SUPPLEMENTAL MATERIAL

Original research article

Observations versus assessments of personality:
A five-method multi-species study reveals numerous biases in ratings
and methodological limitations of standardised assessments

Jana Uher *[a,b], Elisabetta Visalberghi [c]

[a] The London School of Economics and Political Science, United Kingdom
[b] Comparative Differential and Personality Psychology, Free University Berlin, Germany
*[c] Unit of Cognitive Primatology and Primate Centre, Institute of Cognitive Sciences and
Technologies, National Research Council of Italy (ISTC-CNR), Rome, Italy*

Correspondence:

Dr Jana Uher
The London School of Economics and Political Science
Department of Social Psychology
Houghton Street, WC2A 2AE London
United Kingdom
e-mail: uher @ primate-personality.net
www.primate-personality.net

Supplemental Material

Supplemental Information

Tables

Uher, J. & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, 61*, 61-79*, Supplemental Material.*

2/16

Supplemental Information

ad Section 2.3)
Generation of constructs of individual-specific behaviours ("personality") using the Behavioural Repertoire x Behavioural Situations Approach (BR$_x$BS-Approach)

Direct contrasts between observational and assessment-based categorisations of individual-specific behaviours presuppose a suitable approach for generating "personality" constructs and for selecting indicators. The TPS-Paradigm provides a system for classifying the approaches that are used in taxonomic "personality" research on the basis of their methodological rationales (for details, see Uher, 2008a, b, 2011a, b, 2015b). For example, in both human and animal research, many questionnaires and "personality" models were developed with *nomination approaches* in which knowledgeable informants nominate constructs and indicators that they deem important for a given population. But nomination approaches have only limited power to achieve a selection that is representative for the individual differences under study because they rely exclusively on the nominators' pertinent ideas who can therefore introduce all kinds of biases.

The Big Five Model of human "personality", by contrast, was derived from the person-descriptive words in the lexica of everyday language (John et al., 1988) using a so-called lexical manifest system (or bottom-up) approach (Uher, 2015b). This enables representative selections because the universe of all person-descriptors is comprehensively explicitly known and specified in the lexica. But everyday words encode people's everyday knowledge that contains the implicit beliefs, ideas and values that are shared in their particular socio-linguistic community. Lexically derived constructs and items are therefore not suitable to explore the ways in which assessments deviate from observations in a multi-species study.

To minimise anthropocentric biases and to systematically generate constructs of individual-specific behaviours ("personality") for our nonhuman study species (capuchin monkeys), we applied the Behavioural Repertoire x Behavioural Situations Approach (BR$_x$BS-Approach), which is a so-called behavioural manifest system (or bottom-up) approach (Uher, 2008a, b, 2011a, b, 2015b). It allows researchers to systematically generate "personality" constructs on the basis of the behaviour-scientific knowledge that the scientific community has already established about the behavioural repertoire of a study species. Given that the behavioural science literature has no particular focus on "personality" and individual differences, influences of preconceived ideas on the part of the scientists can be minimised. This is essential for unravelling possible biases in assessments and for exploring how raters may interpret "personality" descriptors with regard to observable behaviours.

The BR$_x$BS-Approach was already applied to explore individual-specific behaviours in great apes (Uher et al., 2008) and crab-eating macaques (Uher et al., 2013b) and the socially shared mental representations that human observers have developed of these primate individuals (Uher, 2011b; Uher & Asendorpf, 2008). Systematic contrasts revealed that assessments reflected several attribution biases likely derived from stereotypical beliefs about *human* individuals—thus, anthropomorphic biases (Uher et al., 2013b).

Construct generation procedure
To systematically generate constructs of individual-specific behaviours for capuchin monkeys, we conducted a broad-based review of 68 publications about the behavioural repertoire of wild and captive capuchin monkeys that were available at the start of our study. From these publications, we compiled a large table with all major behavioural categories used in these studies (listed in one column) together with the categories of situations in which these behaviours were described to commonly occur (listed in a second column). Each row of the table thus represents a unit of a particular behavioural category and a particular associated situational category as described in a given publication; this is called a *behaviour$_x$situation-unit* in the BR$_x$BS-Approach. The primary compilation of categories is

Uher, J. & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, 61*, 61-79*, Supplemental Material*.

3/16

designed to be over-inclusive, repeatedly listing the same behavioural and situational categories used in different publications.

Then we reorganised this table by grouping together categories describing the same or functionally similar behaviours; for example, we grouped together different behaviours of aggression such as bite and slap as acts of contact aggression and chase and threaten as acts of non-contact aggression. We also grouped together associated situational categories within each behavioural category; for example, aggressive behaviours in same-sex and opposite-sex interactions in dyads and groups in intra-group and inter-group contexts. Finally, the *behaviour$_x$situation-units* were organised hierarchically, more specific behavioural and situational categories were subsumed under more abstract categories.

