ANIMAL BEHAVIOUR, 2008, **75**, 99–112
doi:10.1016/j.anbehav.2007.04.018

Personality in the behaviour of great apes: temporal stability, cross-situational consistency and coherence in response

JANA UHER*†, JENS B. ASENDORPF* & JOSEP CALL†
*Humboldt-University Berlin, Institute for Psychology, Germany
†Max Planck Institute for Evolutionary Anthropology, Germany

*(Received 22 June 2006; initial acceptance 5 March 2007;
final acceptance 17 April 2007; published online 1 October 2007; MS. number: 9010R2)*

Using a multidisciplinary approach, the present study complements ethological behaviour measurements with basic theoretical concepts, methods and approaches of the personality psychological trait paradigm. Its adoptability and usefulness for animal studies are tested exemplarily on a sample of 20 zoo-housed great apes (five of each of the following species): bonobos, *Pan paniscus*; chimpanzees, *Pan troglodytes verus*; gorillas, *Gorilla gorilla gorilla*; and orang-utans, *Pongo pygmaeus abelii*. Data on 76 single trait-relevant behaviours were recorded in a series of 14 laboratory-based situations and in two different group situations. Data collection was repeated completely after a break of 2 weeks within a 50-day period. All behaviour records were sufficiently reliable. Individual- and variable-oriented analyses showed high/substantial temporal stability on different levels of aggregation. Distinctive and stable individual situational and response profiles clarified the importance of situations and of multiple trait-relevant behaviours. The present study calls for a closer collaboration between behavioural biologists and personality psychologists to tap the full potential of animal personality research.

Keywords: behaviour prediction; bonobo; bottom-up approach; chimpanzee; gorilla; individual differences; orang-utan; personality; traits

Measuring temporally stable variants of normal behaviour (i.e. personality: Funder 2004; Nettle 2006) is becoming increasingly a matter of interest to scientific studies in domestic and wild animals. Given the dimensionality of personality variation, traits are discussed within evolutionary frameworks as ecologically adaptive trade-offs of different fitness costs and benefits (Tooby & Cosmides 1990; Buss 1991; Dall et al. 2004). Therefore, personality investigations in animals can broaden our understanding of the evolutionary origins of interindividual variation in behaviour in human and nonhuman animals.

The study of stable individual differences in the empirical structure of quantitative behavioural measures entails a number of methodological difficulties that are caused by fluctuations in manifest behaviour. As already stated by Donald O. Hebb in the 1940s, finding 'meaningful order and consistency' in 'endless series of specific acts' (Hebb 1946, page 88, 1949) that are stable enough to permit predictions of the individual's future behaviour is rather challenging. Therefore, ethological methods of measuring behaviour have to be complemented with theories, methods and statistics specifically designed for the analysis of stable individual differences developed in human personality psychology (Gosling 2001; Gosling et al. 2003; Sih et al. 2004; Nettle 2006). The present paper shows how methods and theories from neighbouring disciplines can complement each other successfully in a study on personality differences in great apes.

Among many rather different theoretical approaches to personality, the psychological trait paradigm is the most appropriate for research on animal personality. The trait

Correspondence: J. Uher, Lehrstuhl für Persönlichkeitspsychologie, Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, D-12489 Berlin, Germany (email: uher@primate-personality.net). J. Call is at the Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.

0003–3472/08/$30.00/0

paradigm assumes that individuals display stability and consistency in their behaviour across time and situations, and that they differ from each other in the pattern of traits that constitute their personality (Stern 1911; Allport 1937; Funder 2004). Unlike many other conceptualizations of personality, the trait paradigm focuses primarily on measuring and cataloguing lasting behavioural tendencies without assuming cognitive components that may be uniquely human, or specifying the relative contributions of nature or nurture, or their interaction over the course of life.

Personality traits are meant to be latent dimensional variables along which individuals differ from one another in the degree to which they possess a particular trait. They are distinct from states that are, in contrast, externally caused and only temporary (Funder 2004). Temporal stability is the crucial criterion to infer traits from interindividual differences, and personality from individual trait profiles. Traits, furthermore, create stable relations between situations and the responses of an individual across time. However, a situation may have a different impact on different individuals. This results in a low cross-situational consistency of behaviour, yet situational profiles that are stable across time and distinctive between individuals (Funder & Colvin 1991, Mischel et al. 2002). Similarly, a situation can induce a specific behaviour in one kind of individual, and a different behaviour in another, which results in a low coherence between responses within a situation and stable individual response profiles (Asendorpf 1988).

A systematic approach to personality should, therefore, include a variety of trait-related situations and behaviours that are collected repetitively. In fact, the problem of fluctuations in manifest behaviour can only be solved with sufficient aggregation of behaviour scores on a given trait dimension across several trait-relevant situations, or across several observations within the same situation if (and only if) the scores are sufficiently consistent across situations or time (Epstein 1979, 1980; Asendorpf 1988, 1992; Mischel et al. 2002).

In nonhuman primate studies, the trait paradigm has only rarely been applied explicitly despite its suitability and usefulness (e.g. Rouff et al. 2005). For example, many studies lack the important criterion of establishing temporal stability (for exceptions see Hebb 1949; Stevenson-Hinde et al. 1980; Suomi et al. 1996). Temporal stability is, however, implicated in studies reporting on the high heritability of interindividual differences (e.g. Weiss et al. 2000). To date, situational or response profiles have not been investigated in nonhuman primates, although they are crucial for tackling methodological problems derived from cross-situational consistency and coherence in response.

The present empirical study applies the trait paradigm's theory and methods to a personality study on a sample of zoo-housed great apes. A number of trait constructs were selected bottom-up from the species' behavioural repertoires (Uher 2005). They were operationalized in a large number of trait-related behaviours in a variety of situations. Behavioural data were then subjected to analyses of temporal stability, cross-situational consistency and coherence in response.

METHODS

Subjects

Twenty great apes housed in the Wolfgang Köhler Primate Research Center (WKPRC) in the Leipzig Zoo, Germany, participated in this study from January to March 2005. We included five adolescent or adult bonobos, *Pan paniscus*; chimpanzees, *Pan troglodytes verus*; gorillas, *Gorilla gorilla gorilla*; and orang-utans, *Pongo pygmaeus abelii*. Subjects ranged in age from 7 to 31 years, with a median of 16 years (see Table 1). Data on all apes were collected for the observations; in the behavioural tests four subjects in each species could be tested. The subjects were housed in social groups consisting of 5–18 animals in spacious, naturally designed indoor (230–430 m^2) and outdoor enclosures (1680–4000 m^2), and in special testing rooms, each with a number of interlinked cages (each 5.1–7.3 m^2). The subjects were mostly tested individually and always treated in accordance with ethical principles of noninvasive research; testing was stopped if the subjects showed acute signs of distress (e.g. diarrhoea), which happened just once throughout data collection. During the period of testing, all apes received their complete daily diet consisting of various fresh fruits, vegetables, leaves, cereals, eggs and meat, and were never deprived of food or water at any time.

Design

Trait-related behaviour was recorded in a variety of situations, each narrowly defined by situational features. The specificity of some situations allowed measuring just

Table 1. Species, sex, age and rearing history of the subjects

Species	Subject	Name	Sex	Age (years)	Rearing history
Bonobo	B-Jo	Joey	M	22	Nursery
	B-Ku	Kuno	M	8	Nursery
	B-Li	Limbuko	M	9	Nursery
	B-Ul*	Ulindi	F	11	Mother
	B-Ya	Yasa	F	7	Mother
Chimpanzee	C-Do*	Dorien	F	24	Nursery
	C-Fd	Frodo	M	11	Mother
	C-Fk	Fraukje	F	28	Nursery
	C-Ro	Robert	M	29	Nursery
	C-Sa	Sandra	F	11	Mother
Gorilla	G-Be	Bebe	F	25	Mother/peer
	G-Go*	Gorgo	M	23	Nursery
	G-Nd	Ndiki	F	27	Mother/peer
	G-Ru	Ruby	F	7	Mother
	G-Vi	Viringika	F	9	Mother/peer
Orang-utan	O-Bi*	Bimbo	M	24	Nursery
	O-Dk	Dokana	F	16	Mother
	O-Du	Dunja	F	31	Nursery
	O-Pd	Padana	F	7	Mother
	O-Pi	Pini	F	16	Mother

F: female, M: male.

