I know a dog when I see one: <mark>**dogs (Canis familiaris) recognize dogs from videos**</mark>

Paolo Mongillo, Carla Eatherington, Miina Lõoke, Lieta Marinelli*

Laboratory of Applied Ethology, Department of Comparative Biomedicine and Food Science, University of Padua, Italy. Viale dell'Università 16, 35020, Legnaro.

*Correspondance to: lieta.marinelli@unipd.it (LM), +39 04941862

ORCID:

Carla J. Eatherington: 0000-0003-1209-8116

Paolo Mongillo: 0000-0001-8599-4005

Miina Lõoke: 0000-0002-3981-1551

Lieta Marinelli: 0000-0003-0609-2172

Abstract

Several aspects of dogs visual and social cognition have been explored using bidimensional representations of other dogs. It remains unclear, however, if dogs do recognize as dogs the stimuli depicted in such representations, especially with regard to videos. To test this, 32 pet dogs took part in a cross-modal violation-of-expectancy experiment, during which dogs were shown videos of either a dog and that of an unfamiliar animal, paired with either the sound of a dog barking or of an unfamiliar vocalization. While stimuli were being presented, dogs paid higher attention to the exit region of the presentation area, when the visual stimulus represented a dog than when it represented an unfamiliar species. After exposure to the stimuli, dogs' attention to different parts of the presentation area depended on the specific combination of visual and auditory stimuli. Of relevance, dogs paid less attention to the central part of the presentation area and more to the entrance area after being exposed to the barking and dog video pair, than when either was paired with an unfamiliar stimulus. This indicates dogs were surprised by the latter pairings and not by the former and interested in where the barking and dog pair came from, implying recognition of the two stimuli as belonging to a conspecific. The study represents the first demonstration that dogs can recognize other conspecifics in videos.

Kewords

Dogs, Cross-modal, Expectancy violation, Recognition, Species, Videos

Declarations

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: The study was conducted in accordance with relevant legislation for research involving animals, and according to the type of procedure used, no formal ethical approval was required.

Data availability: data are publicly available in the data repository of the University of Padua.

Authors' contributions: the study was conceived by L.M., executed by C.J.E., data was collected by C.J.E. and M.L., analysed by P.M., the manuscript was drafted by C.J.E. and P.M. and revised by L.M.

Recognition is the ability to identify an item based on previous experience or knowledge and it is crucial for animals to perform appropriate social behaviour towards known others. Recognition is an umbrella term, under which abilities with different degrees of complexity and specificity are grouped: from the univocal identification of individuals, to the relatively simpler classification into meaningful groups (Gherardi et al. 2012). Kin recognition is an example of the latter which attracted substantial attention, having been investigated in a variety of species (see Holmes and Sherman 1983; Mateo 2004), including dogs (Hepper 1994). However, a crucial form of recognition is possibly the ability to recognize individuals as belonging to one's own species, or conspecifics recognition.

A number of studies have looked at different aspects of recognition abilities in dogs. The vast majority looked specifically at visually-based recognition of cues, provided through 2D static stimuli, i.e. photographs. For instance, Adachi and collaborators (2007) showed that dogs looked longer at a picture of their owner's face when preceded by an incongruent voice, suggesting that dogs had not expected to see their owner. Eatherington and collaborators (2020) provided further evidence of individual humans' face recognition, by showing that dogs were more likely to approach a picture of their owner's face compared to that of a stranger's. Several other studies provide indications about dogs' recognition of conspecifics, intended here as the ability to identify pictorial representations of dogs as belonging to a group of animals sharing some common features, not to individually recognize other dogs. An early study (Fox 1971) showed that dogs made socially appropriate responses to a life-sized painting of a dog, spending more time sniffing certain regions of the body (e.g. ear, tail or groin). More recently, Range and collaborators (2008) showed that dogs trained to discriminate pictures of dogs could transfer such learning to novel dog pictures; similarly, Autier-Dérian and collaborators (2013) showed that dogs trained to respond to photographs of dogs faces could transfer this to other dog faces, regardless of their phenotype, when presented amongst human and other animal faces. Finally, cross-modal paradigms show that dogs appropriately match dog vocalizations and pictorial representations of dogs under various circumstances (Faragó et al. 2010; Albuquerque et al. 2016; Gergely et al. 2019). Collectively, the evidence suggests that dogs may be able to correctly recognise pictorial representations of conspecifics. However, most of the abovementioned studies compared dogs' responses to conspecifics representations to their response towards very different-looking classes of stimuli, including humans or inanimate objects. The lack of comparison with response to representation of more similar stimuli makes these suggestive evidences not conclusive to this regard.

