

Chapter 2

Comparison of Methods for Assessing Personality in Nonhuman Primates

Hani Freeman, Samuel D. Gosling, and Steven J. Schapiro

Abstract A review of the field of primate personality revealed that two main methods have been used to study primate personality: behavioral codings and observer trait ratings. Both of these methods can be broken down by the conditions in which they are used – naturalistic observation, testing context, or cumulative observation. The most commonly used method-condition combination to assess primate personality has been the behavioral coding/naturalistic observation combination. This combination was used in more than half (60%) of the studies reviewed. Strengths and weaknesses of each of the method-condition combinations are discussed. A review of empirical studies comparing the behavioral-coding and trait-rating methods revealed that the trait-rating method seems to be better at assessing primate personality. However, only three scales have been used for 66% of the primate personality studies using a trait-rating method. Two procedures are suggested in order to develop new ratings scales for studying primate personality. Recommendations are made for the best method/condition combination to use and how to improve the use of each combination when assessing primate personality.

2.1 Introduction

Empirical research on nonhuman primate personality began in earnest in the 1930s. In 1938, for example, Crawford conducted a study of chimpanzees, showing that a reliable rating scale could be developed to assess personality. In this groundbreaking study, personality was assessed by having individuals who were familiar with the animals rate them on a list of 22 items. Soon afterwards, Robert Yerkes (1940) assessed personality as a factor in dominance behavior in chimpanzees, but instead of using a

H. Freeman (✉)

Department of Psychology, Georgia State University, PO Box 5010,
Atlanta, GA 30302-5010, USA
e-mail: hani.freeman@gmail.com

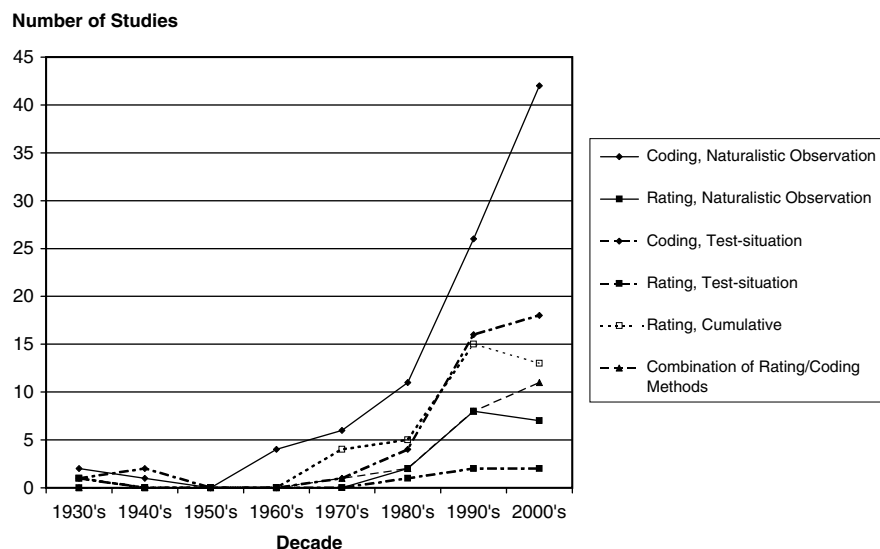


Fig. 2.1 Number of primate personality articles published each decade using each method of personality assessment

rating method, Yerkes assessed personality by recording the frequency and duration of different behaviors in a natural setting (Yerkes 1940). Thus, even in the early days of primate research, diverse methods were being used to assess personality.

Despite this important early work, studies of primate personality were still infrequent. As shown in Fig. 2.1, it was not until the 1960s and 1970s that primate personality research began to gain momentum (Stevenson-Hinde and Hinde 2011). The increased attention continued to be marked by methodological diversity, a trend that has persisted to the present day. However, only a few studies have directly compared assessment methods in terms of their accuracy and efficiency in capturing an animal's personality (Capitanio 1999; Gosling 2001; Itoh 2002; Gosling et al. 2003; Rouff et al. 2005; Vazire et al. 2007).

In this chapter, we will examine and evaluate different methods to assess primate personality. Our analyses will draw on an in-depth review of the literature. We shall discuss the strengths and weaknesses of each method, paying particular attention to the studies that have directly compared assessment methods. We shall conclude by offering some concrete steps that researchers may wish to consider as they measure personality in their own research.

2.2 Literature Review

The review of methods presented in this chapter draws on a much more extensive review of the primate personality literature (Freeman and Gosling 2010). The present review included full-length research articles published in journals or books.

In some cases the latest research is reported as an abstract published in conference proceedings and may never become a published article. Therefore, we included in our review abstracts not published in full-length elsewhere, if they reported sufficient substantive details to contribute to our review.

We based our literature search on keyword searches (or their equivalent) in AnthroPlus, Biosis, PrimateLit, Psychinfo, and Web of Science databases. We searched for all articles containing the keywords “primate and temperament” or “primate and personality.” In addition, some relevant articles mentioned the species by name without mentioning the term “primate” in the search fields. Therefore, we repeated the searches replacing “primate” with terms that would allow us to capture all primate species and subspecies (e.g., “Monkey,” “Lemur,” “Gibbon,” “Bushbaby”). The only primate species excluded from our search was humans. The resulting pool of reports was supplemented with articles cited in the papers identified in the keyword search and from nominations by experts in the field. Articles not directly relevant to the present review were eliminated from the pool (for details see Freeman and Gosling 2010).

The 154 studies identified in our review served as the basis for our analyses of the different methods used to assess primate personality and their strengths and weaknesses. In addition, we consider several arguments drawn from the human personality literature concerning the reliability and efficiency of the different methods.

2.3 Types of Assessment Methods

Two main methods of data collection have been used in primate-personality studies: behavioral codings and observer trait ratings. These two methods also characterize almost all research on nonprimate personality as well (Gosling 2001).

Behavioral coding (used in 87% of studies) involves observing animals and recording their behavior, typically in terms of bout frequency and/or duration. For example, one behavioral-coding study investigated the relationship between behavioral style, dominance rank, and cortisol levels in chimpanzees (Anestis 2005, 2011); the study used an all-occurrences sampling technique (i.e., one in which behavior was recorded anytime one chimpanzee interacted with another) to record social interactions of various kinds (e.g., agonistic, affiliative, etc.).

Trait rating (used in 40% of studies) involves having people who are familiar with the subjects rate them on a set of predefined traits or adjectives. Typically a rating scale is used in which high numbers indicate the trait is strongly characteristic of the target animal and low numbers indicate the trait is strongly uncharacteristic of the animal. For example, King and Figueredo (1997) assessed personality in 100 chimpanzees across 12 US zoos using rating data from an instrument that contained 43 adjectives, such as “playful” and “affectionate.” These two methods (behavioral coding and trait rating) can further be broken down by the conditions in which they are used – naturalistic observation, testing contexts, or cumulative observation. Naturalistic observation (used in 71% of studies) entails coding or rating the animals

over a specific period of time based on their ordinary daily behavior. French's (1981) study of Japanese macaques and individual differences in play behavior would fall into this category because the method of data collection was based on sessions in which the animals were observed in their home cages.