*Subjects dropped from data collection in the series of behavioural tests.

one trait, but most traits were measured in several situations to consider variances in response to different situations (i.e. cross-situational consistency). To further factor individual differences in behavioural expressions (i.e. coherence in response), multiple trait-related behaviours were observed wherever possible.

Behaviour observations in two different group situations in the spacious indoor enclosures and in 14 different laboratory-based tests were each repeated several times in a test period of about 15 consecutive days (except for two tests with a previously unknown potential to elicit fear). These repetitions allowed meeting the fluctuations of manifest behaviour. To further reduce the impact of states on the aggregated data, the laboratory situations were administered in a reasonably random sequence that avoided testing a subject for the same trait more than once a day. Possible after-effects were considered carefully; therefore, mildly disturbing situations were always tested at the end of each subject's test session and at maximum once a day.

After a break of about a fortnight, the whole data collection process was repeated completely again in a second test period of 15 consecutive days following the same scheme of repetitions and randomization described above for the analyses of temporal stability. The two nonoverlapping periods of data collection took place within 50 days.

Behaviour Observations

Situations and behaviour records were highly standardized in all respects. Observations and behaviour tests were always carried out for all the species by the same person (JU) using the same test apparati and test materials. Behavioural tests were carried out in special testing rooms that allowed us to control the test conditions. Cages in these rooms were made of a combination of mesh and glass with exchangeable panels (90 × 70 cm) where a variety of apparati corresponding to the different tests (see below) could be installed. Group observations and the series of behavioural tests are described in detail in the following sections. All traits and their operationalizations in terms of single behaviours and observational or test situations are listed in Table A1 in the Appendix.

Prefeeding observation

Feeding-related arousability in a group situation was recorded by videotaping the focal apes in 15-s time intervals while they could hear and see the keepers approaching right before they were fed in the afternoon at 1330 hours. The behaviour was coded later from tape with one–zero sampling (Altmann 1974) for frequency estimations of intervals that included any amount of time spent in the observed behaviours. Recorded behaviours were rocking, pacing, wrist shaking, pleasure grin (De Waal 1988), vocalization, scratching and changing position (defined as rising from its resting position and sitting down again or staying within 1.5 m from the original place). For each species, data were collected on 10 days in total.

Afternoon observation

Thereafter, trait-relevant behaviour was recorded using scan sampling (Altmann 1974) with 10-min intervals starting from the end of the last scan to estimate time distributions of behavioural states. Group observations included social behaviours, such as the proximity to conspecifics (within 2 m from the focal subject), body contact, self- or allo-grooming and social or solitary play (with or without objects). Physical activity was recorded with the categories resting, moving and changing location, and persistency with dealing with one of the enrichment boxes, which were installed permanently in every enclosure of the primate centre. And finally, whether or not the subject was feeding was recorded. For each species, behaviour was recorded on 24 days in total with seven scan sample points per day.

Button box test

A grey opaque box (breadth was 68 cm, height was 45 cm, depth was 29 cm) with 20 large yellow buttons having a diameter of 3.5 cm and length of 3 cm spaced by 10 cm and forming a 5 × 4 matrix was placed on a table flush against the mesh so that the ape could reach the buttons by sticking its fingers through the mesh. Pressing a button dislodged a reward that came out into an opening below the box, which the ape could take through the mesh. An elastic band repositioned the pressed button to prevent the ape from seeing which of the buttons he had pressed already or what rewards had been delivered. Subjects received four trials. In the first two trials, the ape faced only four baited buttons at a time; a trial ended when he had pressed all the buttons successfully. The third trial constituted the actual test situation in which the ape was presented with all the 20 buttons but none was baited. The trial ended after 5 min had elapsed. The fourth trial was identical to the first two trials. Thus, the ape faced a situation in which rewards failed to materialize despite constant or even increased effort. Persistent apes should continue their effort, give up later, and try again more often than less persistent apes, therefore showing longer durations of pressing buttons in the test trial. Between sessions, the position of the four baited buttons in the first, second and fourth trials was varied among the 20 buttons available three times within each test period. This test was repeated six times in total.

Cage intruder test

The experimenter entered one of the cages adjacent to the cage occupied by the ape and for the next minute watched him silently from a distance of 1 m through a large window that separated the two cages. Then, the experimenter sat down close to the window and threw some raisins into the ape's cage and again after a minute (20 raisins in total). The test was repeated four times, twice in each data collection period. In this potentially mildly disturbing situation, the duration of being in proximity to the experimenter in the neighbouring cage was measured as friendly behaviour, and the frequency of spitting, attempting to grab, jumping or banging against the gate to the experimenter was measured as quasi-aggressive or teasing behaviour, respectively. Additionally, not taking the raisins was considered a measure of anxiety.

Food box test

Food of different preferences was placed successively in a small transparent Plexiglas box (breadth was 17 cm, height was 12 cm, depth was 12 cm) fixed on a platform (45 × 80 cm) flush against a Plexiglas panel. The ape could open this box through a hole in the Plexiglas panel. After an initial trial without food, the box was baited in 10 consecutive trials behind an opaque panel with the following items in the given order: a banana piece, one grape, one monkey pellet, a carrot piece, four raisins, a celery piece, four single quarters of a medium-sized grape, one raisin, one single quarter of a grape and three monkey pellets. If an ape did not open the box, the food remained in the box and was removed after 1 min. This test was repeated six times, three times in each test period. We coded how long it took them to touch the box from the moment they spied the item(s) (not touching was equated with 60 s) and whether they ate or rejected the food after having extracted it out of the box. Since the retrieval of each food item required the same effort, and all subjects were physically and cognitively capable to open the box, reaching for different quantities and qualities of food was operationalized as a measure of food orientation.

Blocked food box test

In four out of the six sessions of the food box test, two additional 2-min trials with highly preferred food items followed in which the box still looked the same but was blocked by a screw. This potentially frustrating situation was used to record two frequency-dependent categories of impulsive reactions. Emotional impulsivity was indicated by clapping and knocking at the Plexiglas panel, walls or floor, whereas motor impulsivity was indicated by attempts to open the box by reaching for the apparatus with the fingers through the Plexiglas panel.

Food competition test

Two apes faced each other across neighbouring cages from where each of them could reach inside a transparent Plexiglas tunnel (92 × 32 cm and 33 cm) that sat on a platform flush against the metal frames of both cages. During testing the tunnel was baited centrally with one piece of banana so that it lay at the same distance from both apes. Each ape could reach inside the tunnel with one arm through a hole (14 × 16 cm) in the Plexiglas panel. Between trials, these openings could be blocked by transparent panels at each ape's side. In the middle of the tunnel, a 1-cm thick transparent Plexiglas plate separated each ape's reaching area except for a small gate (7 × 10 cm) at the bottom. Each ape got access to the baited box separately in the initial trial of every session. Five competitive test trials followed in which the blocking transparent panels were removed simultaneously so that both apes could grab into the tunnel at the same time. All six possible combinations between the four tested apes in each species were run and repeated four times. This test was used to measure dominance (percentage of gained banana pieces, latency to enter the last quarter of the compartment, frequency to enter the box) and competitiveness (grab the fingers of the opponent).

Hidden food test

The ape entered the test room (5.1–7.3 m^2) in which 10 small food items (green grapes, raisins, small pellets) were hidden on the rims of the cage's frame or stuck with honey to the variegated walls. A dab of honey (about 4 cm^2) was also smeared at the wall. No ape was allowed to observe either the baiting or the other conspecifics being tested in this situation. One grape was placed in the middle of the floor in each of the six sessions to determine whether the subjects were willing to collect food items. There was no indication that food was hidden inside the cage except the food itself. Thus, this test measured vigilance, which was operationalized as the number of items recovered within the 10-min test period. The latency to find each item was also scored; nonrecovered items were assigned a 600-s latency. This test was also used to code physical activity in the limited space available in a cage. Therefore, coding categories were restricted to resting and changing the location.

Honey grid test

A small Plexiglas panel (30 × 45 cm) smeared with honey was attached to the mesh and so the ape had to stick its fingers 5 cm through the mesh to get the honey. After the ape started recovering the honey, the experimenter knocked continuously on a Plexiglas panel located 1.5–2 m away from the honey grid with a rubber tool that produced moderate noise for five consecutive minutes. Using a sufficient amount of honey ensured that the ape was occupied with this task for the whole duration of the test. This test was repeated six times distributed evenly over the two data collection periods. The total time spent recovering the honey was used as a measure of the ape's distractibility when occupied with a simple task for which they were motivated greatly.