All of the aforementioned studies employed photographs as stimuli. Although methodologically simple, and appropriate to the aims of such studies, a drawback of this approach is that it confines the assessment to dogs' responses towards static, morphological features of the stimuli being represented. To overcome this limit, animals can be presented with moving visual representations of others (i.e. videos), allowing to incorporate information about motion and, more generally, behaviour. The use of these stimuli is certainly not a novelty in the ethological field. For example, Plimpton and colleagues (1981) showed juvenile bonnet macaques videos of socially diverse behaviours performed by other macaques. They found that the juvenile macaques behaved in a socially appropriate way, acting submissive and seeking contact with their mother when viewing a threatening male, but approaching a passive female. Another demonstration was put forward by Herzog and Hopf (1986) who showed that videos of predators elicited alarm responses by squirrel monkeys, but videos of non-predators did not. The monkeys also reacted to videos of humans as if they were real people, whilst watching them prepare food or when seeing a caretaker who had recently removed a dead neonate and was therefore viewed as a threat. While these are only few examples, the use of videos would have countless applications for the study of dogs' behaviour, especially in response to social stimuli. Surprisingly, however, the field of dog behaviour has not yet seen an extensive use of videos as stimuli. Pongrácz and collaborators (2003) proved that dogs performed above chance in a classical pointing task, where they were shown a projection of an experimenter performing the pointing gesture, implying that dogs perceived the stimulus as a human being. A replica of the same paradigm, in which real-size video of dogs were projected instead of humans, represents another recent example of the use of videos in dogs' behavioural research (Balint et al. 2015). Another recent study reported dogs' differential physiological and behavioural responses to videos of dogs showing asymmetrical tail wagging associated with specific emotional states (Siniscalchi et al. 2013). The dogs' responses were coherent with such states, suggesting that dogs had recognised the video as representing a dog. Other studies used animated representations of dog motion, in the form of dot displays, not of fully informative videos, to assess dogs' reactions to the biological motion of conspecifics and of humans (Ishikawa et al. 2018; Eatherington et al. 2019). Relevant to our aim, the study by Eatherington and collaborators (2019) proved that dogs' looked longer at random dot displays depicting the motion of conspecifics, even when the dots composing the display were randomly rearranged in space, rather than at inverted manipulations of the same stimuli; the same effect was not observed when human stimuli were projected. The finding suggest that dogs are particularly attracted by representations of motion of a quadrupedal animal; however, the lack of control with a non-dog quadrupedal animal species prevents any conclusion about

89 the dogs' ability to recognise these stimuli at the species level. The same holds true for the previously

90 mentioned paper by Siniscalchi and collaborators (2013).

91   A first, necessary step towards the use of videos in the study of dogs' social behaviour is the

92 demonstration that dogs are able to recognize the stimuli being represented. Therefore, the aim of the present

93 study was to assess whether dogs are able to recognize a video representing a dog as a dog. To this aim, we

94 employed a classical cross-modal expectancy violation paradigm, where videos of dogs or of another unfamiliar

95 quadrupedal species were presented after either a dog or another unfamiliar vocalization. According to the

96 expectancy violation paradigm, a non-surprised reaction (i.e. shorter looking time to the area were the stimuli

97 appeared), when matching dog auditory and visual stimuli where presented, than in other conditions, would

98 support dogs' ability to recognize conspecifics in videos.

