

1 **Aggressiveness as a latent personality trait of domestic**

2 **dogs: testing local independence and measurement**

3 **invariance**

4 Conor Goold<sup>1\*</sup>, Ruth C. Newberry<sup>1</sup>

5 <sup>1</sup> Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian

6 University of Life Sciences, Ås, Akershus, Norway

7 \* Corresponding author: conor.goold@nmbu.no (CG)

## 8 **Abstract**

9 Studies of animal personality attempt to uncover underlying or 'latent' personality traits  
10 that explain broad patterns of behaviour, often by applying latent variable statistical  
11 models (e.g. factor analysis) to multivariate data sets. Two integral, but infrequently  
12 confirmed, assumptions of latent variable models in animal personality are: i)  
13 behavioural variables are independent (i.e. uncorrelated) conditional on the latent  
14 personality traits they reflect (*local independence*), and ii) personality traits are  
15 associated with behavioural variables in the same way across individuals or groups of  
16 individuals (*measurement invariance*). We tested these assumptions using observations  
17 of aggression in four age classes (4 - 10 months, 10 months - 3 years, 3 - 6 years, over 6  
18 years) of male and female shelter dogs (N = 4,743) in 11 different contexts. A structural  
19 equation model supported the hypothesis of two positively correlated personality traits  
20 underlying aggression across contexts: aggressiveness towards people and  
21 aggressiveness towards dogs (comparative fit index: 0.96; Tucker-Lewis index: 0.95; root  
22 mean square error of approximation: 0.03). Aggression across contexts was moderately  
23 repeatable (towards people: intraclass correlation coefficient (ICC) = 0.479; towards  
24 dogs: ICC = 0.303). However, certain contexts related to aggressiveness towards people  
25 (but not dogs) shared significant residual relationships unaccounted for by latent levels  
26 of aggressiveness. Furthermore, aggressiveness towards people and dogs in different  
27 contexts interacted with sex and age. Thus, sex and age differences in displays of  
28 aggression were not simple functions of underlying aggressiveness. Our results illustrate  
29 that the robustness of traits in latent variable models must be critically assessed before

30 making conclusions about the effects of, or factors influencing, animal personality. Our  
31 findings are of concern because inaccurate 'aggressive personality' trait attributions can  
32 be costly to dogs, recipients of aggression and society in general.

33

34 *Key words:* animal personality assessment; agonistic behaviour; shelter dogs;

35 measurement bias; behavioural phenotyping

## 36 Introduction

37 Studies of non-human animal personality demonstrate that animals show relatively  
38 consistent between-individual differences in behaviour, and that the behavioural  
39 phenotype is organised hierarchically into broad behavioural dimensions or personality  
40 traits (e.g. sociability, aggressiveness or boldness) that further exhibit inter-correlations  
41 to form behavioural syndromes (e.g. boldness with aggression; [1–5]). To interpret the  
42 complexity inherent in behavioural phenotypes, personality traits and behavioural  
43 syndromes are frequently inferred using latent variable statistical models [6], which  
44 reduce two or more measured variables (the *manifest* variables) into one or more  
45 lower-dimensional variables (the *latent* variables), following work in human psychology  
46 [7–10].

47

48 Many animal personality studies use *formative* models, such as principal components  
49 analysis, that construct composite variables comprised of linear combinations of  
50 manifest variables. However, formative models impose only weak assumptions about  
51 the relationships between latent variables and manifest variables [6,11]. For instance,  
52 formative models do not require manifest variables to be correlated with one another  
53 or illustrate internal consistency [11]. Because behavioural variables comprising  
54 personality traits are expected to correlate with each other [4], the utility of formative  
55 models for studying personality traits has been criticised in both animals [12,13] and  
56 humans [10,11,14,15]. Instead, researchers are increasingly using *reflective* models,

57 such as factor analysis, including confirmatory approaches such as structural equation  
58 modelling (see [1,16–18]). Reflective models regress measured behaviours on one or  
59 more latent variables, incorporating measurement error and possibilities to compare *a*  
60 *priori* competing hypotheses [1,16,19].

61

62 Whilst reflective models offer a powerful framework to examine the latent variable  
63 structure of animal behaviour [19], they impose certain assumptions on the  
64 interpretation and modelling of latent variables that have received scrutiny in human  
65 psychology but are rarely discussed in studies of animal personality. Two foundational  
66 assumptions are *local independence* and *measurement invariance*. Local independence  
67 implies that manifest variables should be independent of each other conditional on the  
68 latent variables [20,21]. For example, given a continuous latent variable  $\theta$  (e.g.  
69 boldness) and two binary manifest variables  $Y_1$  and  $Y_2$  that can take the values 0 and 1,  
70 the item response theory model asserts that  $P(Y_1 = 1, Y_2 = 1|\theta) = P(Y_1 = 1|\theta)P(Y_2 =$   
71  $1|\theta)$ . As such, the latent variables should ‘screen off’ any covariance between manifest  
72 variables. Measurement invariance implies that the latent variables function the same  
73 (i.e. are invariant or equivalent) in different subsets of a population or in the same  
74 individuals through time [21–25]. In the previous example, this means that the expected  
75 values of the manifest variables  $Y_1$  and  $Y_2$  should remain the same across different  
76 groups,  $\pi$  (e.g. sex or different populations), for any fixed value of the latent variable  $\theta_x$   
77 e.g.  $E(Y_1 | \theta_x) = E(Y_1 | \theta_x, \pi)$ . For studies of personality, violations of local  
78 independence or measurement invariance highlight instances where the personality

79 traits do not completely explain variation in the manifest variables, which may lead to  
80 misleading conclusions about the differences between individuals as a function of trait  
81 scores [25–27].

82

83 The goal of this study was evaluate local independence and measurement invariance in  
84 behavioural data on domestic dogs (*Canis lupus familiaris*). Dog personality has been of  
85 scientific interest for decades [28–30], both to predict the behaviour of dogs at future  
86 time points [31] and to elucidate behavioural traits pertinent to dogs' domestication  
87 history [32–35]. Research on personality in dogs has led to different numbers and  
88 composition of hypothesised personality traits with little consensus on how such traits  
89 should be compared within and between studies [36–38]. Dog personality studies  
90 frequently involve collection of data on a wide range of behaviours and, as a result,  
91 latent variable models are popular to reduce behavioural data into personality traits or  
92 dimensions [29]. Importantly, the predictive value of personality assessments in dogs  
93 has been inconsistent [31,39–43], perhaps most prominently in shelter dog personality  
94 assessments (e.g. see [31] for a review). Assessments of aggression are of particular  
95 concern, where aggression has been divided into different aggressiveness traits, such as  
96 owner-, stranger-, dog- or animal-directed factors [29,37,44,45]. Improving inference  
97 about aggressiveness in dogs is important because dog bites are a serious public health  
98 concern [46], especially for animal shelters rehoming dogs to new owners, and  
99 aggressive behaviour is undesirable to many organisations using dogs for various  
100 working roles [47].

101

102 Evaluating local independence and measurement invariance could help refine applied  
103 personality assessments on dogs. Local independence may be violated in standardised  
104 test batteries (a common assessment method; [48–50]) because the sequential  
105 administration of different behavioural subtests means that how dogs responds to one  
106 sub-test may influence their subsequent behavioural responses, as well as the responses  
107 of the dog handlers [31]. Identifying local independence could, thus, highlight which  
108 sub-tests can be interpreted as providing independent information. Local independence  
109 is also relevant to the development and analysis of dog personality questionnaires  
110 completed by dog owners, because the order in which the questions are presented or  
111 redundancy in the content of questions can lead to dependencies between participant  
112 responses not explained by the questionnaire’s intended focus on the dog’s behaviour  
113 [51].

114

115 Dog scientists are also concerned with understanding personality differences in dogs  
116 across a variety of conditions, including ontogeny, age, sex, breed and neuter status  
117 (e.g. [37, 42, 52–54]). Evaluating measurement invariance in personality assessments  
118 would allow researchers to confirm whether differences between individuals or groups  
119 of individuals in personality assessments reflect credible differences in personality trait  
120 scores or whether additional, unaccounted for factors are driving the differences. While  
121 it may be unrealistic for measurement invariance to hold in all instances, it is important

122 to establish whether it holds for personality traits across basic biological variables such  
123 as age and sex, which are generally applicable to dog populations undergoing  
124 personality assessment and have previously been found to show interactions with  
125 personality traits, including playfulness, sociability, curiosity and aggressiveness [33, 55].  
126 However, apart from van den Berg *et al.* [18] who assessed measurement invariance  
127 across breed groups, no studies have confirmed measurement invariance or local  
128 independence for personality traits.

129

130 In this paper, we assessed local independence and measurement invariance of  
131 aggressiveness in shelter dogs using a large sample of data on inter-context aggressive  
132 behaviour. First, we decomposed observations of aggression towards people and dogs  
133 across contexts into separate aggressiveness traits. Secondly, we assessed whether  
134 aggression in different contexts remained associated beyond that explained by latent  
135 levels of aggressiveness, testing local independence. Thirdly, we investigated whether  
136 the probability of aggression in different contexts assumed to be underpinned by the  
137 same aggressiveness trait was measurement invariant with respect to sex and age  
138 groups.

