Appeared in Acta Biologica Hungarica 2015, 66:27-40

A comparison of rating and coding behavioural traits in dogs

Enikő Kubinyi[1*], S. D. Gosling[2,3], Á. Miklósi[4]

[1]MTA-ELTE Comparative Ethology Research Group, Budapest

[2]Department of Psychology, University of Texas, Austin

[3]School of Psychological Sciences, University of Melbourne, Parkville VIC 3010, Australia

[4]Department of Ethology, Eötvös University, Budapest

Running title: Rating and coding in dogs

Abstract

The aim of the present study was to examine the links between independent rating and coding approaches to assessing activity-impulsivity and inattention in dogs. Fifty-six adult Belgian shepherd dogs were videotaped performing in behavioural tests. 17 behavioural variables were measured by coders (video coding).

Raters watched the same videotapes and then rated the activity-impulsivity and inattention of each dog (video rating). Owners filled out the Dog ADHS-RS questionnaire measuring activity-impulsivity and inattention.

Video rating of activity-impulsivity correlated with the scale scores of the owner, but video codings did not. The results suggest that the owner ratings and video ratings are tapping the same constructs, but behavioural variables assessed in the present study were not appropriate for mirroring the owners' assessments. The findings suggest that if consistent individual differences in broad behavioural traits are the primary focus of analyses, then ratings seem to capture information not easily captured in coding approaches designed to assess the same constructs.

Keywords:

dog - behavioural trait – activity – coding - rating

INTRODUCTION

The core empirical task for animal-behaviour researchers is to capture how animals behave. Two main methods are used for recording information about the behaviour of individual animals: Behavioural coding using detailed ethograms [18, 16] and subjective ratings [7, 32, 22, 25]. The two methods reflect different resolutions to the supposed trade-off between quantifying behaviour in terms of objective acts and using humans to record and collate information more subjectively.

Behavioural-coding approaches, rooted in the tradition of Ethology, aim to capture what an animal does on a particular occasion in terms of discrete well-defined behavior units, without reference to their function; for example, researchers might count the number of times an animal performs an act (e.g., charges at another), the latency to do something (e.g., time taken to approach a novel object), or the duration of a behaviour (e.g., time spent looking at another animal). Coding approaches are widely thought to be objective because they are based on observed motor patterns alone, and in theory at least, are not influenced by observers' perceptual and interpretational biases.

Rating approaches, rooted in the tradition of Psychology, aim to capture what an animal does at a higher level of abstraction than specific behaviours. Two types of rating exist: Behaviour rating and adjective rating [27]. Behaviour-rating items describe actions, without using adjectives and require observers to make frequency assessments (e.g., "bites conspecifics when threatened" could be rated from "rarely" to "often"). Adjective-rating items, are even more abstract, requiring observers to use the adjectives' implicit meanings to summarize a range of behaviours (e.g., rating an animal's behavioural history on a scale ranging from "unaggressive" to "aggressive").

Both types of rating intrinsically rely on the experience and judgment of observers. Therefore they are widely considered to be less objective than coding approaches; indeed, they are often referred to as "subjective ratings" (e.g., [24]). However, several researchers have argued that aggregated ratings of multiple observers are reliable and independent of the peculiarities of individual observers (for a review see [31]). In fact, based on psychometric grounds, some researchers have even argued that subjective ratings should be superior to behavioural codings in terms of reliability [30, 34]. Additionally, collating information about animals from experienced observers via broad ratings is relatively efficient compared to behavioural codings, which can be very time consuming.

Both rating methods often correlate with behavioral codings (e.g 12, 20, 10]. What is less well known is whether coding or rating differs in their predictive validity with regards to personality traits.

By definition, personality refers to broad trends—consistency in behaviours across time and situations. So in addition to the psychometric and pragmatic arguments, conceptual arguments suggest that rating approaches are well suited to measuring personality because they capture behaviour at a higher level of abstraction than is found in behaviour codings, which take a more molecular approach.

To our knowledge, no studies have previously compared the predictive validity of subjective rating and behavioural coding. Such findings are important especially regarding to dogs (*Canis familiaris*). The domestic dog is popular in personality studies [see 11, 4, 19 for reviews], and measuring personality in dogs has significant practical implications (e.g., for animal welfare and for selecting working dogs). It is useful for personality assessments to know whether the more objective but time consuming behavioural coding or adjective rating that includes greater levels of observer aggregation is superior in capturing variance between dogs in their personality traits.