Testing contexts (used in 29% of studies) involve rating or coding the behavior of target animals in response to a particular stimulus or experiment designed to elicit differing reactions from animals with different personalities; Fairbanks' (2001) study of social impulsivity in vervet monkeys would fall into this category because the behaviors of the animals were recorded in response to a change in their typical daily routine, i.e., a new individual being shown to, but not introduced to, the group (Fairbanks and Jorgensen 2011).

The third condition, cumulative observation, (used in 25% of studies) applies only to the rating method and refers to the studies in which ratings are based on the knowledge and experience that each rater has accumulated since he or she has known the target animal; Martin's (2005) study of chimpanzees would fall into this category because the observers rated the animals based on their memories of experiences accumulated over the course of their acquaintance, which ranged from 1 to over 20 years.

The various combinations of assessment methods and assessment conditions can be combined into a five-cell grid shown in Table 2.1. Note, the cell representing the combination of the behavioral-coding method/cumulative-experience condition is absent because implementing such a method would be beyond the capacity of human memory; in practice, the missing cell would be equivalent to the cell that combines the rating method/cumulative experience. Also note that the percentages add up to more than 100 because some studies used more than one assessment method (14%).

2.4 How Often Are Each of These Methods Used?

In terms of the five-cell grid depicted in Table 2.1, the most commonly used method-condition combination was behavioral codings/naturalistic observation. This combination was used in more than half (60%) of the studies reviewed. The next most commonly used method-condition combination was behavioral codings/testing contexts, which was used in 27% of studies. This was closely followed by trait-rating/cumulative observation method-condition, which was used in 25% of the studies. Many fewer studies used the remaining combinations. Trait rating/naturalistic observation was used in only 11% of the studies and the trait-rating/testing combination was used in a mere 4% of the studies. To better understand the relative prevalence of these method-condition combinations, we next consider the strengths and weaknesses of the different methods. This analysis will help guide researchers in choosing the method or methods best suited to their research questions.

Table 2.1 Method of collecting/recording data

	Rating (40% of articles)	Coding (87% of articles)
	Pros	Pros
	<ul style="list-style-type: none"> – Faster than coding – Takes variability due to noise into account – Takes cross-situational consistency into account 	<ul style="list-style-type: none"> – Requires less subjective judgment by observer – Easier to make direct comparisons between animals
	Cons	Cons
	<ul style="list-style-type: none"> – Requires subjective judgment by observer – Comparisons between animals can be hard to interpret – Raters may weigh salient events more 	<ul style="list-style-type: none"> – Time consuming – Hard to account for cross-situational consistency – Hard to account for variability due to noise
Naturalistic behavior (71% of articles)	Example	Example
Pros	Capitanio (1999). Trained observers rated 42 adult male rhesus macaques on adjectives (e.g., sociable, aggressive) along a 7-point scale. Ratings were completed after 4 weeks of observations of the animals in their natal groups	Kaplan et al. (2002). Twenty-five adult male cynomolgus macaques were coded using a combination of all-occurrences and scan sampling to record behaviors (e.g., fights wins/losses, grooming) from a predetermined ethogram. The behaviors were recorded while the subjects were in their home cage
Cons		
<ul style="list-style-type: none"> – Based on only observable animals and behaviors – Observations based on behaviors in only one context 		
Test context (29% of articles)	Example	Example
Pros	Bard and Gardner (1996). Twenty-nine infant chimpanzees were rated on 30 items (e.g., responsiveness to objects, cooperativeness) along a 9-point scale, following tests of infant behavior	Hebb (1949). Thirty captive adult chimpanzees were coded using all-occurrence sampling to record behaviors (e.g., grooming, gesturing with fingers through mesh) from a predetermined ethogram. The behaviors were recorded while the subjects were involved in various test situations
Cons		
<ul style="list-style-type: none"> – Requires manipulating the subjects – Based on specific behaviors during a limited time 		

(continued)

Table 2.1 (continued)

Cumulative experience (25% of articles)	Example
Pros	Stevenson-Hinde and Zunz (1978). Trained observers with a minimum of 200 h experience, rated 46 rhesus macaques on a list of 33 adjectives (e.g., confident, eccentric) along a 7-point scale
– Fastest of all methods	
– Based on years of observa- tions of the subjects (i.e., takes into account many behaviors and contexts)	
Cons	
– Not based on direct data collection	

Note: The cell representing the combination of the behavioral-coding method and the cumulative-experience context is absent because implementing such a method would be beyond the capacity of human memory; in practice, the missing cell would be equivalent to the cell that combines the rating method and cumulative experience

2.5 Range of the Present Analysis

Before beginning our discussion about the strengths and weaknesses of the different methods of primate personality assessment we want to be clear about the range of applications covered by our analysis. Our discussion specifically focuses on rating and coding methods in the context of measuring personality. Many, or most, of the points we make in this context may have absolutely no bearing on the use of these methods in other contexts, such as observing behaviors in experimental contexts.

2.6 Strengths and Weaknesses of the Different Methods in Personality Assessment

A summary of the pros and cons of the different method-condition combinations along with examples of each is given in Table 2.1. Several designs have been used because different methods reflect different solutions to the trade-offs that arise when measuring behavior. For example, in an effort to combat the apparent subjectivity of trait ratings, behavioral codings may be used, which rely on direct observations such as recording the frequency of agonistic encounters. However, the added objectivity of behavioral codings comes with its own costs, such as the fact that codings take much longer to collect than do ratings. In addition to methodological considerations, some methods are preferred over others purely because they have been traditionally associated with a particular research tradition. Here we outline the strengths and weaknesses of the different methods so that researchers can make informed decisions about the design most suited to the circumstances of their own research. It should be noted that the present authors’ backgrounds span a variety of disciplines.

2.7 Trait Ratings

2.7.1 *Advantages of Trait Ratings*

As shown in the cell labeled *Trait Rating* in Table 2.1, the first advantage of trait ratings is that they allow data to be collected more quickly and efficiently than is possible with coding systems. Typically, a judge uses a rating scale to indicate how well an adjective or behavioral description reflects the behavior of an animal; the rating can be based on recent observations, testing situations, or past experiences with the subject. By summarizing in just one rating many behaviors performed over time and across situations, ratings are an exceptionally efficient way of capturing behavioral information and distilling it down to a single number in a matter of moments.