Keeper interaction test

A familiar keeper sat in front of the ape's cage and for 2 min encouraged the ape to approach and play with him, but without offering any food. Thereafter, he fed the ape apple slices for 1 min. Then, the keeper resumed his positive interaction with the ape as before for two more minutes, again without offering any food. The test was repeated six times with different keepers. The ape's interaction with the keeper throughout the 5-min period was recorded to assess friendly behaviour measured as the total time spent in proximity (defined by the quadrant of the cage next to him) or in close contact with the keeper (defined as passively permitting or actively initiating contact by sticking their fingers or lips through the mesh). The frequency of spitting, attempting to grab, jumping or banging against the mesh directed at the keeper was coded as quasi-aggressive or teasing behaviour, respectively.

Masked human test

The experimenter entered the test room silently, disguised with a rubber face mask of an elderly man, a black short-haired wig, a thick dark green zoo parka, and large yellow rubber boots. Wearing yellow opaque rubber gloves and sticking the right glove's stiffed fingers through the mesh into the ape's cage, she offered 20 grapes consecutively to the ape. After 4 min, the experimenter left the room. With the unexceptionally friendly and nonthreatening behaviour of the disguised experimenter towards the apes, the potential strength of the situation was reduced to permit the expression of individual differences. With a few exceptions, all apes finally took the offered grapes from the masked human's hand, thus suggesting an ethically acceptable level of arousal and anxiety caused by the uncertainty of the situation. The apes' behaviour was analysed in regard to friendliness (time spent in proximity), aggression (attempts to injure, e.g. biting, grabbing and vigorous banging against the mesh towards the experimenter), quasi-aggression/teasing behaviour (spitting, cage shaking at the side or rear walls of the cage), anxiety (latency to take the first grape offered by the masked human and initially climbing off the floor in response to the masked human's appearance) and arousal (occurrence of pilo-erection). For ethical reasons, this test was repeated just once in the second test period after a break of 4 weeks. To ensure unfamiliarity in the second session, another mask showing another face and a blond-haired wig were used.

Novel food test

The ape received, in turn, apple slices and novel food items (approximately 5–6 cm in size) that consisted of either natural or artificial food changed in shape and colour. To ensure a sufficient degree of novelty, in each of the four administered sessions, two different shapes and two different colours were combined in a two-by-two design resulting in four different food items per session. There were red and green star- and bell-shaped apple slices, yellow and brown cubic marshmallows and meringues, blue and brown heart- and tree-shaped pear slices, and green striped and red spotted wafer reels and white mouse-shaped marshmallows. In each session, every ape got all four combinations of shape and colour twice, so that eight novel food items were presented per session. Food colouring did not add any flavour; and all kinds of novel food were as sweet as normal apples or even sweeter to allow the apes to have the same or even a stronger preference for the novel food compared to normal apple slices. This is contrary to the unfamiliarity these items had for the apes in shape, colour and partly in their artificial quality. The differences in the apes' behaviour towards familiar and novel items in the latencies to touch, the time spent manipulating the items and the percentage of items eaten can be interpreted as curiosity or interest in novel food.

Novel object test

A small object (maximum 10 cm in diameter) was placed in the corner of the cage such that the ape could not see it unless it entered the test cage. The small size was chosen to reduce the potential to elicit fearful reactions that are known to depend on item size (Vochteloo et al. 1991). The items included objects like blue, red, green, purple or yellow small ducks, fishes, turtles, frogs or fluffy balls made of various materials such as plastic, paper or burlap. Subjects were given 10 min to explore them. As all apes touched the objects or at least inspected them closely from a distance of 5 cm at most, the influence of anxiety towards unknown objects was negligibly small. A new object was used in each of the four administered sessions to maintain the crucial criterion of novelty. Finding a small novel object inside their cage was an extremely rare event for the apes in the research centre, therefore the duration of any activity directed to the object was operationalized as curiosity and interest in novel objects, respectively. Additionally, self-sexual activity could be measured in this test.

Pile of food test

The experimenter sat at a table flush against the Plexiglas panel for 3 min in full view of the subject while cutting an apple and a banana into small pieces and piling them up inside a bowl (10 cm in diameter and 3 cm in height). To maintain motivation, the apes were given this pile of food after testing in all cases. This test was repeated six times, three times in each test period. Considering the range of behavioural expressions for arousability in these four species, a number of categories were coded including the durations of rocking, pacing, wrist shaking, pleasure grin, high-pitched vocalization, and the frequencies of scratching and changing the position (see Prefeeding Observation). Furthermore, this situation permitted us to code impulsive responses such as knocking at the Plexiglas panel, walls or floor and clapping.

Food out of reach test

The pile of food from the previous test (or a whole banana in the second session of each test period) was placed in front of the ape but still out of its reach on a table that was flush against the ape's cage. In one trial, the experimenter left the room, and in the other she sat at the table doing nothing. Each trial lasted 2 min; their order was alternated over the four administered sessions. In all cases the ape was given food after testing. Seeing a substantial amount of food out of reach by a few centimetres bore the frustrating potential to elicit impulsive reactions, which were again measured with the frequency of knocking and clapping.

Sudden noise test

After at least 5 min of sitting in the cage without anything happening, a 30-s recording of a French news programme was played for the ape in moderate volume independent of the experimenter's activity. The apes were neither familiar with this language nor with these particular voices. The sudden onset of unfamiliar human sounds inside the test room constituted a situation of uncertainty that elicited a range of trait-related behaviours during playback and five subsequent minutes. Climbing off the ground in response to the radio was measured again as anxiety, pilo-erection and scratching as arousal, and cage shaking and banging as quasi-aggressive behaviour. Due to its (previously unknown) potential to elicit fear, this test was only repeated once after a time-out of 4 weeks. To maintain unfamiliarity, a different record was used in the second session.

Procedure

Data were always collected in two species in parallel. The test periods were then alternated so that the second pair of species was tested during the data collection break of the first pair of species and vice versa. Eight apes in two species were presented individually with a succession of two to four of the behavioural tests described above every morning from 0830 to 1230 hours. When testing was complete, the apes went into the indoor enclosures to join their social groups for the rest of the day. The observations of all five subjects per species started with the recording of prefeeding behaviour shortly before 1330 hours when the apes awaited their afternoon feeding in their social groups. As all the species were fed simultaneously, only one species could be recorded per day, and the observations were alternated between the two species on a day-to-day basis. The afternoon observations started at about 1345 hours and lasted till 1530 hours approximately. To collect data in both species, scanning was alternated between them using the 10-min break between two scans in each species.

Data Analyses

Behaviour recording in the afternoon observation was online; the prefeeding observations and all behavioural tests were videotaped and coded later using the coding software Interact (Rel. 7.2.4., www.behavioural-research.com; Mangold 2006). Within the 50 day study period each ape was recorded for a total of 67.3 h distributed between 425 min of detailed coded video corresponding to the behavioural tests, 40–50 video-recorded 15-s intervals in the prefeeding observation and 168 instantaneous points recorded online in 60 h in the afternoon observation. A second, independent person coded, in parallel to the main coder (JU), all behaviours in 20% of each species' session for each of the 14 behavioural tests and for the prefeeding observations from video. Likewise, 20% of the afternoon observations of each species were recorded online in parallel and independently from the main observer (JU) by this second person. In doing so, intercoder and interobserver reliability, respectively, could be analysed for all 71 raw behavioural variables. The median Cronbach's α was 0.96 ($N = 71$; for aggregations over time, zero–one coded data were treated as metric variables). Five behavioural measures did not exceed $\alpha > 0.80$, but with a range of $\alpha = 0.71–0.77$ their reliability was sufficiently high.

Considering their significance for trait-related behaviour, some of the 71 behavioural raw variables were additionally transformed. For example, latency measures were equated with the maximum test time if the target behaviour did not eventuate, difference scores were computed between different conditions in the novel food test, and activity categories or different numbers of conspecifics in proximity were treated as separate variables, so that finally 76 variables were included in the subsequent analyses.

