tends to be too large for empirical investigations in the same study (John, Angleitner, & Ostendorf, 1988).

Lexically derived constructs and their judgment tools are developed by selecting variables that yield empirical data structures with high internal consistency and that thus measure redundancies (Block, 2010). While this is possible for analyses of everyday language terms about individual differences, redundancies in behaviour may be rare. The pervasive finding that individual behaviour is only moderately consistent across situations (see above) already indicates that behaviourally derived constructs cannot be as internally consistent as lexically derived constructs. Even within the same situation, behaviours that are assigned to the same construct in everyday psychology and in lexical research often show low to zero internal consistency; as this is the case for the behaviours gaze aversion, long pauses in speech, hesitant speaking, and restricted gestures that are all assigned to the construct "shyness" (Asendorpf, 1988). Analogous to cross-situational consistency, individual-specificity emerges in distinct, but stable configurations of individual probabilities in different behaviours. These configurations can be illustrated with stable individual-specific profiles across functionally related behaviours (Fahrenberg, 1986). This so-called individual response specificity was also shown in nonhuman primates (Uher, 2011a, 2011b; Uher et al., 2008).

Ultimately, it must be considered that behavioural data have different properties than judgment data because they capture individual behaviours, not representations of individual behaviour. Behaviours and representations are different types of phenomena (Uher, 2013) and, moreover, studied with different scales. The concept of time-relative probabilities allows for ratio-scaled measurements of individual behaviour. This is not possible for the predefined scales that are being applied in the judgment-based methods of lexical research. For these reasons, the structural patterns of individual differences in behaviour may differ from those that can be found in data reflecting people's subjective (re)constructions and lexically encoded representations of observable individual differences. Specifically, behavioural data need not fulfil the psychometric standards as established in judgment-based research. The BR_xES-Approach considers these peculiarities of behaviour and allows researchers to employ a two-step procedure to explore the taxonomic structures of individual differences in behaviour. In the first step—regardless of potentially low internal consistencies among specific measurement variables—behavioural data can be reduced to functionally defined constructs of individual-specificity based on the studied behaviours' scientifically established functions and meanings. In the second step, these theoretically derived construct measures, rather than the behavioural variables themselves, can be subjected to statistical reduction techniques.

The BR_xES-Approach was already tested in methodological studies on zoo-housed great apes (e.g., Uher, 2011b; Uher et al, 2008). Comprehensive arrays of major behavioural and situational categories were investigated in 14 behavioural tests and 2 group situations. High temporal reliability of individual-specific behaviours over 4-6 weeks was shown at the level of 76 contextualised behavioural measurements and at the level of 17 functionally defined constructs derived from them. Issues of cross-situational consistency and of individual response specificity were highly similar as described for humans.

1.6 The present research

The first aim of the present study is to demonstrate the application of this new meta-theoretical and methodological framework to capuchin monkeys (*Cebus apella*), a South American primate species. The peculiarities of behavioural approaches, in particular regarding the consistency of behavioural measurements and functionally defined constructs are explored systematically. The second aim is to provide the first comprehensive behavioural study on personality differences in this species and to evaluate effects of age, sex, and early life experiences. Early life experiences are well known to have an impact on adult individual behaviour in humans and nonhuman primates (Suomi, 2005). Effects of age and gender/sex on personality differences are frequently reported in judgment-based studies on humans (e.g., Soto, John, Gosling, & Potter, 2011) and nonhuman primates (e.g., King, Weiss, & Sisco, 2008; Kuhar, Stoinski, Lukas, & Maple, 2006). But judgments reflect sociocultural (and anthropocentric) perspectives that need not accurately reflect observable differences between age and sex/gender groups. Behavioural investigations on such differences in a broad range of personality constructs are still scarce.

Capuchin monkeys are highly interesting for personality research. This neo-tropical species lives in multi-male–multi-female groups typically led by one dominant adult male. Their social system is rather egalitarian and the dominance hierarchy is not as rigid as in many other primates, such as rhesus macaques. Despite 35 million years of independent evolution from hominids, this species shows many striking analogies with humans (Fragaszy, Visalberghi, & Fedigan, 2004). Young capuchins develop at a slower rate than most other monkey species and have an extended period of maternal dependency (Byrne & Suomi, 1998; Fragaszy et al., 2004). Capuchin monkeys have larger brains than expected for their body size and a long life span in comparison with other primate species of similar size; captive individuals can live on average 40 years. Capuchins are opportunistic omnivores with highly adaptive foraging skills that allow them to minimise intra-species and inter-species competition. They are very manipulative, strongly motivated to act on and combine objects, and use several types of tools both in captivity and in the wild. Their flexible use of tools is comparable to that reported for chimpanzees (Visalberghi & Fragaszy, 2012, chap. 40).

Although the capuchins' behaviour and cognition has been studied intensely in the last three decades, individual differences have not. Early anecdotal reports from field observations explicitly mentioned pronounced individuality among the alpha-males of different groups that obviously influenced their group structures (Izawa, 1980). Individual differences were first investigated empirically by Byrne and Suomi in a series of captive studies with infant and juvenile capuchins. They found considerable individual differences in both the timing of developmental changes and the individuals' average levels in exploratory and social behaviours in the first year of life (Byrne & Suomi, 1995). Group experiments yielded considerable individual differences in the levels of object manipulation and consistency across situations that involved food and different numbers of objects (Byrne & Suomi, 1996). A study on individual responses to a 2-h separation period from the mother showed substantial stability of individual differences from 6 months to 1 year of age (Byrne & Suomi, 1999). Another study found that personality judgments by human observers on adjectives describing inhibited and fearful behaviours were associated with high cortisol reactivity in capuchins up to 6-year-olds (Byrne & Suomi, 2002). Despite these promising results on individual differences among capuchins in subsets of behaviours, comprehensive investigations in a large number of behaviours in capuchins are still missing.

2. Methods

2.1 Individuals

We studied 26 adult tufted capuchin monkeys (*Cebus apella*) at the Primate Centre of the ISTC-CNR hosted by the Bioparco of Rome in Italy. The individuals were 8 to 33 years old; median age was 15.5 years. Fifteen individuals were mother-reared; the others were hand-reared and introduced in a group when on average 7 months old (range 4-12 months; for details see Table S1 in the Supplemental material). The monkeys lived in four groups of 5-11 individuals in naturally designed outdoor enclosures (18-127 m^2) in which they spent most of their time and in indoor enclosures each with two interlinked cages (4.3-4.5 m^2 per cage). The indoor enclosures also served as test rooms. The monkeys were treated in accordance with the Guidelines for the Treatment of Animals in Behavioural Research and Teaching (2006). All monkeys received a balanced diet consisting of fresh fruits, vegetables, cereals, and dairy products; water was always available *ad libitum*.

2.2 Systematic non-lexical generation of emic personality constructs

We applied the Behavioural Repertoire x Environmental Situations Approach to systematically generate non-lexical emic personality constructs that have high ecological validity for capuchin monkeys (see Section 1, Uher, 2008a, 2008b, 2011a, 2011b). First, we conducted a broad-based review of 68 publications about the behavioural repertoire of captive and wild capuchin monkeys (all references are listed in Section S2 of the Supplemental material). From this review, we compiled a large table in which one column contains the major behavioural categories and a second column the categories of environmental situations in which the behaviours are reported to commonly occur. Each row of the table thus represents a unit of a particular behavioural category and a particular situational category as described in a given study; this is called *behaviour$_x$situation-unit* in the

BR_xES-Approach. The literature references from which each unit was taken were also noted in a further column for documentary purposes.

Behaviours and situations are typically categorised at different degrees of abstraction depending on the focus of the particular study; all functionally related categories compiled in the review can therefore be organised hierarchically in ascending degree of abstraction. More specific categories can be assigned to moderately abstract categories and those turn to even more abstract, broader categories. In the table established for the category compilation, this can be organised by adding columns in which the more abstract categories are listed next to several more specific categories to which they refer. Major rows help to organise the broadest levels of behavioural categories. For example, the specific behavioural categories of Grooming, Touching, Scalp-lifting, and Lip-smacking can be assigned to the more abstract categories of Prosocial contact behaviours (the first two) and of Affiliative facial displays (the latter two). These two categories in turn can be both assigned to the more general category of Prosocial behaviours. This category can be assigned to the very broad and abstract behavioural category of Social behaviours together with other categories of social behaviour, such as of Aggressive, Dominant, or Sexual behaviours. Similarly, more specific environmental situations investigated in a particular study can be assigned to more abstract situational categories that can also be organised in a separate column.

It is important to note that, depending on research focus, many behaviours and situations can be assigned to more than one category, such as co-feeding can be assigned to Feeding, Social contact, and Physical Activity. Because in the table, the behavioural categories are always listed together with the categories of situations in which the behaviours are described to commonly occur (i.e., in $behaviour_x$situation-units), situational categories were often listed repeatedly with several behavioural categories. For example, the three situational categories Encountering novel food, Encountering novel objects, and Encountering novel environments can be listed each together with the behavioural categories of Visual exploration, Olfactory exploration, Oral exploration, and Tactile exploration. This repeated inclusion of behaviours and situations in both the primary compilation of categories and their hierarchical organisation differs from the disjunctive categorisation schemes commonly used in behavioural research. This overinclusive category compilation is the crucial prerequisite with which the BR_xES-Approach enables comprehensive non-lexical generations of emic personality constructs (Uher, 2008a, 2008b, 2011a, 2011b).

From all $behaviour_x$situation-units on moderately abstract levels of categorisation, we then systematically generated personality constructs by hypothetically assuming individual-specific patterns in the behaviours described. These constructs are therefore called *working constructs*. They are defined by individual-specific probabilities in displaying the particular behaviours in the particular situations of the respective categories from which the constructs were derived. Using categories of moderate levels of abstraction ensured that the generation of constructs and their operational definitions were based on relatively homogeneous and still identifiable subsets of behaviours and situations. The working constructs themselves were provisionally labelled with mostly decontextualised terms to facilitate the generation process. These preliminary construct labels were taken down in a new column in each row, that is, for each $behaviour_x$situation-unit of moderate level of abstraction. The overinclusive compilation of behavioural and situational categories and the emic/bottom-up reasoning underlying this methodological approach entails that the same working constructs were repeatedly generated in different parts of the category system. From the table used for category compilation, a second table was then derived with identical content, yet in which the rows were sorted by the generated constructs. By eliminating redundant enumerations of the same behavioural and situational categories, a third table was produced that contained a comprehensive overview of all generated constructs and the major behavioural and situational categories in which they describe individual-specificity.