99

100 **METHODS**

101 *Subjects*

102 Thirty-two dogs with their owners were recruited via the database of volunteers at the Laboratory of Applied

103 Ethology in the University of Padua. Seventeen dogs were pure-breeds (1 American Staffordshire Terrier, 1

104 American Pitbull Terrier, 3 Australian Shepherds, 1 Bracco Italiano, 3 Border Collies, 1 Boxer Dog, 1 Bulldog,

105 1 Golden Retriever, 1 Labrador Retriever, 1 Maremma Sheepdog, 1 Poodle, 1 Yorkshire Terrier, 1 Miniature

106 Pincher) and 15 were mixed-breed dogs (7 small, ≤ 35 cm at the withers; 7 medium, > 35 and < 55 cm; 1 large ≥

107 55 cm). The sample consisted of 17 females and 15 males (mean age±SD: 5.2±3.2 years). Requirements for

108 recruitment were that dogs were in good health, including no apparent sight problems, and at ease in unfamiliar

109 contexts. Also, to ensure that the cow and horse videos, as well as the frog croaking were unfamiliar (see later),

110 dogs with known experience with any of such species were not permitted to take part.

111

112 *Stimuli*

113 Dogs were exposed to pairs of auditory and visual stimuli, which belonged to either a dog or to another species

114 to which subjects were unfamiliar. The dog vocalization was a recording of a barking bout, composed of two

115 barks. The unfamiliar vocalization was a frog croaking bout, composed of two croaks; such sound was chosen

116 for its similarity with the barking in terms of overall development of the dynamics and noisiness. Both

117 vocalizations had the exact same duration of 0.5 s. The sounds were presented so to produce an average sound

118 pressure of about 58 dB at the site where the dog's head was, when the sound was played.

The dog video was a black and white recording of a medium sized, mixed breed, light-coated dog walking laterally across a black, rectangular background area. The animal entered the area from one side and walked across it, taking about two and a half complete leg cycles before completely disappearing on the opposite side. The unfamiliar (non-dog) species video was also a black and white recording of either a light-coated cow or horse, walking across the black background area with the same number of strides as the dog video. The size of the animals was reduced so to match the size of the dog. Both videos had the same duration (3.0 s), from the first to the last frame in which part of the animal was visible. When projected to the presentation area, the black background area had a height of 150 cm and a width of 190 cm, whereas the animal portrayed in the video had a height of about 75 cm (from ground to the topmost part of the animal) corresponding to the actual, real-life size of the dog portrayed in the video.

Experimental setting

The experiment was conducted in a quiet, dimly lit room (see scheme in Fig. 1). Along one of the short sides, at approximately 60 cm from it, was a large white plastic screen (150 cm high, and 200 cm wide), which represented the area on which the visual stimuli were projected. Two smaller screens (150 x 100 cm) were placed at the sides of the large one and 10 cm in front of it. During the presentation of stimuli, the side screen created the impression that the animal portrayed in the video appeared from behind a wall. Behind each of the two smaller panels two active speakers (Hercules XPS 2.0, Hercules Computer Technology, CA, USA) were placed. On the opposite side of the room to the screens was a Toshiba TDP T100 projector mounted 207 cm high on a shelf on the wall. Both the projector and speakers were connected to a MacBook Air laptop (Apple Computers Inc., Cupertino, CA, USA), which was used to control the presentation of the stimuli by an experimenter sitting behind the central panel. During testing, dogs sat or stand at a distance of 240 cm from the screen, and between the legs of their owner who was seated on a small stool behind them. Owners were instructed to gently hold the dog in place and look down at their lap so as not to influence the dog's behaviour. Two CCTV cameras mounted on the ceiling captured respectively a view of the dog from behind, including the projection area, and a detailed view of the dog from straight above the dog. A Canon XA20 (Canon, Tokyo, Japan) camcorder was mounted over the top of the screen via a tripod and pointed towards the dog's face; this camera was set in infrared recording mode, allowing to clearly detect the contour of the pupils and determine eye orientation. The experimenter sitting behind the screen used this camera to see when the dog was looking forward and therefore start the trials.