To reduce the impact of fluctuations in manifest behaviours and to increase their reliability, every single behavioural variable, each of which represented a specific behaviour in a circumscribed situation, was averaged over all sessions within each of the two test periods. The resulting sets of data were analysed for temporal stability on different levels of aggregation starting from single behaviour measures to scores aggregated on the trait level. On each level, two different perspectives were studied. Variable-centred analyses tested for temporal stability in the relative order of the subjects on each variable to characterize the variables themselves. And individual-centred analyses checked temporal stability of the individual profiles, and therewith the stability of the individuals' characteristics (Furr & Funder 2004). Finally, temporal stabilities of situational and response profiles were analysed in the present data to show the principles of cross-situational consistency and coherence in response empirically.

RESULTS

Stability on the Single Behaviour Level

Test–retest reliabilities of all 76 behavioural variables were analysed with Cronbach's α to study the temporal stability of single behaviours from a variable-oriented view. The mean α (using Fisher's r-to-Z transformation) was 0.86 ($N = 76$). For a more direct comparison, Pearson correlations were also computed and showed a mean of $r = 0.78$ using Fisher's r-to-Z transformation ($N = 76$). All reliability scores are listed in Table A1 in the Appendix. A total of 69 variables showed considerable stability in the relative order of the subjects between test block one and two, indicating that for most individuals trait scores were the same in both test periods. Because temporal stability characterizes the behavioural variable in the sample, it does not exclude clear changes in single individuals (see Fig. 1). Seven variables did not exceed $\alpha > 0.50$.

To analyse the data from an individual-centred view, the subjects' profiles consisting of all 76 single behavioural variables were analysed for temporal stability. As such individual behavioural profiles reflect both the mean profile in the given species and the individual deviations from this mean profile, test–retest reliability of the mean profiles was analysed first. The mean profile of all apes showed a test–retest reliability of $\alpha = 0.98$ ($r = 0.97$); the species-specific mean profiles showed similarly high stabilities from $\alpha = 0.95$ to 0.99 ($r = 0.90–0.98$). High cross-species stabilities from $\alpha = 0.85$ to 0.96 ($r = 0.74–0.92$) tentatively indicated the absence of species differences in these profiles, but the small sample sizes ($N = 5$ for each species) did not permit analyses of species differences.

Therefore, we analysed consistency in the average deviation of every subject from the universal great ape

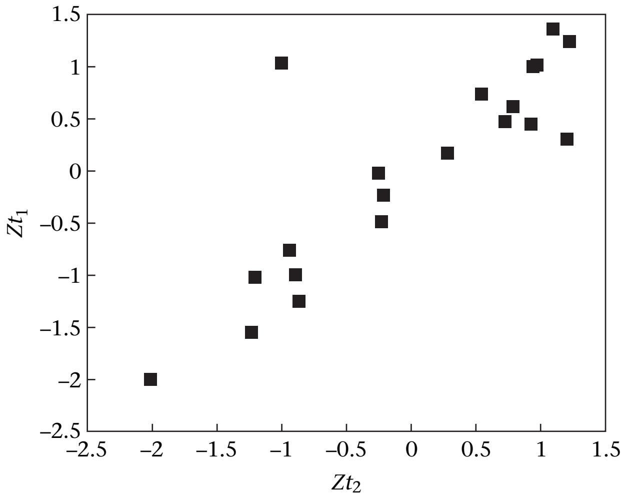

Figure 1. Example of the test–retest reliability of individual differences on the single behaviour level (activity category 'move' in afternoon observation) in the present sample of 20 great apes. Data are *z* scored within each test period.

mean profile computed with *z* scores of all behavioural variables, each across all individuals within a test period. The mean test–retest reliability in these individual profiles (computed with Fisher's *r*-to-*Z* transformation) was $\alpha = 0.86$ ($r = 0.76$), individual stabilities ranged from $\alpha = 0.66$ to 0.97 ($r = 0.49$–0.94, $N = 20$; see first two columns in Table 2). Note that the profiles of one subject per species were restricted to observational data only (see subject overview in Table 1).

Stability on the Trait Level

In a further step, the *z* scores of all single measures that were facets of the same trait were treated as items and aggregated into one trait score. To share the same meaning, some of these *z* scores were reversed, for example, resting that indicates a low score on physical activity (these variables are marked with an asterisk in Table A1 in the Appendix). The resulting aggregated trait scores were subjected to variable-oriented analyses. Mean test–retest reliability (computed with Fisher's *r*-to-*Z* transformation) of the 19 traits was $\alpha = 0.87$ ($r = 0.77$) with a range from $\alpha = 0.40$ to 0.98 ($r = 0.29$–0.97, see Table 3). These results indicated, for most traits, considerable agreement in the relative order of the subjects between both test periods.

Individual-oriented test–retest reliability analyses of the aggregated trait profiles showed a moderate to high temporal stability. Mean α was 0.89 ($r = 0.80$; computed with Fisher's *r*-to-*Z* transformation) and the individual profile stability ranged from $\alpha = 0.56$ to 0.99 ($r = 0.39$–0.97, $N = 20$; see last two columns in Table 2).

Table 2. Test–retest reliability α and *r* of individual profiles consisting of all single behaviours and of scores aggregated on the trait level

		Single behaviour profile		Aggregated trait profile	
Species	Subject	α	r	α	r
Bonobo	B-Jo	0.86	0.76	0.85	0.74
	B-Ku	0.75	0.60	0.71	0.51
	B-Li	0.78	0.64	0.76	0.68
	B-Ul*	0.97	0.94	0.97	0.94
	B-Ya	0.75	0.60	0.67	0.51
	$\text{Total}_{\text{Bonobo}}$†	0.86	0.75	0.84	0.73
Chimpanzee	C-Do*	0.95	0.91	0.98	0.96
	C-Fd	0.77	0.63	0.56	0.38
	C-Fk	0.84	0.72	0.89	0.81
	C-Ro	0.78	0.63	0.80	0.67
	C-Sa	0.79	0.65	0.92	0.88
	$\text{Total}_{\text{Chimpanzee}}$†	0.85	0.74	0.89	0.81
Gorilla	G-Be	0.83	0.72	0.83	0.71
	G-Go*	0.96	0.92	0.98	0.97
	G-Nd	0.86	0.76	0.95	0.91
	G-Ru	0.66	0.49	0.57	0.39
	G-Vi	0.93	0.88	0.99	0.97
	$\text{Total}_{\text{Gorilla}}$†	0.88	0.79	0.94	0.89
Orang-utan	O-Bi*	0.89	0.80	0.81	0.68
	O-Dk	0.82	0.70	0.73	0.58
	O-Du	0.90	0.82	0.90	0.82
	O-Pd	0.80	0.66	0.83	0.70
	O-Pi	0.86	0.75	0.92	0.88
	$\text{Total}_{\text{Orang-utan}}$†	0.86	0.75	0.86	0.72
	Total sample†	0.86	0.76	0.89	0.80

Cronbach's α and Pearson correlation *r* are reported.

*Subjects dropped from data collection in the series of behavioural tests.

†Mean test–retest reliabilities were computed using Fisher's *r*-to-*Z* transformation.

Situational Profiles

Comparing the subjects' trait scores between similar, but nonidentical, situations showed that the temporal stability of trait scores within the same situation was higher than their covariation between different situations. For example, the mean correlation of aggressiveness scores (computed with Fisher's *r*-to-*Z* transformation) between four different situations (cage intruder, keeper interaction, masked human and sudden noise tests) in the first test period was $r = 0.25$ (range $r = -0.09$ to 0.68). This means that the subjects' scores varied considerably between these situations. Recall, that these were scores of the same behaviour in the same subjects and that it was only the situation that differed. Cross-situational consistency is similarly low in humans (Mischel 1968; Mischel & Peake 1982; Funder 2001). This finding, however, does not contradict the concept of stability in individual differences. Instead, personality traits become manifest in stable situational profiles that reflect systematic interactions between individuals and situations. Obviously, these four situations elicit aggressiveness in different subjects to a varied degree, independent of their general tendency to respond with aggression, which is reflected in their aggregated trait scores. For example, there were subjects who reacted more aggressively to keepers than to masked humans in comparison to other subjects, and others in turn who showed the opposite behavioural pattern (see e.g. the

Table 3. Test–retest reliability α and *r* of individual differences in aggregated trait-relevant behaviour

Trait	α	*r*
Aggressiveness to humans	0.92	0.85
Arousability	0.74	0.59
Anxiousness	0.93	0.87
Competitiveness	0.70	0.53
Curiosity	0.82	0.70
Distractibility	0.85	0.74
Dominance	0.86	0.76
Food orientation	0.90	0.83
Friendliness to youngsters	0.98	0.97
Friendliness to conspecifics	0.86	0.76
Friendliness to humans	0.93	0.78
Gregariousness	0.92	0.85
Impulsiveness	0.45	0.29
Persistency	0.88	0.78
Physical activity	0.86	0.76
Playfulness	0.95	0.91
Self care	0.81	0.68
Sexual activity	0.40	0.25
Vigilance	0.94	0.89

Cronbach's α and Pearson correlation *r* are reported; scores depict *z* scores aggregated on the trait level (for the list of behaviours assigned to each trait see Table A1 in Appendix).

shapes of situational profiles in the subjects C-Fd and B-Jo in Fig. 2c, f).