For example, the construct of Arousability was derived from the moderately abstract behavioural categories of Self-grooming in situational contexts occurring in the wild and in captivity some of which are more specifically characterised by social inequity; of further Selfdirected behaviours, such as Scratching, in social group contexts in general and more specifically during social separation, in post-conflict situations and delay of gratification tasks; and of Urine washing (a capuchin-specific behaviour) in the situational contexts of social aggression, social inequity, and of potential stress or danger in both captivity and the wild. Accordingly, this working construct was defined as "individual-specific probabilities to show behaviours of arousal (e.g., scratch, urine wash, pace, body shake, pilo erection) in social situations in general, specifically of uncertainty or tension, such as during social separation or in agonistic group situations, post-conflict situations, as well as in non-social situations characterised by tension, potential stress and danger."

The emic/bottom-up reasoning of the $BR_x ES$-Approach also entailed that all working constructs were a priori constructed to be unipolar describing individual-specific behaviours of the *same* function (see functional definition of behaviour above), such as the construct (low to high) Aggressiveness. Whether several working constructs can be constructed as representing opposite poles of a few, more abstract taxonomic constructs that are bipolar and that describe behavioural patterns of *distinct* functions, such as the hypothetical taxonomic constructs Aggressiveness—Anxiousness or Aggressiveness—Social Orientation, is left to empirical investigation in each study population.

For the present study on captive capuchins, this procedure of the $BR_x ES$-Approach yielded 20 non-lexical emic working constructs (see below). Constructs that describe individual-specific behaviours in situations occurring only in the wild, such as Territoriality and Travelling, or that require the presence of offspring in all studied groups, such as Mothering and Social orientation to youngsters, could not be not considered.

2.3 Operationalisations in contextualised behavioural measurements

We operationalised each of the 20 working constructs with contextualised behavioural measurements. These operationalisations were based directly on the behavioural and situational categories that were used to generate and to define the constructs. To study situations with specific properties, we developed 15 laboratory-based behavioural tests. To study social behaviour, we used two group situations in the monkeys' outdoor enclosures that occurred in their normal daily routines. This combination of behavioural tests and social observations enabled comprehensive investigations of individual behaviour with both Test-data and Life-data sensu Cattell (1957) in the same study. For most constructs, we could obtain behavioural measurements in several test or group situations. Vice versa, most situations allowed us to measure behaviours that are related to various constructs (for an overview, see Table 1). To increase the reliability of measurement, wherever this was possible we measured multiple behaviours that are related to a given construct within the same situation. This procedure also enabled analyses of cross-situational consistency and of the internal consistency of composite construct measures.

2.3.1 Behavioural tests

The behavioural tests correspond to Cattell's concept of standardised and experimental "miniature situations" (Cattell & Warburton, 1967). Each test situation was narrowly defined by particular situational properties that—given the literature reports—were relevant to particular behaviours and thus expected to elicit them with high probability (see Situational relevance above). Based on these reports and our own experiences with this species, we tried to adjust the situational strength (see above) of each test such that individual differences were likely to emerge. For example, we covered the eyes of the soft toys that we used in the Furry animal test (S 1.1.12) to reduce the degree of threat that soft toys constitute for capuchins and to avoid that all individuals likely behave the same. We followed high ethical standards when designing test situations that constituted potential threats (for details see the Ethical note in Section S 1.2).

All behavioural tests were carried out in highly standardised ways in the indoor cages, where we could minimise the impact of random external influences and could test the monkeys individually or in dyads (as in the Food competition test, S 1.1.3). Between the cages and the experimenter area was a mesh from floor to ceiling. This mesh had exchangeable panels near the floor at which we could fix the apparati that we built for some of the tests (see S 1.1.1 – S 1.1.15). One main experimenter (AG) assisted by a second experimenter (JU in most sessions) tested all individuals; AG carried out all tests in which we recorded behaviour towards humans (S 1.1.4, S 1.1.6, S 1.1.7, S 1.1.9–11).

2.3.2 Behavioural observations

In the *Prefeeding observations* (S 1.3.1), we observed the monkeys of one group prior to their daily main feeding, while they could hear the keepers preparing the food in the nearby kitchen and the neighbouring groups being fed. In the subsequent *Social observations* (S 1.3.1), we observed the individuals in situations that occurred naturally in their groups in the outdoor enclosures. All observations were carried out in highly standardised ways based on a detailed ethogram, the Social observations by AG, and the Prefeeding observations by AG and JU. All observers had been previously trained to recognise the individuals and their behaviours reliably. Table 1 briefly describes the test and group situations; more details are provided in Section S1 of the Supplemental material.

Table 1

2.4 Methods of behavioural measurement

All *Behavioural tests* were videotaped for detailed coding in terms of behavioural latencies, durations, and frequencies using the coding software INTERACT (Rel. 9.2.1, www.behavioural-research.com; Mangold, 2010). This allowed us to obtain measures reflecting time-relative probabilities (see Section 1.3). To study the individuals as comprehensively as possible, we coded all behaviours that occurred regularly in the tests and that could be coded reliably from the given camera positions, including some behaviours that were initially not planned to be scored.

In the *Prefeeding observations*, the capuchins showed behaviours of brief duration and highly fluctuating (e.g., Scratching, Food vocalisation). Thus, we used one-zero sampling with 10-s time intervals to estimate frequencies that included any amount of time spent in the respective behaviours (Altmann, 1974). We recorded the behaviours on check sheets with paper and pencil and observed the individuals in a predetermined order. We collected two sample points per individual and observation day.

In the *Social observations*, we combined three methods of behavioural observation to estimate time distributions of frequency and duration behaviours: (a) focal individual sampling continuous recording, in which we followed a single individual for 10 min-periods and recorded continuously its frequency and duration behaviours; (b) scan sampling instantaneous recording every 10 min in which we recorded the presence (or absence) of duration behaviours for all individuals of the group; and (c) event recording of rare frequency and duration behaviours, such as aggressive or sexual behaviours. When recording event behaviours, we interrupted the focal and scan sampling (Altmann, 1974). The individuals were observed in a predetermined order in the focal sampling and in random order in the scan sampling. For each individual, one 10-min focal sample and 7 scan sample points were collected per observation day. For behaviour recording in all three methods, we used an interactive computer software programmed by JU that logged all data entries with a precise timestamp.

In summary, we aimed to obtain $N = 154$ contextualised behavioural raw measurements. However, for 6 variables, the behaviours did not occur at all in both blocks of data collection, for further 5 variables behaviours did not occur in one of the two blocks. Therefore, we obtained $N = 143$ analysable variables. All working constructs, their operationalisations in terms of the specific behavioural measurements and the experimental and group situations respectively in which they were obtained are listed in Table S2 in the Supplemental material.

2.5 Study design

2.5.1 The individuals' participation in tests and observations

The monkeys came indoors for tests voluntarily. Because not all monkeys at the Primate Centre were used to being tested alone, we could study 14 individuals in all tests and observations, 11 individuals only in the observations, 5 only in the Prefeeding observation, and 1 female in almost all behavioural tests, but in none of the observations (because when outdoor she was out of view for most of the time possibly due to social tension; see Table S1).

2.5.2 Measurement repetitions

To reduce the impact of day-to-day fluctuations on the behavioural data, we studied each individual in 10 of the 15 laboratory-based test situations twice within a block of about 2-2.5 weeks. For ethical reasons (see S 1.2), 5 tests that could potentially elicit fear (S 1.1.7, S 1.1.10, S 1.1.11, S 1.1.12, S 1.1.15) were administered for only 1 session. To further reduce the impact of transient states on the aggregated data, the laboratory-based tests were carried out in a pseudo-random sequence to avoid studying individual behaviours that are related to the same construct several times a day. Wherever possible, behavioural tests were presented to all individuals in the same order. We carefully considered possible after-effects, thus mildly disturbing situations were always presented at the end of each individual's test session and no more than once a day. In each group situation, the behavioural observations were carried out for 10 days.

2.5.3 Complete repetition in a second block of data collection

To analyse test-retest reliability, we repeated the entire data collection of block one in a second block; the two blocks were separated by 2-2.5 weeks. In the second block, we followed the same scheme of repetitions and randomisation as in the first block. Overall, 10 tests were repeated four times, 5 tests were repeated twice, and both group observations were carried out for 20 days in two non-overlapping blocks of data collection within a 60-day period.

2.5.4 Total test and observation times. Within this study period, for most individuals (see 2.4) we recorded behaviours in (a) 320 min in the experimental situations; (b) 40 zero/one coded 10-s intervals in the Prefeeding observations distributed over 50 min observation time in total; (c) 200 min of continuous recording in the focal individual sampling; and (d) 140 instantaneous points collected by scan sampling (c and d were collected in 25 h during the Social observations). Thus, each individual participating in all tests and observations was recorded for 31.2 h.

2.6 Study procedure

2.6.1 Daily routine

Between 9.30 a.m. and 3.00 p.m., individual capuchins were presented with 2-4 behavioural tests (see 2.3; Table 1; Section S1). After each test, the monkey joined its group members and came back for further tests later. Prefeeding observations occurred at about 3.15 p.m. while the monkeys awaited their main feeding. The Social observations started after the feeding (at about 4 p.m.) and lasted till 6.30 p.m., approximately.

2.6.2 Schedule of data collection

Tests and observations were carried out in parallel, that is we tested the individuals during the same time block in which we carried out the observations of their group. For each group, we scheduled the two 2-2.5 week blocks of data collection so that there was a break of about 2-2.5 weeks between them. The entire data collection lasted 16 weeks for the 4 groups; it took place from May to September 2011.

2.7 Data analyses

Technical terminology: The philosophy-of-science paradigm applied in this study adopts a more technical terminology than this is common in lexical research (Uher, 2013). Such a precise terminology is needed to refer unambiguously to the different phenomena (e.g., behaviours, situations, representations) studied and to the different concepts (see Section 1) with and the different levels at which these are being described and analysed (e.g., contextualised composite construct measures, see Section 2.7.2). A glossary of the terms relevant to the present analyses is provided at the end of this article. Table 2 gives an overview of the analytical perspectives taken, the different levels of aggregation studied, and the interrelations within and among the four sets *B, S, T,* and *I* (see Section 1.3) that the different analyses explore.

Table 2

2.7.1 Inter-coder and inter-observer reliability of behavioural measurements

Inter-coder and inter-observer reliability was calculated for all behavioural raw measurements obtained in the tests and observations. Two persons (AG, FDP) coded 20% of the sessions of each test from video independently from one another to calculate inter-coder reliability. Two other persons (MS, CM) coded activity behaviours in the tests with 10-min sessions (S 1.1.4, S 1.1.9 – S 1.1.11); to calculate inter-coder reliability, 20% of these sessions were also coded by AG and FDP. To calculate inter-observer reliability for the group observations, 15% of the Prefeeding observations (carried out by AG and JU) and of the Social observations (carried out by AG) were also coded by independent observers (JU, EPS). The median Cronbach's α was .89 (range .71 – .98). For aggregations over time, zero-one coded data were treated as metric variables.

