

1 Full Title

2 Does training method matter?: Evidence for the negative impact of aversive-based methods on  
3 companion dog welfare

4

5 Short Title

6 Training methods and companion dog welfare

7

8 Author names and affiliations

9 Ana Catarina Vieira de Castro<sup>1,2,3\*</sup>, Danielle Fuchs<sup>1,2,4</sup>, Stefania Pastur<sup>1,2,5</sup>, Liliana de Sousa<sup>3</sup>, I  
10 Anna S Olsson<sup>1,2</sup>

11

12 <sup>1</sup>Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

13 <sup>2</sup>i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

14 <sup>3</sup>Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

15 <sup>4</sup>Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, United  
16 Kingdom

17 <sup>5</sup>University of Trieste, Trieste, Italy

18

19 \*Corresponding author:

20 E-mail: [ana.castro@ibmc.up.pt](mailto:ana.castro@ibmc.up.pt) (ACVC)

22 ABSTRACT

23 There is a growing number of dogs kept as companion animals, and the methods by which they  
24 are trained range broadly from those using mostly positive punishment and negative  
25 reinforcement (aversive-based methods) to those using primarily positive reinforcement (reward-  
26 based methods). Although the use of aversive-based methods has been strongly criticized for  
27 negatively affecting dog welfare, these claims do not find support in solid scientific evidence.  
28 Previous research on the subject lacks companion dog-focused research, investigation of the entire  
29 range of aversive-based techniques (beyond shock-collars), objective measures of welfare, and  
30 long-term welfare studies. The aim of the present study was to perform a comprehensive  
31 evaluation of the short- and long-term effects of aversive- and reward-based training methods on  
32 companion dog welfare. Ninety-two companion dogs were recruited from three reward-based  
33 (Group Reward, n=42) and four aversive-based (Group Aversive, n=50) dog training schools. For  
34 the short-term welfare assessment, dogs were video recorded for three training sessions and six  
35 saliva samples were collected, three at home (baseline levels) and three after the training sessions  
36 (post-training levels). Video recordings were then used to examine the frequency of stress-related  
37 behaviors (e.g., lip lick, yawn) and the overall behavioral state of the dog (e.g., tense, relaxed),  
38 and saliva samples were analyzed for cortisol concentration. For the long-term welfare  
39 assessment, dogs performed a cognitive bias task. Dogs from Group Aversive displayed more  
40 stress-related behaviors, spent more time in tense and low behavioral states and more time panting  
41 during the training sessions, showed higher elevations in cortisol levels after training and were  
42 more ‘pessimistic’ in the cognitive bias task than dogs from Group Reward. These findings  
43 indicate that the use of aversive-based methods compromises the welfare of companion dogs in  
44 both the short- and the long-term.

45

## 46 1. Introduction

47 To fulfil their increasingly important role as companion animals, dogs need to be trained  
48 to behave in a manner appropriate for human households. This includes, for example, learning to  
49 eliminate outdoors or walk calmly on a lead [1,2]. The fact that dog behavior problems are the  
50 most frequently cited reason for rehoming or relinquishment of dogs to shelters and for euthanasia  
51 [2] suggests that such training is often missing or unsuccessful.

52 Dog training most often involves the use of operant conditioning principles, and dog  
53 training methods can be classified according to the principles they implement: aversive-based  
54 methods use positive punishment and negative reinforcement and reward-based methods rely on  
55 positive reinforcement and negative punishment [3]. There is a heated debate surrounding the use  
56 of aversive-based training methods, as studies have linked them to compromised dog welfare  
57 (e.g., [4-9]). Some aversive-based tools, such as shock collars, have indeed been legally banned  
58 in some countries [10]. However, a recent literature review by [3] concluded that, because of  
59 important limitations, existing studies on the topic do not provide adequate data for drawing firm  
60 conclusions. Specifically, the authors reported that a considerable proportion of the studies relied  
61 upon surveys rather than on objective measures of both training methods and welfare; that they  
62 focused on sub-populations of police and laboratory dogs which only represent a small portion of  
63 dogs undergoing training; and, finally, that the empirical studies have concentrated mainly on the  
64 effects of shock-collar training, which is only one of several tools used in aversive-based training.  
65 In summary, limited scientific evidence exists on the effects of the entire range of dog training  
66 techniques on companion dog welfare.

67 Furthermore, previous empirical studies have focused on the short-term effects of training  
68 methods on dog welfare. Behavioral and physiological indicators of welfare, such as the  
69 frequency of stress-related behaviors and the concentration of salivary cortisol, have been  
70 collected in and around the training situation (e.g., [8, 11]; see also [3]). However, the long-term  
71 welfare implications of training methods have not yet been examined. To our knowledge, only  
72 one study evaluated the long-term effects of training on welfare. Christiansen et al (2001) [12]

73 found no effect of shock collar training on dog fear or anxiety; however this was based on dog  
74 owner reports of behavior and temperament tests rather than on objective and animal-based  
75 welfare indicators. Importantly, a suitable assessment of the effects of training methods on dog  
76 welfare should comprise an evaluation of both their short- and long-term effects.

77 Long-term (or chronic) stress can arise from the cumulative exposure to aversive  
78 experiences [13], which may reflect the experience of dogs trained with aversive-based methods.  
79 A body of research has shown that long-term stress is associated with changes in the long-term  
80 affective state of animals (e.g., [14-16]). One way to assess affective states is through the  
81 cognitive bias paradigm (e.g., [16]). The cognitive bias task has been validated as an effective  
82 tool to evaluate the affective states of non-human animals and has been extensively used with  
83 several species, including dogs [17-19]. The rationale behind the paradigm is based on theoretical  
84 and empirical findings that an individual's underlying affective state biases its decision-making  
85 and, specifically, that individuals experiencing negative emotional states make more 'pessimistic'  
86 judgements about ambiguous stimuli than individuals experiencing more positive emotional states  
87 [16, 18].

88 Therefore, the aim of the present study was to perform a comprehensive evaluation of the  
89 short- and long-term effects of aversive- and reward-based training methods on companion dog  
90 welfare. By performing an objective assessment of training methods (through the direct  
91 observation of training sessions) and by using objective measures of welfare (behavioral and  
92 physiological data to assess short-term effects, and a cognitive bias task to assess long-term  
93 effects), we addressed the question of whether aversive-based methods actually compromise the  
94 well-being of companion dogs. We hypothesized that dogs trained using aversive-based methods  
95 would display higher levels of stress during training, as determined by behavioral and  
96 physiological indicators of stress during training sessions, and more 'pessimistic' judgments of  
97 ambiguous stimuli during a cognitive bias task performed outside the training context.

98 Understanding the effects of training methods on companion dog welfare has important  
99 consequences for both dogs and humans. Both determining and applying those training methods

100 that are less stressful for dogs is a key factor to ensure adequate dog welfare and to capitalize on  
101 the human benefits derived from interactions with dogs [20, 21].

102

## 103 2. Materials and methods

### 104 2.1. Ethical Statement

105 All procedures were approved by ICBAS (Abel Salazar Biomedical Sciences Institute)  
106 ORBEA (Animal Welfare Body). All head trainers of dog training schools and owners completed  
107 a consent form authorizing the collection and use of the data.

108

### 109 2.2. Training schools

#### 110 2.2.1. Recruitment

111 The first author (ACVC) searched on the internet for dog training schools in the  
112 metropolitan area of Porto, Portugal. Based on geographical proximity and on the listed training  
113 methods, ACVC posteriorly contacted eight schools through telephone. She approached the head  
114 trainers of the different schools on their willingness to participate in a study that aimed to evaluate  
115 dog stress and welfare in the context of training and explained the entire methodology. However,  
116 it was not directly revealed that the aim of the study was to compare the effects of different  
117 training methods. Seven out of the eight dog training schools agreed to participate.

118 The training schools had different class structures and training sites; however, the types  
119 of behaviors trained was fairly standard across training schools. These characteristics are  
120 described in detail in Appendix S1.

121

#### 122 2.2.2. Classification of training methods

123           After securing seven participating schools, we performed an objective assessment of the  
124 training methods used by each school. We videotaped four training sessions at each training  
125 school using a video camera on a tripod. Afterwards, we analyzed the videos for the frequency of  
126 aversive-based operant conditioning procedures utilized, namely positive punishment and  
127 negative reinforcement (see Appendix S1 for the specific definitions). The analysis was  
128 performed by AVCV using The Observer XT software, version 10.1 (Noldus Information  
129 Technology Inc, Wageningen, The Netherlands). The schools were classified as aversive-based  
130 if they used any positive punishment and/or negative reinforcement training techniques, and as  
131 reward-based if they did not use any of these techniques. Schools A, C, D and F were classified  
132 as aversive-based, and Schools B, E and G were classified as reward-based (see Appendix S1).

133

### 134 2.3. Subjects

135           The head trainer of each training school was asked to indicate at least fourteen dogs fitting  
136 our inclusion criteria, and we then approached the owners to ask if they were willing to participate.  
137 The information about the study given to the owners was the same that was given to the head  
138 trainers of the schools. The inclusion criteria for the dogs were: 1) to have attended the training  
139 school for less than two months, in order to mitigate habituation to training methods, and 2) to be  
140 free of behavioral problems (e.g., aggression, fearfulness and separation anxiety, as determined  
141 by the owner and ACVC), in order to prevent any confounding stress.

142           Over the course of the study, which was conducted between October 2016 and March  
143 2019, the owners of 122 companion dogs agreed to participate. However, 30 dog owners dropped  
144 out of the training schools before any meaningful data could be collected. Specifically, these  
145 subjects dropped out before meeting our requirement that at least two training sessions were video  
146 recorded and that the owner completed a written questionnaire. Hence, our final sample comprised  
147 92 subjects, 50 recruited from aversive-based schools, hereafter ‘Group Aversive’ (Schools A, D  
148 and F: 14 dogs, School C: 8 dogs), and 42 from reward-based schools, hereafter ‘Group Reward’  
149 (School B and G: 15 dogs, School E: 12 dogs).