To generate "personality" constructs, we used *behaviour$_x$situation-units* on moderate levels of abstraction that reflect relatively homogeneous subsets of still identifiable concrete behaviours (e.g., contact and non-contact aggression) and situations (e.g., intra-group, inter-group and inter-specific contexts). We generated "personality" constructs by *hypothetically* assuming individual-specific patterns in the given behaviours and situations described (e.g., in aggressive behaviours to conspecifics, thus aggressive*ness* to conspecifics). Given this hypothetical generation, the thus-generated constructs (listed in a third table column) are therefore called *working constructs*. They serve methodological purposes to systematically guide the researchers' decisions of what to study, but they do not imply a priori empirical usefulness.

Given the over-inclusiveness of the compilation, the same working constructs were generated repeatedly in different parts of the category system. This is the essential prerequisite of the BR$_x$BS-Approach that enables researchers to systematically generate "personality" constructs by considering the entire known behavioural repertoire of the population under study. In a second identical table, we then sorted the rows by the generated constructs and eliminated redundant enumerations of the same constructs to obtain a comprehensive overview of all generated constructs and the major behavioural and situational categories in which they describe individual-specificity (more details of this construct generation process and a list of all 68 capuchin publications reviewed are provided in Uher et al., 2013a). This procedure of the BR$_x$BS-Approach yielded 21 "personality" constructs for capuchin monkeys. The 21st construct describing individual-specific behaviours in relation to youngsters was considered only in the analyses of item analyses because not all capuchin groups had young individuals at the time of our study. Working constructs describing behaviours and situations that occur only in the wild, such as territoriality and travelling, were not considered.

Peculiarities of BR$_x$BS-Approach-generated constructs

Given their origins in the behaviour-scientific knowledge bases, BR$_x$BS-Approach-generated constructs are labelled with terms that are much less colloquial than those derived from the human everyday languages (Uher, 2015b; Uher et al., 2013a, b). For example, individual differences in the tendency to engage in social interactions (e.g., grooming) are labelled Social orientation rather than Extraversion, a term that is increasingly used also for animals. This meets efforts to reduce the impact of implicit meanings and anthropomorphic biases. The concepts of Extraversion versus Introversion, for example, denote individuals' tendency to focus their perception and judgement on the outer world versus their inner private world (Jung, 1921) and therefore do not apply well to animal individuals who cannot report about themselves.

A further difference to other methodological approaches is that the BR$_x$BS-Approach generates constructs and not measurement variables (indices) that are considered only in a second step. Their targeted selection for construct operationalisation helps to keep their number manageable for empirical studies, while still enabling comprehensive investigations within the same theoretical framework (if needed using multiple studies; see Uher, 2008a, b, 2011a, b, 2015b).

Uher, J. & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, 61*, 61-79*, Supplemental Material*.

4/16

ad Section 2.5.1) Scientific quantification of individual and individual-specific behaviours

All behavioural tests were videotaped and coded using a detailed encoding scheme and the coding software INTERACT (Rel. 9.2.1, www.behavioural-research.com; Mangold, 2010). We measured the latencies, frequencies and durations of the behaviours that the individuals showed during the fixed test times, ranging from 2 to 10 min per session. In the Prefeeding observations (i.e., 5 min before the distribution of the main meal), behaviours were recorded live on check sheets using one-zero sampling with 10-sec time intervals to estimate frequencies that included any amount of time spent in the respective behaviours (Altmann, 1974). In the Social observations, carried out after the main feeding, we combined three observational methods to estimate time distributions of frequency and duration behaviours: *(a)* focal individual sampling continuous recording of duration and frequency behaviours for 10 min-periods per day and individual; *(b)* scan sampling instantaneous recording every 10 min to record the presence (or absence) of duration behaviours for all individuals of the group; and *(c)* event recording of rare frequency and duration behaviours (e.g., aggressive behaviours). Behaviours were recorded using an interactive computer software programme developed by JU that logged all data entries with a precise timestamp.

To reduce the impact of day-to-day fluctuations on individuals' scores and to generate data that reflect time-relative behavioural probabilities, 10 laboratory-based test situations were repeated twice within a *study block* of about 2-2.5 weeks; 5 tests that may constitute mildly disturbing situations for capuchin monkeys were repeated just once. To avoid studying individual behaviours that are related to the same "personality" construct several times a day, all tests were carried out in a pseudo-random sequence. Social observations occurred on 10 days: in the Prefeeding observations on two zero-one sample points per day and in the Social observations for one 10-min focal sample per individual and day and 7 scan sample points per observation day.

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49, 229–267.