In the present study, a number of situational profiles were identified and subjected to stability analyses. Cross-situational consistency between situations in terms of Pearson correlation between them can be compared to the test–retest correlation between measures within these situations. The mean temporal stability of eight situational profiles each over all subjects (computed with Fisher's *r*-to-*Z* transformation) was $r = 0.69$, whereas the mean cross-situational correlation was $r = 0.19$. Mean test–retest correlations computed separately for each species were $r = 0.89$ in the bonobos, $r = 0.85$ in the chimpanzees, $r = 0.81$ in the gorillas and $r = 0.51$ in the orang-utans (for details see Table 4). It should be noted that due to the small sample sizes these stabilities are very sensitive to single outliers. The only moderate stability of situational profiles of impulsiveness can be explained by learning effects that probably reduced the induced frustration with increased repetition. This became especially obvious in the blocked food box test.

Response Profiles

Individuals not only differ in how they respond to a situation, but also in how they externalize a trait. Research in humans has revealed only low coherence in response across different individuals (Asendorpf 1988). In the present study, coherence in response was also low. For example, the mean intercorrelation (Pearson correlations) between the seven arousability measures in the pile of food test in the first test period was $r = -0.06$ (computed with *r*-to-*Z* transformation).

This low coherence in response imposes problems on the behavioural assessment of personality differences. First, if measures are restricted to only a few indicators of the trait considered, individuals reacting primarily with the other indicators will be classified wrongly. And second, different responses are not equivalent in their meaning for all subjects (Asendorpf 1988). To overcome these problems, the present study analysed (where it was

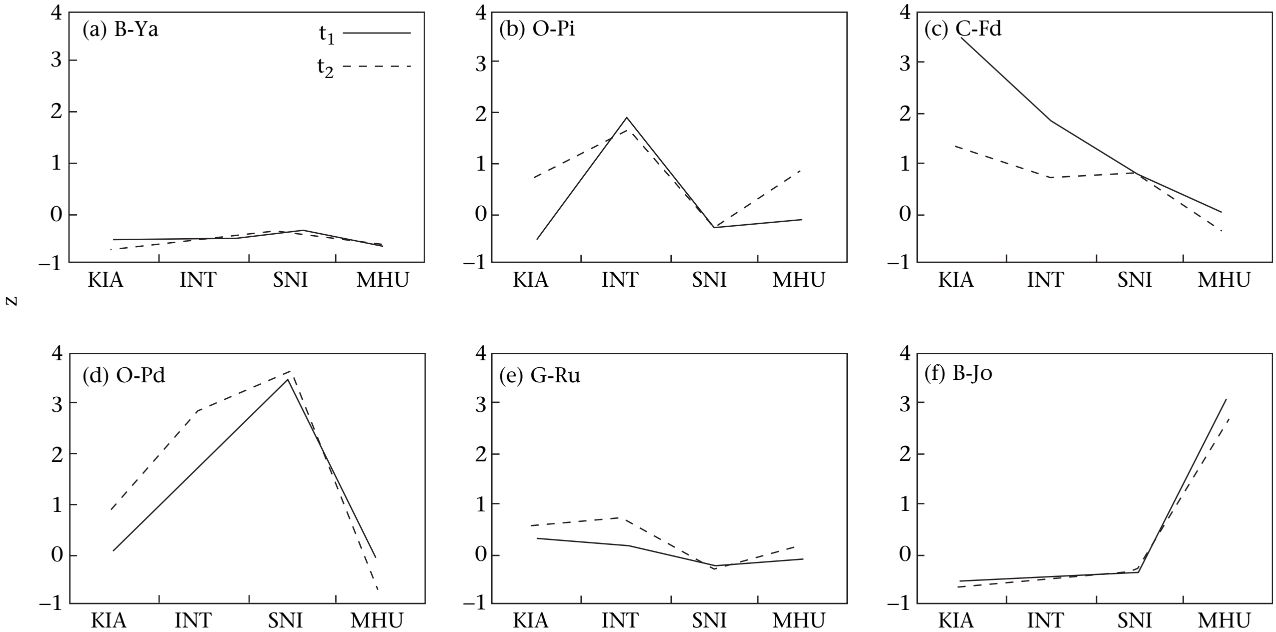

Figure 2. Examples for the temporal reliability of individual situation profiles in aggressiveness. KIA: keeper interaction test, INT: cage intruder test, SNI: sudden noise test, MHU: masked human test; *z* scores aggregated within each test period, six subjects (a) and (f) bonobos, (c) chimpanzee, (e) gorilla and (b) and (d) orang-utans.

Table 4. Mean cross-situational consistency and temporal stability *r* of individual situation profiles

		Mean cross-situational correlation *r**	Temporal stability *r* of individual situation profiles				
Traits	Situations	Total	Total	Bonobo	Chimp	Gorilla	Orang-utan
Aggressiveness	Cage intruder test, keeper interaction test, masked human test, sudden noise test	0.25	0.77	1.00	0.89	0.55	0.77
Arousability	Pile of food test, prefeeding observation, masked human test, sudden noise test	0.07	0.71	0.63	0.81	0.97	0.58
Anxiousness	Cage intruder test, masked human test, sudden noise test	0.35	0.76	0.98	0.92	0.63	−0.07
Curiosity	Artificial novel food, natural novel food, novel object	0.18	0.49	0.33	0.46	0.65	0.62
Friendliness to humans	Cage intruder test, keeper interaction test, masked human test	0.64	0.78	0.69	0.77	0.75	0.68
Impulsiveness	Blocked food box test, pile of food test, food out of reach test: inside condition, outside condition	0.37	0.27	0.27	0.19	0.82	0.50
Persistency	Button test, indoor enclosure	−0.30	0.77	0.80	0.99	0.86	0.51
Physical activity	Hidden food test, indoor enclosure	−0.14	0.75	0.49	0.91	0.88	0.80
Total*		0.19	0.69	0.89	0.85	0.81	0.51

Reported are Pearson correlations *r* of trait scores between different situations within test period t_1, opposed to Pearson correlations across experimental settings between test period t_1 and t_2; trait scores within a situation are composed of the mean of all relevant behavioural measures (see Table A1 in Appendix).

*Mean correlations were computed using Fisher's *r*-to-*Z* transformation.

possible) all behavioural indicators shown by the subjects in the respective situations (see Table A1 in Appendix). Then, the *z* scores of all indicators were averaged for global considerations. Low intercorrelations each between different behavioural measures of the same trait (in the first test period) identified three different types of response profiles in trait–situation units. The mean Pearson intercorrelation was $r = 0.19$. Mean Pearson test–retest correlation between measures within these situations was $r = 0.66$ (computed with Fisher's *r*-to-*Z* transformation) on the level of the total sample, $r = 0.65$ in the bonobos, $r = 0.70$ in the chimpanzees, $r = 0.89$ in the gorillas and $r = 0.43$ in the orang-utans (for details see Table 5).