139

140

141



## 142 **Materials & Methods**

### 143 **Subjects**

144 Observational data on the occurrence of aggression in 4,743 dogs were gathered from  
145 Battersea Dogs and Cats Home's (UK) observational and longitudinal dog behaviour  
146 assessment records (Table 1). The data were from a sample of dogs (N = 4,990) at the  
147 shelter's three rehoming centres during 2014 (including dogs that arrived during 2013 or  
148 left in 2015). We selected the records from all dogs that were at least 4 months old,  
149 excluding younger dogs because they were more likely to be unvaccinated, more limited  
150 in their interactions at the shelter and may have been kennelled in different areas to  
151 older dogs. Although dogs were from a variety of heritages (including purebreds and  
152 mongrels), the analyses here did not explore breed differences because the accurate  
153 visual assessment of breed in dogs with unknown heritage has been refuted [56–58].  
154

**Table 1. Demographic characteristics of the studied dogs.**

<b>Variable</b>	<b>Mean ± SD / N</b>
Average age at shelter (years; all ≥ 4 months of age)	3.75 ± 3.03
Total days at the shelter	25.13 ± 41.53
Weight (average weight if multiple measurements; kg)	19.06 ± 10.26
Rehoming centre: London / Old Windsor / Brands Hatch	2897 / 1280 / 566
Males / females	2749 / 1994
Neutered <sup>1</sup> before arrival / neutered at shelter / not neutered	1218 / 1665 / 1502
Relinquished by owners / returned to shelter / strays	2892 / 260 / 1591

<sup>1</sup>358 dogs had unknown neuter status

155

## 156 **Shelter environment**

157 The shelter was composed of three different UK rehoming centres: a high-throughput,  
158 urban centre based at Battersea, London with capacity for approximately 150-200 dogs;  
159 a semi-rural/rural centre based at Old Windsor with capacity for approximately 100-150  
160 dogs; and a rural centre based at Brands Hatch with capacity for approximately 50 dogs.  
161 All dogs arrived in an intake area of their respective rehoming centre and, when  
162 considered suitable for adoption, were moved to a 'rehoming' area that was partially  
163 open to the public between 1000 h and 1600 h. All kennels were indoors. Kennels varied  
164 in size, but were usually approximately 4m x 2m and included either a shelf and bedding  
165 alcove area, or a more secluded bedding area at the back of the kennel (see [59] for  
166 more details). At different times throughout the day, dogs had access to indoor runs  
167 behind their kennels. In each kennel block area, dogs were cared for (e.g. fed, exercised,  
168 kennel cleaned) by a relatively stable group of staff members, allowing the development  
169 of familiarity with staff members and offering some predictability for dogs after arrival  
170 at the shelter. Although data on the number of dogs in each kennel were incomplete, in  
171 the majority of cases dogs were kennelled singly for safety reasons. The shelter mainly  
172 operated between 0800 h and 1700 h each day. All dogs were socialised with staff  
173 and/or volunteers each day (often multiple times) except on rare occasions when it was  
174 deemed unsafe to handle a dog (when training/behavioural modification proceeded  
175 without physical contact). Dogs were provided water ad libitum and fed commercial

176 complete dry and/or wet tinned food twice daily (depending on recommendations by  
177 veterinary staff). Dogs received daily tactile, olfactory and/or auditory  
178 enrichment/variety (e.g. toys, essential oils, classical music, time in a quiet 'chill-out'  
179 room).

180

## 181 **Data collection**

182 In the observational assessment procedure, trained shelter employees recorded  
183 observations of dog behaviour in a variety of contexts as part of normal shelter  
184 procedures. Behavioural observations pertaining to each context were completed using  
185 an ethogram specific to that context and recorded in a custom computer system.  
186 Multiple observations could be completed each day, although we retained only one  
187 observation in each context per day (the least desirable behaviour on that day; see  
188 below). The ethogram code that best described a dog's behaviour in a particular context  
189 during an observation was recorded by selecting it from a series of drop-down boxes  
190 (one for each context). Although staff could also add additional information in character  
191 fields, a full analysis of those comments was beyond the scope of this study. The  
192 ethogram for each context represented a scale of behaviours ranging from desirable to  
193 undesirable considered by the shelter to be relevant to dog welfare and ease of  
194 adoption. Contexts had between 10 and 16 possible behaviours to choose from, some of  
195 which overlapped between different contexts. Among the least desirable behaviours in  
196 each context was aggression towards either people or dogs (depending on context).

197 Aggression was formally defined as “*Growls, snarls, shows teeth and/or snaps when*  
198 *seeing/meeting other people/dogs, potentially pulling or lunging towards them*”,  
199 distinguished from non-aggressive but reactive responses, defined as “*Barks, whines,*  
200 *howls and/or play growls when seeing/meeting other people/dogs, potentially pulling or*  
201 *lunging towards them*”.

202

203 Observation contexts included both onsite (at the shelter) and offsite (e.g. out in public  
204 parks) settings. For the analyses here, we excluded offsite contexts (which had separate  
205 observation categories) and focused on observations of aggression in nine core onsite  
206 contexts that were most frequently completed by trained staff members: i) *Handling*, ii)  
207 *In kennel*, iii) *Out of kennel*, iv) *Interactions with familiar people*, v) *Interactions with*  
208 *unfamiliar people*, vi) *Eating food*, vii) *Interactions with toys*, viii) *Interactions with*  
209 *female dogs*, ix) *Interactions with male dogs*. For the *In kennel* and *Out of kennel*  
210 contexts, recording of aggression towards both people and dogs was possible. If both  
211 occurred at the same time, aggression towards people was recorded. Therefore, *In*  
212 *kennel* and *Out of kennel* were each divided to reflect aggression shown towards people  
213 and towards dogs only, respectively. This resulted in 11 aggression contexts (Table 2)  
214 used as manifest variables in structural equation models to investigate latent  
215 aggressiveness traits. The average number of days between successive observations  
216 across these contexts and across dogs was 3.27 (SD = 2.08), and dogs had an average of  
217 9.77 (SD = 13.41) observations within each context (N = 416,860 observations in total  
218 across dogs, contexts and days). Observations were recorded in the category that best

219 described the scenario. Nonetheless, certain contexts could occur closely in space and  
220 time, which were investigated for violations of local independence, as explained below.

221

**Table 2. Behavioural observation contexts in which each dog's reactions were analysed for the presence or absence of aggression.**

<b>Context</b>	<b>Definition</b>
Handling	Informal handling by people (e.g. stroking non-sensitive areas, touching the collar, fitting a harness or lead).
In kennel towards people	People approaching or walking past the kennel.
In kennel towards dogs	Dogs in neighbouring kennels or dogs walking past the kennel.
Interactions with familiar people	When outside the kennel and familiar people (interacted with at least once before) approach, make eye contact, speak to or attempt to make physical contact with the dog.
Interactions with unfamiliar people	When outside the kennel and unfamiliar people (never interacted with before) approach, make eye contact, speak to or attempt to make physical contact with the dog.
Out of kennel towards people	When around people outside the kennel who may be a long distance away and who make no attempt to engage with the dog.
Out of kennel towards dogs	When around dogs outside the kennel that may be a long distance away and that are not encouraged to interact with the focal dog.
Eating food	When eating food (e.g. from a food bowl, or toy filled with food) and people approach within close proximity or attempt to touch the food container.
Interactions with toys	When interacting with toys and people approach within close proximity or attempt to touch the toy.
Interactions with female dogs	During structured interaction with a female dog, including approaching each other, walking in parallel, and interacting off-lead. Both dogs are aware of each other's presence and are in close enough proximity to engage in a physical interaction.
Interactions with male dogs	During structured interaction with a male dog, including approaching each other, walking in parallel, and interacting off-lead. Both dogs are aware of each other's presence and are in close enough proximity to engage in a physical interaction.

222 We aggregated behavioural observations across time for each dog into a dichotomous  
223 variable indicating whether a dog had or had not shown aggression in a particular  
224 context at any time while at the shelter (Table S1). This was performed because the  
225 overall prevalence of aggression was low, with only 1.06% of all observations across  
226 days involving aggression towards people and 1.13% towards dogs. Thus, the main  
227 difference between individuals was whether they had or had not shown aggression in a  
228 particular context during their time at the shelter. We interpret aggressiveness here as a  
229 between-individual difference variable.

230

### 231 **Validity of behaviour recordings**

232 Validity of the recording of behaviour was assessed separately from the main data  
233 collection as part of a wider project investigating the use of the observational  
234 assessment method. Ninety-three shelter employees trained in conducting behavioural  
235 observations each watched (in groups of 5 – 10 people) 14 videos, approximately 30  
236 seconds each, presenting exemplars of 2 different behaviours from seven contexts (to  
237 keep the sessions concise and maximise the number of participants). For each context,  
238 behaviours were chosen pseudo-randomly by numbering each behaviour and selecting  
239 two numbers using a random number generator. Experienced behaviourists working at  
240 the shelter filmed the videos demonstrating the behaviours. Videos were shown to  
241 participants once in a pseudo-random order. After each video, participants recorded on  
242 a paper answer sheet the behaviour they thought most accurately described the dog's  
243 behaviour based on the ethogram specific to the context depicted. Two of the videos

244 illustrated aggression: one in a combined *Interactions with new and familiar people*  
245 context (combined because familiarity between specific people and dogs was not  
246 universally known) and one in the *In kennel towards dogs* context. The authors were  
247 blind to the selection of videos shown and to the video coding sessions with shelter  
248 employees.

249

## 250 **Data analysis**

251 All data analysis was conducted in R version 3.3.2 [60].

252

## 253 **Validity of behaviour recordings**

254 The degree to which shelter employees could recognise and correctly record aggressive  
255 behaviour from the videos (chosen by experienced behaviourists at the shelter) was  
256 determined by the percentage of participants who correctly identified the 2 videos as  
257 showing examples of aggression.

258

## 259 **Missing data**

260 Data were missing when dogs did not experience particular contexts while at the  
261 shelter. The missing data rate was between 0.06% and 5% for each context, except for  
262 the *Interactions with female dogs* and *Interactions with male dogs* categories which had



263 17% and 18% of missing values, respectively (because structured interactions with other  
264 dogs did not arise as frequently). Moreover, 16% and 8% of dogs were missing weight  
265 measurement and neuter status data, respectively, which were independent variables  
266 statistically controlled for in subsequent analyses. We created 5 multiply imputed data  
267 sets (using the *Amelia* package; [61]), upon which all following analyses in the sections  
268 below were conducted and results pooled. The multiple imputation took into account  
269 the hierarchical structure of the data (observations within dogs), all independent  
270 variables reported below, and the data types (ordered binary variables for the context  
271 data, positive-continuous for weight measurements, nominal for neuter status; see the  
272 R script). The data were assumed to be missing at random, that is, dependent only on  
273 other variables in the analyses.

274

## 275 **Structural equation models**

276 We used structural equation modelling to assess whether aggression towards people  
277 (contexts: *Handling, In kennel towards people, Out of kennel towards people,*  
278 *Interactions with familiar people, Interactions with unfamiliar people, Eating food,*  
279 *Interactions with toys*) and towards dogs (contexts: *In kennel towards dogs, Out of*  
280 *kennel towards dogs, Interactions with female dogs, Interactions with male dogs*) could  
281 be explained by two latent aggressiveness traits: aggressiveness towards people and  
282 dogs, respectively. Since positive correlations between different aggressiveness traits  
283 have been reported in dogs [55], we compared a model where the latent variables were

284 orthogonal to a model where variables were allowed to covary. Models were fit using  
285 the *lavaan* package [62], with the weighted least squares mean and variance adjusted  
286 (WLSMV) estimator and theta/conditional parameterisation, as recommended for  
287 categorical dependent variables [8,63,64]. The latent variables were standardised to  
288 have mean 0 and variance 1. The results were combined across imputed data sets using  
289 the 'runMI' function in the *semTools* package [65]. The fit of each model was  
290 ascertained using the comparative fit index (CFI) and Tucker Lewis index (TLI), where  
291 values > 0.95 indicate excellent fit, as well as the root mean squared error of  
292 approximation (RMSEA) where values < 0.06 indicate good fit [7]. Parameter estimates  
293 were summarised by test statistics and 95% confidence intervals (CI).