The aim of the present study was to compare the predictive potential of behavioural codings and adjective ratings of the activity-impulsivity and inattention traits in dogs. We used a single breed, Belgian shepherd, in order to control for differential observer biases based on breed-specific expectations.

We chose to focus on individual differences in the configuration of activity-impulsivity and inattention behaviours because these traits have recently generated interest due to their potential for serving as an animal model for attention-deficit hyperactivity disorder (ADHD) in humans [29]. Specifically, [29] used a modified version of the widely applied human ADHD Rating Scale [2] in a large population of Hungarian dogs; they identified two scales (activity-impulsivity and inattention), a finding that was replicated in North American dogs [17].

Thus, our goal here is not to evaluate the coherence and plausibility of the activity-impulsivity and inattention dimensions; instead, our goal is to evaluate the convergence between a previously validated measure of these traits and two alternative methods (codings and ratings) for assessing them. In a narrow sense, these findings will determine whether the activity-impulsivity and inattention scores that are now collected via owner-completed questionnaires can be recovered from codings and ratings of videotaped behaviours. Such findings are important because there could be many cases (e.g., dogs in shelters or research facilities) where owner reports are not available. In a broader sense, the findings will contribute to the fledgling literature on the validity and usefulness of codings and ratings of behaviour in assessing broad traits (e.g. personality) in dogs and other animals [30].

In previous research comparing rating and coding methods, the ratings and codings have usually been made by the same individual, thereby compromising the independence of the two measures (e.g., [3], but see [22]). In our study based on videotapes of dogs' behaviour in a test battery, different individuals, completely

unacquainted with the dogs, were used as coders and raters. Their sets of scores were validated against a criterion measure consisting of a questionnaire scale [29] completed by the dog owners.

MATERIALS AND METHODS

Ethics statement

Non-invasive studies on dogs are currently allowed to be done without any special permission in Hungary by the University Institutional Animal Care and Use Committee (UIACUC, Eötvös Loránd University, Hungary). The currently operating Hungarian law "*1998. évi XXVIII. Törvény"* – the Animal Protection Act – defines experiments on animals in the 9th point of its 3rd paragraph (3. §/9.). According to the corresponding definition by law, our non-invasive observational study is not considered as an animal experiment. The owners volunteered to participate.

Subjects

Subjects were 56 Belgian Shepherd Dogs (two varieties: Tervuerens: 12 males, 14 females, mean age: 5.9 years ± 3.6 SD, 4 dogs were untrained, 5 were beginner (had basic obedience exam) and 17 dogs were advanced (had agility, guarding, etc. exams); Groenendaels: 13 male, 17 female, mean age: 4.4 years ± 2.6 SD, 5 dogs were untrained, 23 were beginner, and 2 dogs were advanced). Behavioural tests were conducted in Budapest in the park surrounding the building of the Eötvös University or in the laboratory of the Department of Ethology. Due to technical reasons, not every test was recorded suitably, so the number of subjects varies from test to test (N-s are always indicated when reporting the results of the statistical models). Tervuerens were better trained ($chi2 = 23.36$, $df = 2$, $p < 0.001$). These subjects were part of a larger sample participating in a test-series conducted both indoors and outdoors on two different days, designed to evaluate the personality characteristics of pet dogs [26].

Procedure

We collected data from three sources: scale scoring, video coding and video scoring.

Scale-scoring

Before the behavioural tests, as a criterion measure, owners completed the validated ADHD-RS questionnaire [29]. Numbers in brackets indicate the order of the questions in the original survey.

Activity-impulsivity scale: (4) Your dog leaves from his/her place when he/she should stay. (5) Your dog cannot be quiet; he/she cannot be easily calmed. (6) Your dog fidgets all the time. (8) Your dog is excessively difficult to control; if he/she lunges, it is hard to hold him/her back, (9) Your dog always wants to play and run. (11) Your dog is likely to react hastily, and that is why he/she is failing tasks. (13) Your dog cannot wait; he/she has no self-control.

Inattention scale: (1) Your dog has a difficult time learning, because he/she is careless, or other things can easily attract his/her attention. (2) It is easy to attract your dog's attention, but he/she loses his/her interest soon. (3) It is difficult for your dog to concentrate on a task or play. (7) It seems that your dog does not listen even if he/she knows that someone is speaking to him/her. (10) Your dog solves simple tasks easily, but he/she often has difficulties with complicated tasks, even if he/she knows them and has practiced them often. (12) Your dog's attention can be easily distracted. (The order of the items in the questionnaire was the same as published in Vas et al., 2007).