150   -- FIGURE 1 ABOUT HERE --

151   **Fig. 1** A schematic representation of the experimental setting, illustrating the position of (A) the projector, (B)

152   the dog, (C) the projection screen, (D) the side screens, (E) the speakers and (F) the experimenter operating on

153   the computer during a presentation (figure elements are not in scale)

154

155   *Experimental procedure and design of the experiment*

156   At the start of each trial, dogs were led into the room by their owner and positioned, facing the screen at the

157   designated location, marked by tape on the floor. When the dog was in place, the experimenter started the

158   presentation of the stimuli, which entailed the reproduction of the vocalization from one of the two speakers,

159   and the simultaneous reproduction of the video of the animal walking in the projection area from the same side

160   the vocalization was played from, and disappearing on the opposite side. After the disappearance of the video,

161   the experimenter waited 30 s before eventually turning on the lights. During this interval the owners were

162   instructed to keep looking at their laps and not to interfere with the behaviour of the dog, except for the gentle

163   restraint. After the 30 seconds had passed, the experimenter turned on the lights, and owner and dog left the

164   room, waiting for 5 minutes before entering for the following trial.

165   All dogs underwent four trials, during which the four possible combinations of dog and non-dog auditory and

166   visual stimuli were presented. The order of presentation of the four combination was balanced within the

167   sample, so that each combination was equally as often presented as 1st, 2nd, 3rd and 4th trial. For half of the dogs

168   in the sample the visual stimulus was represented by the horse, and for another half by the cow.

169

170   *Data collection and analysis*

171   Using Observer XT software (version 12.5, Noldus, Groeningen, The Netherlands) a continuous sampling

172   technique was used to collect data about dogs' orientation, which was coded as either looking centrally (towards

173   the central part screen area), looking at the entrance side, looking at the exit side (respectively, the side of the

174   presentation area where the projected animal came in from, or left at). Data were collected in an interval of time

175   spanning the frame when a part of the animal was first visible, until 30 s after the animal had disappeared from

176   the screen.

177   For the aim of analysis, collected data were split in two different time intervals: one relative to when the

178   projected animal was visible, one relative to the 30 seconds following its disappearance. For each interval, a set

179   of four variables were obtained: the total time spent looking at the entrance, centrally, or at the exit, and at the

7

180 entire presentation area (the latter representing the sum of the first three variables). The rationale for dividing

181 data collection into two intervals, was that we expected dogs' attention to be primarily driven by the presence of

182 the stimuli while the latter were projected; conversely, after the stimuli had disappeared, dogs' attention would

183 be more indicative of possible surprised reactions to expectations induced by the pairing of stimuli. To assess

184 whether dogs' attention was indeed driven by the presence of the stimuli when these were projected, data

185 collected in such interval was further split into three equally long sub-intervals (1 s) corresponding to the

186 stimulus occupying the entrance, central and exit region of the presentation area, respectively.

187 Inter-observer reliability for dogs' head orientation data was assessed using data collected by a second observer

188 on a randomly selected subset of videos (N = 18, ~ 30% of the total number); a Pearson's correlation coefficient

189 of 0.89 was obtained between data collected by the two observers, supporting the reliability of data collection.

190 Data analysis was based on Generalised Estimating Equation (GEE) models. A first model was run to

191 assess whether dogs' overall attention to the presentation area remained stable across the 4 presentations or any

192 decrement in attention was observed. The model included the dogs' name as a random factor accounting for

193 repeated measures taken from each dog, and the order of trials (1 to 4) as a fixed factor. The dependent variable

194 was the total attention to the presentation area. Corrected post-hoc comparisons were run to assess pairwise

195 differences between trials presented at a different place in the sequence. Models were run separately for data

196 collected when the projected animal was present, and after its disappearance.

197 The next analysis assessed whether the type of visual or auditory stimulus, or their combination, had an

198 effect on dogs' orientation, either during stimulus presentation or after the stimulus had disappeared. GEE

199 models were run with the dogs' name as random factor to account for repeated measurement within each dog.