The second advantage of ratings as a method for measuring personality is that they implicitly control for variability in behavior due to nonpersonality factors (Vazire et al. 2007). For researchers trying to assess the aspects of behavior that are consistent across time and situations (i.e., personality), this variance represents noise. Some variability in behavior is due to changes in an animal's situation or environment, including seasonal effects, daily fluctuations, or changes in the social or physical environment (e.g., presence of a predator). For example, an individual might be active at one moment but inactive at another due to situational changes (e.g., feeding time). Observers can reduce the effects of this kind of variability by discounting situational influences on behavior when making their ratings. As Martin and Bateson (1993) noted,

the human rater has played an active role in filtering, accumulating, weighting, and integrating information over a considerable period of time (p 81).

Another portion of variability in behavior can be attributed to random variance. One way to reduce the effects of random variance is to aggregate measures across time so that nonsystematic sources of variance will cancel out. Trait ratings inherently benefit from aggregation because when observers rate an animal they implicitly summarize that animal's behavior across all the days, months, or years they have known it. Thus, data collected from trait ratings are essentially already aggregated across all the times the observer has observed the target animal. This quality of trait ratings drastically improves their reliability.

Another way to reduce the effects of random variance is to aggregate measures across raters so that irrelevant random differences among raters will tend to cancel out each other. Strictly speaking, this advantage could be applied to any of the methods discussed here, e.g., there is nothing to stop behavioral codings being aggregated across coders. However, in practice this step of aggregating across observers has been most widely used for the rating method, almost certainly because the relative efficiency of ratings makes it much easier to collect assessments from multiple raters than from multiple coders.

2.7.2 *Disadvantages of Trait Ratings*

One important disadvantage of trait ratings is that they rely on potentially idiosyncratic trait definitions. Any rating is affected by how the researcher has defined each trait and how the rater interprets that meaning as he or she makes a rating. At both of these stages their subjectivity may enter the process – two different researchers might have different ideas of what constitutes sociability and even if they do agree, two raters using the same scales may apply them differently. By relying on researchers' definitions, ratings of broad traits (e.g., sociability) may make assumptions that have not been verified empirically. For example, behaviors that co-occur in one species also co-occur in another species.

An additional problem with rating scales is that comparisons between animals can be hard to interpret; absolute counts (e.g., number of agonistic encounters, duration of exploration) can readily be compared across time periods, animals, experimental conditions, and labs, but it is harder to interpret the meaning of such differences using rating values. If a common set of raters rate the same individuals, the simple mean ratings of individuals can be compared, but means cannot be compared across species or different settings. For example, direct between-zoo comparisons are not meaningful if different zoos have different animals and different raters. The reason is that ratings are group relative and group means are therefore clustered around the center of the rating scale. Of course, behavior scores may have similar limitations for between-setting comparisons as a result of differences in group density, environmental differences, and other factors.

A further problem with relying on human judgment is that human raters may not appropriately weigh the information they observe when they combine elements of behavior into a single rating. For example, they may put undue emphasis on salient events (e.g., fights) and fail to take into account the degree to which behaviors overlap. A large body of literature in clinical psychology has shown that humans are relatively poor at integrating information in a statistically optimal manner (Meehl 1954).

Another problem with trait ratings is that they rely on the availability of raters experienced in the behavioral repertoire of the target species in various contexts. Raters who are experienced in a wide range of behavior may not always be available. In these situations, it would not be appropriate to use the trait-rating method.

2.8 Behavioral Codings

2.8.1 *Advantages of Behavioral Codings*

As shown in the cell labeled *Behavioral Codings* in Table 2.1, behavioral codings have an advantage over trait ratings in that they involve direct data collection of the duration or frequency of specific behaviors. This method may eliminate much

of the subjectivity associated with ratings and is ideal for certain types of behavioral observation, such as trying to identify the true frequency of behaviors or proportion of time the animal is displaying the behavior.

Another advantage of behavioral codings is that they allow for easy comparison between animals because they are based on frequencies of behavior. For example, if one animal is recorded as playing 50 times and another is recorded as playing 100 times in a year, then a strong case can be made that the second animal is more playful than the first.

2.8.2 *Disadvantages of Behavioral Codings*

There are several disadvantages to behavioral codings when used *as a method for assessing personality*. However, many of these disadvantages are not serious when the coding method is used in nonpersonality contexts (e.g., in experiments). Indeed the rich history of research on animal behavior stands as a testament to the value of coding methods for many research purposes.

One disadvantage is that collecting behavioral codings is often very time consuming compared to collecting trait ratings. Behavioral codings typically require the focal animal to be observed anywhere from 5 to 30 min at a time. This duration is not very long, but the observations are almost always repeated, between 6 and 100 times (more often for the shorter periods of time) in order to capture variations of behavior during different times of day and temperature, resulting in a large amount of total time dedicated to each animal. When conducting behavioral codings, researchers must decide which method to use. Possible methods include all-occurrences, instantaneous, and zero-one sampling. All-occurrences involves recording the frequency and duration all instances of a behavior in a specified observation period. This method has the advantage of capturing data on frequency and duration. The disadvantage to all-occurrences sampling is that it can be difficult to record many different behaviors occurring all at once. Instantaneous sampling involves recording whether a particular behavior is occurring at a specific instance. Zero-one sampling involves recording whether a behavior is occurring or occurred at all during a specific time interval. Both instantaneous and zero-one sampling reduce the time spent focused on recording behaviors. Zero-one sampling captures frequent short duration as well as behaviors that are expressed infrequently for long durations; however, neither frequency nor duration can be accurately estimated from zero-one sampling. In the case of instantaneous sampling, scores give unbiased estimates of the total proportion of time a subject is displaying any recorded behavior. All three methods are not good at capturing very rare behaviors and instantaneous sampling can easily miss behaviors that last a short time (Martin and Bateson 1993). Also, researchers wanting to use behavioral-coding methods face the problem of identifying multiple individuals who are willing to spend hundreds of hours necessary to undergo the requisite training and then code behavior (Vazire et al. 2007). Although raters also need to have spent many hours observing the animals in different

situations and contexts, the observations do not have to be done systematically, in the same way as in behavioral coding. Instead, they could be done in connection with other projects or tasks that the rater was completing at the time (e.g., as a member of the animal care staff).

In addition to being time consuming, behavioral codings do not typically control for as much of the variability in behavior due to nonpersonality factors as trait ratings do. As mentioned earlier in relation to trait ratings, this variability in behavior due to nonpersonality factors can include situational factors and random noise.