DISCUSSION

The present paper showed the viability of the trait paradigm's methods and approaches to analyse stable individual differences in manifest behaviour in a sample of zoo-housed great apes. By sampling the subjects' behaviours repeatedly across varying situations, and by aggregating the data within two nonoverlapping test periods, the present study showed that individual great apes can be differentiated reliably in a wide range of behaviours at least across intermediate periods of time. Individual- and variable-centred views permitted detailed stability analyses of the individuals' characteristics as well as of traits as dimensions describing the sample. Temporal

Table 5. Mean coherence in responses and temporal stability *r* of individual response profiles

		Mean correlation *r** between responses	Temporal stability *r* of individual response profiles				
Trait in test situation	Responses (single measures)	Total	Total	Bonobo	Chimp	Gorilla	Orang-utan
Arousability in pile of food test	Grin, change position, pace, rock, scratch, shake wrist, vocalize	−0.06	0.73	0.79	0.78	0.89	0.56
Arousability in prefeeding observation	Grin, change position, pace, rock, sexual activity, vocalize	0.30	0.67	0.56	0.64	0.97	0.45
Curiosity in novel food test	Reject†, touch novel food†, deal with novel food	0.24	0.55	0.53	0.67	0.61	0.60
Total*		0.16	0.66	0.65	0.70	0.89	0.43

Reported are Pearson correlations *r* between different behavioural responses within the same experimental or observational setting in test period t_1, opposed to Pearson correlation *r* between test periods t_1 and t_2 reflecting temporal stability in individual differences in the externalizations of the same trait within the same setting.

*Mean correlations and test–retest reliability were computed using Fisher's *r*-to-*Z* transformation.

†Inversed score used to share the same meaning.

stability was shown on different levels, starting from the level of single behaviours, over situational and response profiles, to the trait level with scores aggregated over multiple behaviours and situations. Substantial intercoder and interobserver agreement supports the reliability of behavioural measurements in animal personality studies even when recording a very broad range of behaviours with different ethological methods in different species simultaneously.

Although primate researchers repeatedly came across methodological difficulties when analysing personality differences in behavioural data, the trait concept in its theoretical and methodological approach has hardly been considered so far. For example, empirical evidence for temporal stability was reported to be substantial in some behaviour categories, completely absent in others (Hebb 1949; Stevenson-Hinde et al. 1980; Suomi et al. 1996) and has only rarely been established in primate studies up to date. As noted above, appropriate aggregations at least over occasions, if not over different trait-relevant behaviours and situations, are crucial to tackle the methodological problems fluctuations impose on behavioural data. Although some studies successfully dealt with behavioural fluctuations, they failed to analyse differences *between* individuals. Moreover, the methodological challenges of cross-situational inconsistency and coherence in response have not been addressed yet. For instance, in a longitudinal study focusing on continuity and change in the behaviour of seven rhesus macaques, *Macaca mulatta*, aged from 6 to 20 years, Suomi et al. (1996) reported stable individual behavioural profiles. They were composed of various behaviour categories arranged in decreasing order of absolute value. The reported stability of the individual profiles from early to late adulthood (spanning a 15-year period) of at least more than 0.65 in Spearman rank–order correlations is, indeed, remarkable. Profile stability between adjacent 5-year blocks of time is reported to be at least above 0.73 per individual. However, following a mere idiographic approach, these behavioural profiles depict *intra*individual rankings of each subject's most frequently displayed behaviours. Therefore, these profiles reveal only little information about *inter*individual differences in terms of a nomothetic approach. They do not analyse the subjects' deviations from the sample's mean score in each behaviour category. Instead, these profile stabilities measure continuity in manifest behaviours confounded with both differences in the frequency of certain behaviours that are universal in this species and individual variation in relation to peers.

In another study on consistency in behaviours over time and across situations in 25 rhesus macaques, *M. mulatta*, Stevenson-Hinde et al. (1980) were surprised about low cross-situational consistencies in their data despite reliable and meaningfully correlating behavioural measures. The authors argued that their findings would reflect a 'failure to look at appropriate measures rather than a characteristic of the … [individuals] themselves'. They, furthermore, discussed the possibility that the 'lack of any significant correlation … emphasizes the different nature of the two situations' (page 508). Therewith, Stevenson-Hinde and colleagues mirror the issues of the personality versus situation debate personality psychology has been occupied with for the last 40 years (Mischel 1968; Funder 2001). This debate finally resulted in the recognition that cross-situational consistency in individual differences is moderate to low but that the temporal stability of individual personality patterns is high.

In adopting the personality psychological trait paradigm, the present study is the first to analyse cross-situational consistencies and coherence in response systematically in nonhuman primate personality. The results of a much higher temporal stability of situational profiles as compared to the cross-situational consistency of the profiles point to the importance different trait-relevant situations have for personality investigations. Likewise, low consistencies across different behavioural measures of the same trait can be either due to a poor selection of indicators or due to stable individual response profiles that constitute an analogue to the situational profiles. These findings argue for a careful selection of multiple trait-relevant behaviours. This especially applies for studies comparing different species with different behavioural repertoires. Aggregations on the trait level ignore the peculiarities in both individual situational and response profiles. Consequently, studies investigating personality in larger samples should look for empirical evidence of situational or response profiles to identify classes of similar situations or coherent reactions that would define inferior traits or personality profile types. Thus, it is ultimately a matter of empirical evidence whether interindividual differences in different situations reflect the same superior trait or not. If consistency is high, they are probably due to the same trait. If not, they are obviously not due to the same trait. The deeper problem here is that distinct mechanisms may nevertheless result in the same interindividual differences because they may share components that dominate these differences. Low consistency suggests that more than one trait underlies interindividual differences in at least one of the types of situations. In fact, interindividual differences are often determined by multiple traits (Allport 1937).

Despite its small sample size, the results of the present study correspond fairly well to the findings in pertinent studies on human personality (e.g. Asendorpf 1988; Funder 2004). First, it supports the merits of appropriate aggregation to reduce error variance and to increase reliability and validity (Epstein 1979, 1980). Hence, aggregation might be especially effective in studies investigating small samples with high numbers of variables that are referred to as 'intensive' research designs in primate research (Kraemer et al. 1977). Second, cross-situational intercorrelations were similarly low as in human research with situational profiles being distinctive between individuals and stable across time (Mischel 1968; Mischel & Shoda 1995; Funder 2001). And third, the present results provide empirical evidence for substantial interindividual differences in response profiles for the same trait as they have been reported in human studies (e.g. Asendorpf 1988). Thus, for the empirical study of personality differences in great ape behaviour complementing ethological with psychological methods turned out to be very fruitful.

Limitations and Future Directions

The major contribution of the present paper is to provide a substantial body of trait psychological theory and methods for animal personality investigations using a multidisciplinary approach. The broadness of the approach and the diversity of the analyses presented here are inevitably at the expense of a larger sample. Therefore, present empirical findings can only illustrate exemplarily the methods and approaches presented here. The decision for a mixed sample composed of even smaller subsamples in each species was based on multiple considerations. At the time of planning the study, there were already 12 studies on chimpanzees, but only three on gorillas, one on bonobos (Gold & Maple 1994; Murray 1996; Kuhar et al. 2003) and none on orang-utans. Therefore, we decided that it was important to include great apes other than chimpanzees in the sample. Furthermore, all species were available for testing at the research centre although their groups consisted of only five individuals each, except for the chimpanzees of which an equal sample was selected randomly. The priority objective of the study was to show empirically that using exactly the same methods and analyses multiple animal species with different behavioural repertoires can be investigated simultaneously. Some further minor limitations have to be accepted due to practical circumstances. It was not possible to record event behaviours such as intraspecies aggression and allo-sexual activity in nonfeeding contexts of the afternoon observations. Moreover, the number of youngsters living in each group differed between species. Whereas almost 40% of the chimpanzee and orang-utan group members were younger than 7 years, this was true for only 17% in the bonobos and gorillas. Therefore, the possibility to show trait-related behaviour differed between species. For purposes of the present study, the potential impact of these differences was reduced by recoding the absolute number of youngsters appearing as social partners into zero—one. Studies aiming at the comparison of individuals or species, however, should carefully consider differences in the availability of youngsters in the studied samples.

The preliminary empirical evidence for temporal stability, situational and response profiles in the personality traits reported here should be further expanded by investigations of larger samples in each species. Therewith, it would be possible to investigate more fine-grained differences in the great apes' personality structures. Given the high temporal stability over intermediate periods of time reported here, it is of further interest to know more about the long-term stability of personality differences in great apes.