Video coding

Evaluating dogs in a narrow range of contexts could bias the findings by eliciting the kinds of behaviours that were particularly amenable to measurement by just one of the two methods. Therefore, it was important to assess dogs across a range of contexts. The owner, an unfamiliar female experimenter, and a camera-woman were present during the tests.

Test battery procedure

Test 1. Greeting: This test was the same as in [9] for measuring 'social impulsivity'. The owner stood motionless next to the dog and held the leash. An unfamiliar woman approached the dog in a friendly way. She stopped out of reach of the leash and waited for 3 seconds. If the dog was not aggressive, she stepped next to the dog then petted the dog's head and back.

Test 2. Collecting saliva sample 1: If the experimenter could caress the dog, she crouched to the dog and she pretended collecting saliva sample by wiping a little piece of cotton wool next to the lower molars. In case of difficulties the owner was allowed to control and calm the dog or even to collect the sample alone (Figure 1a).

Test 3. Putting on boots and walking (5 min): The owner tried to put 1-1 dog boots on the forelegs of the dog. He or she could try this for 1 min. In case of success he or she tried to walk the dog on a leash for 6 meters. If the dog took off the boots, the owner had to put it back, and they continued walking.

Test 4. Problem solving: A piece of meat was attached to one end of a rope, and was put into a cage out of reach of the dog. However, a 6-7-cm long part of the rope hung out from the cage. The meat could be reached by pulling out the rope. The owner stood 1 m in front of the cage and held the leash of the dog. Trial ended when the dog gets the meat in his mouth, or after 1 min. (Figure 1b).

Test 5. Threatening approach: This was based on [28]. The dog was tethered to a tree. The experimenter greeted the dog as in Test 1. Then she stepped back 10 meters, and approached the dog slowly, by leaning forward her upper body and staring at the eyes of the dog. The owner stood next to the dog. The experimenter stopped approaching when the dog showed signs of aggression or when she reached the dog. Finally, the experimenter stepped back to the starting point, crouched, and asked the owner to let the dog free. Then she started to call the dog in a friendly way.

Test 6. Separation from the owner: The dog was tethered to a tree. The owner was talking with the experimenter 10 m away. The duration of the test was 2 min.

Test 7. Walking with the dog: The dog was walked off leash by the owner in the park. On the experimenter's signal the owner bended down and he or she pretended to "search" for something on the ground for 20 seconds. Then the experimenter held the dog on leash, meanwhile the owner hid behind some landmark object (e.g. a tree) approx. 15-20 meters far from the dog ("hiding"). After 30 seconds the experimenter released the dog and told to it: "You may go!" If the dog did not start moving for 5 seconds, the owner was told to call it (Figure 1c).

Test 8. Collecting saliva sample 2: See Test 2.

Behavioural variables

Seventeen video coding variables were developed to capture elements of activity-impulsivity and inattention. The variables were derived on the basis of our conceptualizations of activity-impulsivity and inattention. Motor activity is generally defined in terms of displacement behaviour. Impulsivity measures can be broadly divided into two categories: impulsive decision-making and motor impulsivity. For assessing inattention (attention-deficit) we measured variables related to the attentional focus of the dogs.

Activity-impulsivity

Behavioural variables:

1-4. Duration of moving the forelegs in Test 1, 4, 5 and 6 (time %).

5-6. Reaction to the experimenter in Test 1 and 5 (greeting episode) (scores between 1 and 6): The dog's behaviour is assessed two times: first before petting, second after petting: (6) The dog approaches the experimenter immediately and shows no signs of aggression; (5) The dog delays the approach but is not aggressive; (4) The dog neither approaches nor avoids the experimenter; (3) The dog shows tendency to avoid the experimenter, (2) The dog barks, growls at the experimenter on loose leash. (1) The dog barks, growls on tight leash. The two scores (before and after patting) were averaged.

7-8. Reluctance in Test 2 and 8 (scores between 0 and 3): (3) The dog does not move during the test; (2) The dog moves its head; (1) The dog moves its body and/or the owner has to help the experimenter in collecting the saliva; (0) The owner has to collect the sample alone (the dog is aggressive or fidgets exceedingly).

9. Reluctance in Test 3 (scores between 0 and 3): (3) The dog walks the usual way; (2) The dog stops and pulls up the legs at least once, (1) The dog tries to get rid of the boots at least once; (0) The walking is not possible with the boots.