Another problem with behavioral-coding systems as they are often used is that the ethograms often implement a rule in which any behavior is assigned to one and only one category. This can be problematic when situations arise where the researcher wants to take into account two behaviors that are occurring simultaneously. For example, if an animal is playing, an observer might want to take into account, not only the fact that play was occurring, but also the type of locomotion (e.g., stationary, walk, run) that was occurring during play to get an idea of energy level. If an ethogram only allows the researcher to assign a behavior to one category, the researcher will not be able to take both the play and locomotion into account for that occurrence. A single behavioral sequence (e.g., rambunctious play) can involve multiple behaviors and could reflect multiple personality traits so a system that assigns behaviors to only one category may not be a valid reflection of all the relevant traits. It should be emphasized that this issue is more a problem with the way that behavior codings are practiced than anything intrinsic to the method itself because there is no reason why behavioral-coding systems need to insist that behaviors are assigned to only a single coding category. One practical reason why behaviors have often been assigned to just a single category in ethograms arises from limitations of observational software programs. These programs are sometimes designed in a way that makes it tricky to assign a behavior to more than one category. However, there are some simple solutions to the issues with using a specific observational software program, Observer version 5.0, which we will discuss later. In addition, a new version, Observer XT has recently been released and may address some of these issues.

Another drawback of behavioral-coding methods is their susceptibility to the effects of random noise because they capture only a narrow snapshot of all the behaviors in which the animal engages throughout the day. Many behaviors relevant to personality only reveal their consistency of performance over time periods too long to be captured by coding systems. For example, even if an animal has 15 fights a year, it is unlikely that any of these events will be recorded in the narrow windows afforded by coding sessions.

In addition, there are some personality traits that would be difficult to translate into a feasible list of behaviors to code. For example, the trait unpredictable would be difficult to operationalize in terms of specific behaviors.

In summary, our analyses of the two methods of assessment suggest that using trait ratings to assess primate personality provides several advantages over behavioral codings. We next consider the strengths and weaknesses of the three types of experimental designs in which these methods are used (i.e., naturalistic observation, testing context, and cumulative observation).

2.9 Naturalistic Observation

2.9.1 *Advantages of Naturalistic Observation*

As shown in the cell labeled *Naturalistic Observation* in Table 2.1, one advantage of naturalistic observation over a testing context is convenience. Naturalistic observation does not require the rater/observer to manipulate the subjects at all.

In studies based on naturalistic observation, each rater spends a standard amount of time observing the subjects before rating them. Therefore, naturalistic observation has the benefit of controlling for differences in familiarity among the observer/raters.

2.9.2 *Disadvantages of Naturalistic Observation*

One disadvantage of naturalistic observation is that the observations are limited to animals that can be seen. If a subject happens to be out of view at any time during the observation period, then its behavior will not be recorded or taken into consideration. In a testing context it is possible, depending on the particular circumstance, that a subject cannot be seen, but unobservability is less likely to be an issue in testing contexts where an effort is usually made to conduct the tests where the subject can be seen.

Another disadvantage of naturalistic observation is that observations are obtained in only one specific context. For example, if the animals are simply observed in their home cage with the same individuals they are typically near, there is no measure of what the animal is like outside the home cage. These other contexts could elicit behavior relevant to understanding the individual's personality (Capitanio 1999; Fairbanks and Jorgensen 2011).

2.10 Testing Context

2.10.1 *Advantages of Testing Context*

As shown in Table 2.1, testing contexts have several advantages. One advantage is that all behaviors performed by the subject during an observation session are seen and recorded because the subjects are often tested individually or, if the testing is done in a group setting, an effort is made to test each subject in the group.

Another advantage of the testing context is that it can examine the effects of context-specific behavior (e.g., novelty seeking) and take the context into account. Testing situations permit subjects to be coded/rated in places other than their home cage and often with different social groupings (e.g., isolated or with only one or two other individuals).

Testing contexts also give researchers control over the situation in which the rating/coding is occurring. For example, a subject can be placed in different social groupings or away from distractions. This contrasts with studies using naturalistic observation where subjects cannot be manipulated at all and where animals cannot be tested separately from other group members or out of their home cages.

Another advantage of testing contexts is the ability to elicit specific types of behavior. For example, Watson and Ward (1996) used the testing context to assess the relationship between temperament and problem solving. Use of this method to create problem-solving opportunities elicited more instances of observed problem solving than would have occurred if the researchers had relied on naturalistic observation.

2.10.2 Disadvantages of Testing Context

There are also disadvantages to assessing personality in testing contexts. One issue is that manipulating subjects can be time consuming or in some situations (e.g., when only one cage is available) impractical.

Another problem with using testing contexts is that they typically focus on only a few specific behaviors during a limited period of time, allowing only a relatively restricted view of the species' behavioral repertoire.

2.11 Cumulative Observation

2.11.1 Advantages of Cumulative Observation

Cumulative observations are recorded exclusively with trait ratings and enjoy several advantages over both naturalistic observations and testing contexts. One benefit of cumulative observation is that raters are familiar with all of the animals they are rating because they have observed and/or interacted with each of the subjects at many different times over an extended period of time, ranging from several months to many years. The raters' familiarity with the animals is an advantage because it suggests that they have typically been able to observe the animals in different contexts and over different periods of time, rather than having just a few interactions on which to base their impressions.

Of the three contexts of assessment, ratings based on cumulative observations require the shortest amount of time to complete. The rater does not have to perform any additional observations or tests before rating the subjects because ratings are based on previous experiences. In turn, the relative ease of obtaining trait ratings reduces the likelihood of errors due to lack of training, misunderstandings, or fatigue on the part of the observer (Vazire et al. 2007).

2.11.2 Disadvantages of Cumulative Observation

One practical limitation of cumulative observation methods is that it may be difficult to identify multiple observers who are sufficiently familiar with the target animals to provide expert ratings.

Another issue in cumulative-observation studies is that raters may vary in their experiences of the contexts and amount of time in which they have observed each subject. This can be problematic for some of the same reasons associated with the behavior-coding methodology. If a person has worked with an animal in only one context at intermittent times, then he or she may not have accumulated a broad impression of the subject's personality. However, if another individual has worked with the animal during regular intervals of time, across many contexts, then he or she might have a more general and accurate sense of the subject's personality.

Another danger with cumulative observations is that the raters almost certainly have talked about the animals with each other, thereby compromising the independence of the observations. Of course, in some instances these conversations may communicate valid information, e.g., reporting that an animal was involved in a fight.

2.12 Empirical Studies Comparing Behavioral-Coding and Trait-Rating Data

The arguments summarized earlier suggest several theoretical strengths and weaknesses of the different methods. How do these strengths and weaknesses play out in practice? Six empirical papers have focused on directly comparing behavioral-coding with trait-rating methods. Although other papers have used both behavioral-coding and trait-rating methods to assess personality, this was not the focus of the studies. If both methods are valid they should converge; assessments of the same animals using both methods should correlate with each other when there is a prediction that they will converge or when the adjective and behavioral meanings are the same, e.g., aggressive the adjective and aggressive behaviors. If they do not converge when predicted to do so, then at least one of them is not valid. Examples of high correlations found in the six studies are summarized in Tables 2.2 and 2.3. The studies in Table 2.2 incorporated the naturalistic observation condition and those in Table 2.3 incorporated the testing context procedure.