Consistency and continuity on the single behaviour level also provide the basis for more effective measurements of personality differences (e.g. by ratings) that are increasingly being studied, however, with only moderate success to predict manifest behaviours so far (Capitanio 1999; Pederson et al. 2005). It is assumed that with the appropriate methods to analyse consistency and continuity in manifest animal behaviour coherence between rating data and manifest behaviour will be much stronger than previously reported.

Among all studies on interindividual differences in nonhuman primates published till date, only three are concerned with wild individuals living in their natural habitat, one in baboons (*Papio anubis*, Buirski et al. 1973) and two in chimpanzees (*Pan troglodytes*, Buirski et al. 1978; Buirski & Plutchik 1991). A further study investigated wild born chimpanzee orphans who had experienced a variety of early-life stressors before arriving at an African sanctuary (King et al. 2005). However, none of these field studies has measured interindividual differences in manifest behaviour. It would be useful, however, to collect such data in wild great ape populations as they are more likely to encounter greater variations and more frequent fluctuations in their natural and social environment. Thus, compared to the present sample of captive apes, temporal stabilities in personality differences in wild great ape populations might not be that pronounced (Sackett et al. 1981). Alternatively, according to the dynamic—interactionistic paradigm in personality psychology, it is conceivable that wild great ape individuals actively choose situations or environments that are best suited to them or that are congruent with their personalities so that stability could even be more pronounced than in captivity. Within the limits set by environmental conditions such as food distribution, geographical borders, etc., wild great apes can freely choose when to join whom in their social community, whether to seek out new groups and new territories or to stay in familiar areas, whether to travel alone or in large companionships, etc. Easier accessibility, more controlled and stable environments and known kinship, in contrast, predestine captive populations of great apes to studies on more fine-grained personality differences, on long-term stabilities and especially on the roles personality differences play in behaviour in general, and in social interactions or group dynamics in particular. Therefore, such investigations could have important practical implications too. For example, measuring personality differences in zoo-housed great apes could provide information needed for decisions on transfers to potential new mates that could complement the genetic selection. Information on the individuals' social compatibility in particular could have a share in reducing distress in captive groups, thus also enhancing breeding and conservation efforts. Therewith, this multidisciplinary approach to sound measurement of animal personality could contribute in significant ways to the conservation and management of highly endangered species.

Acknowledgments

We are indebted to the zoo veterinarian Prof. Klaus Eulenberger and to all zoo-keepers at the Wolfgang Köhler Primate Research Center, Leipzig, for their cooperation and assistance in testing and observing the apes. Special thanks go to Josefine Kalbitz for extensive reliability coding and observation. We also thank Marlene Smith, Alex M. Smith and Annett Hoppe-Cotte for stylistic corrections on the manuscript.

References

Allport, G. W. 1937. *Personality: a Psychological Interpretation*. New York: Macmillan.

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.

Asendorpf, J. B. 1992. A Brunswikean approach to trait continuity: application to shyness. *Journal of Personality*, **60**, 53–77.

Buirski, P. & Plutchik, R. 1991. Measurement of deviant behavior in a Gombe chimpanzee: relation to later behavior. *Primates*, **32**, 207–211.

Buirski, P., Kellerman, H., Plutchik, R., Weininger, R. & Buirski, N. 1973. A field study of emotions, dominance, and social behavior in a group of baboons (*Papio anubis*). *Primates*, **14**, 67–68.

Buirski, P., Plutchik, R. & Kellerman, H. 1978. Sex differences, dominance, and personality in the chimpanzee. *Animal Behaviour*, **26**, 123–129.

Buss, D. M. 1991. Evolutionary personality psychology. *Annual Review of Psychology*, **42**, 459–491.

Capitanio, J. P. 1999. Personality dimensions in adult male rhesus macaques: prediction of behaviors across time and situation. *American Journal of Primatology*, **47**, 299–320.

Dall, S. R. X., Houston, A. I. & McNamara, J. M. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, **7**, 734–739.

De Waal, F. B. M. 1988. The communicative repertoire of captive bonobos (*Pan paniscus*) compared to that of chimpanzees. *Behaviour*, **106**, 183–251.

Epstein, S. 1979. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, **37**, 1097–1126.

Epstein, S. 1980. The stability of behavior: II. Implications for psychological research. *American Psychologist*, **35**, 790–806.

Funder, D. C. 2001. Personality. *Annual Review of Psychology*, **52**, 197–221.

Funder, D. C. 2004. *The Personality Puzzle*. 3rd edn. New York: W.W. Norton.

Funder, D. C. & Colvin, C. R. 1991. Explorations in behavioral consistency: properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, **60**, 773–794.

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

Gold, K. C. & Maple, T. L. 1994. Personality assessment in the gorilla and its utility as a management tool. *Zoo Biology*, **13**, 509–522.

Gosling, S. D. 2001. From mice to men: what can we learn about personality from animal research? *Psychological Bulletin*, **127**, 45–86.

Gosling, S. D., Lilienfeld, S. O. & Marino, L. 2003. Personality. In: *Primate Psychology* (Ed. by D. Maestripieri), pp. 254–288. Cambridge: Harvard University Press.

Hebb, D. O. 1946. Emotions in man and animals: an analysis of the intuitive process of recognition. *Psychological Review*, **53**, 88–106.

Hebb, D. O. 1949. Temperament in chimpanzees: I. Method of analysis. *Journal of Comparative and Physiological Psychology*, **42**, 192–206.

King, J. E., Weiss, A. & Farmer, K. H. 2005. A chimpanzee (*Pan troglodytes*) analogue of cross-national generalization of personality structure: zoological parks and an African sanctuary. *Journal of Personality*, **73**, 389–410.

Kraemer, H. C., Alexander, B., Clark, C., Busse, C. & Riss, D. 1977. Empirical choice of sampling procedures for optimal research design in the longitudinal study of primate behaviour. *Primates*, **18**, 825–833.

Kuhar, C. W., Stoinski, T. S., Lukas, K. E. & Maple, T. L. 2003. Gorilla behavior index: a tool for management of all male western lowland gorilla (*Gorilla g. gorilla*) groups? *American Journal of Primatology*, **60**, 93.

Mangold, P. 2006. *Interact User Guide, V. 7.0. ff.* Arnstorf: Mangold International.

Mischel, W. 1968. *Personality and Assessment*. New York: J. Wiley.

Mischel, W. & Peake, P. K. 1982. Beyond déjà-vu in the search for cross-situational consistency. *Psychological Review*, **89**, 730–755.

Mischel, W. & Shoda, Y. 1995. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, **102**, 246–268.

Mischel, W., Shoda, Y. & Mendoza-Denton, R. 2002. Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science*, **11**, 50–54.

Murray, L. E. 1996. Personality and individual differences in captive African apes. *Primate Eye*, **59**, 40–41.

Nettle, D. 2006. The evolution of personality variation in humans and other animals. *American Psychologist*, **61**, 622–631.

Pederson, A. K., King, J. E. & Landau, V. I. 2005. Chimpanzee (*Pan troglodytes*) personality predicts behavior. *Journal of Research in Personality*, **39**, 534–549.

Rouff, H. J., Sussman, R. W. & Strube, M. J. 2005. Personality traits in captive lion-tailed macaques (*Macaca silenus*). *American Journal of Primatology*, **67**, 177–198.

Sackett, G. P., Sameroff, A. J., Cairns, R. B. & Suomi, S. J. 1981. Continuity in behavioral development: theoretical and empirical issues. In: *Behavioural Development: the Bielefeld Interdisciplinary Project* (Ed. by K. Immelmann, G. W. Barlow, L. Petrinovich & M. Main), pp. 23–57. New York: Cambridge University Press.

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

Stern, W. 1911. *Die differentielle Psychologie in ihren methodischen Grundlagen* (*Methodological Foundations of Differential Psychology*). 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., Novak, M. A. & Well, A. 1996. Aging in rhesus monkeys: different windows on behavioral continuity and change. *Developmental Psychology*, **32**, 1116–1128.

Tooby, J. & Cosmides, L. 1990. On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. *Journal of Personality*, **58**, 17–67.

Uher, J. 2005. Personality in the great apes: methods and approaches. Masters thesis, Humboldt University Berlin and Max Planck Institute for Evolutionary Anthropology.