2.7.2 Steps of data aggregation within each block of data collection

2.7.2.1 Raw measurements

We equated latency scores with the maximum session time if the target behaviour did not occur, and inverted some latency measures considering their meaning for the respective construct. Furthermore, we transformed some measurement variables into new ones, making some of the original variables redundant. Specifically, we computed difference scores between behaviours towards novel versus familiar food (S 1.1.8), and speed of turning the wheel of the belt conveyor from frequency and duration measurements (S 1.1.1; S 1.1.2). Overall, $N = 146$ raw variables each representing a specific behaviour measured in a specific situation were included in the subsequent analyses (Table S2).

2.7.2.2 Contextualised behavioural measurements

Raw data were first aggregated—separately for each time block—across test sessions and observation days respectively to obtain time-relative probabilities of behaviour for each individual. For the behavioural tests, these scores reflect average frequencies or seconds (for frequency or duration behaviours respectively) per test time of one session (e.g., per 10 min in the Large cloth test, S 1.1.11). For the observations, these aggregate scores reflect frequencies or minutes per hour of observation time, and the average number of scans or zero-one records in which the particular behaviour was registered per day. These aggregate scores were then *z*-standardised within each time block. Consequently, the $N = 146$ derived variables reflect differential scores of behavioural probabilities (see Section 1.2; Uher, 2011a); they are labelled contextualised behavioural measurements.

2.7.2.3 Decontextualised and contextualised composite construct measures. From these standardised aggregate scores, three different composite measures were computed. First, we computed for all working constructs two kinds of decontextualised composite construct measures. On the one hand, we aggregated across situations all those construct-related variables that proved to be *test-retest reliable*. These composite measures are thus based on subsets of selected variables that, given their test-retest reliability, are highly informative about individual-specific behaviours. This selection may be useful for psychometric test development, but it inflates estimations of the true temporal stability of individual behaviour exhibited in this species. Therefore, we also computed decontextualised composite construct measures that are based on *all* relevant contextualised behavioural measurements regardless of their empirical test-retest reliability. They constitute conservative measures that reflect more accurately the true temporal stability of individual behaviours that are perceivable for human observers in this species. The differences in the results obtained with these two kinds of decontextualised composite construct measures are systematically explored in all analyses of this study. When computing the composite scores of Physical activity, we excluded five measurement variables (move on the spot in four tests and in the social observation; see Table S2) that represented the middle category of three mutually exclusive behaviour categories reflecting degrees of intensity, because a sum score over these categories would be uninformative about individual differences on the construct level. Thus, $N = 141$ variables were summarised in the composite measures that are based on all variables.

To explore issues of consistency across situations, we additionally computed contextualised composite construct measures by aggregating all (i.e., reliable and non-reliable) contextualised behavioural measurements that are related to a particular working construct within the given study situations.

2.7.3 Variable-oriented and individual-oriented analyses

We explored the data from two complementary perspectives (Stern, 1911): 1) From the *variable-oriented perspective*, we analysed the test-retest reliability of the relative order of individuals. These rank-order reliabilities characterise the measurement variables and composite construct measures on the sample level. 2) From the *individual-oriented perspective*, we analysed the test-retest reliability of each individual's profile across all measurement variables and all construct measures respectively. These profile reliabilities characterise the single individuals (Bergman & Trost, 2006; Furr & Funder, 2004).

2.7.3.1 Test-retest reliability

To explore whether we had in fact captured individual-specificity in behaviour (see 1., Section 1.2) we studied the differential behavioural probabilities (see Section 2.7.2) for their temporal stability between the two time blocks. We analysed test-retest-reliability using Pearson correlations within the 60-day study period from variable-oriented and individual-oriented perspectives on different levels of aggregation: on the level of all $N = 146$ contextualised behavioural measurements and on the level of the two kinds of decontextualised composite measures of the $N = 20$ generated constructs (see 2.7.2.3).

2.7.3.2 Internal consistency

Using variable-oriented analyses, we studied the internal consistency of the *decontextualised* composite measures (see Section 2.7.2.3) of all BR_xES-Approach-generated working constructs. Recall that these constructs are functionally defined rather than empirically derived as in lexical and judgment-based research (see Section 1.5). In these composite construct measures, the contextualised behavioural measurements were fixed, the individuals were random, and the composite scores were derived from aggregating over *z*-standardised measurements that were thus already adjusted for systematic mean-level differences. Therefore, we calculated intra-class correlations in terms of ICC(3,*k*) coefficients among all *k* contextualised behavioural measurements of which each construct was composed (Shrout & Fleiss, 1979). These coefficients reflect the internal reliability of the composite measures. But because the constructs were operationalised with different numbers of measurement variables, this could have affected their internal reliability. For this reason and for more direct comparisons with the test-retest correlations, we also computed the mean inter-measurement correlation for each composite measure. These two internal reliability coefficients were computed for both kinds of decontextualised composite construct measures and separately for each time block.

We also studied the internal consistency of the *contextualised* composite measures (see 2.7.2.3); this was possible for 42 of the 68 measures that were composed of at least 2 variables. We calculated intra-class correlations in terms of ICC(3,k) coefficients among all k contextualised behavioural measurements obtained for a given construct within a given situation (Shrout & Fleiss, 1979). We also computed their mean inter-measurement correlations. To further explore issues of consistency *within* situations in terms of individual response specificity (see above), we computed the test-retest reliability of individual profiles across these measurements between time blocks. We used Pearson correlation r from individual-oriented viewpoints for the contextualised construct measures that were composed of at least $k = 4$ variables. Because the measurements were z-standardised separately for each time block, their within-individual variation reflects the individual's configuration of deviations from average in each of these behavioural measurements.

Given that individual differences in behaviour are known to be only moderately consistent across situations and across different, functionally related behaviours within the same situation, we expected internal consistencies to be moderate rather than high as this is required in lexical and judgment-based research.

2.7.3.3 Cross-situational consistency

In variable-oriented analyses, we studied the consistency of all contextualised composite measures related to a particular working construct across the situations in which they were obtained. For each construct that could be studied in more than three situations, we computed their mean inter-measure correlations in the first time block. In individual-oriented analyses, we studied whether individual-specificity can be found on the level of individual profiles consisting of each individual's composite construct scores across different situations (see Section 1.4). Because these profiles are based on scores z-standardised within each situation (and separately for each time block), the remaining within-individual variation in these profiles reflects the individual's configuration of deviations from average in each of these situations. We computed test-retest reliability using Pearson correlation of all individuals' situation-behaviour profiles between time blocks. This was possible for 10 working constructs that could be measured in at least three different situations.

2.7.3.4 Associations with age, sex, and early life experiences

We explored whether the temporal reliability of individual-specific behaviours was associated with age. In individual-oriented analyses, we computed partial correlations between the individuals' age and their test-retest profile correlation scores on the level of both contextualised behavioural measurements and decontextualised composite construct measures controlling for the number of behavioural variables and construct measures respectively on which these profiles were based. In variable-oriented analyses, we explored associations of age, sex, and early rearing history with the individuals' scores on all working constructs using the two kinds of decontextualised composite construct measures. For more robust results, we used measures that were aggregated across both time blocks. We calculated Pearson correlations with age, and t-tests for independent samples with sex and rearing history respectively as grouping variable. We computed the magnitude of the differences we found with Cohen's effect size d on pooled standard deviations. These effect sizes can also be interpreted in terms of the percent of nonoverlap of the score distributions between the two contrasted groups (Cohen, 1988). Given the small sample sizes, we conducted post-hoc power analyses using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) to explore which of the results on age, sex, and rearing history are likely to replicate.

To compute mean correlations and to test correlations scores for differences between groups and for relations with other parameters, we always used Fisher's *r*-to-*Z* transformation.

3. Results

3.1 Variable-oriented test-retest reliability

On the level of contextualised behavioural measurements, the mean test-retest Pearson correlation for all $N = 146$ variables was $r_m = .60$ (range -.09 to .99). Of these, 86 variables (plus some variables that were redundant or not considered in the computation of composite measures, see Section 2.7.2) met the significance criterion ($p < .05$); the average test-retest correlation of this subset of variables was $r_m = .74$ (range .43 to .99). The test-retest correlations for all measurement variables and their significance levels are listed in Table S2.

On the level of decontextualised composite constructs measures derived from *exclusively temporally reliable* measurements, test-retest correlations were significant ($p < .05$) for 19 constructs showing a mean of $r_m = .76$ (range .47 to .91). One working construct, Competitiveness, had to be discarded. This construct and the constructs Distractibility and Persistency were the only ones for which we could obtain just one measurement. For Competitiveness, this measurement (latency to take a piece of food in a competitive dyadic situation; see S 1.1.3) turned out to be unreliable. On the level of decontextualised composite constructs measures derived from *all* contextualised behavioural measurements, test-retest correlations could be obtained for all 20 working constructs. Of these, 18 were significant showing a mean of $r_m = .66$ (range .22 to .91). As to be expected, *t*-tests for dependent samples showed that these differences between the two kinds of composite measures were significant, $t(18) = 3.172$; $p = .005$. Test-retest correlations for all decontextualised composite measures and their significance levels are provided in Table 3.

Table 3

3.2 Individual-oriented test-retest reliability

For the 26 individuals studied, the average test-retest Pearson correlation of the individual profiles across the $N = 86$ temporally reliable contextualised behavioural measurements was $r_m = .67$ (range -.42 to .98); that of the individual profiles across all $N = 146$ measurements, including some non-significantly reliable ones, was $r_m = .51$ (range -.65 to .94). In *t*-tests for dependent samples, these differences in profile stability were significantly different, $t(25) = 4.473$; $p = .000$. On the level of decontextualised working constructs, the average test-retest correlation of individual profiles across the 19 composite construct measures derived from exclusively reliable variables was $r_m = .78$ (range -.38 to 1.00); that across the 20 composite construct measures derived from on all variables was $r_m = .73$ (range -.38 to 1.00). Interestingly, *t*-tests for dependent samples showed that these differences in profile stability obtained with the two kinds of composite measures were nonsignificant, $t(25) = 0.658$; $p = .517$.

Note that a few individuals could be studied in only a subset of situations (see Section 2.3., Table S1); therefore, their individual profile correlations are based on fewer variables than those of the other individuals. This influenced the profile reliability in some, yet not all cases. All profiles that are non-significantly reliable are based on a much more limited database than the reliable ones. Thus, the lack of temporal reliability of these individual profiles may be accounted for by their insufficient database, rather than reflecting a characteristic of the individuals themselves. In addition, test-retest profile correlation scores for those individuals that we could study sufficiently and whose profiles were significantly test-retest reliable also varied thus indicating between-individual differences in within-individual stability over the medium-term time periods of our study. Individual test-retest profile correlation scores, their significance levels, and the number of contextualised behavioural measurements and decontextualised composite construct measures respectively on which the respective profiles were based are provided in Table 4.