Each of the six studies includes two kinds of assessment method (behavioral codings and trait ratings). For example, the study by Vazire et al. (2007) assessed the personality of 52 chimpanzees using both trait ratings based on cumulative observations and behavioral codings based on naturalistic observation. The studies differ slightly in the level at which the behavioral codings were measured. Five studies coded behavior as aggregates of individual behaviors. For example, the term "aggressive behavior" might include individual behaviors such as biting and attacking. These aggregated measures are listed in the top halves of Tables 2.2 and 2.3 under

Table 2.2 Examples of high correlations in empirical studies comparing behavioral codings and trait ratings through naturalistic observations

Species	<i>r</i>	Dimension	Item label	Study
<i>Aggregated behaviors</i>				
Chimpanzee	0.53	Emotionality	Agonistic	Pederson et al. (2005)
	−0.37	Playfulness	Solitary behavior	Vazire et al. (2007)
Gorilla	−0.54	Dominant	Displacement – receiving	Kuhar et al. (2006)
	0.55	Extroverted	Contact aggression- initiating	
Vervet monkey	0.82	Curious	Play	McGuire et al. (1994)
Stumptailed macaque	−0.46	Excitable	Nonsocial behavior	Mondragon-Ceballos and Santillan-Doherty (1994)
<i>Individual behaviors</i>				
Chimpanzee	−0.48	Extraversion	Idle	Pederson et al. (2005)
	0.59	Playfulness	Social play	Vazire et al. (2007)
Rhesus macaque	0.54	Confident	Groom receive	Capitanio (1999)
Stumptailed macaque	−0.83	Sociable	Resting	Mondragon-Ceballos and Santillan-Doherty (1994)

Table 2.3 Examples of high correlations in empirical studies comparing behavioral codings and trait ratings in a testing situation

Species	<i>r</i>	Dimension	Item label	Study
<i>Aggregated behaviors</i>				
Rhesus macaque	0.41	Confident	Affiliative	Capitanio (1999) video playbacks
<i>Individual behaviors</i>				
Rhesus macaque	0.39	Excitable	Back of the cage	Capitanio (1999) respon- siveness behaviors
	0.70	Sociable	Approach receive	Capitanio (1999) group behaviors
	0.67	Sociable	Lipsmack initiate to nonsocial video	Capitanio (1999) video playbacks
	0.78	Excitable	Fear grimace receive	Capitanio (1999) social dyad

the heading “aggregate behaviors.” Four studies coded, in addition to or instead of aggregate behavior terms, specific behaviors such as “aggress receive.” These behaviors are listed in the bottom halves of Tables 2.2 and 2.3 under the heading “individual behaviors.” Three of the studies used the behavioral codings to code both aggregate and individual behaviors and the other three used the behavioral codings to code either aggregate or individual behaviors.

The first column of Table 2.1 shows the species, and the second through fourth columns show examples of high correlations between ratings and behavior in each of the studies. The last column lists the relevant citation.

Although Table 2.2 gives evidence of some high correlations between ratings and behavior, indicating that convergence is possible, the cross-method convergence correlations are generally quite low indicating that behavioral codings and trait ratings often result in different assessments of the same individuals' personalities. For example, contrary to expectations, the correlation between scratching behavior and emotionality was -0.32 (Pederson et al. 2005). These findings suggest that codings and ratings are not always measuring the same thing. The aggregated codings were substantially less reliable than the aggregated ratings suggesting that the codings were less likely to be measuring any valid construct (including personality) than the ratings (Vazire et al. 2007). The low reliability of the codings reflected the fact that behavioral counts from one coding period correlated poorly with the behavioral counts generated by the same animal in another coding period. The low correlations between the two coding periods may have been a result of taking observations at different times of day, so the animal had a different energy level and as a result behaved differently during the two coding periods. It is not known if poor coding reliabilities could explain the low convergence between trait ratings and behavioral codings in the other studies evaluating validity because reliability was not reported for behavior codings in these studies (McGuire et al. 1994; Mondragon-Ceballos and Santillan-Doherty 1994; Capitanio 1999; Pederson et al. 2005; Kuhar et al. 2006). An alternative explanation for the low instances of convergence between trait ratings and behavioral codings is that only a small fraction of behaviors on an ethogram might be actually captured by a behavioral-coding strategy through naturalistic observation performed in the subject's home cage (Capitanio 1999). Instead, higher validity can be achieved between trait ratings and behavioral codings, when the codings are obtained in a testing situation (Capitanio 1999). This improved validity may occur because the testing situations can be designed to target specific behaviors that might be seen at a much lower frequency under conditions of naturalistic observation.

Table 2.3 shows examples of high correlations when assessing the cross-method convergence correlations obtained under conditions of a testing situation as opposed to naturalistic observation summarized in Table 2.2. The same aggregate/individual trait separation scheme is used in Table 2.3 as was used in Table 2.2. Studies using aggregate terms are shown in the top half of the table and those using individual terms are shown in the bottom half. It should be noted that Table 2.3 is made up of four different studies performed by Capitanio. In most cases, the testing conditions (Table 2.3) did yield stronger cross-method convergence correlations than those obtained in naturalistic observations (Table 2.2). However, comparisons of this sort could not be made in the other five studies because they collected behavior codings exclusively with naturalistic observations so no testing-situation data were available. In addition, it is worth noting that the cross-method convergence correlations obtained in the testing situation were in some cases lower than the convergence correlations obtained in other studies where the behavioral codings were obtained in naturalistic observations. This finding suggests that although collecting behavioral codings in a testing situation might improve the convergence between trait rating and behavioral codings to some degree, it still may not fully account for the poor convergence between the two methods.

The data presented in the six empirical papers that compared trait ratings to behavioral-coding methods suggest that, in some cases, the two methods are not producing the same results. Both Vazire et al. (2007) and Capitanio (1999) have suggested possible explanations for why there may be instances of low convergence between the trait ratings and behavioral-coding methods (i.e., low reliability of the codings, behavioral observations completed through naturalistic observation as opposed to testing situation). Although these potential explanations have been tested by only one paper each, they both point to shortcomings of the behavioral-coding methodology, rather than the trait-rating method for assessing personality. Further studies are needed to fully understand what is driving the disparity between the behavioral codings and the trait ratings.