10. Reaction to the experimenter at the beginning of the threatening approach in Test 5 (scores between 1 and 6): The dog's behaviour is assessed at the beginning and at the end of the test. (6) The dog approaches the experimenter not aggressively (it does not growl or bark); (5) The dog does not approach or avoid the experimenter, and does not show signs of aggressive behaviour; (4) The dog avoids the experimenter and does not show of aggressive behaviour; (3) The dog avoids the experimenter and shows signs of aggressive behaviour; (2) The dog barks, growls on loose leash toward the experimenter; (1) The dog barks, growls on tight leash toward the experimenter.

11. Approach style after hiding in Test 7 (Scores between 0 and 3): (3) The dog gallops to the owner, (2) The dog trotts to the owner, (1) The dog walks to the owner, (0) The dog does not go to the owner.

Inattention

Behavioural variables

12-14. Number of looking at the owner in Test 3, 4 and 5 (scores between 1 and 3): (3): the dog looks at the owner 3 or more times, (2) the dog looks at the owner 2 times, (1) the dog looks at the owner once, (0) the dog does not look at the owner.

15. Duration of orientation toward the cage in Test 4 (time %). The dog's nose is at least 20 cm from the cage.

16. Orientation toward the threatening experimenter in Test 5 (time%).

17. Searching in Test 7 (scores between 0 and 2): (2) The dog goes to the owner and orientates at the investigated point; (1) The dog approaches the owner but does not orientate at the investigated point, (0) The dog does not approach the owner.

All 17 behavioural variables for each dog were measured by Coder 1. Twenty percent of the dogs (N = 11) were also assessed by Coder 2.

Video rating

Two observers rated single-item measures of activity-impulsivity and inattention after watching a videotape of the subjects performing in the test battery. Immediately after watching each videotape, they rated the traits of the dogs on a 5-point Likert-scale, ranging from 1 (not characteristic to the dog) to 5 (very characteristic). Rater 2 rated 20% of the dogs (N = 11).

Statistical analysis

Inter-observer reliability between Coder 1 and Coder 2 and Rater 1 and Rater 2 was computed using intraclass correlation coefficient (ICC 1,1, one-way random single measures). Video-coding variables were not normally distributed. Convergent and discriminant validity were evaluated by computing Spearman

correlations among the variables obtained in the video codings, video ratings, and the scale scoring. Internal consistency of the video-coding variables was measured by computing Cronbach's coefficient alpha.

RESULTS

Inter-observer reliability was strong for both video-coding and video-rating measures. ICCs between Coder 1 and Coder 2 ranged from 0.76 to 0.97. ICCs between Rater 1 and Rater 2 were 0.85 for activity impulsivity and 0.81 for inattention. In this study we did not assess inter-observer agreement for the scale-scorings because [29] reported earlier satisfactory measures of reliability; specifically, they reported test-retest reliability correlations of 0.60 ($p < 0.001$, $N = 48$) and inter-observer agreement correlations of 0.60 ($p < 0.01$, $N = 25$).

Cronbach's alpha was 0.19 for the eleven activity-impulsivity variables and 0.38 for the six inattention variables, suggesting that the codings did not tap into unitary underlying constructs.

We assessed convergent validity by evaluating the extent to which the owner questionnaire criterion measure converged with the video codings and the video ratings. As shown in Table 1, the scale scores (criterion measures) for activity-impulsivity showed significant convergent correlations with video ratings ($r = 0.42$, $p = 0.001$, $N = 56$) but not for the video codings. Similarly, the scale scores for inattention showed significant convergent correlations with video ratings ($r = 0.31$, $p < 0.05$, $N = 50$) but not for the video codings.

Discriminant validity reflects the extent to which a measure is unrelated to measures to which it is theoretically unrelated. Discriminant validity is generally neglected in animal-personality studies [11]. As shown in Table 1, the "off diagonal" correlations between scale scores of activity-impulsivity and video codings and video ratings of inattention and between scale scores of inattention and video codings and video ratings of activity-impulsivity were all low and non-significant.

Table 1 also presents the correlations between the video codings and video ratings. Activity-impulsivity video rating correlated with variable 4: Duration of moving the forelegs in Test 6 ($r = 0.38$, $p < 0.01$, $N = 50$), variable 5: Reaction to the experimenter in Test 1 ($r = 0.30$, $p < 0.05$, $N = 55$), and variable 6: Reaction to the experimenter in Test 5 ($r = 0.33$, $p < 0.05$, $N = 50$).