294

### 295 **Local independence**

296 We tested the assumption of local independence by re-fitting the best-fitting structural  
297 equation model with residual covariances specified between context variables. To  
298 maintain a theoretically driven approach (see [66] regarding the best practice of  
299 including residual covariances in structural equation models) and model identifiability,  
300 we only tested a predefined set of covariances based on which contexts shared close  
301 temporal-spatial relationships. First, we allowed covariances between *Handling* with *In*  
302 *kennel towards people*, *Interactions with familiar people*, *Interactions with unfamiliar*  
303 *people* and *Interactions with toys*, respectively, since the *Handling* context could directly  
304 succeed these other contexts. The residual covariance between *Handling* and *Eating*

305 *food* was not estimated because shelter employees would be unlikely to handle a dog  
306 while the dog ate its daily meals. The residual covariance between *Handling* and *Out of*  
307 *kennel towards people* was not estimated because any association between *Handling*  
308 and *Out of kennel towards people* would be mediated by either the *Interactions with*  
309 *familiar people* or *Interactions with unfamiliar people* context. Therefore, secondly, we  
310 estimated the three-way covariances between *Out of kennel towards people*,  
311 *Interactions with familiar people* and *Interactions with unfamiliar people*. Similarly, and  
312 lastly, we estimated the three-way covariances between *Out of kennel towards dogs*,  
313 *Interactions with female dogs* and *Interactions with male dogs*. No covariances were  
314 inspected between *In kennel towards dogs* and other aggressiveness towards dogs  
315 contexts since large time gaps were more likely to separate observations between those  
316 contexts.

317

### 318 **Measurement invariance**

319 To test for measurement invariance in each of the latent traits derived from the best  
320 fitting structural equation model, we investigated the response patterns across  
321 aggression contexts related to the same latent aggressiveness trait using Bayesian  
322 hierarchical logistic regression models. These models were analogous to the 1-  
323 parameter item response theory model, which represents the probability that an  
324 individual responds correctly to a particular test item as a logistic function of i) each  
325 individual's latent ability and ii) the item's difficulty level. This model can be expressed

326 as a hierarchical logistic regression model [67,68], whereby individual latent abilities are  
327 modelled as individual-specific intercepts (i.e. 'random intercepts'), the propensity for a  
328 correct answer to an item  $i$  is its regression coefficient  $\beta_i$ , and credible interactions  
329 between items and relevant independent variables (e.g. group status) indicate a  
330 violation of measurement invariance. Here, the dependent variable was the binary score  
331 for whether or not dogs had shown aggression in each context and the average  
332 probability of aggression across contexts varied by dog, representing latent levels of  
333 aggressiveness. Context type, dog age, dog sex and their interactions were included as  
334 categorical independent variables. Age was treated as a categorical variable, with  
335 categories reflecting general developmental periods: i) 4 months to 10 months (juvenile  
336 dogs before puberty), ii) 10 months to 3 years (dogs maturing from juveniles to adults),  
337 iii) 3 years to 6 years (adults), and iv) 6 years + (older dogs). Broad age categories were  
338 chosen due to potentially large differences in developmental timing between  
339 individuals. Age was categorised because we predicted that aggression would be  
340 dependent on these developmental periods.

341

342 Models included additional demographic variables (Table 1) that may mediate the  
343 probability of aggression: body weight (average weight if multiple measurements were  
344 taken), total number of days spent at the shelter, the rehoming centre at which dogs  
345 were based (London, Old Windsor, Brands Hatch), neuter status (neutered before  
346 arrival, neutered at the shelter, not neutered) and source type (relinquished by owner,  
347 returned to the shelter after adoption, stray). Categorical variables were represented as

348 sum-to-zero deflections from the group-level intercept to ensure that the intercept  
349 represented the average probability of aggression across the levels of each categorical  
350 predictor. Weight and total days at the shelter were mean-centered and standardised  
351 by 2 standard deviations. Due to the potentially complex relationships between these  
352 variables and aggression (e.g. interactive effects between neuter status and sex; [52]),  
353 which could also include violations of measurement invariance, we decided not to  
354 interpret their effects inferentially. Instead, they were included to make the assessment  
355 of measurement invariance between sexes and age groups conditional on variance  
356 explained by potentially important factors.

357

358 For comparability to other studies in animal personality, behavioural repeatability was  
359 calculated across contexts in each model using the intraclass correlation coefficient  
360 (ICC), calculated as  $\frac{\sigma_{\beta}^2}{\sigma_{\beta}^2 + \sigma_{\epsilon}^2}$ , where  $\sigma_{\beta}^2$  represented the between-individual variance of the  
361 probability of aggression (i.e. the variance of the random intercepts), and  $\sigma_{\epsilon}^2$  was  $\pi^2/3$ ,  
362 the residual variance of the standard logistic distribution [69].

363

#### 364 **Computation**

365 Models were computed using the probabilistic programming language Stan version  
366 2.15.1 [70], using Hamiltonian Monte Carlo, a type of Markov Chain Monte Carlo  
367 (MCMC) algorithm, to sample from the posterior distribution. Prior distributions for all

368 independent variables were normal distributions with mean 0 and standard deviation 1,  
369 attenuating regression coefficients towards zero for conservative inference. The prior on  
370 the overall intercept parameter was normally distributed with mean 0 and standard  
371 deviation 5. The standard deviation of dog-specific intercept parameters was given a  
372 half-Cauchy prior distribution with mean 0 and shape 2. Each model was run with 4  
373 chains of 2,000 iterations with a 1,000 step warm-up period. The Gelman-Rubin statistic  
374 (ideally < 1.05) and visual assessment of traceplots were used to assess MCMC  
375 convergence. We checked the accuracy of the model predictions against the raw data  
376 using graphical posterior predictive checks. For plotting purposes, predicted  
377 probabilities of aggression were obtained by marginalising over the random effects  
378 (explained in the Supporting Information). Regression coefficients were expressed as  
379 odds ratios and were summarised by their mean and 95% Bayesian highest density  
380 interval (HDI), representing the 95% most probable parameter values. To compare levels  
381 of categorical variables and their interactions, we computed the 95% HDI of the  
382 differences between the respective posterior distributions.

383

#### 384 ***Model selection & parameter inference***

385 Models were run on each imputed data set and their respective posterior distributions  
386 were averaged to attain a single posterior distribution for inference. Adopting a  
387 Bayesian approach allowed the estimation of interaction parameters (i.e. testing  
388 measurement invariance) without requiring corrections for multiple comparisons as in

389 null hypothesis significance testing [71]. Nonetheless, models included a large number  
390 of estimated parameters. Two strategies were employed to guard against over-fitting of  
391 models to data. First, we selected the model with the best out-of-sample predictive  
392 accuracy given the number of parameters based on the Widely Applicable Information  
393 Criterion (WAIC; using the R package *loo* [72]). Four variants of each model were  
394 computed: two-way interactions between contexts and age and contexts and sex,  
395 respectively (model 1), a single interaction with sex but not with age (model 2), a single  
396 interaction with age but not with sex (model 3), and no interactions (model 4). All  
397 models included the mediating independent variables above. Second, to avoid testing  
398 point-estimate null hypotheses, the effect of a parameter was only considered credibly  
399 different from zero if the odds ratio exceeded the region of practical equivalence (ROPE;  
400 see [73]) around an odds ratio of 1 from 0.80 to 1.25. An odds ratio of 0.80 or 1.25  
401 indicates a 20% decrease or increase (i.e. 4/5 or 5/4 odds), respectively, in the odds of  
402 an outcome, frequently used in areas of bioequivalence testing (e.g. [74]), which we  
403 here considered to be small enough to demonstrate a negligible effect in the absence of  
404 additional information. If a 95% HDI fell completely within the ROPE, the null hypothesis  
405 of no credible influence of that parameter was accepted; if a 95% HDI included part of  
406 the ROPE, then the parameter's influence was left undecided [73].

407

408 **Ethics statement**

409 Permission to use and publish the data was received from the shelter. Approval from an  
410 ethical review board was not required for this study.

411

412 **Data accessibility**

413 Supporting Information (data, R script, Stan model code, Tables S1-4) can be found at:  
414 [https://github.com/ConorGoold/GooldNewberry\\_aggression\\_shelter\\_dogs](https://github.com/ConorGoold/GooldNewberry_aggression_shelter_dogs).

415

416

417



## 418 **Results**

### 419 **Validity of behaviour recordings**

420 For the video showing aggression towards people, 52% of participants identified the  
421 behaviour correctly as aggression and 42% identified the behaviour as non-aggressive  
422 but (similarly) reactive behaviour (see definitions above). For the video showing  
423 aggression towards dogs, 53% identified the behaviour correctly and 44% identified the  
424 behaviour as non-aggressive but reactive behaviour. For the 12 other videos not  
425 showing aggression, only 1 person incorrectly coded a video as aggression towards  
426 people and 3 people incorrectly coded videos as aggression towards dogs.

427

### 428 **Structural equation models**

429 The raw tetrachoric correlations between the aggression contexts were all positive,  
430 particularly between contexts recording aggression towards people and dogs,  
431 respectively, supporting their convergent validity (Table S2). The model with correlated  
432 latent variables fit marginally better (CFI: 0.96; TLI: 0.95; RMSEA: 0.03) than the model  
433 with uncorrelated variables (CFI: 0.94; TLI: 0.92; RMSEA: 0.04). All regression coefficients  
434 of the model with correlated latent variables were positive and significant (i.e. the 95%  
435 CI did not include zero), and the latent variables shared a significant positive covariance  
436 (Table 3).