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.

Jung, C. G. (1921). *Psychologische Typen*. Zürich: Rascher.

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

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. King, & L. Murray (Eds.), *Personality and Temperament in Nonhuman Primates* (pp. 41-76). New York, NY: Springer.

Uher, J. (2015b). Developing "personality" taxonomies: Metatheoretical and methodological rationales underlying selection approaches, methods of data generation and reduction principles. *Integrative Psychological and Behavioral Science, 49*, 531-589.

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., Addessi, E., & Visalberghi, E. (2013a). 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*, 427-444.

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. (2013b). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality, 47,* 647–667.

Uher, J. & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, 61*, 61-79, Supplemental Material.

5/16

Tables

Table S1 Behavioural tests and observations, their situational description and the working constructs of individual-specific behaviours ("personality") measured

Blocked food tube test. Highly preferred food items (half Cheerios) were dropped one by one into a transparent tube fixed at a 45° angle to the monkey's cage. The tube had a thin slot into which different transparent plastic slides could be inserted. In two first trials, the experimenter inserted a hollowed slide through which the food could fall inside the monkey's cage. In a third trial, she inserted a solid slide on which the food (whole Cheerios) piled up in full view of the monkey but out of his/her reach. Constructs measured: *Arousability, Food orientation, Impulsiveness.*

Conveyor belt test. The experimenter placed food of different desirability and quantity successively on a small conveyor belt fixed to the cage. The monkey could move the conveyor belt by turning a wheel, thereby transporting the food into his/her cage.
Construct measured: *Food orientation.*

Conveyor belt disconnected test. The same conveyor belt was baited with highly preferred food. It still looked the same, but the internal mechanism was disconnected so that the monkey could still turn the wheel yet without any effect on the conveyor belt.
Constructs measured: *Arousability, Impulsiveness.*

Food competition test. Two monkeys, kept in the same cage, were offered one piece of preferred food when they were both at approximately the same distance from it.
Constructs measured: *Aggressiveness, Competitiveness, Dominance.*

Foraging box test. A box on which the monkey could sit was fixed inside the cage. The box was filled with wood shavings, in which three 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 and to retrieve seeds (and substrate).
Constructs measured: *Persistency, Vigilance.*

Furry animal test. A small soft toy attached to a disc was placed in front of the cage with its face away from the monkey. After one minute, the experimenter rotated the disc from a 2 m distance so that the soft toy now faced the monkey for a further minute. In all trials, the eyes of the soft toy were covered with crepe tape to reduce the degree of threat that soft toys generally constitute for capuchin monkeys.
Constructs measured: *Aggressiveness, Anxiousness, Arousability.*

Hidden food test. The monkey entered the test cage in which small food pieces were hidden on different elements or stuck with honey to the variegated walls.
Constructs measured: *Arousability, Anxiousness, Physical activity, Social orientation to conspecifics, Social orientation to humans, Vigilance.*

Human interaction test. The experimenter sat in front of the cage and encouraged the monkey to approach and play without offering any food for two minutes. Then she offered food for one minute before she put the food out of sight and encouraged the monkey again to approach and play for two further minutes.
Constructs measured: *Arousability, Aggressiveness to humans, Social orientation to humans.*

Large cloth test. The monkey encountered a large bed sheet hung up transversally in the test cage and over the wooden perch in the rear end of the cage.
Constructs measured: *Aggressiveness, Anxiousness, Arousability, Curiousness, Creativeness/ inventiveness, Physical activity, Social orientation to humans.*

Masked human test. Disguised with a Venetian mask, a wig, a long dress and boots, the experimenter silently entered the test room and continuously offered highly preferred food in gloved hands. To enable direct contact, she stuck the stiffed fingers of her right's glove through the mesh into the monkey's cage.
Constructs measured: *Aggressiveness to humans, Anxiousness, Arousability, Social orientation to humans.*

Multiple objects test. The monkey encountered six small different objects both familiar and unfamiliar in the middle of the cage.
Constructs measured: *Aggressiveness, Anxiousness, Arousability, Creativeness/ inventiveness, Curiousness, Physical activity.*

Novel food test. The monkey received in alternating order four pieces of normal food and four items of novel food that he/she never had before, in total eight food items.
Constructs measured: *Curiousness, Food orientation.*

Sudden noise test. A foreign news program was suddenly played back to the monkey in moderate volume inside the test room and independent of the experimenter's activities twice for 10 sec with a break of 20 sec.
Constructs measured: *Aggressiveness, Anxiousness, Arousability, Vigilance.*