150 As for subjects' demographics, the average age was 11.9 ( $\pm 9.3$ ) months, 54% were male  
151 and 35% were neutered/spayed. Thirty-four percent were mixed-breed dogs and the remaining  
152 66% belonged to a FCI-recognized breed group: 18% belonged to Group 1: Sheepdogs and  
153 Cattle dogs (except Swiss Cattle dogs), 13% to Group 2: Pinscher and Schnauzer – Molosoid and  
154 Swiss Mountain and Cattle dogs, 5% to Group 3: Terriers, 4% to Group 6: Scent hounds and  
155 related breeds, 2% to Group 7: Pointing dogs, 20% to Group 8: Retrievers – Flushing Dogs –  
156 Water Dogs, and 3% to Group 9: Companion and Toy Dogs.

157

## 158 2.4. Data collection

159 The study had two phases. The goal of Phase 1 was to evaluate the welfare of dogs within  
160 the training context, and the goal of Phase 2 was to evaluate the welfare of these same dogs outside  
161 the training context. These aimed to represent measures of the short- and long-term impact of  
162 training methods on the welfare of dogs.

163

### 164 2.4.1. Phase 1 – Evaluating welfare within the training context

165 In order to evaluate behavioral indicators of welfare during training, each dog was  
166 videotaped for the first 15 minutes of three training sessions using a video camera on a tripod (one  
167 Sony Handycam HDR-CX405 and two Sony Handycam DCR-HC23). Five experimenters were  
168 responsible for data collection (ACVC, Danielle Fuchs – DF, Stefania Pastur – SP, and two  
169 undergraduate students, Margarida Lencastre and Flávia Canastra). The cameras were positioned  
170 to get an optimal view of the specific participant without interfering with training. The day and  
171 time of the training sessions were determined by the training schools and by the participants'  
172 availability.

173 To obtain physiological data on stress during training, six saliva samples were collected  
174 per dog to allow assay of salivary cortisol [8, 22]. Three samples were collected 20 min after each  
175 training session (PT – post-training samples) and three were collected at home on days when no

176 training took place, approximately at the same time as PT samples (BL – baseline samples).  
177 Owners were asked not to give their dog water in the 20 minutes preceding each sample collection,  
178 nor a full meal in the hour preceding each sample collection, respectively. The timing for sample  
179 collections, as well as other recommendations regarding saliva collection for cortisol analysis,  
180 were drawn from previous relevant research on dogs' cortisol responses to potentially stressful  
181 stimuli [8, 22-24]. ACVC collected the first sample of every subject (PT1) while simultaneously  
182 demonstrating proper sample collection to the owners. The following samples were always  
183 collected by the owners. A synthetic swab (Salivette®) was rubbed in the dogs' mouth for about  
184 2 minutes to collect saliva. For samples collected at the training schools (PT), the swab was placed  
185 back into the provided plastic tube and immediately stored on ice. It was then transferred to a -  
186 20°C freezer as soon as possible. For samples collected at home (BL), owners were instructed to  
187 place the swab back into the plastic tube and immediately store it in their home freezer. Owners  
188 were provided with ice-cube plastic makers to transport the BL samples to the training school  
189 during the next scheduled training session without them unfreezing, and they were stored at -20°C  
190 as soon as possible. Owners were also provided with detailed written instructions for saliva  
191 collection and ACVC's cell phone number in case any owners had questions related to sample  
192 collection. For standardization purposes, we ensured that Phase 1 did not last more than three  
193 months for each dog.

194

#### 195 2.4.2. Phase 2 - Evaluating welfare outside the training context

196 After finishing data collection for Phase 1, dogs moved to Phase 2, which consisted of a  
197 spatial cognitive bias task. For standardization purposes, we ensured that 1) dogs had attended the  
198 training school for at least one month prior to moving to Phase 2 and that 2) the cognitive bias  
199 task was conducted within one month of completing Phase 1. Due to limited owner availability,  
200 13 subjects either dropped out of the study or did not meet the criteria for Phase 2, resulting in 79  
201 (44 from Group Aversive and 35 from Group Reward) of the original 92 dogs participating in



202 Phase 2. The tests were scheduled according to owners' availability, both on weekdays and  
203 Saturdays.

204 The test was conducted in an indoor room (7.7 x 3 meters) within a research building at  
205 the Abel Salazar Biomedical Sciences Institute (ICBAS), University of Porto in Portugal. All  
206 dogs were unfamiliar with the room prior to testing. Two experimenters conducted the test while  
207 the dog's owner(s) sat in a chair in a corner area of the room (see Figure 1). Dog owners were  
208 asked not to look into the dog's eyes or to speak to the dog during the test, unless the  
209 experimenters instructed otherwise. The entire test took place over one meeting for each dog. The  
210 room was cleaned with water and liquid detergent at the end of each test.

211

212 Figure 1. Schematic representation of the cognitive bias task.

213

214 2.4.2.1. Familiarization period

215 Prior to the start of the cognitive bias task, the dogs were given the opportunity to  
216 familiarize with the test room and the researchers. This consisted of a 10-min period during which  
217 the dog was allowed to freely explore the room and engage with the researchers and the owner(s).

218

219 2.4.2.2. Training phase

220 The methodology of Phase 2 was based on [19]. During the training phase, dogs were  
221 trained to discriminate between a 'positive' (P) location of a food bowl, which always contained  
222 a food reward, and a 'negative' (N) location, which never contained a food reward. At the start of  
223 each trial, the dog was held by the 'handler' (played by SP, DF, Margarida Lencastre, Flávia  
224 Canastra or Joana Guilherme-Fernandes) behind a barrier (2 x 2 m, see Figure 1), while the  
225 'timer' (ACVC) baited (or did not bait, depending on the type of trial) the bowl with a piece of  
226 sausage (approximately 1.25 g for smaller dogs and 2.5 g for larger dogs). To ensure that the dog,

227 the owner and the ‘handler’ were blind to whether or not the bowl contained food during each  
228 trial, the bowl was baited out of their sight, on the opposite side of the barrier. Additionally, the  
229 food reward was rubbed onto the food bowl before every trial to prevent the influence of olfactory  
230 cues. The height of the food bowl was such that visual cues to the presence or absence of food  
231 could not be judged by the dog at the start position.

232 After baiting (or not baiting) the bowl, the ‘timer’ placed it at one of the two training  
233 locations. The ‘timer’ then determined the start of the trial, by verbally signaling to the ‘handler’,  
234 upon which the ‘handler’ led the dog to the start position and released him. The ‘handler’ always  
235 led the dog to the start position on her left side. Because we found that dogs had some difficulty  
236 noticing the bowl at the end of the room during pilot tests, the ‘handler’ walked towards the bowl  
237 and pointed it out to the dog in the first four trials. For the remaining trials, the ‘handler’ simply  
238 walked the dog to the start position and released him. After the dog reached the food bowl and  
239 (when applicable) ate the reward, the ‘handler’ collected him and led him behind the barrier to  
240 start the next trial. The latency to reach the bowl, defined as the time elapsed between release at  
241 the start position and the dog putting his head in line with the edge of the bowl, was recorded for  
242 each trial by the ‘timer’ using a stopwatch.

243 The position of the ‘positive’ and ‘negative’ locations was counterbalanced across  
244 subjects and training schools, such that for half of the dogs from each training school, the  
245 ‘positive’ location was on the right hand side as they faced the test area, and for the other half it  
246 was on the left. Initially, each dog received two consecutive ‘positive’ trials (bowl placed in the  
247 ‘positive’ location) followed by two ‘negative’ trials (bowl placed in the ‘negative’ location).  
248 Subsequently, ‘positive’ and ‘negative’ trials were presented in a pseudorandom order, with no  
249 more than two trials of the same type being presented consecutively.

250 All dogs received a minimum of 15 training trials to learn the discrimination between  
251 bowl locations. Dogs were considered to have learnt an association between bowl location and  
252 food (the learning criterion) when, after a minimum of 15 trials, the longest latency to reach the  
253 ‘positive’ location was shorter than any of the latencies to reach the ‘negative’ location for the

254 preceding three ‘positive’ trials and the preceding three ‘negative’ trials. Each trial lasted a  
255 maximum of 20 seconds. If the dog did not reach the bowl by that time, the trial automatically  
256 ended and a latency of 20 seconds was recorded.

257 All but two dogs were able to complete the training phase. For the two dogs that failed to  
258 complete training, one did not show any interest in the food reward and the other was food-  
259 motivated but could not focus on the task. These two dogs belonged to Group Aversive.  
260 Therefore, the total number of subjects completing Phase 2 in Group Aversive was 42.

261

#### 262 2.4.2.3. Test phase

263 Testing began once the learning criterion was achieved. Test trials were identical to  
264 training trials except that the bowl (empty) was placed at one of three ambiguous locations equally  
265 spaced along an arc 4 m from the dog’s start position, between the ‘positive’ and ‘negative’  
266 locations. The three test locations were: ‘near-positive’ (NP: one third of the way along the arc  
267 from the ‘positive’ location), ‘middle’ (M: half way along the arc), ‘near-negative’ (NN: one third  
268 of the way along the arc from the ‘negative’ location). Three test trials were presented at each test  
269 location (nine test trials in total) in the following order for all dogs:  
270 M,NP,NN,NP,NN,M,NN,M,NP (each location was presented first, second or third in each block  
271 of three test trials). Each test trial was separated from the next one by two training trials identical  
272 to those conducted in the training phase (one ‘positive’ and one ‘negative’ trials presented in a  
273 random order), in order to maintain the associations between the ‘positive’ and ‘negative’  
274 locations and the presence or absence of food, respectively. Thus, the test phase included a further  
275 sixteen training trials interspersed in blocks of two between the nine test trials.

276 To end the test phase, a final trial was conducted by placing an empty bowl in the  
277 ‘positive’ location to determine whether dogs ran to the empty bowl as quickly as they did to the  
278 baited bowl. This was meant to establish that the dogs were not relying on olfactory or visual cues

279 during the test. During the entire test, each trial was kept as similar as possible in terms of  
280 preparation time and activity, and dogs were handled in the same way throughout the test.

281 Due to circumstances beyond our control, namely people speaking loudly and other dogs  
282 barking in the building during some of the tests, some subjects were clearly distracted and  
283 disengaged from the task during some trials. Whenever this happened, no latency was recorded  
284 for that trial. The experimenters waited for the dog to resettle and moved to the following trial.