**Table 3. Parameter estimates from the best-fitting structural equation model.**

Parameter	Estimate	SE	t value	95% CI
Handling <sup>a</sup>	0.81	0.06	14.25	[0.70, 0.92]
In kennel towards people <sup>a</sup>	1.29	0.09	14.17	[1.12, 1.46]
Out of kennel towards people <sup>a</sup>	0.83	0.07	11.99	[0.69, 0.96]
Interactions with familiar people <sup>a</sup>	0.96	0.07	14.23	[0.83, 1.09]
Interactions with unfamiliar people <sup>a</sup>	1.54	0.12	12.46	[1.23, 1.78]
Eating food <sup>a</sup>	0.70	0.06	12.33	[0.59, 0.81]
Interactions with toys <sup>a</sup>	0.51	0.06	8.32	[0.39, 0.63]
In kennel towards dogs <sup>b</sup>	0.70	0.06	11.94	[0.59, 0.82]
Out of kennel towards dogs <sup>b</sup>	0.47	0.04	10.80	[0.38, 0.55]
Interactions with female dogs <sup>b</sup>	0.87	0.07	12.05	[0.72, 1.02]
Interactions with male dogs <sup>b</sup>	0.88	0.07	12.23	[0.74, 1.03]
Covariance: People ~ Dogs	0.26	0.03	7.94	[0.19, 0.33]

<sup>a</sup> Contexts reflecting aggressiveness towards people

<sup>b</sup> Contexts reflecting aggressiveness towards dogs

437

### 438 **Local independence**

439 Allowing the pre-defined residuals to co-vary in the best-fitting structural equation

440 model resulted in a better fit (CFI = 0.98; TLI = 0.97; RMSEA: 0.03). Significant negative

441 covariances were observed between the *Handling* and *In kennel towards people*  
442 contexts (Table 4) and the *Handling* and *Interactions with unfamiliar people* contexts. A  
443 significant positive covariance was observed between *Out of kennel towards people* and  
444 *Interactions with unfamiliar people* contexts. No significant residual covariances  
445 between contexts reflecting aggressiveness towards dogs were observed.

446

**Table 4. Estimated residual covariances between contexts.**

Residual covariances	Estimate	SE	t value	95% CI
Handling ~ In kennel towards people <sup>a</sup>	-0.60	0.21	-2.86	[-1.01, -0.19]
Handling ~ Interactions with familiar people <sup>a</sup>	0.16	0.09	1.84	[-0.01, 0.33]
Handling ~ Interactions with unfamiliar people <sup>a</sup>	-0.48	0.19	-2.49	[-0.86, -0.10]
Handling ~ Interactions with toys <sup>a</sup>	0.14	0.07	1.85	[-0.01, 0.28]
Out of kennel towards people ~ Interactions with familiar people <sup>a</sup>	0.04	0.08	0.49	[-0.12, 0.20]
Out of kennel towards people ~ Interactions with unfamiliar people <sup>a</sup>	0.24	0.09	2.56	[0.06, 0.42]
Interactions with familiar people ~ Interactions with unfamiliar people <sup>a</sup>	-0.02	0.12	-0.16	[-0.25, 0.21]
Out of kennel towards dogs ~ Interactions with female dogs <sup>b</sup>	-0.55	0.48	-1.15	[-1.50, 0.40]
Out of kennel towards dogs ~ Interactions with male dogs <sup>b</sup>	-0.45	0.40	-1.13	[-1.22, 0.33]
Interactions with female dogs ~ Interactions with male dogs <sup>b</sup>	-0.24	0.50	-0.49	[-1.23, 0.74]

<sup>a</sup> Contexts reflecting aggressiveness towards people

<sup>b</sup> Contexts reflecting aggressiveness towards dogs

## 448 **Measurement invariance**

449 Separate models were run for contexts reflecting aggressiveness towards people and  
450 aggressiveness towards dogs. All models converged. Posterior predictive checks of  
451 model estimates reflected the raw data (Figs 1 and 2). The full measurement invariance  
452 model (model 1) including interactions between contexts and sex and contexts and age  
453 groups had the best out-of-sample predictive accuracy for both the aggressiveness  
454 towards people and aggressiveness towards dogs models, respectively, illustrated by the  
455 lowest WAIC values (Table 5). Since some models included numerous interactions, we  
456 provide an overall summary of the main results below (Figs 1 and 2) with full parameter  
457 estimates provided in Tables S3 and S4.

458

**Table 5. Mean  $\pm$  standard error of the Widely Applicable Information Criteria (WAIC) values (lower is better) per model and aggressiveness variable.**

<b>Model</b>	<b>Aggressiveness towards people</b>	<b>Aggressiveness towards dogs</b>
Model 1	13405.6 $\pm$ 179.0	15257.2 $\pm$ 133.1
Model 2	13506.3 $\pm$ 179.6	15381.4 $\pm$ 133.4
Model 3	13426.3 $\pm$ 179.1	15285.3 $\pm$ 133.0
Model 4	13521.7 $\pm$ 179.5	15407.6 $\pm$ 133.4

459

460

## 461 **Aggressiveness towards people**

462 The odds of aggression towards people, across categorical predictors and for an average  
463 dog of mean weight and length of stay at the shelter, were 0.022 (HDI: 0.021 to 0.024), a  
464 probability of approximately 2%. On average, aggression was most likely in the *In kennel*  
465 *towards people* context (OR = 0.054; HDI: 0.049 to 0.058) and least probable in the  
466 *Interactions with toys* context (OR = 0.008; HDI: 0.007 to 0.009).

467

468 Aggression was less likely across contexts for females than males (OR = 0.719; HDI:  
469 0.668 to 0.770), although there were also credible interactions between sex and  
470 contexts (Fig 1A; Table S3). Whereas males and females had similar odds of aggression  
471 in the *Out of kennel towards people* context, smaller differences were observed  
472 between *Out of kennel towards people* and *Handling* (OR = 0.578; HDI: 0.481 to 0.682),  
473 *Eating food* (OR = 1.812; HDI: 1.495 to 2.152) and *Interactions with familiar people* (OR =  
474 1.798; HDI: 1.488 to 2.126) contexts in females compared to males. Additionally,  
475 whereas aggression in the *Interactions with unfamiliar people* context was similar  
476 between males and females, larger differences were observed between *Interactions*  
477 *with unfamiliar people* and *Handling* (OR = 0.616; HDI: 0.530 to 0.702), *Eating food* (OR  
478 = 0.594; HDI: 0.506 to 0.686) and *Interactions with familiar people* (OR = 0.598; HDI:  
479 0.513 to 0.687) contexts in females compared to males.

480

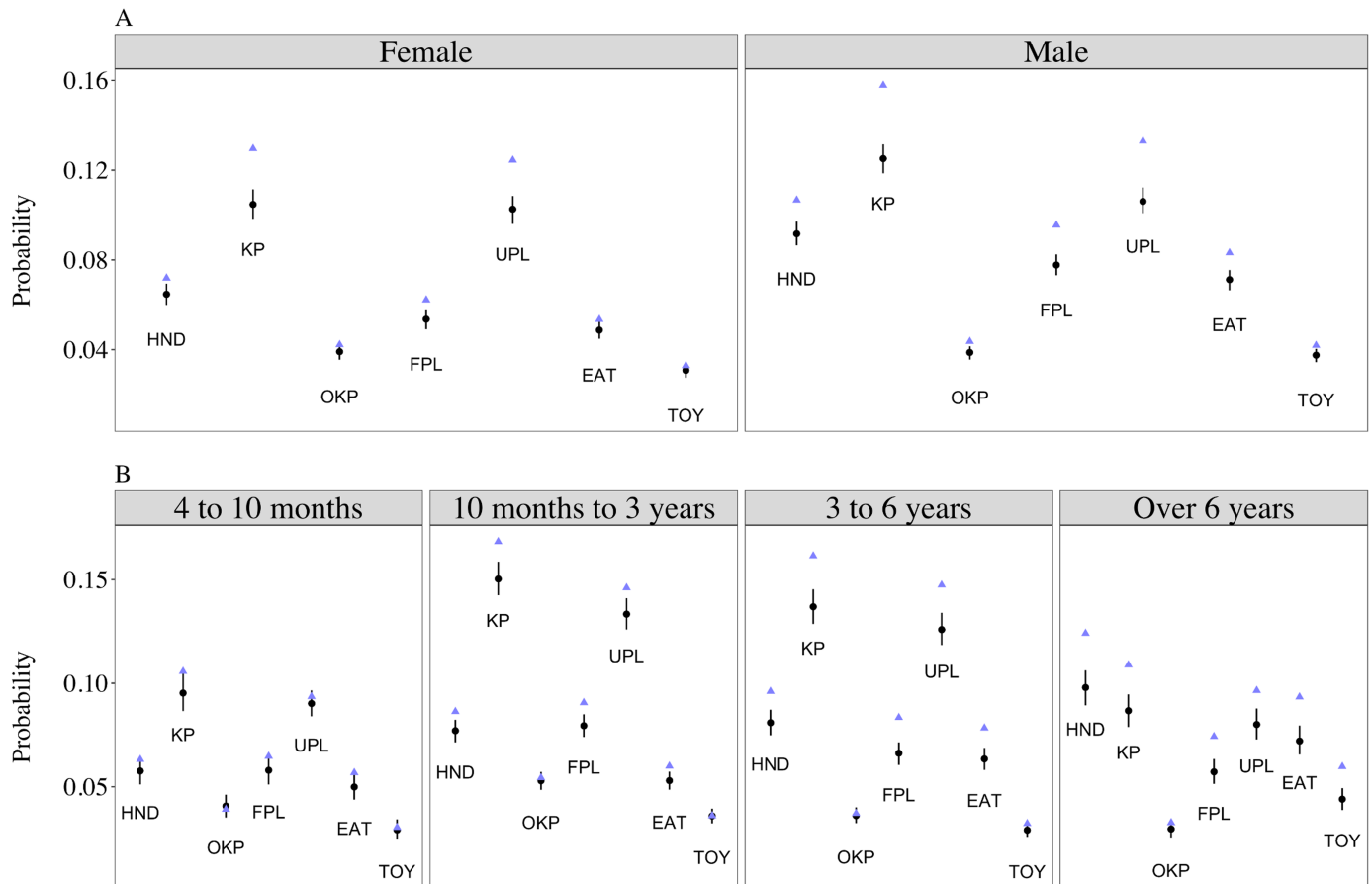
481 Apart from lower odds of aggression in 4 to 10 month olds compared to 10 month to 3  
482 year old dogs (OR = 0.638; HDI: 0.565 to 0.705), there was no simple influence of age  
483 group on aggressiveness. Between the 4 to 10 months old and 3 to 6 years old groups,  
484 differences between the odds of aggression across contexts varied due to an increase of  
485 aggression in certain contexts but not others (Fig 1B; Table S4). Aggression in *In kennel*  
486 *towards people* and *Interactions with unfamiliar people* contexts particularly increased,  
487 leading to larger differences between, for example, *In kennel towards people* and *Eating*  
488 *food* (OR = 0.524; HDI: 0.400 to 0.642) and *Eating food* and *Interactions with unfamiliar*  
489 *people* (OR = 1.721; HDI: 1.403 to 2.059) contexts for 10 month to 3 year olds compared  
490 to 4 to 10 month olds, and between *In kennel towards people* and *Out of kennel towards*  
491 *people* (OR = 0.470; HDI: 0.355 to 0.606) and *Out of kennel towards people* and  
492 *Interactions with unfamiliar people* (OR = 2.051; HDI: 1.608 to 2.543) contexts in 3 to 6  
493 year olds compared to 4 to 10 month olds. In 3 to 6 year old compared to 10 month to 3  
494 year old dogs, aggression increased in the *Handling* and *Eating food* contexts but  
495 decreased in the *Out of kennel towards people* context, resulting in larger differences  
496 between, for instance, *Handling* and *Out of kennel towards people* (OR = 0.526; HDI:  
497 0.409 to 0.631) and *Out of kennel towards people* and *Interactions with unfamiliar*  
498 *people* (OR = 2.349; HDI: 1.891 to 2.925), and smaller differences between *Eating food*  
499 and *Interactions with familiar people* (OR = 0.576; HDI: 0.468 to 0.687).