Table 4

3.3 Internal consistency of functionally defined working constructs

In variable-oriented analyses on data from the first time block, the average internal consistency of the $N = 19$ *decontextualised* composite construct measures derived from exclusively temporally reliable measurements was *ICC(3,k)* = .604 (range -.131 to .859); that of the $N = 20$ constructs measures derived from all measurements was *ICC(3,k)* = .629 (range -.131 to .929). As to be expected, the internal consistency of constructs composed of more measurement variables tended to be higher than those of constructs composed of fewer variables. But although the number k of measurement variables varied considerably among constructs (range 2 to 15 and 2 to 28 respectively); this effect was nonsignificant for both kinds of construct measures ($r = .396$, $p = .084$ and $r = .346$, $p = .136$). Furthermore, construct measures composed of exclusively temporally reliable measurements were *not* more internally consistent than those derived from both temporally reliable and non-reliable measurements, $t(19) = 1.521$; $p = .145$. This is also reflected in the mean inter-measurement correlations that were identical for both kinds of composite measures, $r_m = .27$. The internal consistencies obtained in the second time block were virtually identical (*ICC(3,k)*= .622 and .657; $r_m = .27$ (range -.04 to .75) and $r_m = .28$ (range -.04 to .81); differences between both kinds of composite measures were likewise nonsignificant. This moderate internal consistency of functionally defined composite construct measures contrasts with their substantial variable-oriented test-retest correlations of $r = .76$ and $r = .66$ reported above. The internal consistencies for all composite construct measures are provided in Table 3; the single contextualised behavioural measurements assigned to each construct are listed in Table S2.

The average internal consistency of the $N = 42$ *contextualised* composite construct measures was *ICC(3,k)* = .415 (range -1.369 to .889). The number k of measurement variables obtained in the same situation varied among constructs ($m_k = 2.64$; $mdn_k = 2$; range 2 to 6); however, constructs composed of more variables were not more internally consistent than those composed of fewer variables ($r = .161$, $p = .307$). Mean inter-measurement correlations were on average $r_m = .24$ (range -.43 to .80). In contrast, the individuals' profiles across the k measurements assigned to the same construct within a given situation showed an average test-retest correlation of $r_m = .86$ (range .48 to .99); as demonstrated in the 6 contextualised construct measures that were composed of at least 4 measurements (Table 5). Thus individual-specificity emerged at the fine-grained level of functionally related behaviours obtained within the same situation.

Table 5

3.4 Cross-situational consistency and individual situation-behaviour profiles

The cross-situational consistency of contextualised composite measures that are related to the same working construct was low to moderate, except for Creativeness/inventiveness; the mean inter-measure correlation was $r_m = .28$ (range -.11 to .81). This contrasts with the mean test-retest correlation of the individuals' profiles across these measures of $r_m = .75$ (range .36 to .97). Only for Creativeness/inventiveness and Physical activity, cross-situational consistency was higher than the average situation-behaviour profile stability (see Table 6). The internal consistency of the contextualised composite construct measures was unrelated to their cross-situational consistency ($r = .134$, $p = .437$).

Hence, the individuals differed substantially in their probabilities to show particular, functionally related behaviours in different situations, but their distinct patterns of situation-to-situation variability were relatively stable between time blocks in most constructs. For example, in the first time block, the average variable-oriented intercorrelation among composite measures of Social orientation to humans obtained in four different situations—Human interaction test (S 1.1.6), Masked human test (S 1.1.7), Social observation (S 1.2.2), and Hidden food test (S 1.1.4.)— was only moderate, r_m = .43. But the individuals' unique Social-orientation-to-humans–profiles across these situations showed a high individual-oriented test-retest correlation of on average r_m = .80. Figure 1 illustrates this profile similarity between the two time blocks (indicated by broken and continuous lines) for 8 individuals.

Table 6

The graphs in Figure 1 furthermore illustrate that some individuals (e.g., Pedro and Robinia) were generally more socially oriented to humans than the average individual of this sample in most or all of these four situations. Yet there were also differences in the individuals' profile shapes. For example, Pedro deviated most strongly from average in the Human interaction test and in the Hidden food test, whereas Robinia's deviations from average showed an almost inversed pattern. Moreover, test-retest profile reliability was high for most individuals, yet only moderate for a few of them (e.g., Sandokan and Rucola). This suggests that individual differences in medium-term temporal stability of individual behaviour also emerge on the level of situation-behaviour profiles. Table 6 provides the mean cross-situational correlations and mean individual test-retest profile correlations between blocks for all constructs for which these analyses were possible.

Figure 1

3.5 Analyses of associations with age, sex, and rearing history

3.5.1 Age

In individual-oriented analyses, all correlations between the individuals' age and their test-retest profile correlation scores (based on exclusively test-retest reliable and on all variables respectively) were nonsignificant on the level of both contextualised behavioural measurements (r = .045, p = .831 and r =.051, p = .810) and decontextualised composite construct measures (r = .125, p = .552 and r = .228, p = .237) indicating no association between age and individual profile stability over medium-term periods of time. In variable-oriented analyses, we found significant negative correlations only on the working construct Impulsiveness (r = -.55, p = .034 and r = -.55, p= .035, N = 15), but no further associations. Post-hoc power analyses suggested that the probability of detecting a difference of this size was 60%; with 80% power, the estimated correlation is r = .65.

3.5.2 Sex

Between the sexes, individual score distributions differed significantly only on the constructs Aggressiveness, t(19) = 9.044; p = .007 and 9.035; p =.034, and Dominance t(19) = 11.746; p = .001 (for both kinds of composite measures). Males were more aggressive than females d = 1.20 (37.8% overlap in the score distributions and 72% achieved power of detecting such a difference); with 80% power, the achieved sensitivity is d = 1.33 (33.84% overlap). Males were also substantially more dominant than females d = 1.68 (25.06% overlap; 94% achieved power).

3.5.3 Rearing history

Between hand-reared and mother-reared monkeys, individual score distributions differed significantly only on the constructs Aggressiveness to humans, t(19) = 17.406; p =

.000; Distractibility, $t(13) = 8.876$; $p = .029$; and Gregariousness, $t(19) = 18.182$, $p = .020$ (all scores were identical for both measures). The analyses revealed that hand-reared individuals were less aggressive to humans, $d = -2.01$ (18.5 % overlap in score distributions; 98% achieved power in detecting such a difference). Hand-reared individuals were also more distractible by humans than mother-reared ones, $d = 1.37$ (32.74 % overlap; 64% achieved power); with 80% power, the achieved sensitivity is d = 1.65 (25.75% overlap). Furthermore, hand-reared individuals were less gregarious with their conspecifics than mother-reared individuals, $d = -1.05$ (61.2 % overlap; 60% achieved power); with 80% power, the achieved sensitivity is $d = -1.33$ (33.84% overlap).

Note that in all analyses results obtained from composite measures that were based on exclusively temporally reliable variables and those that were based on all variables were (almost) identical.

4. Discussion

We successfully applied to a New World primate species a novel research paradigm that provides an elaborated meta-theoretical and methodological framework for contextualised research on individual differences in behaviour (Uher, 2013). The present study evidenced fundamental differences in the internal consistency between the personality constructs derived from behavioural approaches and measurements and those derived from lexical approaches and judgment-based measurements. Moreover, our results broaden previous empirical findings (Byrne & Suomi, 1995, 1996, 1999, 2002) showing that capuchins have pronounced and stable individual differences in a broad range of behaviours, comparable to those described in great apes and humans. The relations of capuchins' individual differences with age, sex, and early life experiences provide new insights into comparative and developmental aspects of personality differences.

4.1 A new paradigm for research on individual differences

The new framework enabled objective quantifications of individual differences in the sense that it is explicitly known which particular behaviours, which particular situations, how many occasions and what overall time periods as well as which particular individuals are considered in these measurements. Based on the new concept of time-relative probabilities, this enabled ratio-scaled measurements and computations to estimate the consistency of observable individual behaviour over time, across situations, and among functionally related behaviours (Uher, in preparation, 2013). Our study on capuchin monkeys revealed strong empirical evidence for a broad range of individual-specific behaviours that were stable over intermediate-term time periods. Some individuals were more stable than others in their behavioural patterns both within particular situations and in their overall tendencies mirroring previous findings obtained in humans (Caspi & Roberts, 1999; Mischel et al., 2002) and great apes (Uher et al., 2008). Thus, individual differences at the level of within-individual stability are present also in capuchins, which are phylogenetically less closely related to humans than are great apes.

Our findings on only moderate cross-situational consistency yet temporally reliable individual situation-behaviour profiles show that the person-situation controversy, which has preoccupied human psychologists for decades (Fleeson, 2004; Funder, 2006; Hartshorne & May, 1928; Mischel, 1968, 1977; Mischel et al., 2002), covers issues relevant to other species as well. Stable patterns of individual-situation interaction refer to fundamental, and possibly universal, aspects of individual behaviour and development. Our findings open up promising starting points to further explore these patterns empirically in other species. Future studies may help to unravel underlying psychological processes that contribute to the individuals' unique ways of behaving in different kinds of situations by exploring the properties of situations to which particular groups of individuals respond in similar ways (Bandura, 1986; Lazarus, 1981; Mischel & Shoda, 1995; Rotter, 1981), as demonstrated successfully in children (e.g., Wright & Mischel, 1987; Wright & Zakriski, 2003).

4.1.1 Non-lexical emic construct generation using the BR_xES-Approach

The constructs that we generated for capuchins with the BR_xES-Approach covered individual differences that have already been described in previous studies on juvenile capuchins, specifically those labelled here as Arousability, Anxiousness (Byrne & Suomi, 1999), Aggressiveness, Dominance, (Byrne & Suomi, 2002), Creativeness/Inventiveness (Byrne & Suomi, 1996), Curiousness, Food orientation, Gregariousness, Social orientation to conspecifics, Playfulness, and Physical activity (Byrne & Suomi, 1995, 2002). Additionally, we obtained strong evidence for further individual differences in behaviours labelled here as Aggressiveness to humans, Distractibility, Impulsiveness, Persistency, Self-cleanliness, Social orientation to humans, Sexual activity, and Vigilance. These constructs describe ecologically relevant behaviours not yet identified with previous approaches that are less comprehensive and systematic.