285

## 286 2.5. Questionnaire

287 All owners were asked to complete a brief written questionnaire regarding dog  
288 demographics and background, and owner demographics and experience with dogs and dog  
289 training. The questionnaire was based on [9].

290

## 291 2.6. Data analysis

### 292 2.6.1. Phase 1 – Evaluating welfare within the training context

#### 293 2.6.1.1. Behavior coding

294 We developed two ethograms based on previous literature to record the frequency of  
295 different stress-related behaviors and the time spent in different behavioral states and panting  
296 during the training sessions [8, 9, 25]. The behaviors and their definitions are described in  
297 Appendix S2.

298 Behavior coding was conducted by three observers (ACVC, DF and SP). DF and SP were  
299 blind to how the training schools had been classified according to training methods. Each video  
300 was coded twice, once with the ethogram for stress-related behaviors, using a continuous  
301 sampling technique (by ACVC and DF, see Appendix S2), and a second time with the ethogram  
302 for overall behavioral state and panting, by scan-sampling at 1 minute intervals (by ACVC and  
303 SP, see Appendix S2). The Observer XT software, version 10.1 (Noldus Information Technology

304 Inc, Wageningen, The Netherlands) was used to code for stress-related behaviors and Windows  
305 Movie Player and Microsoft Excel to code for overall behavioral state and panting.

306 Before coding independently, observers DF and SP were trained to become familiar with  
307 the ethograms, and inter-observer reliability was assessed. Inter-observer reliability was tested by  
308 having DF (for the ethogram for stress-related behaviors) and SP (for the ethogram for overall  
309 behavioral state and panting) code sets of four videos in parallel with ACVC. Cohen's Kappa  
310 coefficient was calculated using The Observer XT. Where there was poor agreement ( $r < 0.80$ ),  
311 observers received further training. Values of  $r > 0.80$  were assumed to indicate strong agreement  
312 (see, for example, [8]), and once this level was attained the observers began coding videos  
313 independently. ACVC coded 77% of the videos with the ethogram for stress-related behaviors  
314 and 65% with the ethogram for overall behavioral state and panting. For each ethogram, the videos  
315 were distributed randomly between observers, with the exception that each observer coded a  
316 similar number of videos from Group Aversive and Group Reward.

317

#### 318 2.6.1.2. Cortisol analysis

319 Two dogs (one from School B and one from School E, both from Group Reward) did  
320 not cooperate with the saliva collection procedure and, as such, no saliva samples were extracted  
321 from them. For the remaining 90 dogs, only 23 dog owners (11 from Group Reward and 12 from  
322 Group Aversive) were able to appropriately collect six saliva samples. The samples from these  
323 subjects were selected for analysis. An additional 40 dog owners (15 from Group Reward and 25  
324 from Group Aversive) were able to properly collect at least four saliva samples. From these 40  
325 dogs, eight were randomly selected to have their samples analyzed (four from Group Reward and  
326 four from Group Aversive). In the end, 16 dogs from Group Aversive and 15 dogs from Group  
327 Reward had their samples selected for analysis (Schools A, C, D, E and F:  $n=4$ ; School B:  $n=5$ ;  
328 School G:  $n=6$ ). These samples were sent to the Faculty of Sport Sciences and Physical Education  
329 of the University of Coimbra, Coimbra, Portugal, where they were assayed for cortisol  
330 concentration using standard ELISA kits (Salimetrics®).

331

## 332 2.6.2. Phase 2 – Evaluating welfare outside the training context

333 For each dog, we calculated the average latency to reach the food bowl during each of the  
334 three types of test trials (NP, M, NN) as well as the average latency to reach the ‘positive’ and  
335 ‘negative’ training locations during the test phase. Afterwards, to control for individual variation  
336 in running speeds, we adjusted each dog’s test latencies by taking into account their mean  
337 latencies to get to the ‘positive’ and ‘negative’ training locations during the test phase as follows  
338 [19]:

339

$$340 \text{ adjusted latency} = \frac{\text{mean latency to test location} - \text{mean latency to positive location}}{\text{mean latency to negative location} - \text{mean latency to positive location}} * 100$$

341

342 This adjusted score expresses all test latencies as a percentage of the difference between  
343 each dog’s mean latencies to the ‘positive’ and ‘negative’ locations [19].

344 Seventy-three dogs completed the cognitive bias task. From these 73 subjects, 14  
345 disengaged from the task for some trials due to noise outside the test room. Thirteen disengaged  
346 for one test trial (Group Reward: one dog at location M, three dogs at location NP, and one dog  
347 at location NN; Group Aversive: five dogs at location NP, and three dogs at location NN), and  
348 one (Group Reward) for three test trials (one at each test location). For these dogs, the average  
349 latencies to the test locations were calculated from the remaining test trials. Of the remaining four  
350 dogs, one (from School G) completed the first seven test trials (at locations  
351 M,NP,NN,NP,NN,M,NN), two (one from School A and one from School E) completed the first  
352 five test trials (at locations M,NP,NN,NP,NN), and one (from School G) completed the first three  
353 test trials (at locations M,NP,NN); then they stopped cooperating with the task. Their average  
354 latencies to the test locations were calculated from these trials.

355

## 356 2.7. Statistical analyses

357 All statistics were conducted using the software SPSS® Statistics 25.0. All data were  
358 analyzed using the Shapiro-Wilk test to check for normality. Except for the number of trials  
359 required to reach the learning criterion in the cognitive bias task, the data were not normally  
360 distributed; hence, except for the former variable, non-parametric tests were used for analysis.

361 Namely:

362 - Chi-square tests were used to compare the two groups (Reward and Aversive) for dog  
363 demographics and background, and owner demographics and experience with dogs  
364 and dog training (variables collected with the questionnaire, see Appendix S3), in  
365 order to investigate whether the two groups differed in these factors.

366 - Kruskal-Wallis and Mann-Whitney tests were used to verify if the dog-related  
367 variables age, FCI breed group, and age of separation from the mother (which differed  
368 between Group Reward and Group Aversive, see Appendix S3) affected the  
369 dependent variables measured.

370 - Friedman's ANOVA and Wilcoxon signed-rank tests were used to analyze the  
371 frequency of stress-related behaviors, the percentage of scans in the different  
372 behavioral states and the percentage of scans panting across the three training  
373 sessions, in order to examine whether there was an effect of sampling period. Where  
374 no effect of training session was found, the data for each dog were averaged across  
375 the three sessions.

376 - Mann-Whitney tests were used to compare the frequency of stress-related behaviors,  
377 the percentage of scans in the different behavioral states, the percentage of scans  
378 panting, the difference in cortisol concentrations and the adjusted latencies for the  
379 cognitive bias task between Group Reward and Group Aversive, in order to  
380 investigate whether the groups differed in these different indicators of welfare.

381 - A Wilcoxon signed-rank test was used to compare the post-training with the baseline  
382 levels of cortisol within both Group Reward and Group Aversive, with the aim of

383 examining whether the concentration of salivary cortisol changed as a result of  
384 training for each group.

385 - A Mann-Whitney test was used to compare the number of training classes attended  
386 by the dogs before moving to Phase 2, to verify if these did not differ between groups.

387 - A t-test for independent samples was used to compare the number of trials needed to  
388 reach the learning criterion in the cognitive bias task between Group Reward and  
389 Group Aversive, in order to explore whether the groups differed in learning speed.

390 - A Wilcoxon signed-rank test was used to compare, in the cognitive bias task, the  
391 latency to reach the P location during test trials and during the final trial, when the  
392 bowl contained no food, to verify whether dogs were relying on olfactory or visual  
393 cues to discriminate between bowl locations.

394 - Spearman correlation coefficients were used to examine the correlation between the  
395 number of aversive stimuli used in the different aversive-based schools and the  
396 frequency of stress-related behaviors, the percentage of scans in the different  
397 behavioral states, the percentage of scans panting, the difference in cortisol  
398 concentrations, and the adjusted latencies in the cognitive bias task; in order to  
399 investigate whether there was a correlation between the frequency of aversive stimuli  
400 used in training and the different indicators of welfare.

401

402 With the exception of the analyses conducted to test for the effect of the sampling period  
403 on behavioral data and the analysis performed to verify if dogs were relying on olfactory or visual  
404 in the cognitive bias task, which were within-subjects, all remaining analyses were between-  
405 subjects. Two-tailed tests were used for all but correlations analyses, for which one-tailed tests  
406 were conducted. The level of significance was set at  $\alpha= 0.05$ .

407

408 3. Results



### 409 3.1. Questionnaire

#### 410 3.1.1 Dog demographics and background

411           Concerning dog demographics, the two groups did not differ in sex and neuter status  
412 ratios, but they differed with regards to age [ $X^2(3)=10.361$ ,  $p=0.016$ ] and FCI breed group  
413 [ $X^2(7)=19.586$ ,  $p=0.007$ ]. As for dog background, the groups differed only in the age of separation  
414 from the mother [ $X^2(7)=19.041$ ,  $p=0.003$ , see Appendix S3 for full results]. However, statistical  
415 analysis performed for each training method group revealed that age, FCI breed group and age of  
416 separation from the mother did not affect any of the dependent variables measured in the present  
417 study (Appendix S3).

418

#### 419 3.1.2. Owner demographics, experience with dogs and dog training

420           Regarding owner demographics, the two groups did not differ in owner age, family  
421 household size and whether they had children, but they differed in owner sex [ $X^2(1)=8.360$ ,  
422  $p=0.006$ ]. Regarding owner experience with dogs and dog training, the groups did not differ in  
423 whether owners had attended training classes with a previous dog, but they differed in whether  
424 owners had had other dog(s) in the past [ $X^2(1)=4.658$ ,  $p=0.037$ ] and in the information they relied  
425 on for choosing the dog training school [ $X^2(4)=11.656$ ,  $p=0.016$ , see Appendix S3 for full results].