Based on the evidence presented on the strengths and weaknesses of each of the method combinations and from the six studies directly comparing behavioral-coding with trait-rating methods, we propose that the method combination of trait rating/cumulative observation is the most efficient and reliable method for assessing primate personality.

2.13 “Etic” and “Emic” Approaches to Rating Scale Development

There are two broad strategies to developing rating scales, which are somewhat analogous to the etic and emic approaches used in cross-cultural research.

2.13.1 “Etic” Approach

In human cross-cultural research, the etic approach is one in which personality scales from one culture are imported and applied to another culture (Berry 1999; Gosling and John 1999; Weiss et al. 2006). An etic approach in animals involves using scales that have been developed and used for assessment in one species to inform the development and implementation of a scale for another species. For example, taking a scale that was developed for use in rhesus monkeys and using it to measure personality in gorillas would be an etic approach.

One benefit of using an etic approach is that the scale has usually been shown to be reliable and valid in previous research using that scale. In addition, using a previously used scale allows researchers to easily perform cross-species or cross-colony comparisons because of the commonalities in the adjective names and definitions used in all the samples using those scales.

The main problem with using an etic approach is that the adjectives used to describe one species may not be the best ones for describing another species. As a result,

researchers could include adjectives that do not fit the behavior of a new species. An etic approach might also lead researchers to neglect or omit adjectives that reflect important traits in the new species.

2.13.2 “Emic” Approach

In human cross-cultural research the emic approach is one in which personality scales for a culture are developed within that culture (Berry 1999; Gosling and John 1999; Weiss et al. 2006). An emic approach to rating-scale development in animals involves developing a list of traits based on systematic observations made by trained researchers familiar with the specific species’ behavior. To make sure the adjective set is representative of the full range of the species’ behavior, adjectives would be included based on previous research on the behavior of the species. For example, an emic approach to developing a scale for use in rhesus monkeys would draw both from observations and previous research on rhesus monkeys.

An emic approach to scale development allows a researcher to include an exhaustive list of trait adjectives that are representative of the species examined. Doing so increases the chances of comprehensively capturing the behaviors expressed in that species.

The problem with using an emic approach to scale development is that cross-species comparisons are hindered by cross-study differences in the instruments used.

In sum, there are strengths and weaknesses to both the etic and emic approaches. As we shall see below, we recommend using a combined etic and emic approach. That is, using some items from previous instruments to permit cross-species and cross-study comparisons, and also generating items specifically for use in the study species to ensure the full breadth of the species’ behavioral repertoire is captured.

2.14 Most Commonly Used Rating Scales for Assessing Primate Personality

The present review revealed that nearly all personality rating research has been done using just a few instruments. In fact two thirds of the 59 studies used just three scales.

2.14.1 *Emotions Profile Index*

Buirski et al.’s (1973) Emotions Profile Index (EPI) was used in 10% of the primate rating studies. It was developed using an etic approach and was based on methods designed to assess human personality (Plutchik 1965; Kellerman and Plutchik 1968). The primate version of the scale was first described and assessed in Buirski et al. (1973), where it was used in a field study of 7 olive baboons using 3 raters.

2.14.2 *Trait Descriptive Adjective Set*

King and Figuerdo's (1997) scale was used in 17% of the primate rating studies. It was also designed using an etic approach. It was based on adjectives taken from Goldberg's (1990) adjective list originally developed for use in humans. The scale was first described and assessed in King and Figuerdo (1997), where it was used to rate 100 chimpanzees at 12 zoological parks using 53 raters.

2.14.3 *Stevenson-Hinde and Zunz Instrument*

Stevenson-Hinde and Zunz's (1978) scale was used in 40% of the primate rating studies. It was designed using aspects of an emic process, in which adjectives were generated by three people familiar with the 45 rhesus macaques and their behavior; it did not include a thorough literature search for additional behaviors that might not have been captured initially.

The development and widespread use of these scales has probably played a large role in the acceptance of trait rating as a reliable and valid method for assessing primate personality. A number of benefits accrue from having similar instruments used in multiple studies, most notably, the ability to perform cross-study comparisons. For example, broad comparisons can be made across all 23 studies that have used the Stevenson-Hinde and Zunz (1978) instrument.

However, the reliance on just a few instruments also raises a number of dangers. First, there is a danger that an instrument will be used in a species other than that for which it was specifically developed. For example, although the Stevenson-Hinde and Zunz (1978) instrument was developed for use in rhesus macaques it has also been used in gorillas (Gold and Maple 1994) and tufted capuchins (Byrne and Suomi 2002). Second, if an instrument is created and widely adopted before there has been time to replicate the personality findings in multiple labs then subsequent research could be overly influenced by the idiosyncrasies of the particular group of subjects on which the instrument was based or the perspectives of the particular group of researchers who developed it.

2.15 The Need for New Scale Development

The three scales most commonly used to assess primate personality have helped establish the legitimacy of the trait-rating method. However, it is important that existing scales be augmented by scales tailored to unique features of other primate species. Continued development and improvement of existing scales will also bolster the foundations upon which subsequent primate personality research is built. Below, we outline how such scale development could proceed. We draw heavily on Uher (2008a, b) and Gosling (1998).

Uher (2011) discusses a behavioral repertoire approach to developing a list of traits that can be used to assess personality. A full example of this approach can be seen in recent publications as applied to the great ape species (Uher and Asendorpf 2008; Uher et al. 2008). The behavioral repertoire approach, which can be broken down into a four-step method, was initially designed to develop a list of traits that could be operationalized using behavior codings, behavior ratings, and adjective ratings. However, for this chapter it will just be discussed for the purposes of describing the development of new adjective ratings scales. Gosling (1998) developed a three-step method, originally used to establish a reliable and valid personality assessment scale in spotted hyenas

The choice of whether to use the four- or three-step approach depends on how much previous behavioral research has been conducted on the species of interest and on the availability of humans with necessary experience observing the species. If the amount of previous behavioral research is limited or few experienced researchers are available, the four-step emic approach to adjective trait-rating scale development should be used. If there is an extensive body of previous behavioral research and/or the experienced individuals are available to assist with the research, then the three-step approach is recommended.

2.15.1 Four-Step Approach

This approach is designed to maximize comprehensiveness and explicitly acknowledges the importance of basing the rating categories on behavioral data acquired across a variety of contexts and occasions (Uher 2008a, b).

Step 1: Begin by doing a thorough review of the literature to discover all behavioral aspects of a species' life, both in the wild and captivity. This should encompass only observable and measurable behaviors. The list of behaviors should be noted along with the different situations in which they are found to occur.

Step 2: Merge the behavioral domains and situations into theoretical trait constructs. For example, aggressive behavior towards a novel object may be related to the trait aggressiveness.