Tunnel basket test. A large, open-worked laundry basket of pale blue PVC was placed in the middle of the cage with its larger opening towards the door through which the monkey entered. The basket was open at both ends. But at the smaller end, a dark tubular cloth was attached that looked like a scoop net but that was actually open so that the monkey could pass through it.
Constructs measured: *Aggressiveness, Anxiousness, Arousability, Creativeness/ inventiveness, Curiousness, Physical activity, Social orientation to humans.*

Yoghurt grid test. A platform was fixed outside to the cage and baited with plain yoghurt that the monkey could reach through the mesh with his/her fingers and tongue, while the experimenter continuously produced noise by knocking with a plastic tube on the metal frame of the cage 2 m away from the yoghurt platform.
Constructs measured: *Distractibility.*

Prefeeding observation. Prior to their daily main feeding, the monkeys could hear and see the keepers preparing their food in the nearby kitchen and distributing it in the enclosures of the neighbouring groups before they finally received their own food.
Constructs measured: *Arousability, Food orientation, Social orientation to conspecifics.*

Social observations. The monkeys were all together in their groups in their naturally designed outdoor enclosures without access to their indoor enclosures.
Constructs measured: *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. More details are reported in Uher, Addessi & Visalberghi (2013a).

Table S2 The Capuchin Personality Inventory – Behaviour-Descriptive Verb items (CPI-BV): Inter-rater reliability, temporal reliability and raters' item interpretations

Complete statements of the Behaviour-Descriptive Verb Items (BV)		Inter-rater reliability [c]		Temporal reliability [d]	
Item code [a]	Global interpretations of the contextualised behaviours described [b]	$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] starts agonistic interactions in the group by threatening and/or chasing other group members.		**.810**	**.630**	**.90**	.000
AGCPB1	aggressive (5), despotic (2), dominant (3), at the upper levels of the hierarchy, supervises everything and everyone in the group, chasing, bullying, threatening, nervous, hostile				
When somebody stays in front of the cage, [Name] jumps at the grate, and may also try to grab that person.		**.790**	**.620**	**.68**	.000
AHCPB1	aggressive towards humans (3), tries to catch the person's attention in an assertive and forceful way, not confident with humans, nervous, maybe feels threatened by the person, bold, not fearful toward humans, angry (sometimes), social and extrovert with humans, playful, interactive				
When there are unusual noises outside the cage, [Name] starts pacing and/or scratching.		.480	.220	.36	.063
ARCPB1	anxious (3), reacts nervously (to sudden events) (2), scared (by noises) (2), frightened (2), always in alarm, fearful, agitated, upset				
When awaiting the feeding, [Name] paces restlessly and/or scratches him/herself.		.290	.200	**.58**	.002
ARCPB2	nervous (4), subordinate (3), anxious (2), cannot stand the waiting time, scared by other dominant monkeys, excited about the food, impatient, frightened, stressed, obsessive				
In social conflict situations, [Name] screams quickly and flees from others.		**.790**	**.590**	**.80**	.000
AXCPB1	subordinate (3), nervous (2), fearful (2), screams to obtain others'/ kin support (2), frightened, feels threatened by others, scared, vulnerable, submissive, upset, shy				
[Name] keeps a distance to unknown objects, persons, and/or avoids uncertain situations.		**.650**	.440	.32	.109
AXCPB2	neophobic (3), fearful (2), frightened (of new situations) (2), avoidant, diffident, cautious, shy, subordinate, seems to feel uncomfortable				
In the group, [Name] starts conflicts over food or foraging places with others.		**.730**	**.510**	**.89**	.000
COCPB1	aggressive (4), dominant (4), competitive (2), feels self-confident, intolerant, wants to outbrave dominant subjects, bold, bully, threatening, hostile, territorial				