426

### 427 3.2. Phase 1 – Evaluating welfare within the training context

#### 428 3.2.1. Behavioral data

##### 429 3.2.1.1. Stress-related behaviors

430           For each dog, we first summed all the occurrences of each stress-related behavior as  
431 defined in the ethogram for stress-related behaviors. A comparison between the two groups  
432 revealed that dogs from Group Aversive displayed significantly more stress-related behaviors

433 than dogs from Group Reward [Group Aversive:  $M(\pm SEM)=57.06\pm 5.98$  vs Group Reward:  
434  $M(\pm SEM)=10.56\pm 1.04$ ;  $U=1974.50$ ,  $p<0.001$ ;  $M$  – Mean,  $SEM$  – Standard Error of the Mean].

435 When analyzing each of the stress-related behaviors separately, we found that dogs from  
436 Group Aversive showed a significantly higher frequency of body turn ( $U=1833.50$ ,  $p<0.001$ ),  
437 move away ( $U=1308.50$ ,  $p=0.042$ ), crouch ( $U=1630.50$ ,  $p<0.001$ ), salivating ( $U=1211.50$ ,  
438  $p=0.019$ ), yawn ( $U=1555.50$ ,  $p<0.001$ ) and lip lick ( $U=2004$ ,  $p<0.001$ ) than dogs from Group  
439 Reward. Additionally, there was a tendency for dogs from Group Aversive to exhibit more lying  
440 on side/back ( $U=1134$ ,  $p=0.062$ ), yelp ( $U=1174$ ,  $p=0.072$ ) and paw-lift ( $U=1258.50$ ,  $p=0.094$ )  
441 behaviors. Finally, there were no differences regarding the average frequency of whine  
442 ( $U=1141.50$ ,  $p=0.353$ ), body shake ( $U=1194.50$ ,  $p=0.247$ ), and scratch ( $U=972$ ,  $p=0.482$ )  
443 behaviors. Fear-related elimination was never displayed during this study (Figure 2).

444

445 Figure 2. Number of occurrences of each stress-related behavior averaged across the three training  
446 sessions for Group Reward (filled bars) and Group Aversive (empty bars). Vertical bars show the  
447 SEM. \* stands for statistically significant differences at  $\alpha=0.05$ .

448

449 Because we observed a difference in the frequency of aversive stimuli used during  
450 training among the different aversive-based schools (see Appendix S3), we further analyzed the  
451 relationship between the number of stress-related behaviors displayed by the dogs and the number  
452 of aversive stimuli used in the different schools. We found a strong positive correlation between  
453 the number of aversive stimuli used and the total number of stress-related behaviors shown by the  
454 animals (Spearman correlation coefficient,  $r_s=0.833$ ,  $p<0.001$ ).

455 Moreover, when examining the correlation between the frequency of each individual  
456 stress-related behavior and the number of aversive stimuli used by each training school, we found  
457 positive correlations for body turn ( $r_s=0.653$ ,  $p<0.001$ ), move away ( $r_s=0.200$ ,  $p=0.028$ ), crouch  
458 ( $r_s=0.567$ ,  $p<0.001$ ), lying on side/back ( $r_s=0.273$ ,  $p=0.004$ ), salivating ( $r_s=0.334$ ,  $p=0.001$ ), yawn

459 ( $r_s=0.533$ ,  $p<0.001$ ), paw lift ( $r_s=0.224$ ,  $p=0.016$ ) and lip lick ( $r_s=0.850$ ,  $p<0.001$ ) behaviors. For  
460 the remaining behaviors there were no significant correlations (yelp:  $r_s=0.107$ ,  $p=0.155$ ; whine:  
461  $r_s=0.111$ ,  $p=0.147$ ; body shake:  $r_s=0.127$ ,  $p=0.114$ ; scratch:  $r_s=-0.151$ ,  $p=0.075$ ).

462

### 463 3.2.1.2. Overall behavioral state

464 For each training session of each dog, we calculated the percentage of scans spent in each  
465 overall behavioral state (Tense, Low, Relaxed, Excited). A Friedman's ANOVA conducted on  
466 the percentage of scans spent in each behavioral state across the three training sessions revealed  
467 that there was a significant effect of session in the percentage of scans in the Excited state  
468 [ $\chi^2(2)=12.105$ ,  $p=0.002$ ], but not in the remaining states [Tense:  $\chi^2(2)=3.645$ ,  $p=0.162$ ; Low:  
469  $\chi^2(2)=1.077$ ,  $p=0.584$ ; Relaxed:  $\chi^2(2)=4.244$ ,  $p=0.120$ ]. Specifically, the percentage of scans in  
470 the Excited state decreased across training sessions - Session 1:  $M(\pm SEM)=50.21\pm 3.51$ ,  
471 Session 2:  $M(\pm SEM)=42.94\pm 3.59$ , Session 3:  $M(\pm SEM)=41.13\pm 3.55$ . When analyzing the two  
472 training groups separately, we found that this pattern was statistically significant for Group  
473 Aversive [ $\chi^2(2)=9.490$ ,  $p=0.009$ ] but not for Group Reward [ $\chi^2(2)=3.674$ ,  $p=0.159$ ].

474 When comparing between training groups, dogs from Group Aversive showed a  
475 significantly higher percentage of scans in Tense state than dogs from Group Reward for all three  
476 training sessions [Session 1:  $U=1769$ ,  $p<0.001$ ; Session 2:  $U=1820.50$ ,  $p<0.001$ ; Session 3:  
477  $U=1592$ ,  $p<0.001$ ] and a significantly higher percentage of scans in Low states for Session 1 and  
478 Session 2 [Session 1:  $U=1239$ ,  $p=0.002$ ; Session 2:  $U=1176.50$ ,  $p=0.011$ ; Session 3:  $U=925$ ,  
479  $p=0.211$ ]. Additionally, dogs from Group Aversive showed a significantly lower percentage of  
480 scans in Excited [Session 1:  $U=296.500$ ,  $p<0.001$ ; Session 2:  $U=287.500$ ,  $p<0.001$ ; Session 3:  
481  $U=157$ ,  $p<0.001$ ] and Relaxed [Session 1:  $U=628$ ,  $p<0.001$ ; Session 2:  $U=793$ ,  $p=0.042$ ; Session  
482 3:  $U=525$ ,  $p=0.001$ ] states for all three training sessions than dogs from Group Reward (see  
483 Figure 3).

484

485 Figure 3. Average percentage of scans in the different behavioral states in training sessions 1 (S1),  
486 2 (S2) and 3 (S3) for Group Reward (left) and Group Aversive (right).

487

488 When averaging the results of all three training sessions, dogs from Group Aversive  
489 showed a significantly higher percentage of scans in Tense [Group Aversive:  
490  $M(\pm SEM)=40.50\pm 3.07$  vs Group Reward:  $M(\pm SEM)=4.18\pm 0.74$ ;  $U=2010.50$ ,  $p<0.001$ ] and Low  
491 [Group Aversive:  $M(\pm SEM)=2.18\pm 0.63$  vs Group Reward:  $M(\pm SEM)=0.16\pm 0.12$ ;  $U=1138$ ,  
492  $p<0.001$ ] states and a lower percentage of scans in Relaxed [Group Aversive:  
493  $M(\pm SEM)=10.03\pm 1.95$  vs Group Reward:  $M(\pm SEM)=16.24\pm 2.06$ ;  $U=652$ ,  $p=0.002$ ] and Excited  
494 [Group Aversive:  $M(\pm SEM)=24.92\pm 3.27$  vs Group Reward:  $M(\pm SEM)=68.43\pm 2.81$ ;  $U=170$ ,  
495  $p<0.001$ ] states than dogs from Group Reward.

496 Finally, when examining the correlation between the percentage of scans in each  
497 behavioral state and the number of aversive stimuli used in training, we found positive  
498 correlations for Tense ( $r_s=0.881$ ,  $p<0.001$ ) and Low ( $r_s=0.472$ ,  $p<0.001$ ) states, and negative  
499 correlations for Relaxed ( $r_s=-0.472$ ,  $p<0.001$ ) and Excited ( $r_s=-0.821$ ,  $p<0.001$ ) states.

500

### 501 3.2.1.3. Panting

502 For each training session of each dog, we also calculated the percentage of scans spent  
503 panting. A significant effect of training session was found for Group Reward [ $\chi^2(2)=7.043$ ,  
504  $p=0.030$ ], but not for Group Aversive [ $\chi^2(2)=0.294$ ,  $p=0.863$ ]. However, there was no systematic  
505 increase or decrease in panting across sessions for Group Reward; instead, panting increased from  
506 Session 1 to Session 2 [Session 1:  $M(\pm SEM)=13.93\pm 2.91$  vs Session 2:  $M(\pm SEM)=20.66\pm 3.99$ ;  
507  $T=234$ ,  $p=0.054$ ] and then decreased slightly, although not significantly, from Session 2 to  
508 Session 3 [Session 2:  $M(\pm SEM)=20.66\pm 3.99$  vs Session 3:  $M(\pm SEM)=16.17\pm 3.22$ ;  $T=116.50$ ,  
509  $p=0.213$ ]. When comparing between training groups, dogs from Group Aversive panted more

510 than dogs from Group Reward during all training sessions (Session 1:  $U=1516.500$ ,  $p<0.001$ ;  
511 Session 2:  $U=1359$ ,  $p=0.008$ ; Session 3:  $U=1223.500$ ,  $p=0.001$ , see Figure 4).

512

513 Figure 4. Average percentage of scans spent panting in training sessions 1 (S1), 2 (S2) and 3 (S3)  
514 for Group Reward (left) and Group Aversive (right).

515

516 When averaging the results of the three training sessions, we also found that dogs from  
517 Group Aversive panted more overall than dogs from Group Reward [Group Aversive:  
518  $M(\pm SEM)=37.71\pm 3.56$  vs Group Reward:  $M(\pm SEM)=16.99\pm 2.71$ ;  $U=1540$ ,  $p<0.001$ ].

519 Finally, the percentage of scans spent panting was positively correlated with the number  
520 of aversive stimuli used in training ( $r_s=0.407$ ,  $p<0.001$ ).

521

### 522 3.2.2. Physiological data

523 In order to investigate potential changes in salivary cortisol concentration as a result of  
524 training methods, we averaged the baseline sample values (BL) and the post-training sample  
525 values (PT). Afterwards, we computed the difference between the average post-training  
526 concentration and the average baseline concentration. The average results for each training  
527 method are depicted in Figure 5. There was a significant difference between groups, with dogs  
528 from Group Aversive showing an average increase of  $0.10 \mu\text{g/dL}$  in salivary cortisol  
529 concentration after training and dogs from Group Reward showing, on average, no changes in  
530 cortisol concentration ( $U=196$ ,  $p=0.002$ ).