For checking consistency we run the same analysis separately on four subsamples: Groenedaels, Tervuerens, males and females. The positive correlation between activity-impulsivity scale scoring and video rating is consistent, as it emerged in the Groenendael, Tervueren and female subpopulation. Other correlations were less consistent as they emerged in 0-1 subsamples.

DISCUSSION

In this study our main goal was to examine the relative effectiveness of rating and coding approaches to recovering owners' ratings of their dogs on a validated measure of activity-impulsivity and inattention. Based on videotapes of dogs' behaviour in an approximately 20-min long test battery, independent observers coded the observable behaviours (e.g., duration of leg-moving) and independent raters rated the behaviour of the dogs. These two sets of scores were correlated with a criterion measure consisting of a questionnaire scale [29] completed by the dog owners.

According to [23] correlations stabilize when the number of subjects approaches 250. Although our overall sample size was considerably smaller, the analysis of subpopulations indicated that correlations between scale scoring and video rating is consistent, at least in case of activity-impulsivity trait. Thus our analyses indicate that the activity impulsivity video rating seemed to be measuring the same construct as the owner's scale-scores but the video-coding variables did not.

It is promising to note that merely watching a sequence of a dog's performance in a test-battery predicted the relevant trait scores derived from information gleaned over a much longer period of time. This finding is consistent with [8] who found significant correlations between owners' ratings of their dogs and ratings of dogs made by strangers on the basis of the dogs' performance on a set of behavioural tests.

The result suggests that ratings may be better suited than codings to capturing a construct like activity-impulsivity because the human raters are better equipped to take into account the way the activity-impulsivity is expressed and other indicators of activity-impulsivity that may not have been specified in the coding definition; for example, the rater is not restricted to viewing how much the dogs' legs move but can also consider other elements, like moving the head, the tail, ears, and the frequency of jumping. Of course,

289 it's possible that coding definitions could be expanded to include such additional behaviours but it may be
290 difficult to specify beforehand all the ways in which activity-impulsivity (or any other trait) is going to be
291 expressed and it may not be possible to fully define the variety and configuration of all the subtle behaviours
292 that constitute activity-impulsivity.
293 When completing a questionnaire, owners rely on several years of experiences of seeing their dogs
294 performing across a wide variety of situations, not just those performed during behavioural tests, and they
295 aggregate information from multiple observations. Therefore, we believe the decision to use an owner-
296 completed questionnaire with demonstrated reliability and validity as a criterion was a reasonable one.
297 Although inexperienced dog owners may have little information with which to compare their dog's
298 behaviour, making their ratings unreliable, recent research has suggested that different levels of experience
299 with dogs are not critical in rating the majority of behaviours. For example, [25] compared the ratings of
300 dog-owners, veterinarians, dog-trainers and non-owners, and found that they did not differ in proper labeling
301 of indifferent, fearful, confident, friendly, submissive, defensive, playful and aggressive behaviour. In a
302 similar vein, [22] reported that trained scientists and skilled search dog operatives rated search dogs'
303 behaviour alike, and [21] also found no differences between dog owners and non-owners in regard to their
304 ability to judge the emotional attributes of dog barks.
305 Previous studies comparing ratings with codings have yielded mixed results. [30] found some evidence for
306 strong convergence between coding methods and rating methods; for example, codings of threat behaviours
307 in chimpanzees correlated 0.52 with ratings on the trait belligerence. But there were many instances where
308 expected convergences were not found; for example, codings of fleeing behaviour were correlated only 0.15
309 with ratings of submissiveness. The failure to find convergences across two measures that are theoretically
310 tapping the same underlying construct suggests that at least one of them—possibly both of them—are wrong.
311 The question for researchers in such cases of divergence is, which measure, if any, should be considered the
312 most trustworthy. The fact that most previous studies have used codings to validate ratings, not the other way
313 around, reflects the widely held assumption in the fields of Ethology and Animal Behaviour that codings are
314 more trustworthy than are ratings.
315 Arguments can be advanced in favor of the theoretical superiority of either approach [30]. From an empirical
316 standpoint, the relative superiority of the two measures can be evaluated with regard to a criterion that is

theoretically tapping the same construct as the two measures. That was the approach taken in the present research. Coding and rating methods measures were evaluated side-by-side with regard to criterion measure furnished by dog owners. Clearly, the value of this analysis rides on the appropriateness of the criterion measure.