Complete statements of the Behaviour-Descriptive Verb Items (BV)		Inter-rater reliability [c]		Temporal reliability [d]	
Item code [a]　Global interpretations of the contextualised behaviours described [b]		$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] displaces others that are groomed by others or that huddle together with a social partner.		**.730**	**.520**	**.65**	.000
COCPB2	despotic (2), dominant (2), possessive (2), jealous (2), envious, competitive, aggressive				
In his/her activities, [Name] involves several objects.		**.610**	**.400**	**.80**	.000
CRCPB1	manipulative (6), active (3), curious (3), explorative (2), creative, tool user, smart				
[Name] explores new, potentially edible materials by sniffing, touching, and/or mouthing them.		.350	.290	**.80**	.000
CUCPB1	curious (6), explorative (5), manipulative (2), less neophobic, not fearful, bold, more interested in what the others do, gluttonous				
[Name] readily explores changes in the environment.		.480	.370	**.73**	.000
CUCPB2	curious (4), active (3), vigilant, attentive, experimental, less signs of neophobia, general fear or anxiety, careful, explorative, approaching, dominant,				
[Name] can focus long on activities that take effort and time. (inversed meaning)		**.630**	.410	**.47**	.013
DICPB1*	persistent (3), patient (2), not distractible, tenacious, active, focused (e.g. on experimental tasks or when searching for food), goal-focused, diligent, accurate, precise, concentrate, rugged				
In the group, [Name] can occupy the best places for foraging and resting.		**.870**	**.720**	**.92**	.000
DOCPB1	dominant (6), respected, other members of the group let him doing it because they recognise his position, prestigious, aggressive				
Already at a distance, [Name] makes way for others and/or avoids others that are approaching. (inversed meaning)		**.820**	**.660**	**.89**	.000
DOCPB2*	subordinate/submissive (6), avoidant (e.g., aggression, unsafe encounters) (2), fearful (2), scared by others, therefore always keeps an eye on others, shy				
During daytime, [Name] spends much time searching for food and eating.		**.520**	**.310**	**.74**	.000
FOCPB1	hungry (2), active (2), gluttonous (2), persistent, hard-working, solitary, less interested in social activities, independent, explorative, this is quite usual for capuchins				
When there is food, [Name] is quickly on the spot.		.190	.160	**.74**	.000
FOCPB2	dominant (4), voracious/ gluttonous (2), no fear, hungry, reactive, active, alert, vigilant, fast				

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material.*

Complete statements of the Behaviour-Descriptive Verb Items (BV)	Inter-rater reliability [c]		Temporal reliability [d]	
Item code [a] Global interpretations of the contextualised behaviours described [b]	*ICC*(3,*k*)	*ICC*(3,*1*)	r_{tt}	*p*
[Name] sits close together with other members of the group.	**.770**	**.590**	**.81**	.000
GRCPB1 sociable/ social (4), friendly (3), dominant, less subordinate, has good relationships with dominant individuals, no fear and so stays close together with others, better included in the social context, extravert, cooperative				
During daytime, [Name] spends much time on his/her own. (inversed meaning)	**.640**	.420	**.75**	.000
GRCPB2* solitary (5), isolated (2), subordinate (3), asocial (2), shy, independent, seems to feel uncomfortable with the rest of the group				
When [Name] does not get his/her food or reward immediately, he quickly bangs against the mesh or tries to get it forcefully.	**.740**	**.540**	**.77**	.000
IMCPB1 less able to wait (3), nervous (3), Impulsive (2), impatient, agitated, possessive, gluttonous				
During daytime, [Name] constantly moves around in the enclosure.	**.540**	.340	**.82**	.000
PACPB1 active (2), nervous (2), excited, agitated, anxious, maybe isolated from the rest of the group, maybe subordinate, has to pay attention to others and therefore constantly moves around, explorative, obsessive, upset, seems to feel uncomfortable				
During daytime, [Name] takes rests. (inversed meaning)	**.400**	.170	.31	.119
PACPB2* calm (2), relaxed (2), less active (2), inactive, peaceful, comfortable with situation, lazy				
When trying to obtain food that is scattered or only difficult to get, [Name] can spend considerable time and effort without interrupting his/her activity.	.260	.190	**.83**	.000
PECPB1 hungry (2), goal-directed/ goal-focused (2), persistent (2), determinate to achieve his goal, not distractible, wilful, focused, concentrated, pays little attention to the rest, obstinate, tenacious, stubborn,				
[Name] plays with materials and/or objects in his/her environment on his/her own.	**.820**	**.650**	**.77**	.000
PLCPB1 curious (3), manipulative (2), explorative (2), active (2), playful (2), not very social, less involved in social activities, solitary				
[Name] engages in rough-and-tumble play or play chases with his/her group members.	**.890**	**.750**	**.77**	.000
PLCPB2 playful (5), active (4), sociable/social (3), gregarious, better included in the social context, friendly				