200 The model included as fixed factors the type of visual stimulus (dog, non-dog), the type of auditory stimulus

201 (bark, croak), and their interaction; to assess potential differences between the horse and cow video, the model

202 also included the effect of the type of non-familiar species (cow, horse), as a nested factor within the type of

203 visual stimulus and in interaction with the type of auditory stimulus. Furthermore, to assess whether dogs'

204 allocation of attention to different parts of the presentation area was driven by the movement of the stimulus, the

205 location of the stimulus (at entrance, central, or exit region) was also included as a fixed factor. Different

206 models were run, using as dependent variables the time spent looking centrally, at the entrance or at the exit

207 side, respectively while the stimulus was present and after its disappearance. Corrected post-hoc comparisons

208 were run to assess pairwise differences between trials presented at a different place in the sequence.

209   Analysis was performed with SPSS (ver. 26; IMB, Armonk, NY). Results are reported as mean±SD

210   unless otherwise stated.

211

212   **RESULTS**

213   During the presentation of the stimuli, dogs spent on average 2.8±0.4 s (min: 0.6, max: 3.0) oriented to the

214   presentation area, with no significant difference between trials presented in different order (Wald Chi-square =

215   4.3, P = 0.23, GEE). However, the order of trial presentation had an effect on the length of time dogs were

216   oriented to the presentation area after the stimulus had disappeared (Wald Chi-square = 17.3, P = 0.001);

217   specifically, no difference was found between the 1st (estimated mean±SD = 20.0±1.5 s) and the 2nd trial

218   (20.2±1.5 s), but the time spent looking at presentation area decreased significantly in the 3rd (17.7±1.7 s; P =

219   0.026) and the 4th trial (15.5±1.5 s; P < 0.001). To adopt a conservative approach, we therefore decided to limit

220   further analysis of dogs' orientation after the disappearance of the stimuli to data of the 1st and 2nd trials.

221   Conversely, data from all four trials were analyzed for dogs' orientation while stimuli were projected.

222   During the presentation of the stimuli, dogs spent an average of 1.6±0.8 s looking centrally, 0.7±0.7 s

223   looking at the stimulus entrance side, and 0.5±0.5 s at the stimulus exit side. Table 1 summarizes the results of

224   the GEE indicating the effects of the type of stimuli presented and of the region occupied by the stimulus on the

225   projection area on dog's orientation variables during the presentation of the stimuli. The region where the

226   stimulus was projected significantly affected all orientation variables, as shown through the heatmap in Figure

227   2. As regards the effects of the type of stimuli presented, the time spent oriented centrally was not affected by

228   either the visual or auditory stimulus. The time spent looking at the entrance side was affected by an interaction

229   between the two factors: however, after applying corrections for multiple comparisons, no significant difference

230   was found between different levels of the interaction. The time spent looking at the exit was affected by the type

231   of visual stimulus, with longer looking observed when a video of a dog was presented, than when a non-dog

232   video was presented (Figure 3).

233   -- FIG 2 ABOUT HERE --

234   **Fig. 2** Heat-map representing the percentage of time spent by dogs' oriented to different region of the

235   presentation area, as a function of the region predominantly occupied by the projected visual stimulus. Arrows

236   indicate significant differences in means (P< 0.05, Bonferroni-corrected post-hoc comparisons after Generalized

237   Linear Equation Models)

238

239   Table 1. Generalized Estimating Equations model assessing the effects of the type of visual stimulus (dog/non

240   dog), the species of the non-dog and of auditory stimulus on time spent by dogs looking centrally, at the

241   entrance side or at the exit side, during and after the presentation of the stimuli.

Factor	Looking at entrance side	Looking centrally	Looking at exit side
Region occupied by stimulus	$X^2=90.04$ P<0.001	$X^2=8.69$ P=0.013	$X^2=31.44$ P<0.001
Visual stimulus	$X^2=0.04$ P=0.843	$X^2=3.10$ P=0.078	$X^2=7.98$ P=0.005
Auditory stimulus	$X^2=1.74$ P=0.187	$X^2=1.44$ P=0.231	$X^2=0.017$ P=0.896
Auditory stimulus*Visual stimulus	$X^2=5.71$ P=0.017	$X^2=2.50$ P=0.114	$X^2=1.19$ P=0.290
Auditory stimulus*Species of non-dog visual stimulus (nested within visual stimulus)	$X^2=1.34$ P=0.501	$X^2=4.77$ P=0.098	$X^2=1.51$ P=0.468