The BR_xES-Approach has also been successfully employed for comprehensive taxonomic investigations of representations that human observers develop about individual behaviour, such as of great apes (Uher, 2011b, Uher & Asendorpf, 2008) and crab-eating macaques (Uher, Werner, & Gosselt, 2013). We are currently investigating how human observers perceive and mentally re-construe individual-specific behaviours in capuchins. Applied to humans, the BR_xES-Approach would enable systematic (and still missing) investigations of how people's representations and lexically derived models actually map on observable individual behaviour (Uher, submitted for publication b). This approach therefore constitutes a new and independent alternative to lexical approaches (Uher, 2008a, 2008b).

4.1.2 Moderate internal consistency of behaviourally derived personality constructs.

The new philosophy-of-science paradigm highlights essential differences between individual behaviour and people's representations of individual behaviour as studied with lexical approaches and judgment-based methods (see Section 1; Uher, 2013). Important for taxonomic research is the fact that individual differences in behaviour are not as consistent as are judgments of the respective representations. As expected, the internal consistency of constructs defined on the basis of the behaviours' functions was only moderate both within and across situations. Nevertheless, individual-specificity emerged in terms of distinct, yet stable individual behaviour profiles across situations and across different, functionally related behaviours within a situation. These findings parallel those established in human psychology (e.g., Asendorpf, 1988; Fahrenberg, 1986; Mischel, 1968, 1977). If such individual-specific patterns emerge, then internal consistency and cross-situational consistency cannot be high on the sample level (Uher, 2011a, 2011b). This has important implications for taxonomic research.

Thus, there is a fundamental difference between behavioural research on individual differences and the respective psychometric and lexical researches that largely rely on judgment-based methods and for which high internal consistency among measurement variables is mandatory for construct definition. This important difference may derive from the way people mentally represent observations. Human minds strive to detect recurrent patterns in their environment that could enable predictions of future events (Kelly, 1955). People may therefore develop somewhat coherent representations that are consistent with the logic of the human mind—yet not necessarily with what can actually be observed (Daston & Galison, 2007; Uher, 2013). The first reduction step of the BR_xES-Approach corresponds to these processes of mental abstraction and intuitive reconstruction, but is, in contrast to them, made explicit and based on scientific knowledge—and can therefore be traced to the specific behaviours and situations considered (see Section 1). The patterns of low to moderate consistency in individual behaviour do not become readily apparent in judgment-based and lexical research, which strongly capitalise on people's intuitive mental reconstructions of their observations. The impact of people's mental processes on the structure of judgment-based data on individual differences has received only little consideration so far (Uher, 2013).

Moreover, representations and their lexical encodings can be developed in redundant ways (Block, 2010). In behaviour, however, redundancies are rare possibly because of ecological and evolutionary constraints. This reduces the possibility to obtain behavioural data with internal consistency as high as that found for judgment-based data. Furthermore, the limited and enforced answer format of predefined scales may contribute to their higher internal consistency as compared to the ratio-scaled measurements that can be obtained from behaviour. The small sample size of the present study allowed only for analyses of the internal consistency of functionally defined constructs as generated by the BR_xES-Approach. Much larger samples are needed to empirically study also the second, statistical step of data reduction. It will be illuminating to compare the results obtained from both steps envisaged by the BR_xES-Approach with those obtained from subjecting behavioural measurements directly to statistical techniques.

4.1.3 Is selecting test-retest reliable behavioural measurements really necessary?

In this study, we contrasted the results obtained from considering exclusively test-retest reliable behavioural measurement variables with those obtained from considering all variables regardless their temporal reliability. The selection for reliable variables necessarily affected the variable-oriented temporal reliability of decontextualised constructs. But interestingly, there were no differences in results of stability of individual profiles across decontextualised constructs, of internal consistency of composite construct measures, or of relations to age, sex, and rearing history. This suggests that—provided a representative range of ecologically relevant contextualised behaviours is selected systematically as enabled by the BR_xES-Approach—unselected sets of measurements can reveal important findings on individual behaviour equally well as this is possible for subsets that are restricted to temporally reliable variables. Moreover, the results obtained from such representative sets reflect the individuals' true-to-life patterns of behaviour more accurately than those obtained from restricted sets of variables that may artificially inflate the true consistency of observable individual behaviour. This interesting finding merits further investigation in larger samples in the future.

4.2 Capuchin personality differences and relations with age, sex, and early life experience

The present findings on adult individuals allowed us to analyse relations of personality differences to age, sex, and early life experience. Interestingly, we did not find either age or sex to be associated with most of our 20 personality constructs. These results contradict the common assumption that individual differences in animals simply reflect species-typical differences in age and sex (cf. Sih, Bell, Johnson, & Ziemba, 2004; Yerkes, 1938). In capuchins, and likely in other species as well, classifying individuals by age and sex is clearly insufficient to describe the behavioural diversity among them.

4.2.1 Age differences

We found associations with age only for individual differences in Impulsiveness; older capuchins tended to be less impulsive than younger ones. Interestingly, in the construct Creativeness/inventiveness there were no age differences; this was probably due to the fact that all our subjects were adults. In fact, previous research comparing adults and juveniles found juveniles being more manipulative with food and objects than adults (Byrne & Suomi, 1996; Visalberghi, 1988). Our analyses also showed that age was unrelated to within-individual stability over medium-term period. That is, older individuals were not more stable in their individual behaviours than younger ones. Future research should explore factors in the individuals' developmental trajectories that contribute to individual differences in within-individual stability over time.

4.2.2 Sex differences

We did not find sex differences in the individual score distributions of our personality constructs, except for Aggressiveness and Dominance (for similar results in capuchins see Byrne & Suomi, 1996, 1999). Adult males were more aggressive and dominant than adult females; the overlap of the Dominance score distributions between sexes was only 25%.

Although males are generally more dominant than females in capuchins, given their tolerant social structure we did not expect effect sizes of this magnitude.

Our finding that capuchins, a group-living species with pronounced sex dimorphism in body size, do not exhibit pronounced sex differences in individual behaviour, as frequently hypothesised by evolutionary psychologists for our human ancestors (e.g., Buss, 1999), may shed light on the causes and consequences of sex and gender differences in humans. Future studies should investigate species in which females dominate males, but males are larger than females, such as bonobos. Contrasting species with different social and behavioural systems can provide new insights on the impact of biology and culture on human behaviour. Specifically, such studies may help to understand to which extent human cultures contribute to the development and magnitude of sex and gender differences of modern human societies, and the extent to which biological differences and differences in societal structure inevitably lead to sex and gender differences in various behaviours (Keller, 2007). Judgment-based studies frequently report associations with age and sex/gender in humans and in nonhuman primates (see above). But in lack of systematic contrasts to differences that can actually be observed, the extent to which judgments reflect socioculturally shared (thus anthropomorphic) attributions is still largely unknown. The behavioural research framework including the BR_xES-Approach of Uher's novel philosophy-of-science paradigm opens up unprecedented possibilities for systematic investigations. For example, in a 3-year study, systematic investigations of and comparisons between individual-specific behaviours of 104 crab-eating macaques and the representations that 99 human observers developed of them demonstrated influences of socioculturally shared age- and gender-related stereotypes about *human* individuals on the observers' judgments of these macaques—and thus anthropomorphic attribution biases (Uher et al., 2013). We are currently investigating with judgment-based methods whether human observers construct age and sex differences in capuchin monkeys, and to what extent their perceptions and representations match the individual differences in behaviour reported in this study.

4.2.3 Effects of early life experiences

Hand-reared adult capuchins were less aggressive to humans, less gregarious with their conspecifics and more distractible by humans than mother-reared capuchins. Our findings provide interesting empirical quantifications of the frequent observations that hand-rearing can have long-lasting effects. They indicate that pervasive effects are still present at 8-33 years of age, although hand-rearing lasted on average 7 months (range 4-12) and although after reintroduction all individuals had lived in groups for many years. Interestingly, we did not find effects on individual differences in Arousability and Anxiousness, as shown for peer-reared versus mother-reared rhesus monkeys in laboratory settings (Higley, Hasert, Suomi, & Linnoila, 1991).

Overall, our findings may reflect the profound differences between hand-rearing at home in stimulating and enriched environments with human caregivers (for further information see Visalberghi & Riviello, 1987) versus hand rearing in standard laboratory-conditions, where infants are raised without receiving a rich variety of social and physical stimuli. These differences may also reflect species differences in coping with early life stressors. These and further effects of early life experiences on individual development should be explored in future studies.

4.3 Summary and future directions

This study successfully demonstrated how personality differences can be studied with a behavioural framework in contextualised ways. This framework enabled systematic investigations of the stability of individual-specific behaviour over medium-term periods of time and its consistency within and across various ecologically relevant situations in capuchin monkeys. Our findings, similar to those reported for humans, great apes, and rhesus monkeys, indicate profound similarities in primates' individual behavioural development. Our study also provides substantial empirical evidence that the Behavioural

Repertoire x Environmental Situations Approach (Uher, 2008a, 2008b, 2011a, 2011b) constitutes a viable non-lexical approach for comprehensive taxonomic investigations of individual differences in behaviour in various species, unbiased by constructs of human personality. This approach constitutes a new, independent alternative to lexical approaches established in human personality psychology. It can therefore be used to systematically investigate how the internal consistencies of behaviourally derived and of lexically derived personality constructs differ from one another and how lexically derived personality constructs actually map on individual differences observable in human behaviour.

Finally, this study provides substantial empirical evidence that capuchin monkeys do exhibit pronounced individual differences, as previously reported in anecdotes and in a few empirical studies. Interestingly, the identified personality differences were largely unrelated to differences in age, sex, and early life experiences. Future studies should investigate larger samples to explore the underlying taxonomic and typological structure of the behavioural individual differences that we have identified, and to test the robustness of our results.

***

Acknowledgements

This research was funded by a research grant to Jana Uher (JU) from the Deutsche Forschungsgemeinschaft, DFG (German Research Foundation, Grant Nr. UH249/1-1), and the PNR-CNR Aging Program 2012-2014. We are grateful to the Editor and two anonymous reviewers for their constructive and helpful comments on an earlier draft. We thank our main experimenter Antonella Giordano (AG) for carrying out the behavioural tests and observations, and Eugenia Polizzi di Sorrentino (EPS) for helping with the observational ethogram and for providing some of the reliability observations. We also thank the Fondazione Bioparco for hosting the Primate Centre and Valentina Truppa, Francesco Natale, Massimiliano Bianchi, and Simone Catarinacci for their help while we carried out the study. Finally, we thank Federica Dal Pesco (FDP), Matilde Sauvaget (MS), and Christina Meier (CM) for behavioural coding from video and help with the data organisation, and Lisa Rodova for her help with the literature.

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour, 49,* 227-267.

Asendorpf, J. B. (1988). Individual response profiles in the behavioral assessment of personality. *European Journal of Personality, 2,* 155-167.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396-403.

Bandura, A. (1986). *Social foundations of thought and action.* Englewood Cliffs, NJ: Prentice Hall.

