

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

2 **A comparison of rating and coding behavioural traits in dogs**

3 Enikő Kubinyi^{1*}, S. D. Gosling^{2,3}, Á. Miklósi⁴

4

5 ¹MTA-ELTE Comparative Ethology Research Group, Budapest

6 ²Department of Psychology, University of Texas, Austin

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

8 ⁴Department of Ethology, Eötvös University, Budapest

9

10 Running title: Rating and coding in dogs

11

12 **Abstract**

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

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

19 Video rating of activity-impulsivity correlated with the scale scores of the owner, but video codings did not.

20 The results suggest that the owner ratings and video ratings are tapping the same constructs, but behavioural
21 variables assessed in the present study were not appropriate for mirroring the owners' assessments. The
22 findings suggest that if consistent individual differences in broad behavioural traits are the primary focus of
23 analyses, then ratings seem to capture information not easily captured in coding approaches designed to
24 assess the same constructs.

25

26 *Keywords:*

27 dog - behavioural trait – activity – coding - rating

28

29

30

INTRODUCTION

31

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

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

44 Rating approaches, rooted in the tradition of Psychology, aim to capture what an animal does at a higher
45 level of abstraction than specific behaviours. Two types of rating exist: Behaviour rating and adjective rating
46 [27]. Behaviour-rating items describe actions, without using adjectives and require observers to make
47 frequency assessments (e.g., "bites conspecifics when threatened" could be rated from "rarely" to "often").
48 Adjective-rating items, are even more abstract, requiring observers to use the adjectives' implicit meanings to
49 summarize a range of behaviours (e.g., rating an animal's behavioural history on a scale ranging from
50 "unaggressive" to "aggressive").

51 Both types of rating intrinsically rely on the experience and judgment of observers. Therefore they are
52 widely considered to be less objective than coding approaches; indeed, they are often referred to as
53 "subjective ratings" (e.g., [24]). However, several researchers have argued that aggregated ratings of
54 multiple observers are reliable and independent of the peculiarities of individual observers (for a review see
55 [31]). In fact, based on psychometric grounds, some researchers have even argued that subjective ratings
56 should be superior to behavioural codings in terms of reliability [30, 34]. Additionally, collating information
57 about animals from experienced observers via broad ratings is relatively efficient compared to behavioural
58 codings, which can be very time consuming.

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

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

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

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

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

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

90 In previous research comparing rating and coding methods, the ratings and codings have usually been made
91 by the same individual, thereby compromising the independence of the two measures (e.g., [3], but see [22]).

92 In our study based on videotapes of dogs' behaviour in a test battery, different individuals, completely

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

95

96

97

MATERIALS AND METHODS

98

99 **Ethics statement**

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

106

107 **Subjects**

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

118

119 **Procedure**

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

121

122 *Scale-scoring*

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

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

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

137

138 *Video coding*

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

143

144 *Test battery procedure*

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

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

153 **Test 3. Putting on boots and walking (5 min):** The owner tried to put 1-1 dog boots on the forelegs of the
154 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
155 meters. If the dog took off the boots, the owner had to put it back, and they continued walking.

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

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

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

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

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

175

176 *Behavioural variables*

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

182

183 **Activity-impulsivity**

184 Behavioural variables:

185 1-4. Duration of moving the forelegs in Test 1, 4, 5 and 6 (time %).

186 5-6. Reaction to the experimenter in Test 1 and 5 (greeting episode) (scores between 1 and 6): The dog's
187 behaviour is assessed two times: first before petting, second after petting: (6) The dog approaches the
188 experimenter immediately and shows no signs of aggression; (5) The dog delays the approach but is not
189 aggressive; (4) The dog neither approaches nor avoids the experimenter; (3) The dog shows tendency to
190 avoid the experimenter, (2) The dog barks, growls at the experimenter on loose leash. (1) The dog barks,
191 growls on tight leash. The two scores (before and after patting) were averaged.

192 7-8. Reluctance in Test 2 and 8 (scores between 0 and 3): (3) The dog does not move during the test; (2) The
193 dog moves its head; (1) The dog moves its body and/or the owner has to help the experimenter in collecting
194 the saliva; (0) The owner has to collect the sample alone (the dog is aggressive or fidgets exceedingly).

195 9. Reluctance in Test 3 (scores between 0 and 3): (3) The dog walks the usual way; (2) The dog stops and
196 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
197 possible with the boots.

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

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

207

208 **Inattention**

209 Behavioural variables

210 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
211 owner 3 or more times, (2) the dog looks at the owner 2 times, (1) the dog looks at the owner once, (0) the
212 dog does not look at the owner.