531

532 Figure 5. Average difference in cortisol concentration (PT: post-training average concentration,  
533 BL: baseline average concentration) for Group Aversive and Group Reward. Vertical bars show  
534 the SEM.

535

536 Additional analysis within each group revealed that the increase in cortisol concentration  
537 for Group Aversive was statistically significant [Baseline:  $M(\pm SEM)=0.14\pm 0.01$  vs Post-training:  
538  $M(\pm SEM)=0.24\pm 0.03$   $\mu$ /dL;  $T=134$ ,  $p=0.001$ ], whereas no significant differences existed for  
539 Group Reward [Baseline:  $M(\pm SEM)=0.13\pm 0.02$  vs Post-training:  $M(\pm SEM)=0.13\pm 0.02$   $\mu$ /dL;  
540  $T=51$ ,  $p=0.609$ ]. Moreover, whereas there were no differences between groups regarding baseline  
541 levels ( $U=151.50$ ,  $p=0.216$ ), the two groups differed significantly for post-training levels  
542 ( $U=192.50$ ,  $p=0.003$ ).

543 The average difference in cortisol concentrations also showed a positive correlation with  
544 the number of aversive stimuli used in training ( $r_s=0.512$ ,  $p<0.002$ ).

545

### 546 3.3. Phase 2 – Evaluating welfare outside the training context

547 Before performing the cognitive bias task, dogs from Group Reward had attended, on  
548 average, 6.07 training classes ( $SEM=0.36$ ) and dogs from Group Aversive had attended, on  
549 average, 6.66 ( $SEM=0.39$ ), with no statistically significant differences observed between groups  
550 ( $U=1232.50$ ,  $p=0.147$ ).

551

#### 552 3.3.1. Training phase

553 On average, dogs took 27.14 ( $SEM=0.85$ ) trials to reach the learning criterion. Dogs from  
554 the Group Reward took significantly fewer trials to reach the learning criterion than dogs from  
555 Group Aversive [Group Reward:  $M(\pm SEM)=24.80\pm 1.26$  vs Group Aversive:  
556  $M(\pm SEM)=29.10\pm 1.08$ ,  $t=-2.612$ ,  $p=0.011$ ].

557

#### 558 3.3.2. Test phase

559 Figure 6 shows the average adjusted latencies for the two training stimuli (P, N) and the  
560 three test stimuli (NP, M, NN) for Group Reward and Group Aversive. As noted in the figure,  
561 whereas there were no significant differences between groups for either the NP ( $U=755.50$ ,  
562  $p=0.834$ ) or the NN ( $U=874.50$ ,  $p=0.154$ ) stimuli, there was a statistically significant difference  
563 for the M stimulus, with dogs from Group Aversive taking longer to approach this bowl location  
564 than dogs from Group Reward ( $U=987$ ,  $p=0.01$ ).

565

566 Figure 6. Average adjusted latency to reach the food bowl as a function of location: P - 'positive',  
567 NP - 'near positive', M - 'middle', NN - 'near negative', N - 'negative', for Group Reward  
568 (black circles) and Group Aversive group (white circles). Vertical bars show the SEM. \* stands  
569 for statistically significant differences at  $\alpha=0.017$ .

570

571 Additionally, we found a positive correlation between the number of aversive stimuli used  
572 in training and the adjusted latency for the M stimulus ( $r_s=0.258$ ,  $p=0.012$ ). No correlations were  
573 found between the number of aversive stimuli used in training and the adjusted latencies for the  
574 NP ( $r_s=-0.013$ ,  $p=0.454$ ) and the NN ( $r_s=0.076$ ,  $p=0.256$ ) stimuli.

575 Lastly, an analysis comparing the average latency to reach the P location during test trials  
576 and the latency to reach this same location during the final trial (when the bowl contained no  
577 food) revealed no significant differences ( $T=1295.50$ ,  $p=0.328$ ), confirming that the dogs were  
578 not relying on olfactory or visual cues to discriminate between bowl locations.

579

#### 580 4. Discussion

581 This was the first empirical study to systematically investigate the short- and long-term  
582 effects of aversive- and reward-based training methods on the welfare of companion dogs. We  
583 objectively classified training methods, extended the study of aversive-based methods to other

584 techniques and tools besides shock collars, and used objective and validated measures for the  
585 assessment of both the short- (behavioral and physiological stress responses during training) and  
586 long-term welfare (cognitive bias task outside the training context) of companion dogs. Overall,  
587 our results showed that dogs trained with aversive-based methods displayed poorer indicators of  
588 both short- and long-term welfare as compared to those trained using reward-based methods.

589         Short-term welfare was assessed during training sessions. Here, dogs from Group  
590 Aversive spent more time in Tense and Low behavioral states as well as more time panting than  
591 dogs from Group Reward. Tense and low body postures reflect states of stress and fear in dogs  
592 (e.g., [26]), and panting has also been associated with acute stress in dogs (e.g., [8, 27]). Dogs  
593 from Group Aversive also displayed more stress-related behaviors than dogs from Group Reward.  
594 More specifically, these dogs exhibited more lip licks, yawns, body turns, moves away, crouches  
595 and instances of salivating. There was also a tendency for dogs from Group Aversive to engage  
596 in more yelping, paw-lifting and lying on side/back. In previous studies, high levels of lip licking  
597 and yawning behaviors have been consistently associated with acute stress in dogs (e.g., [9, 23]).  
598 Importantly, lip licking has been associated with stressful social situations [23]. This most likely  
599 explains the large magnitude of this behavior observed in Group Aversive, as aversive-based  
600 training methods comprise social and physical confrontation with the dog. Paw lifting, salivating  
601 and yelping have previously been interpreted as responses to pain and stress (e.g., [23, 25, 28]).  
602 The display of avoidance behaviors such as body turn, move away, crouch and lying on side/back,  
603 specifically in response to training techniques, highlights the aversive nature of the training  
604 sessions at the aversive-based schools. Notably, lying on side/back was only displayed in  
605 aversive-based schools (and mostly in School A, the school employing the highest frequency of  
606 aversive stimuli). Finally, no differences were found between groups for body shake, scratch and  
607 whine. Previous studies on dog training methods have also failed to identify significant  
608 differences regarding these behaviors [8, 9], suggesting that these behaviors may not be reliable  
609 indicators of stress, at least in the context of training. In support of this view, whining has also  
610 been associated with social solicitation, attention seeking and food begging behavior in dogs [29,



611 30]. It is possible that body shaking and scratching may also be associated with excitement and  
612 arousal rather than ‘negative’ stress. Hence, the present study shows a strong association between  
613 the use of aversive-based training methods and an increased frequency of stress-related behaviors  
614 during companion dog training. These results strengthen and extend the findings of previous  
615 studies on companion dogs, which suggested a positive correlation between the use of both shock  
616 collars [8] and other negative reinforcement techniques [9], and stress behaviors in the context of  
617 dog training.

618 With regards to physiological measures of stress, whereas Group Reward showed no  
619 differences in the concentration of salivary cortisol between baseline and post-training samples,  
620 dogs from Group Aversive exhibited a statistically significant increase in the concentration of  
621 salivary cortisol in post-training samples. Previous studies investigating cortisol levels in dogs in  
622 relation to training have yielded contradictory results. Schalke et al (2007) [31] found significant  
623 differences in the cortisol levels of three groups of laboratory dogs trained using shock collars  
624 with different degrees of shock predictability (the lower the predictability, the higher the cortisol  
625 levels). However, studies comparing aversive- and reward-based training methods have found  
626 either no significant differences or the opposite pattern. Namely, [8] and [32] found no elevation  
627 in cortisol levels after the use of shock collars and lemon-spray bark collars when compared to  
628 control groups, and [33] found that a negative punishment training method (a quitting signal)  
629 resulted in higher levels of cortisol than the use of a pinch collar (aversive-based technique).  
630 Hence, the present study is the first to report a significant increase in cortisol levels in dogs trained  
631 with aversive-based methods as compared to dogs trained with reward-based methods.

632 Meanwhile, the increase in cortisol levels observed in the present study ( $M=0.10 \mu\text{g/dL}$ )  
633 was smaller than that reported in other studies that found significant increases after dogs were  
634 exposed to aversive stimuli ( $0.20\text{-}0.30 \mu\text{g/dL}$  in [31] and  $0.30\text{-}0.40 \mu\text{g/dL}$  in [23]). One possible  
635 explanation for this difference in magnitude may be related to the nature of the stimuli used in the  
636 different studies. Whereas the reported elevations in cortisol in [31] and [23] appeared after the  
637 presentation of non-social stimuli (shocks in [31], and shocks, sound blasts and a falling bag in

638 [23]), the stimuli used during training in the present study were mainly of a social nature (i.e.:  
639 leash jerks, physical manipulation or yelling at the dog). Stimuli administered in a social context  
640 may be more predictable or better anticipated and, therefore, generate less acute stress responses  
641 [23]. In support of this view, [23] did not find elevations in cortisol after the presentation of social  
642 stimuli (physical restraint and opening an umbrella).

643         When considering long-term welfare, the present results were also in accordance with our  
644 hypothesis. Dogs from Group Aversive displayed a more ‘pessimistic’ judgment of the ‘middle’  
645 ambiguous test location in the cognitive bias task, revealing less positive underlying affective  
646 states (and hence poorer welfare) than the dogs from Group Reward. The only other study that,  
647 to our knowledge, addressed the long-term welfare effects of training methods in dogs was  
648 performed by Christiansen et al (2001). In this study, hunting dogs were trained not to attack  
649 sheep using shock collars. No general effect of the use of shock collars on dog fear and anxiety  
650 was found one year after training took place. However, unlike the test used by [12], which was a  
651 modified version of a temperament test used by the Norwegian Kennel Club, the cognitive bias  
652 approach used in the current study is a widely established and well-validated method for  
653 evaluating animal welfare (e.g., [17, 18]). Hence, to our knowledge, this is the first study to  
654 reliably assess and report the effects of aversive- and reward-based training methods in the long-  
655 term affective states of dogs.