Step 3: Operationalize the trait constructs into behavioral rating (behavior-descriptive verbs) and adjective rating lists and collect data based on these lists. For example, in Uher and Asendorpf (2008) as a behavioral rating, curiosity was operationalized as "animal often touches new objects at great length." In Uher and Asendorpf (2008), as an adjective rating, curiosity was operationalized as "animal is very curious."

Step 4: Analyze data for stable interindividual differences and perform a multivariate analysis on the data to understand relationships among different traits.

The emic approach endorsed by Uher (2008a, b) is a comprehensive approach to match the behavior of a specific species with the personality traits being assessed. However, in tailoring an instrument to a particular species, the ability to compare

the findings across studies is compromised. Thus, a natural tension exists between the demands of comprehensiveness and comparability (Gosling 1998, 2001). In capturing the idiosyncrasies of a particular species, researchers may be forced to use traits that are not applicable to other species. A balance needs to be reached in which a basic set of standardized descriptors (gleaned from other research and operationally defined in species-appropriate terms) is supplemented by important species-specific descriptors (using procedures along the lines of those proposed by Uher 2008a, b). It should be noted that the emic and etic approaches are not the only approaches to the development of rating scales. A more thorough discussion of various approaches and their benefits and drawbacks can be seen in recent papers by Uher (2008a, b).

2.15.2 Three-Step Approach

This approach is designed to strike a balance between comparability and comprehensiveness in generating trait-rating items (Gosling 1998).

- Step 1:* Generate a comprehensive list of behavioral traits from three sources: “previous research on animal personality, previous research on human personality, and expert nominations” (p. 108 Gosling 1998).
- Step 2:* Eliminate redundancy from the list of behavioral traits suggested in step 1, preferably with the help of experts.
- Step 3:* Define each of the adjectives in terms of species-specific behavior.

2.16 Recommendations

2.16.1 What Is the Best Method for Assessing Primate Personality?

Based on our analysis of the advantages and limitations of each of the methods and conditions, as well as our review of the empirical comparisons between trait-rating and behavioral-coding methods, we conclude that the trait-rating/cumulative observation method-condition is the most reliable and practical method-condition combination.

However, the recommendation of a trait-rating/cumulative observation method-condition does not negate the usefulness of a combination of behavioral-coding and trait-rating methods, when the time and resources are available to include them. In fact, over the past two decades (see Fig. 2.1) a trend has emerged where a combination of rating and coding methods is increasingly being used. A combination of methods offers several advantages over studies that use just one personality assessment method. One benefit to using both methods is that a researcher obtains information about both a context-relevant behavior (from behavioral coding) as well as a

broader view of the animal's personality (from trait rating). In addition, multimethod designs allow better assessment of reliability and validity (McGuire et al. 1994; Capitanio 1999; Pederson et al. 2005; Kuhar et al. 2006).

2.16.2 *What Improvements Can Be Made to Methods of Assessing Primate Personality?*

Several suggestions can be made to improve the current methods of primate personality assessment. Although we suggest the cumulative observation/trait-rating method for use in most assessment contexts, there may be situations in which another method-condition combination must be used. In such cases, there are several ways in which current methods can be improved.

2.16.2.1 All Methods

Both raters and coders should be well experienced in working with or observing the animals they are assessing. They should also be thoroughly trained on the assessment method. In addition, interrater reliability should always be obtained.

2.16.2.2 Behavioral Coding of Naturalistic Observation or Testing Situation conditions

There are circumstances where behavioral codings are useful in personality assessment. For example, an important feature of the emic approach is gathering a broad array of species-specific behaviors from which ratings can be derived. Behavioral coding of naturalistic observation and testing situations can help accomplish this task in order to identify behavioral domains and situations that can serve as the foundation for the development of behavior rating and adjective-based rating scales (Uher 2008a, b).

Another recommendation is to use a behavioral ethogram that allows behaviors to be assigned to more than one category. As mentioned earlier, this is often difficult when using behavioral observation software programs, because of their design. However, there are several ways to work around this issue in some programs, including Observer 5.0 (Noldus 1991). One strategy is to identify behaviors that are likely to be seen along with other behaviors (e.g., screaming). Then identify the different behaviors to which this one behavior is likely to be linked (e.g., running, sitting, climbing). Then, list screaming as a modifier for each of these possible scenarios. This technique can be effective if a researcher is able to accurately predict the behaviors that might co-occur, but it does not allow for situations where the co-occurrence of behaviors has not been anticipated. An alternative method is to turn on the comments feature of the ethogram in Observer. This step will allow the addition of a comment anytime the researcher records an observation. Then, if two behaviors

occur simultaneously, the observer can enter the code for one behavior and just add the other behavior in the comments section. This procedure may add more work during the data-analysis phase because the comments will have to be added to the coded behaviors but it will allow for a more accurate representation of the behavior.

2.16.2.3 Testing Situation

If the goal of a study is to investigate a specific aspect of personality (e.g., novelty seeking, social impulsivity) a behavioral coding/testing situation combination will allow the researcher to capture more of the specific behaviors than is possible with a general rating scale assessment. However, in order to insure the validity of the behavioral coding/testing situation results, a trait-rating method should also be used. The combination of these two methods (behavioral coding/testing situation and trait-rating methods) allows researchers to assess whether the behavioral coding/testing situation is accurately capturing the specific aspects of personality that are the focus of the study. For example, if a behavioral coding/testing situation is used to assess different aspects of impulsivity, there should be a strong correlation between the frequency of impulsive behaviors recorded and the rating of impulsivity of the animals being assessed.

2.17 Conclusions

In this chapter, we examined and evaluated the different methods used to assess primate personality. Our analyses drew on an in-depth review of the literature. We discussed the strengths and weaknesses of each method, paying particular attention to the studies that have directly compared assessment methods. We concluded that in most cases the trait-rating/cumulative observation method combination is the best method for primate personality assessment. However, more research is needed to compare empirically the different methods of personality assessment. We also suggested some improvements for primate personality assessment methods that can be implemented for any chosen method. The suggested steps and improvements provide a blueprint for a solid foundation for future research on primate personality.

Acknowledgments We are grateful to Stephanie Anestis, John Capitanio, Frans de Waal, Lynn Fairbanks, James Dee Higley, James King, Lindsay Murray, Jon Sefcek, and Alex Weiss for their help with this task.