500

501 Dogs over 6 years old demonstrated qualitatively different response patterns across  
502 certain contexts than all other age groups. While aggression was most probable in *In*  
503 *kennel towards people* and *Interactions with unfamiliar people* contexts for dogs aged 4  
504 months through 6 years, dogs over 6 years old were most likely to show aggression in  
505 the *Handling* context, leading to interactions between, for example, *Handling* and *In*  
506 *kennel towards people*, and between *Handling* and *Interactions with unfamiliar people*  
507 contexts compared to the other age groups (Fig 1B; Table S3). Aggression when *Eating*  
508 *food* and in *Interactions with toys* contexts also increased compared to that expressed  
509 by younger dogs, resulting in credible differences between, for instance, *Eating food* and  
510 *Interactions with familiar people* contexts between dogs aged 10 months to 3 years and  
511 over 6 years (OR = 0.379; HDI: 0.300 to 0.465) and between *Out of kennel towards*  
512 *people* and *Interactions with toys* contexts between over 6 year olds and all other age  
513 groups (Table S3).

514





515 **Fig 1. Predicted probabilities of aggression towards people in different contexts by sex**  
516 **(panel A) and age groups (panel B).** Black points and vertical lines show mean and 95%  
517 highest density intervals of model parameter estimates; blue triangles show raw sample  
518 data. Model estimates were obtained by marginalising over the random effects (see the  
519 Supporting Information). Abbreviations used in the figure: HND (*Handling*); KP (*In kennel*  
520 *towards people*); OKP (*Out of kennel towards people*); FPL (*Interactions with familiar*  
521 *people*); UPL (*Interactions with unfamiliar people*); EAT (*Eating food*); TOY (*Interactions*  
522 *with toys*).

523

## 524 **Aggressiveness towards dogs**

525 The odds of aggression towards dogs, across categorical predictors and for an average  
526 dog of mean weight and length of stay at the shelter, was 0.176 (HDI: 0.168 to 0.184),  
527 corresponding to a probability of approximately 15%. Dogs were most likely to show  
528 aggression in the *Interactions with male dogs* context (OR = 0.297; HDI: 0.198 to 0.217)  
529 and least likely in the *In kennel towards dogs* context (OR = 0.099; HDI: 0.094 to 0.104;  
530 Fig 2; Table S4).

531

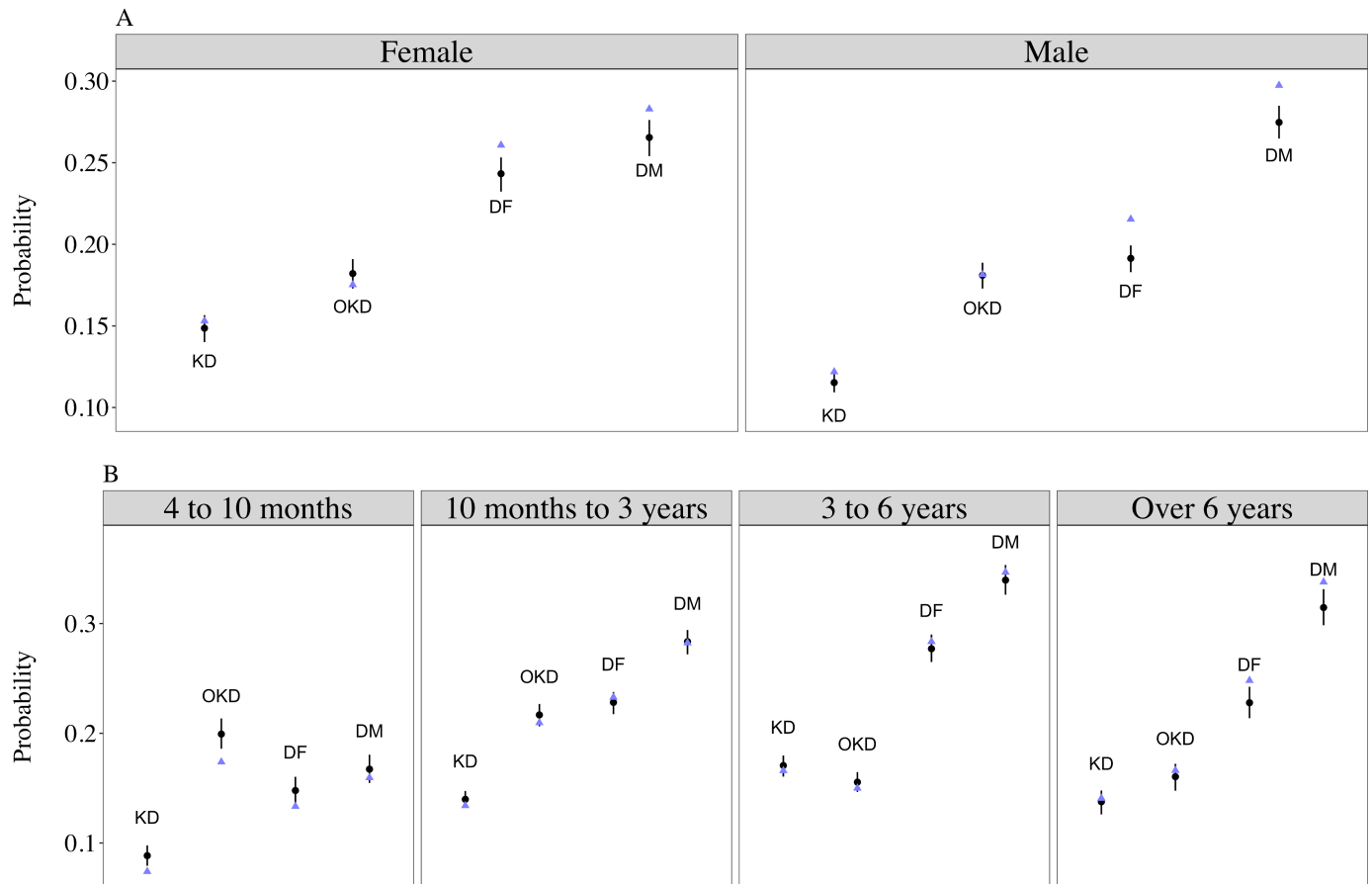
532 No credible mean-level differences existed between females and males (OR = 1.187;  
533 HDI: 1.128 to 1.250). However, the difference in aggression between the *Interactions*  
534 *with female dogs* and *Interactions with male dogs* contexts was smaller for females (OR  
535 = 1.542; HDI: 1.400 to 1.704; Fig 2a; Table S4), as were the differences between  
536 *Interactions with male dogs* and *In kennel towards dogs* (OR = 0.661; HDI: 0.590 to  
537 0.732) and *In kennel towards dogs* and *Out of kennel towards dogs* (OR = 1.420; HDI:  
538 1.269 to 1.587). Females were also more likely to show aggression in *Interactions with*  
539 *female dogs* than *Out of kennel towards dogs* compared to males (OR = 1.444; HDI:  
540 1.301 to 1.603).

541

542 Dogs aged 4 to 10 months old had credibly lower odds of aggression towards dogs than  
543 older dogs across contexts (Fig 2b; Table S4). However, contexts and age also showed

544 interactive effects. In particular, aggression in *Interactions with female dogs* and  
545 *Interactions with male dogs* contexts tended to increase relative to other contexts. For  
546 instance, the relationship between *Interactions with female dogs* and *Out of kennel*  
547 *towards dogs* contexts reversed in direction between 4 to 10 month and 10 month to 3  
548 year olds (OR = 0.595; HDI: 0.495 to 0.688) as did the relationship between *Interactions*  
549 *with male dogs* and *Out of kennel towards dogs* contexts (OR = 0.499; HDI: 0.422 to  
550 0.575). The relationship between *In kennel towards dogs* and *Out of kennel towards*  
551 *dogs* contexts also changed across age groups (Fig 2b; Table S4). Four to 10 months old  
552 were more likely to show aggression in *Out of kennel towards dogs* than *In kennel*  
553 *towards dogs* contexts, but the difference was smaller in 10 months to 3 year olds (OR =  
554 0.608; HDI: 0.505 to 0.728) and in over 6 year olds (OR = 0.396; HDI: 0.316 to 0.481). The  
555 latter relationship was reversed in 3 to 6 year olds compared to 4 to 10 month old dogs  
556 (OR = 0.277; HDI: 0.227 to 0.331) and 10 month to 3 year old dogs (OR = 0.456; HDI:  
557 0.396 to 0.516).

558



559 **Fig 2. Predicted probabilities of aggression towards dogs in different contexts by sex**  
560 **(panel A) and age groups (panel B).** Black points and vertical lines show mean and 95%  
561 highest density intervals of model parameter estimates; blue triangles show raw sample  
562 data. Model estimates were obtained by marginalising over the random effects (see the  
563 Supporting Information). Abbreviations used in the figure: KD (*In kennel towards dogs*);  
564 OKD (*Out of kennel towards dogs*); DF (*Interactions with female dogs*); DM (*Interactions*  
565 *with male dogs*).

566

## 567 **Repeatability**

568 Both aggressiveness towards people and dogs showed moderate repeatability across  
569 contexts ( $ICC_{people} = 0.479$ ; HDI: 0.466 to 0.491;  $ICC_{dogs} = 0.303$ ; HDI: 0.291 to 0.315),  
570 although aggressiveness towards people was more repeatable than aggressiveness  
571 towards dogs ( $ICC_{difference} = 0.176$ ; HDI: 0.158 to 0.192).

572

## 573 Discussion

574 In this study, we have examined local independence and measurement invariance of  
575 aggressiveness traits in shelter dogs. Observational recordings of aggression directed  
576 towards people and dogs across different shelter contexts were explained by two  
577 positively correlated latent variables, and behaviour across contexts was moderately  
578 repeatable. These results are consistent with the definition of animal personality as  
579 behaviour that shows moderately consistent between-individual differences across time  
580 or contexts, and characterised by multiple observed behaviours being decomposed into  
581 lower-dimensional behavioural traits [4]. However, we found violations of local  
582 independence between contexts with close temporal-spatial relationships and  
583 measurement invariance with respect to sex and age groups, highlighting potential  
584 measurement biases.