656         Additionally, dogs from Group Reward learned the cognitive bias task faster than dogs  
657 from Group Aversive. Similar findings were observed previously by [34], who found a positive  
658 correlation between the reported use of reward-based training methods and a dog’s ability to learn  
659 a novel task (touching a spoon with the nose). In another study, [35] found that dogs with high-  
660 level training experience were more successful at opening a box to obtain food than dogs that  
661 have received either none or only basic training. Although the authors reported that all subjects’  
662 training included positive reinforcement methods, they did not specify whether positive  
663 punishment and/or negative reinforcement were used in combination. Altogether, previous  
664 research suggests that training using positive reinforcement may improve the learning ability of

665 dogs. However, in all previous studies cited above, animals were required to perform a given  
666 behavior in order to obtain a positive reinforcer. It is unclear whether the same effect would stand  
667 if the dogs had to learn a task whose goal was, for example, to perform a behavior to escape from  
668 an unpleasant situation. It may be the case that dogs trained with positive reinforcement develop  
669 a specific ‘learning set’ [36] for tasks involving positive reinforcement, but that dogs trained with  
670 aversive-based methods perform better in tasks involving some sort of aversive stimuli. Further  
671 research is needed to clarify the relationship between training methods and learning ability in  
672 dogs.

673         Notably, we found that the frequency of aversive stimuli used in training (which differed  
674 among the aversive-based schools) was correlated with all the welfare indicators measured in the  
675 present study. Specifically, we found that the higher the frequency of aversive stimuli used in  
676 training, the greater the impact on the short- and on the long-term welfare of dogs. This result is  
677 in line with the findings of a previous survey study by [7], whose results showed that a higher  
678 frequency of punishment was correlated with higher anxiety and fear scores. However, even if  
679 only comparing the schools with the lowest frequency of aversive stimuli used during training  
680 (C and F) to Group Reward, differences in welfare were found (Appendix S4). Dogs trained at  
681 schools C and F showed more stress-related behaviors, spent more time in Tense and Low  
682 behavioral states and more time panting during training than dogs trained at reward-based schools,  
683 and they also took more trials to learn the cognitive bias task. However, the difference in cortisol  
684 concentration between post-training and baseline samples was only marginally significant when  
685 dogs from schools C and F were compared with dogs trained with reward-based methods, and no  
686 differences were found in the latencies to reach the ambiguous stimuli in the cognitive bias task.  
687 This seems to suggest that, although dogs trained in ‘least aversive’ schools show poorer  
688 indicators of welfare than dogs trained in reward-based schools, these are not as poor as those of  
689 dogs trained in ‘most aversive’ schools. Notably, schools C and F also used positive reinforcement  
690 frequently during their training sessions alongside aversive-based methods, which was rarely the  
691 case in schools A and D. Nonetheless, caution is required when interpreting these results due to  
692 the reduced sample size for analysis of only two training schools. Future studies should further

693 compare the effects on dog welfare of reward-based training and what can be called ‘balanced’  
694 training, or the use of reward-based methods alongside aversive-based methods. Despite the fact  
695 that the results of the present study show that reward-based methods are better for dog welfare,  
696 the efficacy of training should also be examined. Presently, there is a lack of scientific evidence  
697 regarding the efficacy of different training methods (see also [3]), and thus further research in the  
698 topic is required. If reward-based methods are, as the current results show, better for dog welfare  
699 than aversive-based methods, and also prove to be more effective or equally effective to  
700 aversive-based methods, there is no doubt that owners and dog professionals should use  
701 reward-based training practices. If, on the other hand, aversive-based methods prove to be more  
702 effective, we would advise using aversive stimuli as infrequently as possible during training, and  
703 use them in combination with reward-based techniques, due to the implications for dog welfare.

704         Ultimately, some limitations of the present study must be considered. Firstly, because this  
705 was an empirical rather than an experimental study, we cannot infer a true causal relationship  
706 between training methods and dog welfare. To do so would require a randomized control trial.  
707 Because we did not randomly allocate dogs to the two treatments (training methods), we cannot  
708 discard the possibility that dogs enrolling training at aversive-based schools had higher stress  
709 levels *a priori*, or that there are other significant differences between dog-owner pairs that lead  
710 some owners to choose an aversive-based school and others to choose a reward-based school. The  
711 fact that the baseline levels of cortisol did not differ between groups weakens the former  
712 possibility. Regarding the latter, however, we did indeed find differences between groups in  
713 owner sex, previous experience with dogs and in the information they relied on for choosing the  
714 dog training school. However, conducting an experimental study would raise ethical concerns  
715 (but see [8]), as previous studies have already suggested an association between the use of  
716 aversive-based methods and indicators of stress in dogs (see [3] for a review), as well as with the  
717 quality of dog-owner attachment [37]. Secondly, we need to consider the possibility for a  
718 volunteer bias and hence any generalization of the present results must take this in account.

719

## 720 5. Conclusions

721 Our results show that companion dogs trained using aversive-based methods experienced  
722 poorer welfare as compared to companion dogs trained using reward-based methods, at both the  
723 short- and the long-term level. Specifically, dogs attending schools using aversive-based methods  
724 displayed more stress-related behaviors and body postures during training, higher elevations in  
725 cortisol levels after training, and were more ‘pessimistic’ in a cognitive bias task than dogs  
726 attending schools using reward-based methods. Moreover, we found that the higher the frequency  
727 of aversive stimuli used in training, the greater the impact on the short- and the long-term welfare  
728 of dogs. To our knowledge, this is the first comprehensive and systematic study to evaluate and  
729 report the effects of dog training methods on companion dog welfare. Critically, our study points  
730 to the fact that the welfare of companion dogs trained with aversive-based methods appears to be  
731 at risk.

732

## 733 Acknowledgements

734 We are grateful, first and foremost, to all dogs and their owners who participated in this  
735 study; without them this research would never have been possible. A very special  
736 acknowledgment to the dog training schools and their trainers that opened their doors for our  
737 participant recruitment and data collection.

738 We would also like to thank Joana Guilherme-Fernandes for the support provided in the  
739 development of the setup for the cognitive bias task, for helping with data collection and  
740 especially for all the invaluable discussions during study planning and data interpretation. A  
741 special acknowledgment also for Margarida Lencastre and Flávia Canastra, who also helped in  
742 data collection. Finally, we want to thank Jennifer Barrett for input given during data collection  
743 and analysis.

744 The current research study was supported by FCT - Fundação Portuguesa para a Ciência  
745 e Tecnologia (Fellowship SFRH/BPD/111509/2015) and UFAW – Universities Federation for

746 Animal Welfare (Grant 14-16/17). Stefania Pastur was supported by PIPOL - Regione Friuli  
747 Venezia Giulia.

748

749 References

- 750 1. Reid P. Learning in dogs. In: Jensen P, editor. *The Behavioural Biology of Dogs*. CABI;  
751 2007. pp. 120–144.
- 752 2. Reisner I. The learning dog: A discussion of training methods. In: Serpell J, editor. *The*  
753 *Domestic Dog: Its Evolution, Behavior and Interactions with People*, 2nd Edition. Cambridge  
754 University Press; 2017. pp. 211-226.
- 755 3. Guilherme-Fernandes J, Olsson IAS, Vieira de Castro AC. Do aversive-based training  
756 methods actually compromise dog welfare?: A literature review. *Appl Anim Behav Sci*. 2017;  
757 196, 1–12.
- 758 4. Hiby EF, Rooney NJ, Bradshaw JWS. Dog training methods: their use, effectiveness and  
759 interaction with behaviour and welfare. *Anim Welf*. 2004; 13, 63–69.
- 760 5. Blackwell EJ, Twells C, Seawright A, Casey RA. The relationship between training methods  
761 and the occurrence of behavior problems, as reported by owners, in a population of domestic  
762 dogs. *J Vet Behav*. 2008; 3: 207–217.
- 763 6. Herron ME, Shofer FS, Reisner IR. Survey of the use and outcome of confrontational and  
764 non-confrontational training methods in client-owned dogs showing undesired behaviors.  
765 *Appl Anim Behav Sci*. 2009; 117, 47–54.
- 766 7. Arhant C, Bubna-Littitz H, Bartles A, Futschik A, Troxler J. Behaviour of smaller and larger  
767 dogs: effects of training methods, inconsistency of owner behavior and level of engagement  
768 in activities with the dog. *Appl Anim Behav Sci*. 2010; 123: 131–142.
- 769 8. Cooper JJ, Cracknell N, Hardiman J, Wright H, Mills D. The welfare consequences and  
770 efficacy of training pet dogs with remote electronic training collars in comparison to reward  
771 based training. *PLoS One*. 2014; 9, e102722.

- 772 9. Deldalle S, Gaunet F. Effects of 2 training methods on stress-related behaviors of the dog  
773 (Canis familiaris) and on the dog-owner relationship. *J Vet Behav.* 2014; 9, 58–65.
- 774 10. Companion Animal Welfare Council, 2012. The Use of Electric Pulse Training Aids (EPTAs)  
775 in Companion Animals. Available from:  
776 <http://eprints.lincoln.ac.uk/14640/1/CAWC%20ecollar%20report.pdf> (Accessed 12 August  
777 2019).
- 778 11. Haverbeke A, Laporte B, Depiereux E, Giffroy JM, Diederich C. Training methods of military  
779 dog handlers and their effects on the team’s performances. *Appl Anim Behav Sci.* 2008; 113,  
780 110–122.
- 781 12. Christiansen FO, Bakken M, Braastad BO. Behavioural changes and aversive conditioning in  
782 hunting dogs by the second-year confrontation with domestic sheep. *Appl Anim Behav Sci.*  
783 2001; 72, 131–143.
- 784 13. Wiepkema PR, Koolhaas JM. Stress and animal welfare. *Anim Welfare.* 1993; 2: 195–218.
- 785 14. Conrad CD. A critical review of chronic stress effects on spatial learning and memory. *Prog*  
786 *Neuro-Psychoph.* 2010; 34: 742–755.
- 787 15. Destrez A, Deiss V, Lévy F, Calandreau L, Lee C, Chaillou-Sagon E. Chronic stress induces  
788 pessimistic- like judgment and learning deficits in sheep. *Appl Anim Behav Sci.* 2013; 148:  
789 28–36.
- 790 16. Harding EJ, Paul ES, Mendl M. Animal behaviour: cognitive bias and affective state. *Nature.*  
791 2004; 427: 312–312. PMID: 14737158.
- 792 17. Boissy A, Erhard HW. How studying interaction between animal emotion, cognition, and  
793 personality can contribute to improve farm animal welfare. In: Grandin T, Deesing MJ,  
794 editors. *Genetics and the Behavior of Domestic Animals*, 2nd Edition. Academic Press. 2014.  
795 pp. 81–113.
- 796 18. Mendl M, Burman OHP, Parker RMA, Paul ES. Cognitive bias as an indicator of animal  
797 emotion and welfare: emerging evidence and underlying mechanisms. *Appl Anim Behav Sci.*  
798 2009; 118, 161–181.