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.

Block, J. (2010). The five-factor framing of personality and beyond: Some ruminations. *Psychological Inquiry, 21,* 2–25.

Buss , D. M. (1999). *Evolutionary psychology*: *The new science of the mind*. Boston: Allyn & Bacon.

Byrne, G., & Suomi, S. J. (1995). Development of activity patterns, social interactions, and exploratory behavior in infant tufted capuchins (*Cebus apella*). *American Journal of Primatology, 35,* 255–270.

Byrne, G., & Suomi, S. J. (1996). Individual differences in object manipulation in a colony of tufted capuchins. *Journal of Human Evolution, 31,* 259–267.

Byrne, G. & Suomi, S. J. (1998). Relationship of early infant state measures to behavior over the first year of life in the tufted capuchin monkey (*Cebus apella*). *American Journal of Primatology, 44,* 43-56.

Byrne, G., & Suomi, S. J. (1999). Social separation in infant *Cebus apella*: Patterns of behavioral and cortisol response. *International Journal of Developmental Neuroscience, 17,* 265–274.

Byrne, G., & Suomi, S. J. (2002). Cortisol reactivity and its relation to homecage behavior and personality ratings in tufted capuchin (*Cebus apella*) juveniles from birth to six years of age. *Psychoneuroendocrinology, 27,* 139-154.

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* (pp. 300-326, 2nd ed). New York, NY: Guilford Press.

Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. New York, NY: World Book.

Cattell, R. B. & Warburton, F. W. (1967). *Objective Personality and Motivation Tests.* Urbana: University of Illinois Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Lawrence Erlbaum.

Daston, L. & Galison, P. (2007). *Objectivity.* New York, NY: Zone Books.

Fahrenberg, J. (1986). Psychophysiological individuality: A pattern analytic approach to personality research and psychosomatic medicine. *Advances in Behaviour Research and Therapy, 8*, 43-100.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. (2004). *The Complete Capuchin*. Cambridge University Press.

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.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations and behaviors. *Journal of Research in Personality, 40,* 21-34.

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality, 23,* 369-401.

Furr, R. M. & Funder, D. C. (2004). Situational similarity and behavioural consistency: subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality, 38,* 421-447.

Guidelines for the Treatment of Animals in Behavioural Research and Teaching (2006). Issued by Elsevier and incorporated by reference into the Animal Behaviour Guide for Authors. *Animal Behaviour, 71,* 245-253.

Gunthert, K. C., Conner, T. S., Armeli, S., Tennen, H., Covault, J., & Kranzler, H. R. (2007). Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: A daily process approach to gene-environment interaction. *Psychosomatic Medicine, 69,* 762-768. doi: 10.1097/PSY.0b013e318157ad42

Hartshorne, H. & May, M. A. (1928). *Studies in the nature of character. Vol 1. Studies in deceit*. New York, NY: Macmillen.

Higley, J. D., Hasert, M. L., Suomi, S. J., & Linnoila, M. (1991). A new nonhuman primate model of alcohol abuse: Effects of early experience, personality, and stress on alcohol consumption. *Proceedings of the National Academy of Science, 88,* 7261-7265.

Izawa, K. (1980). Social behavior of the wild black-capped capuchin (*Cebus apella*). *Primates, 21,* 443–467.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2,* 171-203.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Erlbaum.

Kelly, G. A. (1955). *A theory of personality: The psychology of personal constructs.* New York: Norton.

King, J. E., Weiss, A., & Sisco, M. M. (2008). Aping humans: Age and sex effects in chimpanzee (*Pan troglodytes*) and human (*Homo sapiens*) personality. *Journal of Comparative Psychology, 122,* 418-27.

Kubinger, K. D. (2009). The technique of objective personality-tests sensu R. B. Cattell nowadays: The Viennese pool of computerized tests aimed at experiment-based assessment of behavior. *Acta Psychologica Sinica, 41,* 1024-1036.

Kuhar, C. W., Stoinski, T. S., Lukas, K. E., & Maple, T. L. (2006). Gorilla behavior index revisited: age, housing and behavior. *Applied Animal Behavior Science, 96,* 315-326.

Lazarus, A. A. (1981). *The practice of multimodal therapy*. New York, NY: McGraw-Hill.

Mangold, P. (2010). *Interact User Guide*, V. 9.0. ff. Arnstorf: Mangold International.

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality, 60,* 329-361.

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

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857–870.

Millikan, R. (1993). *White Queen Psychology and Other Essays For Alice*. Cambridge, MA: Bradford Books/ MIT 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., & 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., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science, 11,* 50-54.

Rotter, J. B. (1981). The psychological situation in social learning theory. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective*. Hillsdale, NJ: Lawrence Erlbaum.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioural syndromes: an integrative overview. *The Quarterly Review of Biology, 79*, 241-277.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big-Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100,* 330-348.

Stern, W. (1911). *Die Differentielle Psychologie in ihren methodischen Grundlagen [Differential Psychology in its methodological foundations]*. Leipzig: 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. (2005). How gene-environment interactions shape the development of impulsive aggression in rhesus monkeys. In D. M. Stoff & E. J. Sussman (Eds.) *Developmental Psychobiology of Aggression*, (pp. 252–268). New York: Cambridge University Press.

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.

Uher, J. (2013). Personality psychology: Lexical approaches, assessment methods, and trait concepts reveal only half of the story—Why it is time for a paradigm shift. *Integrative Psychological and Behavioral Science, 47*, 1-55. http://dx.doi.org/10.1007/s12124-013-9230-6

Uher, J. (in preparation). Meta-theoretical foundations of objectivity versus subjectivity in quantifications of behaviour and personality.

Uher, J. (submitted for publication-a). An integrative meta-theoretical framework for research on individual behaviour in context—situations, populations, species.

Uher, J. (submitted for publication-b). Methodological approaches to taxonomies of human personality differences: The Behavioural Repertoire x Environmental Situations Approach - A non-lexical alternative.

Uher J. (2008a). Comparative personality research: Methodological approaches. *European Journal of Personality, 22,* 427-455.

Uher J. (2008b). Three methodological core issues of comparative personality research. *European Journal of Personality, 22,* 475-496.

Uher, J. (2011a). Individual behavioral phenotypes: An integrative meta-theoretical framework. Why 'behavioral syndromes' are not analogues of 'personality'. *Developmental Psychobiology, 53*, 521–548.

Uher, J. (2011b). 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*. New York, NY: Springer.

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.

Uher, J., Werner, C. S., & Gosselt, K. (2013). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality*. Available online: http://dx.doi.org/10.1016/j.jrp.2013.03.006

Visalberghi, E. (1988). Responsiveness to objects in two social groups of tufted capuchin monkeys *(Cebus apella). American Journal of Primatology,* 15, 349-360.

Visalberghi, E. & Fragaszy, D. (2012). What is challenging about tool use? The Capuchin's Perspective. In T. R. Zentall & E. A. Wasserman (Eds). *The Oxford Handbook of Comparative Cognition*, Chapter 40. Oxford University Press.

Visalberghi, E. & Riviello, M.C. (1987). The integration into a social group of a hand-reared capuchin monkey*. International Zoo Yearbook*, *26,* 232-236.

Wright, J. C. & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology, 55,* 454-469.

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.

Yerkes, R. M. (1939). The life history and personality of the chimpanzee. *American Naturalist, 73,* 97-112.

Glossary of terms from the meta-theoretical and methodological framework for behavioural research on individual differences relevant to the present analyses

Contextualised behavioural measurements: Measurement variables reflecting differential scores of individual probabilities to display specific behaviours in specific situations. They were derived from aggregating the *raw measurements* across test sessions and observation days per time block respectively to obtain time-relative probabilities for each individual. These individual aggregate scores were then *z*-standardised across all individuals within each time block to obtain scores reflecting differential scores in these probabilities. If temporal reliability can be shown for these measurements, then they reflect *individual-specific patterns of behaviour*.

Contextualised composite construct measures: Composite variables reflecting differential scores of individual behavioural probabilities on the level of *working constructs* measured in specific situations. They were derived from aggregating within each studied situation all *contextualised behavioural measurements* that were assigned to a particular construct based on the behaviours' scientifically established functional relatedness. Composite measures for which temporal reliability can be evidenced reflect summary scores of *individual-specific patterns of behaviour*.

Decontextualised composite construct measures: Composite variables reflecting individual differential probabilities on the level of the decontextualised *working constructs*. They were derived from aggregating across all studied situations all *contextualised behavioural measurements* that were assigned to a particular construct based on the behaviours' scientifically established functional relatedness. Two kinds of composite measures were computed: One based on exclusively test-retest reliable measurements, and another one based on all measurements regardless of their test-retest reliability. The former measure inflates the true temporal stability of individual behaviour that can be observed in this species; the latter reflects it more accurately (i.e., true-to-life).

Individual-oriented analyses: Used to explore on different levels of abstraction the single individuals' distinct configurations of deviations from average across various variables. These configurations can be illustrated as profiles across variables. The profiles and their test-retest reliabilities characterise the single individuals.

Individual-specific patterns of behaviour: Because behaviours are highly fluctuating, individuals can be characterised only in behavioural probabilities that differ among individuals in relatively stable ways. To reflect individual-specificity, differential patterns of behavioural probabilities must be stable across time periods longer than those in which the probabilities were first ascertained and in ways considered to be meaningful (e.g., defined via test-retest correlation scores). Accordingly, measurements reflecting individual-specificity are derived from *1)* aggregation over repeated measurement occasions; *2)* standardisation of these aggregate scores; and *3)* evidence of the temporal reliability of the standardised aggregate scores.

Individual-specific situation-behavior profiles: These profiles illustrate *individual-specific patterns* of situation-to-situation variability in behaviour. They reflect the individual's unique configuration of deviations from average in particular behaviours across various situations. These profiles were based on *contextualised composite construct measures* that were *z*-standardised within each situation and separately for each time block. Their test-retest profile reliability between time blocks was studied in *individual-oriented analyses* and contrasted to cross-situational consistency in the first time block as studied with *variable-oriented analyses*.

Non-lexical emic personality constructs: Constructs of individual-specificity that are derived not from within the lexica of everyday language terms of a given population as in lexical research, but from within the scientifically described behavioural-ecological system of a given population using the BR_xES-Approach.

Raw measurements: Behavioural variables of the individuals' raw scores obtained in tests and observations; each variable reflects a specific behaviour measured in a specific situation. Note that these variables are of different type of measure (e.g., frequencies, durations; see Table S2). For subsequent analyses, some of these variables were reversed and/or transformed considering their meaning for a particular construct. Inter-coder and inter-observer reliability was analysed in these raw measurements.