585

586 Local independence implies that the association between manifest variables is greater  
587 than that explained by the latent variable. For aggressiveness towards people,  
588 aggression in the *Handling* context was negatively related with the *In kennel towards*  
589 *people* and *Interactions with unfamiliar people* contexts, while positive covariances were  
590 present between *Out of kennel towards people* and *Interactions with unfamiliar people*  
591 contexts. Violations of local independence may arise through shared method variance  
592 [75–78] or unmodelled latent variables influencing manifest variables [79,80]. If a dog  
593 showed aggression when an unfamiliar person approached, it may be less likely to be

594 handled by that person, which may explain the negative residual covariations between  
595 the *Handling* and *In kennel towards people* and *Interactions with unfamiliar people*  
596 contexts, respectively. These contexts were, in fact, positively correlated when latent  
597 levels of aggressiveness were not accounted for (Table S4). In addition, the positive  
598 residual correlation between *Out of kennel towards people* and *Interactions with*  
599 *unfamiliar people* may be mediated by additional traits of interest to personality  
600 researchers, such as fearfulness or anxiety [29,81], if dogs who are fearful of interacting  
601 with unfamiliar people are more likely to show aggression beyond that described by a  
602 latent aggressiveness trait.

603

604 While authors have argued that greater standardisation and validation of personality  
605 assessments is key to ensuring the accurate measurement of underlying traits  
606 [36,48,49], it may be untenable to avoid dependencies between testing contexts.  
607 Displays of aggression in one sub-test will likely change how people conduct future sub-  
608 tests with the same dog, regardless of test standardisation. Human psychologists have  
609 argued that violations of local independence are a natural consequence of the  
610 organisation of behaviour as a complex dynamic system [82,83], which unfolds with  
611 respect to time- and context-dependent constraints [84]. Thus, awareness of local  
612 independence and its violation could facilitate closer understanding of the dynamics  
613 driving personality test responses beyond explanations purely based on personality  
614 traits.

615

616 While different subsets of a population may differ in mean levels of trait expression,  
617 interactions between behavioural responses and those subsets indicate that the same  
618 phenomenon is not under measurement across groups [23,24]. We found that the  
619 probability of aggression across contexts was dependent on sex and age conditional on  
620 latent levels of aggressiveness (Figs 1 and 2; Tables S3 and S4). Female dogs, for  
621 example, were more likely than males to show aggression in *Out of kennel towards*  
622 *people* and *Interactions with unfamiliar people* contexts relative to other contexts (Fig  
623 1A). Females also demonstrated similar odds of aggression during *Interactions with*  
624 *female dogs* and *Interactions with male dogs*, whereas males were more likely to show  
625 aggression towards male than female dogs (Fig 2a). As with local independence,  
626 different behavioural variables unaccounted for in this study may result in violations of  
627 measurement invariance. While dogs up to 6 years old were most likely to show  
628 aggression in *In kennel towards people* and *Interactions with unfamiliar people* contexts,  
629 dogs over 6 years old demonstrated aggression most commonly in the *Handling* context.  
630 Dogs over 6 years old also showed an increase in aggression in the *Eating food* and  
631 *Interactions with toys* contexts relative to other age groups. These results suggest that  
632 older dogs in shelter populations may be less tolerant during close interactions with  
633 people (i.e. handling, people in the vicinity of their food and toys) compared to other  
634 contexts, which may driven by other quantifiable factors such as pain or sensitivity (e.g.  
635 [29]).

636



637 Local independence and measurement invariance are assumptions of reflective latent  
638 variable models, but are not assumptions for formative models, which posit that the  
639 latent variable is simply a linear composite of manifest variables, rather than a causal,  
640 underlying variable [11,21]. While formative models such as principal components  
641 analysis may, as a result, appear attractive and are popular in dog personality studies  
642 (e.g. [32,33,85]), their use has been discouraged. Principal components analysis will  
643 result in lower-dimensional variables explained by linear combinations of manifest  
644 variables even when those manifest variables are uncorrelated random variables (e.g.  
645 see [12]). In other words, principal components are a function of the behavioural  
646 variables, whereas personality traits are normally assumed to influence the expression  
647 of behavioural variables [16]. Consequently, principal components that underlie  
648 behavioural data should be interpreted cautiously. As we have shown here, reflective  
649 models represent latent variables or personality traits as causal variables [10,86], and  
650 confirmatory approaches (e.g. structural equation modelling or item response theory) in  
651 particular could be used to evaluate the reproducibility of dog personality traits across  
652 existing studies. A number of articles demonstrate how these models can be applied to  
653 animal personality data [1,16,19].

654

655 Although we have identified violations of both local independence and measurement  
656 invariance, we remain cautious about hypothesising *a posteriori* about their causes.  
657 Personality traits in animal behaviour are typically defined operationally, based on the  
658 statistical repeatability of quantifiable behaviour [77,87,88]. As discussed in human

659 personality psychology, operational definitions can be ontologically ambiguous [89,90].  
660 That is, while operational definitions facilitate experimentation in animal personality [4],  
661 they do not necessarily designate biological mechanisms underlying trait expression. For  
662 example, Budaev and Brown remark that boldness, defined as a propensity to take risks,  
663 could encompass a range of distinct personality traits, each with a different biological  
664 basis [75]. Whilst reflective latent variable models allow researchers to test hypotheses  
665 about the relatedness of measured behaviours via one or more underlying traits, they  
666 have also been criticised as ambiguous [82]. For example, it is uncertain what reflective  
667 latent variables may represent in biological organisation [89] or even whether they are  
668 features individuals possess or simply emergent features of between-individual  
669 differences [91,92]. Such considerations highlight the importance of research on the  
670 proximate mechanisms of personality [87] and longitudinal data analyses to separate  
671 between- from within-individual behavioural variation [93,94].

672

673 A number of authors have emphasised the poor predictive value of aggression tests in  
674 shelter dogs [39–41,50] and that low occurrence of aggression specifically can make its  
675 accurate measurement difficult [40]. The probability of observing aggression on any  
676 particular day was low in this study (approximately 1%), and the number of dogs who,  
677 on average, showed aggression to people at least once while at the shelter was much  
678 lower than the number that showed aggression towards dogs, on average (Figs 1 and 2).  
679 Nonetheless, evaluations indicated that shelter employees might mistake observations  
680 of aggression for non-aggressive responses (e.g. over-excitement and frustration when

681 seeing people/dogs), meaning that the true probability of aggression was potentially  
682 under-estimated (although incorrectly coding other behaviours as aggression also  
683 occurred, albeit rarely). At the same time, our assessments of validity were based on  
684 shelter staff evaluations of brief video recordings that may be less reliable than the live,  
685 spontaneous behavioural recordings upon which our main analyses were based,  
686 resulting in a lower percentage of correctly identified instances of aggression.  
687 Nevertheless, infrequent occurrence and/or recording of aggression may also limit  
688 accurate predictions of future behaviour. Patronek and Bradley [50] demonstrate using  
689 simulation that the low prevalence of aggression inflates the chance that aggression  
690 shown in a shelter assessment represents a false positive. In general, our results support  
691 this conclusion in the sense that aggression may be shown differentially across contexts  
692 not explained by latent levels of aggressiveness. Violations of local independence and  
693 measurement invariance as found here indicate, further, that it is not only the  
694 difference between false and true positives and negatives, but the validity of inferring  
695 homogeneous personality traits by which to compare individual dogs, that needs careful  
696 consideration. Consequently, we agree with recommendations to establish the efficacy  
697 of longitudinal, observational assessments rather than relying on a single assessment  
698 made using a traditional test battery [31,40,50]. This approach will prioritise the  
699 cumulative understanding of a dog's context-dependent behaviour and help to guide  
700 decisions about the potential risk a dog poses to humans and other animals.  
701

## 702 **Conclusion**

703 This study has tested the assumptions of local independence and measurement  
704 invariance of personality traits in shelter dogs. Using structural equation modelling,  
705 aggression across behavioural contexts was explained by two correlated latent variables  
706 and demonstrated repeatability. Nevertheless, significant residual covariances remained  
707 between certain behavioural contexts related to aggressiveness towards people,  
708 violating the assumption of local independence. In addition, aggression in different  
709 contexts showed differential patterns of response across sex and age, indicating a lack  
710 of measurement invariance. Violations of local independence and measurement  
711 invariance imply that the aggressiveness towards people and dogs traits did not  
712 completely explain patterns of aggression in different contexts, or that inferences based  
713 on these hypothesised personality traits may in fact be misleading. We encourage  
714 researchers to more closely assess the measurement assumptions underlying reflective  
715 latent variable models before making conclusions about the effects of, or factors  
716 influencing, personality.

717

## 718 **Acknowledgements**

719 The authors are extremely grateful to Battersea Dogs and Cats Home for allowing us to  
720 access the data for this study.

## 721 **References**

- 722 1. Dochtermann NA, Jenkins SH. Behavioural syndromes in Merriam’s kangaroo rats  
723 (*Dipodomys merriami*): a test of competing hypotheses. Proc R Soc B. 2007;274:  
724 2343–2349. doi:10.1098/rspb.2007.0622
- 725 2. Dochtermann NA, Dingemanse NJ. Behavioral syndromes as evolutionary  
726 constraints. Behav Ecol. 2013; art002. doi:10.1093/beheco/art002
- 727 3. Westneat DF, Wright J, Dingemanse NJ. The biology hidden inside residual within-  
728 individual phenotypic variation. Biol Rev Camb Philos Soc. 2015;90: 729–743.  
729 doi:10.1111/brv.12131
- 730 4. Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ. Integrating animal  
731 temperament within ecology and evolution. Biol Rev Camb Philos Soc. 2007;82: 291–  
732 318. doi:10.1111/j.1469-185X.2007.00010.x
- 733 5. Sih A, Bell AM, Johnson JC, Ziemba RE. Behavioral syndromes: an integrative  
734 overview. Q Rev Biol. 2004;79: 241–277. doi:10.1086/422893
- 735 6. Budaev DS. How many dimensions are needed to describe temperament in animals:  
736 a factor reanalysis of two data sets. Int J Comp Psychol. <http://cogprints.org/5478/>;  
737 1998.
- 738 7. Spearman C. “General intelligence,” objectively determined and measured. Am J  
739 Psychol. 1904;15: 201–292. doi:10.2307/1412107