- 799 19. Mendl M., Brooks J, Basse C, Burman O, Paul E, Blackwell E, Casey R. Dogs showing  
800 separation-related behaviour exhibit a ‘pessimistic’ cognitive bias’. *Curr Biol.* 2010; 20,  
801 R839 - R840.
- 802 20. Barker SB, Wolen AR. The benefits of human-companion animal interaction: a review. *J Vet*  
803 *Med Educ.* 2008; 35(4): 487–495.
- 804 21. Crawford EK, Worsham NL, Swinehart ER. Benefits derived from companion animals, and  
805 the use of the term “attachment”. *Anthrozoos.* 2006; 19 (2), 98-112
- 806 22. Dreschel NA, Granger DA. Methods of collection for salivary cortisol measurement in dogs.  
807 *Horm Behav.* 2009; 55(1): 163-188.
- 808 23. Beerda B, Schilder MBH, van Hooff JARAM, de Vries HW, Mol JA. Behavioural, saliva  
809 cortisol and heart rate responses to different types of stimuli in dogs. *Appl Anim Behav Sci.*  
810 1998; 58, 365–381.
- 811 24. Kobelt A, Hemsworth PH, Barnett JL, Butler KL. Sources of sampling variation in saliva  
812 cortisol in dogs. *Res Vet Sci.* 2003; 75: 157–161.
- 813 25. Schilder MBH, van der Borg JAM. Training dogs with help of the shock collar: short and  
814 long term behavioural effects. *Appl Anim Behav Sci.* 2004; 85: 319-334.
- 815 26. Beerda B, Schilder MBH, van Hooff JARAM, de Vries HW. Manifestations of chronic and  
816 acute stress in dogs. *Appl Anim Behav Sci.* 1997; 52:307–319.
- 817 27. Voith VL, Borchelt PL. Fears and phobias in companion animals. In: Voith VL, Borchelt PL,  
818 editors. *Readings in companion animal behaviour.* Trenton, NJ: Veterinary Learning  
819 Systems; 1996. pp. 140–152.
- 820 28. De Palma C, Viggiano E, Barillari E, Palme R, Dufour AB, Fantini C, Natoli E. Evaluating  
821 the temperament in shelter dogs. *Behaviour.* 2005; 142: 1307-1328.
- 822 29. Mills DS, Beral A, Lawson S. Attention seeking behavior in dogs—what owners love and  
823 loathe! *J Vet Behav.* 2010; 5: 60.
- 824 30. Pongrácz P, Molnár C, Miklósi A, Csanyi V. Human listeners are able to classify dog (*Canis*  
825 *familiaris*) barks recorded in different situations. *J Comp Psy.* 2005; 119: 136–144.



- 826 31. Schalke E, Stichnoth J, Ott S, Jones-Baade R. Clinical signs caused by the use of electric  
827 training collars on dogs in everyday life situations. *Appl Anim Behav Sci.* 2007; 105: 369–  
828 380.
- 829 32. Steiss JE, Schaffer C, Ahmad HA, Voith VL. Evaluation of plasma cortisol levels and  
830 behavior in dogs wearing bark control collars. *Appl Anim Behav Sci.* 2007; 106, 96–106.
- 831 33. Salgirli Y, Schalke E, Hackbarth H. Comparison of learning effects and stress between 3  
832 different training methods (electronic training collar, pinch collar and quitting signal) in  
833 Belgian Malinois Police Dogs. *Rev Méd Vét.* 2012; 163, 530–535.
- 834 34. Rooney NJ, Cowan S. Training methods and owner-dog interactions: links with dog  
835 behaviour and learning ability. *Appl Anim Behav Sci.* 2011; 132, 169–177.
- 836 35. Marshall-Pescini S, Valsecchi P, Petak I, Accorsi PA, Prato-Previde E. Does training make  
837 you smarter? The effects of training on dogs' performance (*Canis familiaris*) in a problem  
838 solving task. *Behav Process.* 2008; 78: 449-454.
- 839 36. Harlow HF. The formation of learning sets. *Psychol Rev.* 1949; 56: 51-65.
- 840 37. Vieira de Castro AC, Barrett J, de Sousa L, Olsson IAS. Carrots versus stick: The relationship  
841 between training methods and dog-owner bond. *Appl Anim Behav Sci.* 2019; 219: 104831.

842

843 Supporting information captions

844

845 Appendix S1. Information on training schools and training methods.

846 Table S1a. Characterization of the training schools by types of behaviors trained, training sites  
847 and class structure.

848 Table S1b. Definition of the aversive-based operant conditioning procedures used to classify the  
849 dog training schools as aversive-based or reward-based. The schools were classified as aversive-  
850 based if they used some sort of positive punishment and/or negative reinforcement and as reward-  
851 based if they did not use any of these techniques.

852 Table S1c. Frequency (mean  $\pm$  standard deviation) of positive punishment and negative  
853 reinforcement used during the four training sessions videotaped at the dog training schools.  
854 Schools A, C, D and F were classified as aversive-based and Schools B, E and G were classified  
855 as reward-based.

856

857 Appendix S2. Ethograms used for the analysis of the video recordings of dog training sessions.

858 Table S2a. Ethogram for stress-related behaviors.

859 Table S2b. Ethogram for overall behavioral state and panting.

860

861 Appendix S3. Questionnaire data and relationship with the dependent variables measured in the  
862 present study.

863 Table S3a. Variables obtained from the questionnaire (dog demographics and background, and  
864 owner demographics and experience with dogs and dog training). Chi-square tests were used to  
865 compare the two groups (Reward and Aversive).

866 Table S3b. Statistical results for the effects of the dog-related variables that differed between the  
867 Reward and the Aversive groups (age, FCI breed group and age of separation from the mother)  
868 on the different dependent variables measured in the current study. The variable age comprised  
869 five categories (< 6 months; 6 -11 months; 1-3 years; 4-7 years; >7 years), the FCI breed group  
870 variable comprised eleven (Mixed breed; Sheepdogs and Cattle dogs, except Swiss Cattle dogs;  
871 Pinscher and Schnauzer – Molossoid and Swiss Mountain and Cattle dogs; Terriers; Daschunds;  
872 Spitz and primitive types; Scent hounds and related breeds; Pointing dogs; Retrievers, Flushing  
873 Dogs and Water Dogs; Companion and Toy Dogs; Sight Hounds), and the variable age of  
874 separation from the mother comprised nine [less than 1 month; 1 – 1.5 months (inclusive); 1.5 –  
875 2 months (inclusive); 2 – 2.5 months (inclusive); 2.5 – 3 months (inclusive); 3 – 4 months  
876 (inclusive); 4 – 5 months (inclusive); more than 5 months, don't know]. Kruskal-Wallis and  
877 Mann-Whitney tests were used to compare more or less than two categories, respectively.

878 Significant differences were found for the frequency of paw-lift and percentage of scans in Tense  
879 state in Group Aversive and for the percentage of scans panting in Group Reward; however, post-  
880 hoc pairwise comparisons with adjusted Bonferroni's correction revealed no differences between  
881 the different categories.

882

883 Appendix S4. Statistical results for the comparison of the different welfare indicators between  
884 Group Reward and the two schools from Group Aversive that used the lowest frequency of  
885 aversive stimuli during training (C and F).

886

887 Appendix S5. Raw data underlying all the analyzes performed in the current research paper.

'handler'



baiting area

'timer'