The value of the analysis also rides on the selection of appropriate behaviors to code. So it is possible that the failure to find convergence between the codings and the scale scores was due to a poor selection of coding variables and the use of a poor coding scheme. However, the procedures adopted here to identify suitable coding schemes were not dissimilar from those widely used in behavioral ecology and applied settings.

It is important to acknowledge that every measure has its faults and limitation. For example it is known that owners can be biased in their views of their dogs and their ratings may be vulnerable to anthropomorphic projections [13]. Moreover, the fact that both the video ratings and the owner-provided scale scores were based on rating-scales results in shared method variance that could have contributed to the relative superiority of the video ratings over the video codings. Thus, rather than viewing these findings as the final arbiter of which method is best, we instead view them merely as a single piece of evidence in the broader construct validation endeavor [1].

Obviously, we would not like to suggest that personality traits should not be measured by ethologically based coding methods. Indeed, activity-impulsivity scale-scores of German Shepherd Dogs performing in a modified version of the test-battery used in this study correlated with a behavioural scale with high internal consistency [15]. Still, it is remarkable from the present result that finding the right test and ethological variables for assessing a personality trait proved exceedingly difficult.

The preferred method used for describing the behaviour could also depend on the goals of the particular study, and also on the means that are available for research. The utilization of ratings seems to be a convenient way to get first hand information on behaviour, without a need for understanding the detailed structure of it. Thus ratings could be used when testing for effects of independent factors (e.g., age, breed) or when examining associations between traits displayed by dogs and their owners. However, behaviour coding cannot be avoided if the goal is to develop a functional model of behaviour, in which the model determines or predicts specific occurrences of behavior in space and time (e.g., to develop virtual and robotic agents that mimic the behavior of animals; [14]).

In sum, the pattern of findings suggests, that at least in the case of activity-impulsivity, rating is a viable method for assessing behaviour; in fact, our findings suggest that rating approaches are, under some circumstances, superior to coding approaches. However, every method has its own set of advantages and disadvantages so it is unlikely that any one method is optimal in all situations. For example, in the case of carefully controlled studies where the specific frequencies of behaviour in that situation are of interest, coding methods may be appropriate. But if consistent individual differences in higher level traits are the focus of the analysis, then it seems that raters use valid information that goes beyond mere motor activity. Given that codings are typically considerably more time-consuming than ratings, we suggest that researchers consider using rating method for measuring personality traits in dogs.

ACKNOWLEDGEMENTS

We are thankful for Orsolya Dán, Márta Gácsi, Bianca Jansen, Anita Kurys, Éva Péch, József Topál and Lilla Tóth for their help in designing and conducting the tests, and collecting the data. This research was supported partly by the Hungarian Scientific Research Fund (K84036), the Bolyai Foundation and the Hungarian Academy of Sciences (MTA 01 031) of the Hungarian Academy of Sciences for E.K. and Á.M., and the National Science Foundation for S. D. G. (0731216).

363 REFERENCES

364

365 1. Cronbach, L.J., Meehl, P.E. (1955) Construct validity in psychological tests. *Psychol. Bull. 52,* 281-302.

366 2. DuPaul, G.J. (1998) ADHD Rating Scale-IV: Checklist, Norms and Clinical Interpretations. Guilford
367 Press, New York.

368 3. Feaver, J., Mendl, M., Bateson, P. (1986) A method for rating the individual distinctiveness of domestic
369 cats. *Anim. Behav. 34,* 1016-1025

370 4. Fratkin, J. L., Sinn, D. L., Patall, E. A.,Gosling, S. D. (2013) Personality consistency in dogs: a meta-
371 analysis. *PloS One 8*, e54907.

372 5. Freeman, H.D., Gosling, S.D. (2010) Personality in nonhuman primates: A review and evaluation of
373 past research. *Amer. J. of Primatol. 71,* 1-19.

374 6. Gosling, S.D. (2001) From mice to men: What can we learn about personality from animal research?
375 *Psychol. Bull. 127,* 45-86.

376 7. Gosling, S.D., Bonnenburg, A.V. (1998) An integrative approach to personality research in
377 anthrozoology: Ratings of six species of pets ant their owners. *Anthrozoös, 11,* 148-156.

378 8. Gosling, S.D., Kwan, V.S.Y., John, O.P. (2003) A dog's got personality: A cross-species comparative
379 approach to evaluating personality judgments. *J. Pers. Soc. Psychol. 85*, 1161-1169.