Variable-oriented analyses: Used to explore the individuals' rank-order on various variables for their test-retest reliability, internal consistency, or cross-situational consistency. The results obtained characterise the sample, not the single individuals.

Working constructs: Constructs generated with the BR_xES-Approach based on the hypothetical assumption that *individual-specific patterns* can be found in behaviours of particular categories displayed in situations of particular categories. These hypothetical assumptions must be substantiated empirically; otherwise the particular working construct has to be discarded. In contrast to lexically derived constructs, BR_xES-Approach-generated constructs are based on the scientifically known functions of the behaviours described, not on redundancy in the information encoded in the measurement variables. In accordance with well-established findings on only moderate consistency of individual differences in behaviour across different situations (i.e., cross-situational consistency) and across different, functionally related behaviours within the same situation (i.e., individual response specificity), these constructs' internal consistency is likewise often only moderate. This differs fundamentally to lexical and judgment-based research on individual differences.

Tables

Table 1 Behavioural tests and observations, their situational description, and the personality constructs measured.

Behavioural tests and observations	Situational description	Personality constructs measured
Conveyor belt test (S 1.1.1)	Food of different desirability and quantity was placed successively on a small conveyor belt fixed to the cage. By turning a wheel, the monkey could move the conveyor belt thereby transporting the food items into his/her cage.	Food orientation
Conveyor belt disconnected test (S 1.1.2)	The same conveyor belt was baited with highly preferred food. It still looked the same, but was disconnected by an internal mechanism. The monkey could still turn the wheel, yet with no effect on the conveyor belt.	Arousability Impulsiveness
Food competition test (S 1.1.3)	The experimenter offered one piece of preferred food to two individuals kept in the same cage, when they were both at approximately the same distance from it.	Aggressiveness Competitiveness Dominance
Hidden food test (S 1.1.4)	The monkey entered the test room in which small food pieces were hidden on the cage elements or stuck with honey to the variegated walls.	Arousability Anxiousness Physical activity Social orientation to conspecifics Social orientation to humans Vigilance
Yoghurt grid test (S 1.1.5)	The monkey could reach through the mesh for plain yoghurt spread on a platform outside the cage, while the experimenter continuously produced noise by knocking with a plastic tube on the metal cage frame located at the other end of the cage.	Distractibility
Human interaction test (S 1.1.6)	The experimenter sat in front of the monkey's cage and encouraged him/her to approach and to play without offering any food. Then she fed him/her for a limited time before she put the food out of sight and encouraged the monkey again to approach and to play.	Arousability Aggressiveness to humans Social orientation to humans
Masked human test (S 1.1.7)	Disguised with a Venetian mask, a wig, a long dress and boots, the experimenter entered the test room silently and offered continuously food in gloved hands. She thereby stuck the right's glove stiffed fingers through the mesh into the monkey's cage to enable direct contact.	Aggressiveness to humans Anxiousness Arousability Social orientation to humans
Novel food test (S 1.1.8)	The monkey received four pieces of normal food and four novel food items he/she never had before in alternating order, in total eight food items.	Curiousness Food orientation

Multiple objects test (S 1.1.9)	The monkey found six small different objects familiar and unfamiliar in the middle of his/her cage.	Aggressiveness Anxiousness Arousability Creativeness/Inventiveness Curiousness Physical activity
Tunnel basket test (S 1.1.10)	The monkey encountered a large, open-worked laundry basket of PVC placed in the middle of the cage with its larger opening towards the door through which he/she entered. The basket was open at both ends, but at the smaller end a tubular dark cloth was fixed that looked like a scoop net, yet that was actually open so that the monkey could go through it.	Aggressiveness Anxiousness Arousability Creativeness/Inventiveness Curiousness Physical activity Social orientation to humans
Large cloth test (S 1.1.11)	The monkey encountered a large bed sheet that was hung up transversally in the test cage and over the wooden perch in the rear end of the cage.	Aggressiveness Anxiousness Arousability Curiousness Creativeness/Inventiveness Physical activity Social orientation to humans
Furry animal test (S 1.1.12)	When the monkey was inside the test cage, a small soft toy was placed in front of the cage facing away from the monkey. It was attached to a disc that the experimenter rotated from a 2 m distance after one minute so that the soft toy then faced the monkey for a further minute. The eyes of the soft toy were covered with crepe tape in all trials to reduce the degree of threat that soft toys constitute for capuchins.	Aggressiveness Anxiousness Arousability
Blocked food tube test (S 1.1.13)	Half items of highly preferred food (Cheerios) were dropped one by one into a transparent tube that was fixed at a 45° angle to the monkeys cage. The tube had a thin slot in which the experimenter inserted transparent plastic slides. In two first trials, she inserted a hollowed slide allowing the food to pass and to fall inside the monkey's cage. In a third trial, she inserted a solid slide above that the food piled up in full view, yet out of reach of the monkey.	Arousability Food orientation Impulsiveness
Foraging box test (S 1.1.14)	The monkey encountered a box inside the cage on which he/she could sit. The box was filled with wood shavings, in which 3 pumpkin seeds of the same colour were hidden. Through an opening at the top partially covered by a transparent Plexiglas panel, the monkey could peer and reach into the box with one arm to search in the substrate.	Persistency Vigilance

Sudden noise test (S 1.1.15)	A foreign news program was suddenly played back to the monkey in moderate volume independent of the experimenter's activities inside the test room twice for 10 sec with a break of 20 sec.	Aggressiveness Anxiousness Arousability Vigilance
Prefeeding observation (S 1.2.1)	The monkeys could hear and see the keepers preparing the food for their daily main feeding in the nearby kitchen and distributing it in the enclosures of the neighbouring groups before they finally received their own food.	Arousability Food orientation Social orientation to conspecifics
Social observations (S 1.2.2)	The monkeys were all together in their groups in their naturally designed outdoor enclosures; access to the indoor enclosures was blocked for the time of the observations.	Aggressiveness Aggressiveness to humans Anxiousness Arousability Dominance Food orientation Gregariousness Physical activity Playfulness Self-cleanliness Sexual activity Social orientation to conspecifics Social orientation to humans

Note. All behavioural test and observations are described in more details in the Supplemental Material S1.

Table 2 Analytical perspectives, levels of aggregation, and the interrelations within and among the elements of the four sets of behaviours *B,* situations *S,* times *T,* and individuals *I* explored in the respective analyses

Analytical Perspective[1]	Variables on different levels of aggregation[2]	Analyses[1]	Elements and their interrelations[3] considered in the sets of Behaviours *B*	Situations *S*	Times *T*	Individuals *I*
Variable-oriented	Contextualised behavioural measurements	Test-retest reliability	Within elements	Within elements	Between aggregates (i.e., blocks) of elements (i.e., occasions)	Over elements
	Contextualised composite construct measures	Internal consistency	Between elements of functionally related subsets	Within elements	Within aggregates of elements	Over elements
		Cross-situational consistency	Between aggregates of functionally related subsets	Between subsets of elements	Within aggregates of elements	Over elements
	Decontextualised composite construct measures	Internal consistency	Between elements of functionally related subsets	Between subsets of elements	Within aggregates of elements	Over elements
		Test-retest reliability	Within aggregates of functionally related subsets	Within aggregates of subsets of elements	Between aggregates of elements	Over elements
Individual-oriented	Contextualised behavioural measurements	Test-retest profile reliability	Over elements	Over elements	Between aggregates of elements	Within elements
	Contextualised composite construct measures	Individual behaviour (response) profile reliability	Over elements of functionally related subsets	Within elements	Between aggregates of elements	Within elements
		Individual situation-behaviour profile reliability	Over aggregates of functionally related subsets	Over subsets of elements	Between aggregates of elements	Within elements
	Decontextualised composite construct measures	Test-retest profile reliability	Over aggregates of functionally related subsets	Over aggregates of subsets of elements	Between aggregates of elements	Within elements

[1] see 2.7.3; [2] see 2.7.2.; [3] Behaviourally derived constructs of personality always refer to interrelations within and among these four essential sets of elements. See 1., Section *Objective quantifications and comparisons of individual behaviour*

Table 3 Variable-oriented internal consistency and temporal reliability of decontextualised working constructs

Working constructs	Decontextualised composite construct measures[1] based on										
	Exclusively test-retest reliable contextualised behavioural measurements					All contextualised behavioural measurements					Individuals
	Internal consistency t_1			Test-retest reliability		Internal consistency t_1			Test-retest reliability		N
	ICC (3,k)	N_{cbm}	r_m	r	p	ICC (3,k)	N_{cbm}	r_m	r	p	
Aggressiveness	-.004	4	.23	**.63**	.002	.490	7	.11	**.50**	.020	21
Aggressiveness to humans	-.131	4	.00	**.68**	.000	-.131	4	-.04	**.68**	.000	21
Arousability	.508	7	.04	**.80**	.000	.639	23	.07	**.47**	.010	26
Anxiousness	.741	15	.09	**.86**	.000	.784	28	.11	**.51**	.017	21
Competitiveness	-	0	-	--		-	1	-	.33	.250	14
Creativeness/inventiveness	.929	3	.20	**.70**	.000	.852	4	.59	**.70**	.000	15
Curiousness	.683	9	.27	**.90**	.000	.627	13	.11	**.91**	.000	15
Distractibility	-	1	-	**.64**	.010	-	1	-	**.64**	.010	14
Dominance[2]	.705	4	.27	**.79**	.000	.705	4	.29	**.79**	.000	21
Food orientation	.497	10	.03	**.61**	.000	.398	12	.05	**.60**	.000	26
Gregariousness	.422	2	.78	**.85**	.000	.422	2	.27	**.85**	.000	20
Impulsiveness	.493	3	.41	**.74**	.000	.768	5	.38	**.60**	.020	14
Physical activity	.655	6	.12	**.86**	.000	.712	10	.20	**.67**	.001	21
Persistency	-	1	-	**.70**	.010	-	1	-	**.70**	.010	14
Playfulness	.859	2	.96	**.82**	.000	.859	2	.75	**.82**	.000	21
Self-cleanliness	-	1	-	**.47**	.030	.475	2	.31	.23	.310	21
Social orientation to conspecifics	.772	7	.35	**.64**	.000	.819	10	.31	**.64**	.000	25
Social orientation to humans	.497	3	.01	**.91**	.000	.449	5	.14	**.86**	.000	21
Sexual activity	-	1	-	**.83**	.000	.736	3	.48	.22	.330	21
Vigilance	.571	2	.34	**.81**	.000	.373	4	.13	**.65**	.010	15
Means	.604		.27	**.76**		.629		.27	**.66**		

Note. Internal consistency computed as intra-class correlation $ICC(3,k)$ and as mean inter-measurement correlation r_m based on N_{cbm} contextualised behavioural measurements per construct. Test-retest reliability between time blocks computed as Pearson-correlations r; significant correlations are bold. [1]For the assignment of contextualised behavioural measurements to particular working constructs and their test-retest reliability see Table S2 in the Supplemental Material. [2] We studied Dominance as personality construct, that is, as individual-specific patterns in dominant-submissive behaviours in which all individuals can be quantified and compared, not as social status that refers to only a few individuals (such as the alpha male, see Table S1).