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material.*

Complete statements of the Behaviour-Descriptive Verb Items (BV)	Inter-rater reliability [c]		Temporal reliability [d]	
Item code [a] Global interpretations of the contextualised behaviours described [b]	$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] cleans himself/herself intensely by self-licking or self-grooming.	.260	.080	.07	.730
SCCPB1 nervous (2), uncomfortable, has dirty fur, anxious, frightened, not necessarily more anxious, in a vulnerable position or stands a controversial unrewarding situation restless, possibly more alone, solitary, clean, accurate, precise				
To persons approaching or standing in front of the cage, [Name] approaches, lip-smacks, and/or shows scalp lifts.	**.570**	.380	**.81**	.000
SHCPB1 friendly with people (3), wants to interact with human (therefore lip-smacks and shows scalp lifts) (2), sociable/social (2), tolerant to humans, tries to catch the person's attention in an affiliative way, confident with humans, always happy to see them, may also desire something to eat, approaching				
[Name] approaches and lip-smacks to other group members.	.250	.130	**.64**	.000
SOCPB1 sociable/social (3), friendly (3), willing to be accepted by the other group members, has good relationships with other group members, gregarious, extrovert, wants to interact with others, cooperative, approaching, reconciling (if after a fight), sexually receptive (when in sexual context)				
[Name] touches and also grooms other group members.	**.640**	.440	**.76**	.000
SOCPB2 sociable/ social (4), friendly (2), has good relationships with others, maybe not high in the hierarchical scale, grooms others to reinforce his/her social relationships, better included in a social context, less fearful, approaching, extrovert, cooperative				
[Name] feeds together with others in close proximity.	**.650**	.430	**.65**	.000
SOCPB3 sociable/social (4), well integrated in his/her group (2), friendly (2), has good relationships with others, confident, cooperative, probably they are relatives				
[Name] tries to contact others sexually with scalp lifting, head cocking, grinning, and/or chest rubbing.	**.710**	.400	**.76**	.000
SXCPB1 (easily sexually) aroused, more often but not always dominant, courtships, socially active, sexually active, female in oestrus, approaching				
[Name] watches everything around him/her closely.	**.560**	.220	**.48**	.011
VICPB1 curious (4), vigilant (4), explorative (2), pays more attention to the surrounding, more careful in her explorative behaviour, alert, attentive, aroused, maybe less secure				

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material.*

Complete statements of the Behaviour-Descriptive Verb Items (BV)	Inter-rater reliability [c]		Temporal reliability [d]	
Item code [a] Global interpretations of the contextualised behaviours described [b]	$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] quickly spots small food items, potential prey or changes in the environment.	**.540**	.190	**.54**	.004
VICPB2 attentive (3), vigilant (3), active (2), alert (2), interested in/ careful to what is happening around (2), curious, always aroused less "sleepy", accurate				
[Name] observes youngsters with interest, tries to get close to them, and/or initiates playful interactions with them. [d]	.407	.146	**.85**	.000
YOCPB1 playful (3), approaches youngsters (2), friendly with youngsters (2), shows attachment to youngsters, takes care of youngsters, maternal, collaborative, tolerant, sociable, extrovert				
[Name] takes care of youngsters, carries, embraces, and/or grooms them. [d]	**.571**	.250	**.93**	.000
YOCPB2 friendly (with youngsters) (4), sociable/social (2), shows attachment to youngsters, maternal, probably is a relative, amenable, better included in the social environment, approaching, extrovert				

Note. [a] The first two digits of the item code abbreviate the construct, the second two the species (CP = capuchin) and the last two the item format (AD = trait-adjective items, B1 to B3 = behaviour-descriptive verb items). Construct abbreviations: AG Aggressiveness to conspecifics, AH Aggressiveness to humans, AR Arousability, AX Anxiousness, CO Competitiveness, CR Creativeness/ Inventiveness, CU Curiousness, DI Distractibility, DO Dominance, FO Food orientation, GR Gregariousness, IM Impulsiveness, PA Physical activity, PE Persistency, PL Playfulness, SC (Self-)Cleanliness, SH Social orientation to humans, SO Social orientation to conspecifics, SX Sexual activity, VI Vigilance, YO Social orientation to youngsters. [b] Numbers in parentheses indicate the frequencies by which a given lexical element occurred in the six raters' item interpretations. [c] Mean inter-rater reliabilities of all institution-specific scores. [d] Determined only for the ISTC-CNR sub-sample. Bold *ICC* and r_{tt} scores >.50. * items with inversed meaning.

Descriptions of the scales are provided in the main article.

Table S3 The Capuchin Personality Inventory – Trait-Adjective Items (CPI-TA): Inter-rater reliability, temporal reliability and raters' item interpretations

Complete item statement Item code [a] Interpretations in terms of typical behaviours and situations [b]	Inter-rater reliability [c]		Temporal reliability [d]	
	$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] is aggressive to conspecifics. AGCPAD threatens others (4), chases others (3), starts conflicts (3), attacks physically, becomes aggressive in conflict situations (over food, mates, etc.)	**.780**	**.580**	**.89**	.000
[Name] is aggressive to humans. AHCPAD threatens humans (4), tries to grab humans (4), tries to bite humans (2), throws objects against humans, jumps at the grate, bangs toward it, usually comes to the mesh, shows teeth	**.730**	**.520**	**.74**	.000
[Name] is excitable. ARCPAD gets nervous (2), produces loud vocalizations (when expecting food) (2), reacts in exaggerated ways in the presence of food and/or of human friends, immediately reacts to noises and movements inside and outside the enclosure, attentive to noises and movements in and outside the enclosure, active, displays emotions to others, paces restlessly around the enclosure, easily excited, gets pleased, easily gets cheerful, tries to court others, usually shows them scalp-lifting and chest-rubbing, can also be very playful in some situation (especially with youngsters)	**.730**	**.530**	**.59**	.001
[Name] is anxious. AXCPAD scratches (4), vigilant (also toward others' behaviour) (3), paces restlessly around the enclosure (2), moves more around in the enclosure, seems to be constantly aroused, cautious to environment and others, continuously monitors the environment in search for potential threats, watches everything around him with attention, typically reacts to unusual or sudden noises outside the cage or actions of others or persons, is always in alarm, alert, alarm calls, easily scared by unusual noises and situations, signs of distress, self-licks, self-grooms	.420	.250	**.70**	.000