- 740 8. Beaujean AA. Latent variable modeling using R: A step-by-step guide. New York:  
741 Routledge; 2014.
- 742 9. Bollen KA. Latent variables in psychology and the social sciences. *Annu Rev Psychol.*  
743 2002;53: 605–634. doi:10.1146/annurev.psych.53.100901.135239
- 744 10. Borsboom D. The attack of the psychometricians. *Psychometrika.* 2006;71: 425–440.  
745 doi:10.1007/s11336-006-1447-6
- 746 11. Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation  
747 perspective. *Psychol Bull.* 1991;110: 305–314. doi:10.1037/0033-2909.110.2.305
- 748 12. Budaev SV. Using principal components and factor analysis in animal behaviour  
749 research: caveats and guidelines. *Ethology.* 2010;116: 472–480. doi:10.1111/j.1439-  
750 0310.2010.01758.x
- 751 13. Dingemanse NJ, Dochtermann N, Wright J. A method for exploring the structure of  
752 behavioural syndromes to allow formal comparison within and between data sets.  
753 *Anim Behav.* 2010;79: 439–450. doi:10.1016/j.anbehav.2009.11.024
- 754 14. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of  
755 exploratory factor analysis in psychological research. *Psychol Methods.* 1999;4: 272–  
756 299. doi:10.1037/1082-989X.4.3.272
- 757 15. Preacher KJ, MacCallum RC. Repairing Tom Swift’s electric factor analysis machine.  
758 *Understanding Statistics.* 2003;2: 13–43. doi:10.1207/S15328031US0201\_02

- 759 16. Araya-Ajoy YG, Dingemanse NJ. Characterizing behavioural “characters”: An  
760 evolutionary framework. *Proc R Soc B*. 2014;281: 20132645.  
761 doi:10.1098/rspb.2013.2645
- 762 17. Arden R, Adams MJ. A general intelligence factor in dogs. *Intelligence*. 2016;55: 79–  
763 85. doi:10.1016/j.intell.2016.01.008
- 764 18. van den Berg SM, Heuven HCM, van den Berg L, Duffy DL, Serpell JA. Evaluation of  
765 the C-BARQ as a measure of stranger-directed aggression in three common dog  
766 breeds. *Appl Anim Behav Sc*. 2010;124: 136–141.  
767 doi:10.1016/j.applanim.2010.02.005
- 768 19. Martin JS, Suarez SA. Personality assessment and model comparison with behavioral  
769 data: a statistical framework and empirical demonstration with bonobos (*Pan*  
770 *paniscus*). *Am J Primatol*. 2017;9999: e22670. doi: 10.1002/ajp.22670
- 771 20. Bartholomew DJ. Factor analysis for categorical data. *J R Stat Soc Series B Stat*  
772 *Methodol*. 1980;42: 293–321. <http://www.jstor.org/stable/2985165>
- 773 21. Markus KA, Borsboom D. *Frontiers of test validity theory: Measurement, causation,*  
774 *and meaning*. 1st ed. New York, N.Y: Routledge; 2013.
- 775 22. Drasgow F. Study of the measurement bias of two standardized psychological tests.  
776 *Journal of Applied Psychology*. 1987;72: 19–29. doi:10.1037/0021-9010.72.1.19
- 777 23. Mellenbergh GJ. Item bias and item response theory. *Int J Educ Res*. 1989;13: 127–  
778 143. doi:10.1016/0883-0355(89)90002-5

- 779 24. Meredith W. Measurement invariance, factor analysis and factorial invariance.  
780 Psychometrika. 1993;58: 525–543. doi:10.1007/BF02294825
- 781 25. Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response  
782 theory: two approaches for exploring measurement invariance. Psychol Bull.  
783 1993;114: 552–566. doi:10.1037/0033-2909.114.3.552
- 784 26. Jansen BRJ, van der Maas HLJ. Statistical test of the rule assessment methodology by  
785 latent class analysis. Dev Rev. 1997;17: 321–357. doi:10.1006/drev.1997.0437
- 786 27. Wicherts JM, Dolan CV. Measurement invariance in confirmatory factor analysis: an  
787 illustration using IQ test performance of minorities. Educ Meas Issues Pract.  
788 2010;29: 39–47. doi:10.1111/j.1745-3992.2010.00182.x
- 789 28. Goddard ME, Beilharz RG. Genetic and environmental factors affecting the suitability  
790 of dogs as Guide Dogs for the Blind. Theor Appl Genet. 1982;62: 97–102.  
791 doi:10.1007/BF00293339
- 792 29. Hsu Y, Serpell JA. Development and validation of a questionnaire for measuring  
793 behavior and temperament traits in pet dogs. J Am Vet Med Assoc. 2003;223: 1293–  
794 1300. doi:10.2460/javma.2003.223.1293
- 795 30. Scott JP, Fuller JL. Genetics and the social behavior of the dog. University of Chicago  
796 Press; 2012.



- 797 31. Rayment DJ, Groef BD, Peters RA, Marston LC. Applied personality assessment in  
798 domestic dogs: limitations and caveats. *Appl Anim Behav Sci.* 2015;163: 1–18.  
799 doi:10.1016/j.applanim.2014.11.020
- 800 32. Persson ME, Wright D, Roth LSV, Batakis P, Jensen P. Genomic regions associated  
801 with interspecies communication in dogs contain genes related to human social  
802 disorders. *Sci Rep.* 2016;6. doi:10.1038/srep33439
- 803 33. Sundman A-S, Johnsson M, Wright D, Jensen P. Similar recent selection criteria  
804 associated with different behavioural effects in two dog breeds. *Genes Brain Behav.*  
805 2016;15: 750–756. doi:10.1111/gbb.12317
- 806 34. Barnard S, Marshall-Pescini S, Pelosi A, Passalacqua C, Prato-Previde E, Valsecchi P.  
807 Breed, sex, and litter effects in 2-month old puppies' behaviour in a standardised  
808 open-field test. *Sci Rep.* 2017;7: 1 – 8. doi:10.1038/s41598-017-01992-x.
- 809 35. Svartberg K. Breed-typical behaviour in dogs – historical remnants or recent  
810 constructs? *Appl Anim Behav Sci.* 2006;96: 293 – 313. doi:  
811 10.1016/j.applanim.2005.06.014
- 812 36. Jones AC, Gosling SD. Temperament and personality in dogs (*Canis familiaris*): a  
813 review and evaluation of past research. *Appl Anim Behav Sci.* 2005;95: 1–53.  
814 doi:10.1016/j.applanim.2005.04.008
- 815 37. Fratkin JL, Sinn DL, Patall EA, Gosling SD. Personality consistency in dogs: a meta-  
816 analysis. *PLOS ONE.* 2013;8: e54907. doi:10.1371/journal.pone.0054907

- 817 38. Posluns JA, Anderson RE, Walsh CJ. Comparing two canine personality assessments:  
818 convergence of the MCPQ-R and DPQ and consensus between dog owners and dog  
819 walkers. *Appl Anim Behav Sci*. doi:10.1016/j.applanim.2016.12.013
- 820 39. Bennett SL, Litster A, Weng H-Y, Walker SL, Luescher AU. Investigating behavior  
821 assessment instruments to predict aggression in dogs. *Appl Anim Behav Sci*.  
822 2012;141: 139–148. doi:10.1016/j.applanim.2012.08.005
- 823 40. Mohan-Gibbons H, Weiss E, Slater M. Preliminary investigation of food guarding  
824 behavior in shelter dogs in the United States. *Animals*. 2012;2: 331–346.  
825 doi:10.3390/ani2030331
- 826 41. Mornement KM, Coleman GJ, Toukhsati SR, Bennett PC. Evaluation of the predictive  
827 validity of the Behavioural Assessment for Re-homing K9's (B.A.R.K.) protocol and  
828 owner satisfaction with adopted dogs. *Appl Anim Behav Sci*. 2015;167: 35–42.  
829 doi:10.1016/j.applanim.2015.03.013
- 830 42. Riemer S, Müller C, Virányi Z, Huber L, Range F. The predictive value of early  
831 behavioural assessments in pet dogs: a longitudinal study from neonates to adults.  
832 *PLOS ONE*. 2014;9: e101237. doi:10.1371/journal.pone.0101237
- 833 43. Wilsson E, Sundgren P-E. Behaviour test for eight-week old puppies: heritabilities of  
834 tested behaviour traits and its correspondence to later behaviour. *Appl Anim Behav*  
835 *Sci*. 1998;58: 151–162. doi:10.1016/S0168-1591(97)00093-2

- 836 44. Goodloe LP, Borchelt PL. Companion dog temperament traits. *J Appl Anim Welf Sci*.  
837 1998;1: 303–338. doi:10.1207/s15327604jaws0104\_1
- 838 45. Jones AC. Development and validation of a dog personality questionnaire. Doctoral  
839 Thesis. University of Texas at Austin. 2008. [http://gosling.psy.utexas.edu/wp-](http://gosling.psy.utexas.edu/wp-content/uploads/2014/10/Amanda-Claire-Jones-Diss-2008.pdf)  
840 [content/uploads/2014/10/Amanda-Claire-Jones-Diss-2008.pdf](http://gosling.psy.utexas.edu/wp-content/uploads/2014/10/Amanda-Claire-Jones-Diss-2008.pdf)
- 841 46. Orritt R. Dog bites: a complex public health issue. *Vet Rec*. 2015;176: 640–641.  
842 doi:10.1136/vr.h3215
- 843 47. Svartberg K. Shyness-boldness predicts performance in working dogs. *Appl Anim*  
844 *Behav Sci*. 2002;79: 157–174. doi:10.1016/S0168-1591(02)00120-X
- 845 48. Taylor KD, Mills DS. The development and assessment of temperament tests for  
846 adult companion dogs. *J Vet Behav*. 2006;1: 94–108. doi:10.1016/j.jveb.2006.09.002
- 847 49. Haverbeke A, Plujimakers J, Diederich C. Behavioral evaluations of shelter dogs:  
848 literature review, perspectives, and follow-up within the European member state’s  
849 legislation with emphasis on the Belgian situation. *J Vet Behav*. 2015;10: 5 – 11. doi:  
850 10.1016/j.jveb.2014.07.004
- 851 50. Patronek GJ, Bradley J. No better than flipping a coin: reconsidering canine behavior  
852 evaluations in animal shelters. *J Vet Behav*. 2016;15: 66–77.  
853 doi:10.1016/j.jveb.2016.08.001