Table 4 Individual-oriented temporal reliability on the level of contextualised behavioural measurements and decontextualised working constructs

Individuals	Level of contextualised behavioural measurements based on						Level of decontextualised composite construct measures based on					
	Exclusively test-retest reliable contextualised behavioural measurements			All contextualised behavioural measurements			Exclusively test-retest reliable contextualised behavioural measurements			All contextualised behavioural measurements		
	r	*p*	*N*	*r*	*p*	*N*	*r*	*p*	*N*	*r*	*p*	*N*
Brm	**.64**	.001	24	**.58**	.000	36	**.81**	.001	12	**.65**	.016	13
Cam	.67	.212	5	.67	.099	7	.98	.141	3	.98	.136	3
Cog	**.81**	.000	85	**.67**	.000	141	**.92**	.000	19	**.88**	.000	20
Cta	-.42	.482	5	-.65	.114	7	-.38	.752	3	.42	.726	3
Gal	**.95**	.013	5	**.94**	.002	7	.99	.083	3	**1.00**	.036	3
Hod	**.48**	.000	85	**.41**	.000	141	**.86**	.000	19	.13	.578	20
Pat	**.72**	.000	85	**.49**	.000	141	**.53**	.020	19	**.49**	.030	20
Pch	**.50**	.013	24	**.50**	.002	36	.41	.184	12	**.58**	.039	13
Pdr	**.77**	.000	85	**.59**	.000	141	**.86**	.000	19	**.84**	.000	20
Peo	**.76**	.000	24	**.64**	.000	36	**.80**	.002	12	**.77**	.002	13
Pip	**.73**	.000	69	**.58**	.000	77	**.23**	.023	17	.26	.318	17
Pja	**.61**	.001	24	**.47**	.004	36	.03	.930	12	-.06	.835	13
Pka	**.98**	.004	5	.66	.105	7	**1.00**	.029	3	.44	.709	3
Pnl	**.69**	.000	85	**.57**	.000	141	**.68**	.002	18	**.74**	.000	19
Ppe	**.51**	.000	85	**.32**	.000	141	**.71**	.001	19	**.46**	.043	20
Qui	**.54**	.000	85	**.35**	.000	141	**.72**	.001	19	.33	.163	20
Ram	.46	.439	7	.61	.145	7	.64	.562	3	**1.00**	.037	3
Rbn	**.70**	.000	85	**.62**	.000	141	**.87**	.000	19	**.83**	.000	20
Rla	**.63**	.000	85	**.53**	.000	141	**.75**	.000	19	**.73**	.000	20
Rot	**.81**	.000	85	**.66**	.000	141	**.74**	.000	19	**.48**	.034	20
Rta	**.66**	.000	85	**.36**	.000	141	**.76**	.000	19	**.48**	.034	20
Ruc	**.54**	.000	85	**.39**	.000	141	**.49**	.034	19	**.70**	.001	20
San	**.78**	.000	85	**.66**	.000	141	**.91**	.000	19	**.83**	.000	20
Sar	.22	.956	24	-.01	.956	36	.33	.300	12	-.18	.566	13
Vig	.34	.060	24	.32	.060	36	.52	.082	12	.36	.221	13
Vis	**.65**	.000	85	**.43**	.000	141	**.59**	.008	19	**.59**	.007	20
Means	**.76**			**.51**			**.78**			**.73**		

Note. Pearson-correlations *r* between both time blocks within a 60-day period. Individual names are abbreviated according to Table S1. The different number of contextualised behavioural variables and composite construct variables depends on the comprehensiveness of the collected data that differed among individuals (see 2.5, Table S1). Significant correlations are in bold.

Table 5 Internal consistency of functionally related behavioural measurements within situations and temporal reliability of individual-specific profiles across these measurements

Construct	Situation	Internal consistency among *k* behavioural measurements in time block t_1		Test-retest reliability of individual profiles across the *k* behavioural measurements		
		k	r_m	r_{tt} *mean*	r_{tt} *min*	r_{tt} *max*
Anxiousness	Furry animal test	4	-.09	.83	-.40	.99
	Masked human test	5	.32	.86	.00	.99
	Multiple objects test	4	.23	.48	-.90	.99
	Tunnel basket test	6	.26	.55	-.95	.99
Curiousness	Novel food test	6	.26	.59	-.97	.99
Impulsiveness	Conveyor belt disconnected test	4	.29	.99	-.91	.96
Mean			.21	.86		

Note. Internal consistency computed with inter-measurement correlation r_m based on $N = 15$ individuals in all cases; test-retest reliability between time blocks computed with Pearson correlation *r*. The profiles are based on data *z*-standardised separately for each time block; thus the remaining within-individual variation reflects the individual's configuration of deviations from average in the measurements studied.

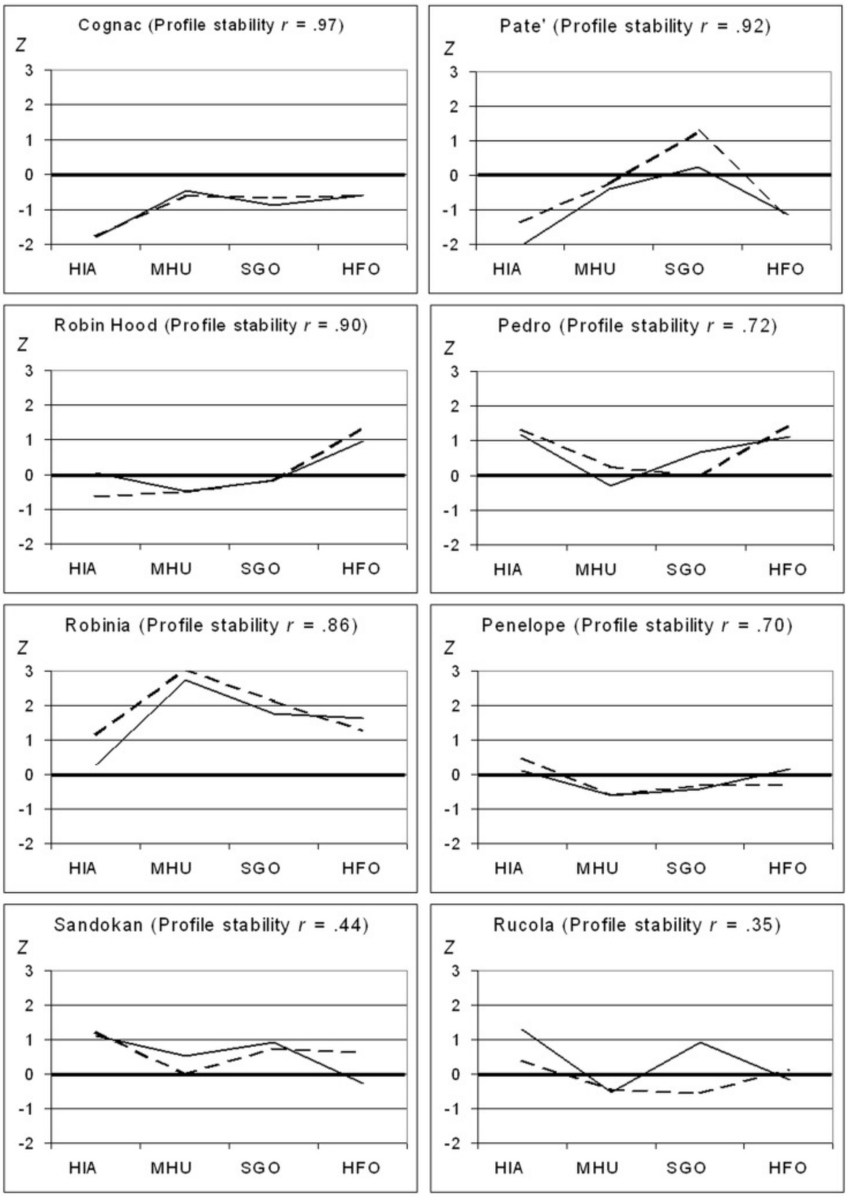
Table 6 Cross-situational consistency and temporal reliability of individual-specific situation-behaviour profiles

Construct	Situations	Mean cross-situational correlation in time block t_1	Mean individual test-retest profile correlation	*N*
Aggressiveness	Furry animal test, Social observation, Tunnel basket test	.04	.97	14
Aggressiveness to humans	Human interaction test, Masked human test, Social observation	-.11	.52	14
Arousability	Blocked food tube test, Conveyor belt disconnected test, Hidden food test, Human interaction test, Prefeeding observation, Social observation	.10	.60	15
Anxiousness	Furry animal test, Hidden food test, Masked human test, Multiple object test, Social observation, Sudden noise test, Tunnel basket test	.17	.54	15
Curiousness	Large cloth test, Multiple objects test, Novel food test, Tunnel basket test	.29	.83	14
Creativeness/ inventiveness	Large cloth test, multiple objects test, tunnel basket test	.81	.41	14
Food orientation	Blocked food tube test, Conveyor belt test, Hidden food test, Novel food test, Prefeeding observation, Social observation	.12	.81	14
Physical activity	Hidden food test, Large cloth test, Multiple objects test, Tunnel basket test, Social observation	.40	.36	15
Social orientation to conspecifics	Hidden food test, Prefeeding observation, Social observation,	.25	.92	13
Social orientation to humans	Hidden food test, Human interaction test, Masked human test, Social observation	.43	.80	15
Mean		.28	.75	

Note. Variable-oriented cross-situational consistency computed as mean intercorrelations of contextualised composite construct measures across the situations in which they were obtained; individual-oriented test-retest reliability of individual situation-behaviour profiles computed with Pearson correlation *r*.

Figures

Figure 1 Individual-specific situation-behaviour profiles in Social orientation to humans across four different situations



Note. Contextualised composite measures of the construct Social orientation to humans in four different situations: HIA = Human interaction test (S 1.1.6), MHU = Masked human test (S 1.1.7), SGO = Social group observation (S 1.3.2), and HFO = Hidden food test (S 1.1.4). The profiles are based on data *z*-standardised within each situation separately for the two blocks of time; thus the remaining within-individual variation reflects the individual's configuration of deviations from average in the situations studied. — Continuous lines indicate scores from study block t_1, - - - broken lines indicate scores from study block t_2. The average cross-situational correlation of these scores in the sample was $r = .43$ in the first block of time, whereas individual test-retest profile correlations averaged $r = .80$ ($N = 15$).