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material.*

| Complete item statement | Inter-rater reliability [c] | | Temporal reliability [d] | |
Item code [a] Interpretations in terms of typical behaviours and situations [b]	ICC(3,k)	ICC(3,1)	r_{tt}	p
[Name] is competitive. COCPAD tries to have the best places (for resting, to stay) (3), tries to have best food/ best places for foraging (3), contests for best resources (2), starts conflicts (over food or foraging places with others) (2), responds to aggressions, does not avoid conflicting situations, could also threaten other group members, can become aggressive toward others and also attack them to obtain preferred things, able to grab food, tries to get the best mates	**.830**	**.660**	**.93**	.000
[Name] is inventive. CRCPAD able to adopt new strategies/ to invent new methods/ tools to solve problems/ tasks (4) manipulates objects (2), uses and combines objects/ materials in creative/ innovative ways (2), able to solve problems using tools, curious, explorative, smart	**.740**	**.530**	**.72**	.000
[Name] is curious. CUCPAD interested in, readily approaches, explores (touch, sniff, study, manipulate), uses novel objects (6), readily explores novel situations, (changes in the) environment (4), interested in/ approaches easily novel food (2), less/not neophobic (2), observes everything and everyone with interest, explorative, plays with novel objects, not scared by the new environmental elements, moves around the environment more than the others	**.770**	**.560**	**.73**	.000
[Name] is distractible. DICPAD when involved in activities or during an experiment, he/she easily interrupts this activity because of unusual noises or movements of group members in the enclosure (2) cannot focus (his attention) long on a task/ an object (2), not able to focus on something without interrupting his/her activity, easy attracted to what is around him (e.g., if she/he hears a noise during a task), he stops and looks around, his/her attention is captured by environmental sounds and uncertain social situations, usually scans the disclosure in search for potential threats	.320	.200	.38	.052

| Complete item statement | Inter-rater reliability [c] | | Temporal reliability [d] | |
| Item | | | | |
code [a] Interpretations in terms of typical behaviours and situations [b]	ICC(3,k)	ICC(3,1)	r_{tt}	p
[Name] is dominant.	.930	.830	.92	.000
DOCPAD takes the best/ better food before the others or occupies the best locations to feed (4), reaches for the food/ can feed before the others (3), occupies the best places (for rest and sleep) (3), received grooming (3), is the first to mate (2), almost never is the victim in conflicts or receives threats or aggressions, sometimes more aggressive, chasing, threatening others, aggressive to others, displays sometimes bullying behaviours moves freely in the environment, when he/she approaches others leave, stays often with other group members, usually approaches first humans outside the cage				
[Name] is gluttonous.	.770	.570	.89	.000
FOCPAD spends more time and efforts foraging and eating (4), tries to reach for food quickly/ is immediately on the spot (2), first collects as much food as possible (2) and then consumes it away from others, ingests big amounts of food, is used to eat also when he is not hungry, tries to get the better food; when there is food, also tries to displace other group members; vocalizes when something good is given, when persons stay near the cage jumps at the grate begging for food, tries always to get food from humans				
[Name] is gregarious.	.420	.260	.77	.000
GRCPAD spends more time in close proximity with other group members (4), grooms (with) others (2), feeds near/ together with others (2), plays with others (2), spends much close to other group members (whether directly interacting with them or not), receives grooming, affiliated, social, usually is engaged in positive and cooperative situations with the other group members, likes to play also with the youngsters				
[Name] is impulsive.	.620	.280	.78	.000
IMCPAD not able/ does not like to wait to receive/ do something (e.g., food) (3), less able to refrain from reaching for food or other interesting things (2), nervous (2), during experiments, tries to solve the task quickly, if the reward arrives not immediately or not at all, bangs against the mesh, throws tools or other things toward the experimenter, acts by instinct, has no patience, hurried, tries to get the reward forcefully, in social relationships, becomes competitive to have better access to resources, is usually dominant				