start position

4 m

4 m

4 m



P



NP



M



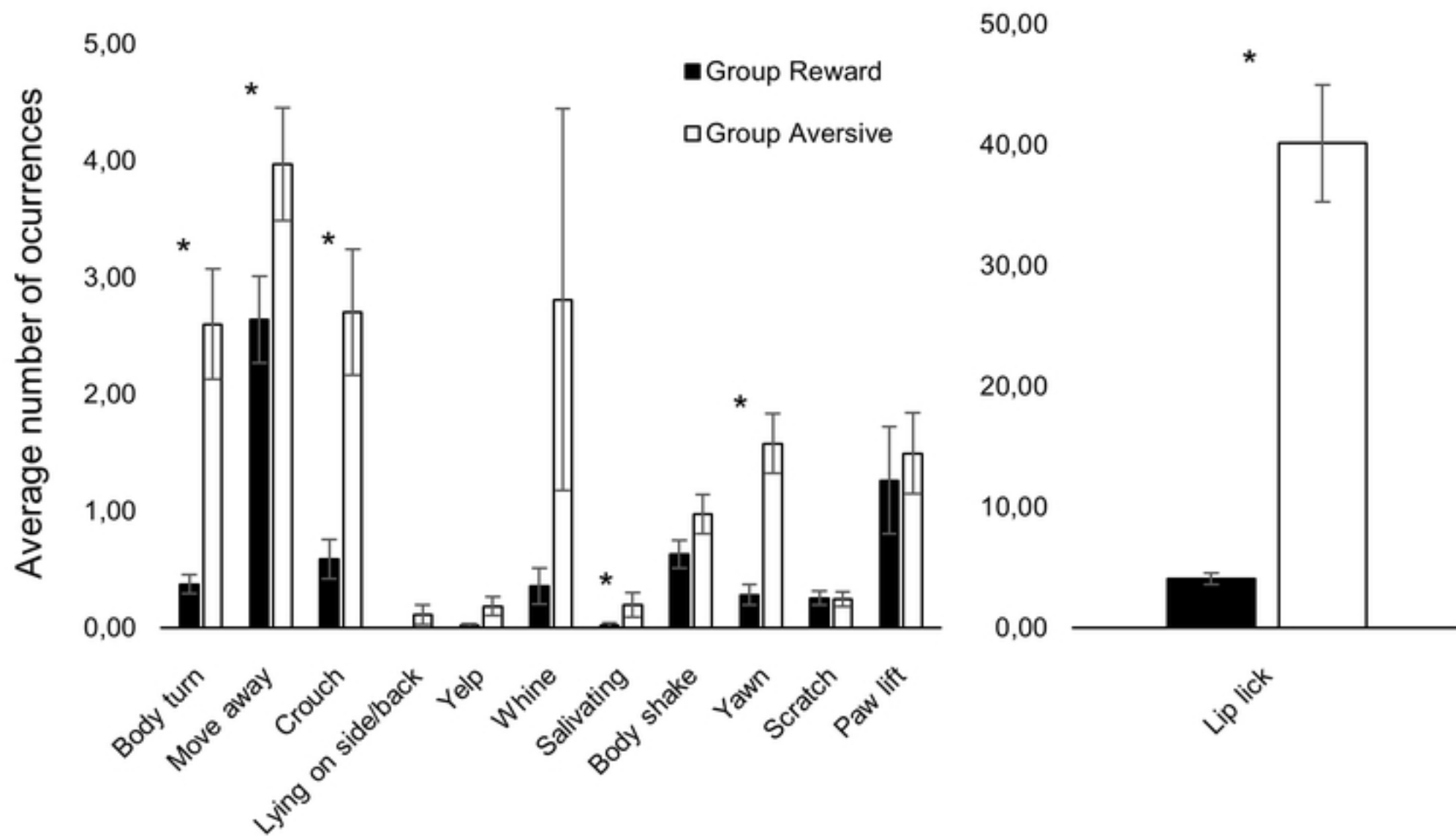
NN

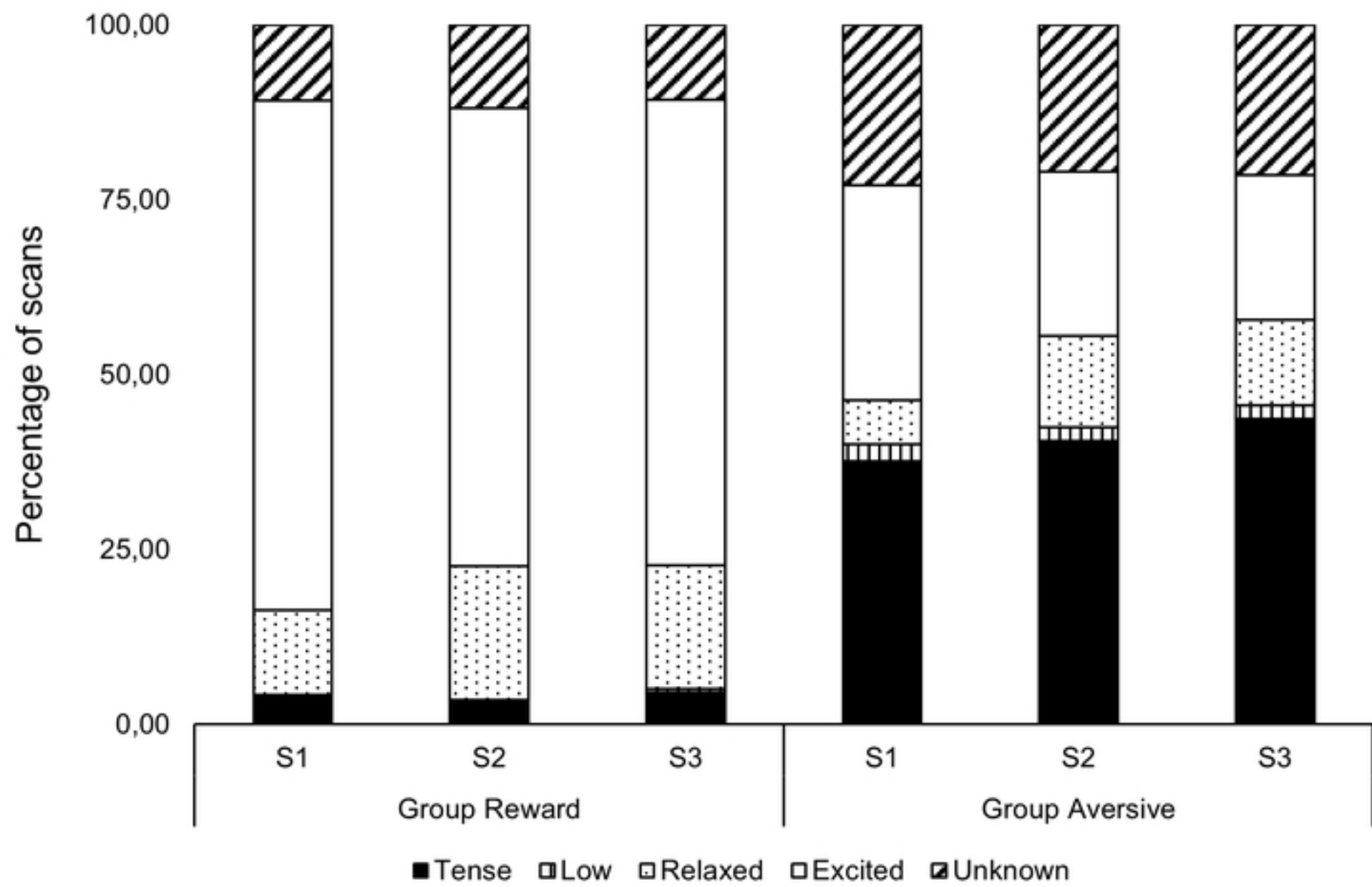


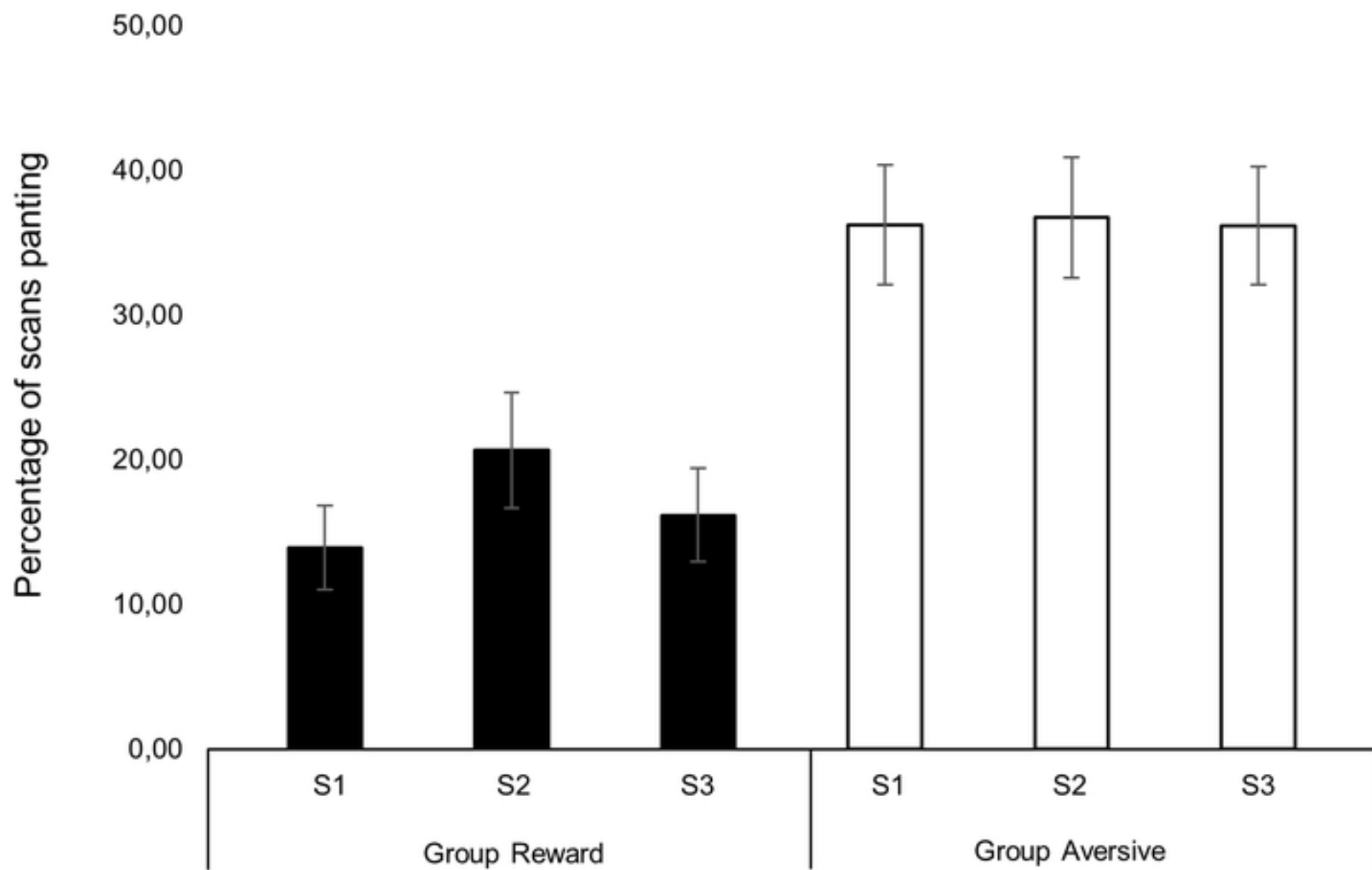
N

owner





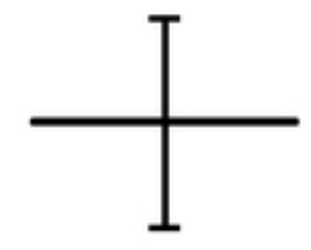




Difference in cortisol concentraion (PT-BL)

0,15  
0,10  
0,05  
0,00  
-0,05

Group Reward



Group Aversive