References

- Anestis SF (2005) Behavioral style, dominance rank, and urinary cortisol in young chimpanzees (*Pan troglodytes*). *Behaviour* 142:1245–1268
- Anestis SF (2011) Primate personality and behavioral endocrinology. In: Weiss A, King JE, Murray L (eds) *Personality and temperament in nonhuman primates*. Springer, New York

- Bard KA, Gardner KH (1996) Influences on development in infant chimpanzees: Enculturation, temperament, and cognition. In: AE Russon, KA Bard (eds) *Reaching into thought: The minds of the great apes*. Cambridge University Press, Cambridge
- Berry JW (1999) Emics and etics: A symbiotic conception. *Cult Psychol* 5:165–171
- Buirski P, Kellerman H, Plutchik R et al. (1973) A field study of emotions, dominance, and social behavior in a group of baboons (*Papio anubis*). *Primates* 14:67–78
- Byrne G, Suomi SJ (2002) Cortisol reactivity and its relation to homecage behavior and personality ratings in tufted capuchin (*Cebus paella*) juveniles from birth to six years of age. *Psychoneuroendocrinology* 27:139–154
- Capitanio JP (1999) Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. *Am J Primatol* 47:299–320
- Crawford MP (1938) A behavior rating scale for young chimpanzees. *J Comp Psychol* 26:79–91
- Cronbach LJ, Meehl PE (1955) Construct validity in psychological tests. *Psychol Bull* 52:281–302
- Fairbanks LA (2001) Individual differences in response to a stranger: Social impulsivity as a dimension of temperament in vervet monkeys (*Cercopithecus aethiops sabaeus*). *J Comp Psychol* 115:22–28
- Fairbanks LA, Jorgensen MJ (2011) Objective behavioral tests of temperament in nonhuman primates. In: Weiss A, King JE, Murray L (eds) *Personality and temperament in nonhuman primates*. Springer, New York
- Freeman H, Gosling SD (2010) Personality in nonhuman primates: A review and evaluation of past research. *Am J Primatol* 71:1–19
- French JA (1981) Individual differences in play in *Macaca fuscata*: The role of maternal status and personality. *Int J Primatol* 2:237–246
- Gold KC, Maple TL (1994). Personality assessment in the gorilla and its utility as a management tool. *Zoo Biol* 13:509–522
- Goldberg LR (1990). An alternative ‘description of personality’: The Big-Five factor structure. *J Pers Soc Psychol* 59:1216–1229
- Gosling SD (1998) Personality dimensions in spotted hyenas (*Crocuta crocuta*). *J Comp Psychol* 112:107–118
- Gosling SD (2001) From mice to men: What can we learn about personality from animal research. *Psychol Bull* 127:45–86
- Gosling SD, John OP (1999) Personality dimensions in nonhuman animals: A cross-species review. *Curr Dir Psychol Sci* 8:69–74
- Gosling SD, Kwan VSY, John O P (2003). A dog’s got personality: A cross-species comparative approach to evaluating personality judgments. *J Pers Soc Psychol* 85:1161–1169
- Hebb DO (1949) Temperament in chimpanzees: I. Method of analysis. *J Comp Physiol Psychol* 42:192–206
- Itoh K (2002) Personality research with non-human primates: theoretical formulation and methods. *Primates* 43:249–261
- Kaplan JR, Manuck SB, Fontenot MB et al. (2002) Central nervous system monoamine correlates of social dominance in cynomolgus monkeys (*Macaca fascicularis*). *Neuropsychopharmacology* 26:431–443
- Kellerman H, Plutchik R (1968) Emotion-trait interrelations and the measurement of personality. *Psychol Rep* 23:1107–1114
- King JE, Figueredo AJ (1997) The Five-Factor Model plus Dominance in chimpanzee personality. *J Res Pers* 31:257–271
- Kuhar CW, Stoinski TS, Lukas KE, Maple TL (2006) Gorilla Behavior Index revisited: Age, housing and behavior. *Appl Anim Behav Sci* 96:315–326
- Martin JE (2005) The influence of rearing on personality ratings of captive chimpanzees (*Pan troglodytes*). *Appl Anim Behav Sci* 90:167–181
- Martin P, Bateson P (1993) *Measuring Behaviour*. Cambridge University Press, Cambridge
- McGuire MT, Raleigh MJ, Pollack DB (1994) Personality features in vervet monkeys: the effects of sex, age, social status, and group composition. *Am J Primatol* 33:1–13

- Meehl PE (1954) Clinical vs. statistical prediction: A theoretical analysis and review of the evidence. University of Minneapolis Press, Minneapolis
- Mondragon-Ceballos R, Santillan-Doherty AM (1994) The relationship between personality and age, sex and rank in captive stump-tail macaques. In: Roeder JJ, Thierry B, Anderson JR, Herrenschmidt N (eds) Current primatology, Vol. II: Social development, learning and behavior. University of Louis Pasteur, Strasbourg
- Noldus LPJJ (1991) The Observer: A software system for collection and analysis of observational data. Behav Res Methods Instrum Comput 23:415–429
- Pederson AK, King JE, Landau VI (2005) Chimpanzee (*Pan troglodytes*) personality predicts behavior. J Res Pers 39:534–549
- Plutchik R (1965) What is an emotion? J Psychol 61:295–303
- Rouff JH, Sussman RW, Strube MJ (2005) Personality traits in captive lion-tailed macaques (*Macaca silenus*). Am J Primatol 67:177–198
- Stevenson-Hinde J, Zunz M (1978) Subjective assessment of individual rhesus monkeys. Primates 19:473–482
- Stevenson-Hinde J, Hinde CA (2011) Individual characteristics: Weaving psychological and ethological approaches. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York
- Uher J (2008a) Comparative personality research: Methodological approaches. Eur J Pers 22:427–455
- Uher J (2008b) Three methodological core issues of comparative personality research. Eur J Pers 22:475–496
- Uher J, Asendorpf JB (2008) Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. J Res Pers 42:821–838
- Uher J, Asendorpf JB, Call J (2008) Personality in the behaviour of the great apes: Temporal stability, cross-situational consistency and coherence in response. Anim Behav 75:99–111
- Uher J (2011) Personality in nonhuman primates: What can we learn from personality psychology. In: Weiss A, King JE, Murray L (eds) Personality and temperament in nonhuman primates. Springer, New York
- Vazire S, Gosling SD, Dickey AS et al. (2007) Measuring personality in nonhuman animals. In: Robins RW, Fraley RC, Krueger RF (eds), Handbook of research methods in personality psychology. Guilford Press, New York
- Watson S, Ward JP (1996) Temperament and problem solving in the small-eared bushbaby (*Otolemur garnetti*). J Comp Psychol 110:377–385
- Weiss A, King JE, Perkins L (2006) Personality and subjective well-being in orangutans (*Pongo pygmaeus* and *Pongo abelii*). J Pers Soc Psychol 90:501–511
- Yerkes RM (1940) Social behavior of chimpanzees: Dominance between mates, in relation to sexual status. J Comp Psychol 30:147–186

Personality and Temperament in Nonhuman Primates

Weiss, A.; King, J.E.; Murray, L. (Eds.)

2011, XIV, 346 p. With online files/update., Hardcover

ISBN: 978-1-4614-0175-9