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material.*

Complete item statement Item code [a]	Interpretations in terms of typical behaviours and situations [b]	Inter-rater reliability [d]		Temporal reliability [c]	
		$ICC(3,k)$	$ICC(3,1)$	r_{tt}	p
[Name] is physically active. PACPAD	moves around for long periods (5), explores the environment (2), jumps (on the environmental enrichments) (2), climbs the environmental enrichments or mesh (2), spends less time resting or being inactive (2), plays with others (2), manipulates and combines objects, forages for a long time	**.700**	.490	**.81**	.000
[Name] is persistent. PECPAD	when engaged in an activity, able to spend much time doing/ exploring the same thing/ for longer time without being distracted (6), can spend a lot of time/ energies/ effort doing the same thing to obtain his aim (2), pursues a goal until he/she reaches it (e.g., when a food is difficult to get), goal-focused, motivated, accurate, does not give up even when facing difficulties	**.600**	.270	**.80**	.000
[Name] is playful. PLCPAD	engages in rough-and-tumble play/ play chases with others (e.g., gently grabbing and biting them like in a simulated fight, runs peacefully after others (6), plays/ wants to play with others (3), plays with objects (2), plays with the youngsters, plays alone jumping, climbs trunks and ropes, plays with humans gently grabbing and biting their hands and clothes, runs around in the enclosure, always active, social, cooperative	**.820**	**.660**	**.83**	.000
[Name] is cleanly with him/herself. SCCPAD	spends a lot of time grooming and licking himself (3), does not present incrustation on the fur possibly because others groom him/her or because he/she self-grooms, usually scared or upset, shows this behaviour to calm down and console himself, his/her behaviour can become quite obsessive, I don't know capuchins cleaner than others	.010	-.100	.16	.419
[Name] is friendly to humans. SHCPAD	displays affiliative expressions like lip-smacking and scalp-lifting to a person near the cage (6), approaches humans (4), tries to (gently) play with humans (2), tries to groom humans, is often close to the mesh, allows you to touch him/her easily, not fearful towards humans	**.550**	.330	**.76**	.000

Uher, J. & Visalberghi, E. (in press 2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality, Supplemental Material*.

Complete item statement Item code [a] Interpretations in terms of typical behaviours and situations [b]	Inter-rater reliability [c]		Temporal reliability [d]	
	ICC(3,k)	ICC(3,1)	r_{tt}	p
[Name] is friendly to group members. SOCPAD grooms others (4), sits together with other monkeys (4), displays affiliative expressions/ lip-smacks, shows scalp lifts to others (3), plays with others (2), spends a lot of time interacting in positive ways with other group members, doesn't start conflicts in the group, does not threaten other group members, approaches others, social, playful, cooperative, likes to eat together with others	.440	.230	**.84**	.000
[Name] is sexually active. SXCPAD courts others through lips making, scalp lifting, grinning, head cocking (6), copulates (2), sexually aroused, spends lots of time trying to find mates	**.800**	**.620**	**.84**	.000
[Name] is vigilant. VICPAD watches everything around him/her closely (5), attentive to noises and movements in and outside the enclosure/ group (3), less focussed on activities, can get easily distracted by other things happening around, alert, seems frightened when there is a strong sound, begins to move around the cage in an obsessive manner, can also scream	.340	.210	**.60**	.001
[Name] is friendly to youngsters. [d] YOCPAD plays with youngsters (6), shows attachment toward youngsters, takes care of youngsters, shows cuddling or other affiliative behaviours, spends part of his time staying in contact with youngsters, behaves with youngsters in a very friendly manner, assumes sort of maternal role, game is very important and a way to learn new skills for the youngsters	**.813**	**.521**	**.83**	.000

Note. [a] The first two digits of the item code abbreviate the construct, the second two the species (CP = capuchin) and the last two thw item format (AD = trait-adjectives, B1 to B3 the behaviour-descriptive verb items). Construct abbreviations: AG Aggressiveness to conspecifics, AH Aggressiveness to humans, AR Arousability, AX Anxiousness, CO Competitiveness, CR Creativeness/ Inventiveness, CU Curiousness, DI Distractibility, DO Dominance, FO Food orientation, GR Gregariousness, IM Impulsiveness, PA Physical activity, PE Persistency, PL Playfulness, SC (Self-)Cleanliness, SH Social orientation to humans, SO Social orientation to conspecifics, SX Sexual activity, VI Vigilance, YO Social orientation to youngsters. [b] Numbers in parentheses indicate the frequencies by which a given lexical element occurred in the six raters' item interpretations. [c] Mean inter-rater reliabilities of all institution-specific scores. [d] Determined only for the ISTC-CNR sub-sample. Bold *ICC* and r_{tt} scores >.50. Descriptions of the scales are provided in the main article.