380 9. Hejjas, K., Kubinyi, E., Ronai, Zs., Szekely, A., Vas, J., Miklósi, Á., Sasvari-Szekely, M., Kereszturi, E.
381 2009. Molecular and behavioural analysis of the intron 2 repeat polymorphism in canine dopamine D4
382 receptor gene. *Genes Brain Behav. 8,* 330-336.

383 10. Horback, K. M., Miller, L. J., Kuczaj II, S. A. (2013) Personality assessment in African elephants
384 (Loxodonta africana): Comparing the temporal stability of ethological coding versus trait rating. *Appl.*
385 *Anim. Behav. Sci. 149,* 55-62.

386 11. Jones, A.C., Gosling, S.D. (2005) Temperament and personality in dogs (Canis familiaris): A review
387 and evaluation of past research. *Appl. Anim. Behav. Sci. 95,* 1-53.

388 12. Ijichi, C., Collins, L. M., Creighton, E., & Elwood, R. W. (2013). Harnessing the power of personality
389 assessment: subjective assessment predicts behaviour in horses. *Behav. Proc. 96*, 47-52.

390 13. Kwan, V.S.Y., Gosling, S.D., John, O.P. (2008) Anthropomorphism as a special case of social

perception: A cross-species comparative approach and a new empirical paradigm. *Soc. Cogn. 26,* 129-142.

14. Kubinyi, E., Miklósi, A., Kaplan, F., Gácsi, M., Topál, J., Csányi, V. (2004) Social behaviour of dogs encountering AIBO, an animal-like robot in a neutral and in a feeding situation *Behav. Proc. 65,* 231-239.
15. Kubinyi, E., Vas, J., Héjjas, K., Ronai, Zs., Brúder, I., Turcsán, B., Sasvári-Székely, M., Miklósi, Á. (2012) Polymorphism in the tyrosine hydroxylase (TH) gene is associated with activity-impulsivity in german shepherd dogs. *PLoS One*, 7, e30271.
16. Lehner, P.N. (1996) *Handbook of ethological methods.* Cambridge University Press.
17. Lit, L., Schweitzer, J.B., Iosif, A.M., Oberbauer, A.M. (2010) Owner reports of attention, activity, and impulsivity in dogs: a replication study. *Behav. Brain Functions. 6,* 1.
18. Martin, P., Bateson, P. (1993) *Measuring Behaviour: An introductory guide.* Cambridge University Press, Cambridge
19. Miklósi, Á. (2007) *Dog behaviour, evolution, and cognition.* Oxford University Press, Oxford, United Kingdom.
20. Mirkó, E., Dóka, A., Miklósi, Á. (2013) Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (Canis familiaris). *Appl. Anim. Behav. Sci.* 149, 45-54.
21. Pongrácz, P., Miklósi, Á., Molnár, Cs., Csányi, V. (2005) Human listeners are able to classify dog barks recorded in different situations. *J. Comp. Psychol. 119,* 136-144.
22. Rooney, N.J., Gaines, S.A., Bradshaw, J.W.S., Penman, S. (2007) Validation of a method for assessing the ability of trainee specialist search dogs. *Appl. Anim. Behav. Sci. 103,* 90-104.
23. Schönbrodt, F.D., Perugini, M. (2013) At what sample size do correlations stabilize? *J. Res. Pers.* 47, 609-612.
24. Stevenson-Hinde, J., Zunz, M. (1978) Subjective assessment of individual rhesus monkeys. *Primates. 19,* 473-482.
25. Tami G., Gallagher A. (2009) Description of the behaviour of domestic dog (Canis familiaris) by experienced and inexperienced people. *Appl. Anim. Behav. Sci. 120,* 159-169.