Vochteloo, J. D., Timmermans, P. J., Duijghuisen, J. A. & Vossen, J. M. 1991. Responses to novelty in phobic and non-phobic cynomolgus monkeys: the role of subject characteristics and object features. *Behaviour Research and Therapy*, **29**, 531–538.

Weiss, A., King, J. E. & Figueredo, A. J. 2000. The heritability of personality factors in chimpanzees (*Pan troglodytes*). *Behavior Genetics*, **30**, 213–221.

Appendix

Table A1. Means, standard deviation and temporal reliability α and *r* of individual differences in all single behaviours listed by trait and situation

Trait	Situation	Single behaviour and method of its measurement		Mean t_1	SD t_1	α	*r*
Aggressiveness	Cage intruder test	Quasi-aggressive	*F*	0.80	1.42	**0.95**	0.90
	Keeper interaction test	Quasi-aggressive	*F*	0.87	1.56	0.48	0.32
	Masked human test	Aggressive	*F*	1.87	4.26	**0.97**	0.93
		Quasi-aggressive	*F*	4.22	8.11	**0.94**	0.89
	Sudden noise test	Quasi-aggressive	*F*	3.73	11.02	**1.00**	1.00
Arousability	Masked human test	Pilo-erection	*Z*	0.73	0.46	**0.85**	0.74
	Pile of food test	Change position	*F*	4.13	3.01	**0.90**	0.80
		Grin	*D*	1.99	5.40	**0.86**	0.75
		Pace	*D*	13.82	20.74	**0.96**	0.93
		Rock	*D*	12.93	29.11	**0.92**	0.84
		Scratch	*F*	17.69	31.49	**0.86**	0.76
		Shake wrist	*D*	0.17	0.37	0.40	0.25
		Vocalize	*D*	2.02	4.60	**0.87**	0.78
	Prefeeding observation	Change position	*Z*	12.36	8.77	**0.79**	0.66
		Grin	*Z*	2.91	5.71	**0.65**	0.48
		Pace	*Z*	33.91	15.15	**0.73**	0.65
		Rock	*Z*	7.36	15.89	**0.95**	0.90
		Sexual activity	*Z*	6.45	11.63	**0.57**	0.41
		Vocalize	*Z*	18.96	21.67	**0.95**	0.91
	Sudden noise test	Pilo-erection	*Z*	0.27	0.46	**0.79**	0.65
		Scratch	*F*	7.67	14.42	**0.94**	0.89
Anxiousness	Cage intruder test	Take raisin from intruder*	*P*	96.88	12.50	**1.00**	1.00
	Masked human test	Climb off initially	*Z*	0.40	0.51	**0.73**	0.58
		Take grapes from masked human	*L*	129.50	106.64	**0.73**	0.58
	Sudden noise test	Climb off initially	*Z*	0.40	0.51	**0.77**	0.63
Competitiveness	Food competition test	Grab finger of opponent inside box	*P*	24.13	23.71	**0.70**	0.53
Curiosity	Novel food test	Eat novel food	*P*	75.17	34.87	**0.76**	0.61
		Reject novel food*	*P*	17.53	24.23	**0.67**	0.51
		Touch novel versus normal food, difference score, not touching equated 60 s* (raw data)	*L*	4.51 (2.01)	12.95 (2.00)	**0.64**	0.47
		Deal with novel versus normal food, difference score	*D*	7.95	12.56	**0.81**	0.68
	Novel object test	Deal with novel object	*D*	152.28	168.91	**0.75**	0.60
Distractibility	Honey grid test	Recover honey*	*D*	272.11	22.94	**0.85**	0.74
Dominance	Food competition test	Gain banana piece	*P*	50.57	23.45	**0.92**	0.84
		Reach inside box	*P*	78.29	23.76	**0.78**	0.64
		Reach inside box, not reaching equated 5 s* (raw data)	*L*	1.87 (1.00)	0.93 (0.30)	**0.74**	0.59
Food orientation	Afternoon observation	Feed	*S*	33.02	14.01	**0.87**	0.78
	Food box test	Touch box, not touching equated 60 s* (raw data)	*L*	4.23 (3.38)	2.53 (2.21)	**0.84**	0.63
		Eat bait	*P*	93.66	8.69	**0.83**	0.72
		Reject bait*	*P*	3.42	6.23	**0.63**	0.46
Friendliness to youngsters	Afternoon observation	Groom youngster	*S*	1.07	2.91	**0.96**	0.92
		Physical contact to ≥1 youngster(s)	*S*	15.05	25.94	**0.99**	0.98
		Play with youngster	*S*	2.20	3.12	**0.89**	0.80
		Proximity to ≥1 youngster(s)	*S*	36.51	39.64	**0.98**	0.96
Friendliness to conspecifics	Afternoon observation	Body contact to ≥1 conspecific (s)	*S*	9.01	7.54	**0.85**	0.70
		Give grooming	*S*	3.51	5.23	**0.90**	0.82
Friendliness to humans	Cage intruder test	Sit in quadrant next to experimenter	*D*	98.77	51.22	**0.90**	0.81
	Keeper interaction test	Close contact to keeper	*D*	69.06	71.39	**0.89**	0.80
		Active close contact	*D*	22.50	49.97	**0.99**	0.98
		Passive close contact	*D*	46.55	63.33	**0.89**	0.84
		Sit in quadrant next to keeper	*D*	202.24	78.84	**0.86**	0.80
	Masked human test	Sit in quadrant next to experimenter	*D*	104.38	92.38	**0.89**	0.72

(*continued on next page*)

Table A1 (*continued*)

Trait	Situation	Single behaviour and method of its measurement		Mean t_1	SD t_1	α	r
Gregariousness	Afternoon observation	Proximity to 0 adult/adolescent conspecific*	*S*	61.67	17.38	**0.93**	0.87
		Proximity to $\geq$1 adult/adolescent conspecifics	*S*	27.50	10.77	**0.78**	0.64
		Proximity to $\geq$2 adult/adolescent conspecifics	*S*	10.60	10.42	**0.97**	0.94
Impulsiveness	Blocked food box test	Reach for food box	*F*	21.86	9.90	**0.63**	0.46
		Knock against panel or floor, clap	*F*	1.17	3.71	0.07	0.03
	Food out of reach test	Knock against panel or floor, clap	*F*	0.85	1.43	0.41	0.26
	Experimenter inside	Knock against panel or floor, clap	*F*	1.23	1.91	**0.78**	0.64
	Experimenter outside	Knock against panel or floor, clap	*F*	0.47	1.81	0.33	0.20
	Pile of food test	Knock against panel or floor, clap	*F*	5.38	12.94	**0.54**	0.37
Persistency	Afternoon observation	Deal with enrichment box	*S*	2.20	3.10	**0.85**	0.80
	Button box test	Press buttons	*D*	105.56	38.85	**0.89**	0.76
Physical activity	Afternoon observation	Rest*	*S*	33.86	17.08	**0.93**	0.87
		Move	*S*	55.92	16.01	**0.92**	0.84
		Change location	*S*	10.22	6.34	**0.75**	0.60
	Hidden food test	Rest*	*S*	26.46	19.79	**0.92**	0.84
		Change location	*S*	16.88	13.19	0.29	0.17
Playfulness	Afternoon observation	Social play with adolescent or adult conspecifics	*S*	2.95	4.44	**0.94**	0.88
		Solitary play without object	*S*	0.98	1.90	**0.92**	0.85
		Solitary play with object	*S*	4.47	6.09	**0.85**	0.74
Self care	Afternoon observation	Self-groom	*S*	2.69	3.29	**0.89**	0.81
		Self care	*S*	2.62	2.40	**0.71**	0.56
Sexual activity	Novel object test	Self-sexual activity	*D*	7.26	22.93	0.39	0.24
	Prefeeding observation	Self- and allo-sexual activity	*Z*	6.45	11.63	**0.57**	0.40
Vigilance	Hidden food test	Find hidden objects, not found equated 600 s* (raw data)	*L*	306.13 (120.44)	185.25 (111.80)	**0.95**	0.91
		Items found	*P*	58.75	35.61	**0.92**	0.86

Cronbach's α and Pearson correlation r are reported; for aggregations over time, zero–one coded data were treated as metric variables. *D*: duration (in s), *F*: frequency, *L*: latency (in s), *P*: percentage, *S*: scan sampling, *Z*: zero–one sampling.

*Inversed score used for further analyses to share the same meaning.