- 854 51. Edelen MOE, Reeve BB. Applying item response theory (IRT) modeling to  
855 questionnaire development, evaluation, and refinement. *Qual Life Res.* 2007;16: 5 –  
856 18. doi: 10.1007/s11136-007-9198-0
- 857 52. Casey RA, Loftus B, Bolster C, Richards GJ, Blackwell EJ. Human directed aggression  
858 in domestic dogs (*Canis familiaris*): occurrence in different contexts and risk factors.  
859 *Appl Anim Behav Sci.* 2014;152: 52–63. doi:10.1016/j.applanim.2013.12.003
- 860 53. Hsu Y, Sun L. Factors associated with aggressive responses in pet dogs. *Appl Anim*  
861 *Behav Sci.* 2010;123: 108–123. doi:10.1016/j.applanim.2010.01.013
- 862 54. Sherman CK, Reisner IR, Taliaferro LA, Houpt KA. Characteristics, treatment, and  
863 outcome of 99 cases of aggression between dogs. *Appl Anim Behav Sci.* 1996;47: 91–  
864 108. doi:10.1016/0168-1591(95)01013-0
- 865 55. Asp HE, Fikse WF, Nilsson K, Strandberg E. Breed differences in everyday behaviour  
866 of dogs. *Appl Anim Behav Sci.* 2015;169: 69–77. doi:10.1016/j.applanim.2015.04.010
- 867 56. Voith VL, Trevejo R, Dowling-Guyer S, Chadik C, Marder A, Johnson V, et al.  
868 Comparison of visual and DNA breed identification of dogs and inter-observer  
869 reliability. *Sociology.* 2013;3: 17–29. doi: 10.5923/j.sociology.20130302.02
- 870 57. Olson KR, Levy JK, Norby B, Crandall MM, Broadhurst JE, Jacks S, Barton RC,  
871 Zimmerman MS. Inconsistent identification of pit bull-type dogs by shelter staff. *Vet*  
872 *J.* 2015;206: 197 – 202. doi:10.1016/j.tvjl.2015.07.019

- 873 58. Simpson RJ, Simpson K, VanKavage L. Rethinking dog breed identification in  
874 veterinary practice. *J Am Vet Med Assoc.* 2012;241: 1163 – 1166. doi:  
875 10.2460/javma.241.9.1163
- 876 59. Owczarczak-Garstecka, SC, Burman OHP. Can sleep and resting behaviour be used as  
877 indicators of welfare in shelter dogs (*Canis lupus familiaris*)? *PLOS ONE.* 2016;11:  
878 e0163620. doi:10.1371/journal.pone.0163620
- 879 60. R Development Core Team. R: a language and environment for statistical computing.  
880 Vienna, Austria: R Foundation for Statistical Computing; 2016. [https://www.r-](https://www.r-project.org/)  
881 [project.org/](https://www.r-project.org/)
- 882 61. Honaker J, King G, Blackwell M. Amelia: A program for missing data. 2015. Version  
883 1.7.4. <https://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf>
- 884 62. Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., et al,  
885 Lavaan: Latent Variable Analysis. 2016. Version 0.5-22. [https://cran.r-](https://cran.r-project.org/web/packages/lavaan/)  
886 [project.org/web/packages/lavaan/](https://cran.r-project.org/web/packages/lavaan/)
- 887 63. Muthén B, Christoffersson A. Simultaneous factor analysis of dichotomous variables  
888 in several groups. *Psychometrika.* 1981;46: 407–419. doi:10.1007/BF02293798
- 889 64. Muthén B. A general structural equation model with dichotomous, ordered  
890 categorical, and continuous latent variable indicators. *Psychometrika.* 1984;49: 115–  
891 132. doi:10.1007/BF02294210

- 892 65. Jorgensen TD, Pornprasertmanit S, Miller P, Schoemann A, Rosseel Y, Quick C, et al.  
893 semTools: useful tools for structural equation modeling. 2016.  
894 <https://rdrr.io/cran/semTools/>
- 895 66. Hermida R. The problem of allowing correlated errors in structural equation  
896 modeling: concerns and considerations. *Comp Method Soc Sci.* 2015;3: 05–17.
- 897 67. Kamata A. Item analysis by the hierarchical generalized linear model. *J Educ Meas.*  
898 2001;38: 79–93. doi:10.1111/j.1745-3984.2001.tb01117.x
- 899 68. Van den Noortgate W, De Boeck P. Assessing and explaining differential item  
900 functioning using logistic mixed models. *J Educ Behav Stat.* 2005;30: 443–464.
- 901 69. Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: a  
902 practical guide for biologists. *Biol Rev Camb Philos Sci.* 2010;85: 935–956.  
903 doi:10.1111/j.1469-185X.2010.00141.x
- 904 70. Stan Development Team. Stan modeling language users guide and reference  
905 manual. 2016. Version 2.15.0. [https://github.com/stan-](https://github.com/stan-dev/stan/releases/download/v2.14.0/stan-reference-2.14.0.pdf)  
906 [dev/stan/releases/download/v2.14.0/stan-reference-2.14.0.pdf](https://github.com/stan-dev/stan/releases/download/v2.14.0/stan-reference-2.14.0.pdf)
- 907 71. Kruschke JK. Bayesian data analysis. *Wiley Interdiscip Rev: Cogn Sci.* 2010;1: 658–  
908 676. doi:10.1002/wcs.72
- 909 72. Vehtari A, Gelman A, Gabry J, Pironen J, Goodrich B. Loo: Efficient leave-one-out  
910 cross-validation and WAIC for Bayesian models. 2016. Version 1.0.0. [https://cran.r-](https://cran.r-project.org/web/packages/loo/loo.pdf)  
911 [project.org/web/packages/loo/loo.pdf](https://cran.r-project.org/web/packages/loo/loo.pdf)

- 912 73. Kruschke J. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. 2nd ed.  
913 Academic Press; 2014.
- 914 74. Chen JJ, Tsong Y, Kang S-H. Tests for equivalence or noninferiority between two  
915 proportions. *Drug Inf J*. 2000;34: 569–578. doi:10.1177/009286150003400225
- 916 75. Budaev S, Brown C. Personality traits and behaviour. In: Brown C, Laland K, Krause J,  
917 editors. *Fish cognition and behavior*. Wiley-Blackwell; 2011. pp. 135–165.
- 918 76. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-  
919 multimethod matrix. *Psychol Bull*. 1959;56: 81–105. doi:10.1037/h0046016
- 920 77. Carter AJ, Feeney WE, Marshall HH, Cowlshaw G, Heinsohn R. Animal personality:  
921 what are behavioural ecologists measuring? *Biol Rev Camb Philos Soc*. 2013;88: 465–  
922 475. doi:10.1111/brv.12007
- 923 78. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in  
924 behavioral research: a critical review of the literature and recommended remedies. *J*  
925 *Appl Psychol*. 2003;88: 879–903. doi:10.1037/0021-9010.88.5.879
- 926 79. Cole DA, Ciesla JA, Steiger JH. The insidious effects of failing to include design-driven  
927 correlated residuals in latent-variable covariance structure analysis. *Psychol*  
928 *Methods*. 2007;12: 381–398. doi:10.1037/1082-989X.12.4.381
- 929 80. Yen WM. Scaling performance assessments: strategies for managing local item  
930 dependence. *J Educ Meas*. 1993;30: 187–213. doi:10.1111/j.1745-  
931 3984.1993.tb00423.x

- 932 81. Carter AJ, Marshall HH, Heinsohn R, Cowlshaw G. How not to measure boldness:  
933 novel object and antipredator responses are not the same in wild baboons. *Anim*  
934 *Behav.* 2012;84: 603–609. doi:10.1016/j.anbehav.2012.06.015
- 935 82. Cramer AOJ, van der Sluis S, Noordhof A, Wichers M, Geschwind N, Aggen SH, et al.  
936 Dimensions of normal personality as networks in search of equilibrium: you can't  
937 like parties if you don't like people. *Eur J Pers.* 2012;26: 414–431.  
938 doi:10.1002/per.1866
- 939 83. Schmittmann VD, Cramer AOJ, Waldorp LJ, Epskamp S, Kievit RA, Borsboom D.  
940 Deconstructing the construct: a network perspective on psychological phenomena.  
941 *New Ideas Psychol.* 2013;31: 43–53. doi:10.1016/j.newideapsych.2011.02.007
- 942 84. Hamaker EL, Nesselroade JR, Molenaar PCM. The integrated trait-state model. *J Res*  
943 *Pers.* 2007;41: 295–315. doi:10.1016/j.jrp.2006.04.003
- 944 85. Fadel FR, Driscoll P, Pilot M, Wright H, Zulch H, Mills D. Differences in trait  
945 impulsivity indicate diversification of dog breeds into working and show lines. *Sci*  
946 *Rep.* 2016;6. doi:10.1038/srep22162
- 947 86. Schimmack U. What multi-method data tell us about construct validity. *Eur J Pers.*  
948 2010;24: 241–257. doi:10.1002/per.771
- 949 87. Duckworth RA. Neuroendocrine mechanisms underlying behavioral stability:  
950 implications for the evolutionary origin of personality. *Ann N Y Acad Sci.* 2015;1360:  
951 54–74. doi:10.1111/nyas.12797



- 952 88. Koski SE. How to measure animal personality and why does It matter? Integrating  
953 the psychological and biological approaches to animal personality. In: Inoue-  
954 Murayama M, Kawamura S, Weiss A, editors. From Genes to Animal Behavior.  
955 Springer Japan; 2011. pp. 115–136. doi:10.1007/978-4-431-53892-9\_5
- 956 89. Borsboom D. Measuring the mind: Conceptual issues in contemporary  
957 psychometrics. Cambridge University Press; 2005.
- 958 90. Maul A, Torres Iribarra D, Wilson M. On the philosophical foundations of  
959 psychological measurement. *Measurement*. 2016;79: 311–320.  
960 doi:10.1016/j.measurement.2015.11.001
- 961 91. Molenaar PCM. A manifesto on psychology as idiographic science: bringing the  
962 person back into scientific psychology, this time forever. *Measurement*. 2004;2:  
963 201–218. doi:10.1207/s15366359mea0204\_1
- 964 92. Adolf J, Schuurman NK, Borkenau P, Borsboom D, Dolan CV. Measurement  
965 invariance within and between individuals: a distinct problem in testing the  
966 equivalence of intra- and inter-individual model structures. *Front Psychol*. 2014;5.  
967 doi:10.3389/fpsyg.2014.00883
- 968 93. Stamps JA, Briffa M, Biro PA. Unpredictable animals: individual differences in  
969 intraindividual variability (IIV). *Anim Behav*. 2012;83: 1325–1334.  
970 doi:10.1016/j.anbehav.2012.02.017

971 94. Biro PA, Stamps JA. Using repeatability to study physiological and behavioural traits:

972 ignore time-related change at your peril. *Anim Behav.* 2015;105: 223–230.

973 doi:10.1016/j.anbehav.2015.04.008

974