419 26. Tóth L., Gácsi, M., Topál, J., Miklósi, Á. (2008) Playing styles and possible causative factors in dogs'
420 behaviour when playing with humans. *Appl. Anim. Behav. Sci. 114,* 473-484.
421 27. Uher, J., Asendorpf, J.B. (2008) Personality assessment in the Great Apes: Comparing ecologically
422 valid behaviour measures, behaviour ratings, and adjective ratings. *J. Res. Pers. 42,* 821-838.
423 28. Vas, J, Topál, J, Gácsi, M, Miklósi, Á, Csányi, V. (2005) A friend or an enemy? Dogs' reaction to an
424 unfamiliar person showing behavioural cues of threat and friendliness at different times. *Appl. Anim.*
425 *Behav. Sci., 94,* 99-115.
426 29. Vas J., Topál J., Péch, É., Miklósi, Á. (2007) Measuring attention deficit and activity in dogs: A new
427 application and validation of a human ADHD questionnaire. *Appl. Anim. Behav. Sci. 103,* 105–17.
428 30. Vazire, S., Gosling, S.D., Dickey, A.S., Schaprio, S.J. (2007) Measuring personality in nonhuman
429 animals. In R.W. Robins, R.C. Fraley, R. Krueger (Eds.), *Handbook of research methods in personality*
430 *psychology.* New York, NY: Guilford, pp. 190-206.
431 31. Weinstein, T.A.R., Capitanio, J.P., Gosling, S.D. (2008) Personality in Animals. In John, O.P., Robins,
432 R.W., Pervin, L.A. (Eds.), *Handbook of Personality Theory and Research*. New York: Guilford, pp.
433 328-348.
434 32. Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B. (2001) Assessing the 'whole animal': a
435 Free-Choice-Profiling approach. *Anim. Behav. 62,* 209-220.
436 33. Wilsson, E., Sundgren, P-E. (1998) Behaviour test for eight-week old puppies—heritabilities of tested
437 behaviour traits and its correspondence to later behaviour. *Appl. Anim. Behav. Sci. 58,* 151-162.
438 34. Wilsson, E., Sinn, D. L (2012) Are there differences between behavioral measurement methods? A
439 comparison of the predictive validity of two ratings methods in a working dog program. *Appl. Anim.*
440 *Behav. Sci. 141,* 158–172.
441
442

Table 1. Correlations between scale scorings, video codings and video ratings (Spearman's rho). Convergent correlations (which are expected to be significant for valid assessment methods) have a gray background. Significant correlations are highlighted with asteriks (** = p<0.01, * = p < 0.05).

Variables		Activity-impulsivity scale scoring	Inattention scale scoring	Activity-impulsivity video rating	Inattention video rating
Activity-impulsivity video rating		**0.42****	-0.04	1	-0.03
Inattention video rating		0.00	**0.31***	-0.03	1
Activity-impulsivity video-codings	1. Duration of moving the forelegs in Test 1	**0.06**	0.05	**0.21**	0.15
	2. Duration of moving the forelegs in Test 4	**0.01**	-0.14	**-0.05**	0.00
	3. Duration of moving the forelegs in Test 5	**0.00**	-0.12	**-0.25**	0.12
	4. Duration of moving the forelegs in Test 6	**0.24**	0.07	**0.38****	0.24
	5. Reaction to the experimenter in Test 1	**0.00**	-0.20	**0.31***	-0.17
	6. Reaction to the experimenter in Test 5	**0.14**	-0.14	**0.33***	0.06
	7. Reluctance in Test 2	**0.04**	0.21	**0.19**	0.05
	8. Reluctance in Test 8	**-0.07**	-0.19	**0.02**	-0.11
	9. Reluctance in Test 3	**-0.13**	-0.23	**0.04**	0.00
	10. Reaction to the experimenter at the beginning of the threatening approach in Test 5	**-0.06**	-0.02	**-0.25**	-0.25
	11. Approach style after hiding in Test 7	**-0.03**	-0.05	**0.21**	-0.29
Inattention video-codings	12. Number of looking at the owner in Test 3	0.15	**0.04**	0.08	**0.07**
	13. Number of looking at the owner in Test 4	-0.01	**0.08**	0.10	**-0.05**
	14. Number of looking at the owner in Test 5	-0.06	**-0.21**	-0.22	**0.00**
	15. Duration of orientation toward the cage in Test 4	0.07	**-0.20**	0.25	**-0.25**
	16. Orientation toward the threatening experimenter in Test 5	0.04	**0.02**	0.14	**0.00**
	17. Searching in Test 7	0.17	**-0.06**	0.10	**-0.28**

Figure 1. Illustrations for the subtests. A) Collecting the saliva sample: The experimenter uses a small clot of cotton to collect saliva from the inner side of the mouth. B) Problem solving test: The dog can pull out a rope from a cage in order to get a treat attached to the rope. C) Hiding of the owner: The experimenter holds the dog on leash, meanwhile the owner hides behind a tree) approx. 15-20 meters far from the dog. After 30 seconds the experimenter releases the dog.