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

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

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

219

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

222

223 *Video rating*

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

227 Rater 2 rated 20% of the dogs (N = 11).

228

229

Statistical analysis

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

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

235

236

237

RESULTS

238

239 Inter-observer reliability was strong for both video-coding and video-rating measures. ICCs between Coder 1
240 and Coder 2 ranged from 0.76 to 0.97. ICCs between Rater 1 and Rater 2 were 0.85 for activity impulsivity
241 and 0.81 for inattention. In this study we did not assess inter-observer agreement for the scale-scorings
242 because [29] reported earlier satisfactory measures of reliability; specifically, they reported test-retest
243 reliability correlations of 0.60 ($p < 0.001$, $N = 48$) and inter-observer agreement correlations of 0.60 ($p <$
244 0.01 , $N = 25$).

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

247 We assessed convergent validity by evaluating the extent to which the owner questionnaire criterion measure
248 converged with the video codings and the video ratings. As shown in Table 1, the scale scores (criterion
249 measures) for activity-impulsivity showed significant convergent correlations with video ratings ($r = 0.42$, p
250 $= 0.001$, $N = 56$) but not for the video codings. Similarly, the scale scores for inattention showed significant
251 convergent correlations with video ratings ($r = 0.31$, $p < 0.05$, $N = 50$) but not for the video codings.

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

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

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

265

266

267

DISCUSSION

268

269 In this study our main goal was to examine the relative effectiveness of rating and coding approaches to
270 recovering owners' ratings of their dogs on a validated measure of activity-impulsivity and inattention.

271 Based on videotapes of dogs' behaviour in an approximately 20-min long test battery, independent observers
272 coded the observable behaviours (e.g., duration of leg-moving) and independent raters rated the behaviour of
273 the dogs. These two sets of scores were correlated with a criterion measure consisting of a questionnaire
274 scale [29] completed by the dog owners.

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

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

284 The result suggests that ratings may be better suited than codings to capturing a construct like activity-
285 impulsivity because the human raters are better equipped to take into account the way the activity-
286 impulsivity is expressed and other indicators of activity-impulsivity that may not have been specified in the
287 coding definition; for example, the rater is not restricted to viewing how much the dogs' legs move but can
288 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

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

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

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

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

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

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

354

355 ACKNOWLEDGEMENTS

356

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

362

- 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 and 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

- 391 perception: A cross-species comparative approach and a new empirical paradigm. *Soc. Cogn.* 26, 129-
392 142.
- 393 14. Kubinyi, E., Miklósi, A., Kaplan, F., Gácsi, M., Topál, J., Csányi, V. (2004) Social behaviour of dogs
394 encountering AIBO, an animal-like robot in a neutral and in a feeding situation *Behav. Proc.* 65, 231-
395 239.
- 396 15. Kubinyi, E., Vas, J., Héjjas, K., Ronai, Zs., Brúder, I., Turcsán, B., Sasvári-Székely, M., Miklósi, Á.
397 (2012) Polymorphism in the tyrosine hydroxylase (TH) gene is associated with activity-impulsivity in
398 german shepherd dogs. *PLoS One*, 7, e30271.
- 399 16. Lehner, P.N. (1996) *Handbook of ethological methods*. Cambridge University Press.
- 400 17. Lit, L., Schweitzer, J.B., Iosif, A.M., Oberbauer, A.M. (2010) Owner reports of attention, activity, and
401 impulsivity in dogs: a replication study. *Behav. Brain Functions.* 6, 1.
- 402 18. Martin, P., Bateson, P. (1993) *Measuring Behaviour: An introductory guide*. Cambridge University
403 Press, Cambridge
- 404 19. Miklósi, Á. (2007) *Dog behaviour, evolution, and cognition*. Oxford University Press, Oxford, United
405 Kingdom.
- 406 20. Mirkó, E., Dóka, A., Miklósi, Á. (2013) Association between subjective rating and behaviour coding
407 and the role of experience in making video assessments on the personality of the domestic dog (*Canis*
408 *familiaris*). *Appl. Anim. Behav. Sci.* 149, 45-54.
- 409 21. Pongrácz, P., Miklósi, Á., Molnár, Cs., Csányi, V. (2005) Human listeners are able to classify dog barks
410 recorded in different situations. *J. Comp. Psychol.* 119, 136-144.
- 411 22. Rooney, N.J., Gaines, S.A., Bradshaw, J.W.S., Penman, S. (2007) Validation of a method for assessing
412 the ability of trainee specialist search dogs. *Appl. Anim. Behav. Sci.* 103, 90-104.
- 413 23. Schönbrodt, F.D., Perugini, M. (2013) At what sample size do correlations stabilize? *J. Res. Pers.* 47,
414 609-612.
- 415 24. Stevenson-Hinde, J., Zunz, M. (1978) Subjective assessment of individual rhesus monkeys. *Primates.*
416 19, 473-482.
- 417 25. Tami G., Gallagher A. (2009) Description of the behaviour of domestic dog (*Canis familiaris*) by
418 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

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

446

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

447

448

449

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



455