242

243

244   -- FIG 3 ABOUT HERE --

245 **Fig. 3** Mean±SD time (s) spent looking at the different regions of the presentation area while any part of the

246 stimulus was visible on it, as a function of the stimuli pair (*P<0.05, Bonferroni-corrected post-hoc comparisons

247 after Generalized Linear Equation Models)

248

249   Table 2 summarizes the results of the GEE indicating the effects of the type of visual stimulus,

250 vocalization, and their interaction on dog's orientation variables during the presentation of the stimuli. After the

251 stimulus had disappeared, dogs looked centrally for a mean±SD of 6.9±6.6 s, at the entrance side for 5.3±4.9 s

252 and at the exit side for 8.0±7.3s. The time spent looking centrally was affected by an interaction between the

253 type of visual and the type of auditory stimulus, with shorter time spent when a video of a dog was paired with

254 barking, than when either of the dog-stimuli was paired with a non-dog counterpart (P < 0.05). The pairing of

255 non-dog stimuli resulted in intermediate amounts of attention, not different from any other stimulus combination

256 (Figure 4). The interaction between visual and auditory stimulus also affected the time spent looking at the

257 entrance side, which was longer in the case of matching pairs (dog + barking, or non-dog + croaking), than

258 when the auditory and visual stimuli did not match (P < 0.05) (Figure 3). No effect of the type of visual or of

259 auditory stimulus was found for the time spent looking at the exit.

260

261 Table 2. Generalized Estimating Equations model assessing the effects of the type of visual and of auditory

262 stimulus on time spent by dogs looking centrally, at the entrance side or at the exit side, after the presentation of

263 the stimuli.

Factor	Looking at entrance side	Looking centrally	Looking at exit side
Visual stimulus	$X^2=0.14$	$X^2=1.90$	$X^2=0.09$
	P=0.708	P=0.168	P=0.768
Auditory stimulus	$X^2=0.00$	$X^2=0.85$	$X^2=2.00$
	P=0.985	P=0.355	P=0.157
Visual stimulus * auditory stimulus	$X^2=19.13$	$X^2=12.09$	$X^2=2.30$
	P<0.001	P=0.001	P=0.129
Auditory stimulus *Species of non-dog visual stimulus (nested within visual stimulus)	$X^2=1.18$ P=0.880	$X^2=4.553$ P=0.336	$X^2=5.91$ P=0.206

264

265

266

267   -- FIG 3 ABOUT HERE --

268 **Fig. 4** Mean±SD time (s) spent looking at the different regions of the presentation area after the stimuli had

269 disappeared, as a function of the stimuli pair (*P<0.05, Bonferroni-corrected post-hoc comparisons after

270 Generalized Linear Equation Models)

271

272 DISCUSSION

273 In this study we employed a cross-modal, expectancy violation paradigm to assess whether dogs can recognize

274 the species of conspecifics from videos. Dogs were presented with pairs of auditory and visual stimuli, which

275 could be any combination of dog-related on non-dog related vocalization and video. Dogs' orientation towards

276 the presentation area, as a function of the presented pair of stimuli, was analysed during two time intervals, in

277 which different mechanisms were most likely at play.

278   The first interval spanned from the onset of the vocalization to the last frame in which the video of the animal

279   crossing the screen was visible. Dogs orientation in this interval therefore reflected a proximate reaction to the

280   presence of the stimuli, rather than an after-effect of the pairing.

281   Dogs spent almost the entire interval oriented toward the projection area. Moreover, dogs' attention to specific

282   regions of the projection area roughly followed the stimulus occupation of such regions. This is most likely a

283   direct result of the capacity of motion stimuli to elicit orientation responses, an effect that is particularly relevant

284   for stimuli abruptly appearing within the visual field (Hillstrom and Yantis 1994) and for stimuli depicting

285   animate entities (Pratt et al. 2010) two features that characterised the visual stimuli that were presented in this

286   experiment.

287   A breakdown analysis of dogs' orientation to the different parts of the projection area revealed that dogs spent

288   longer time looking at the exit area when a dog video was projected than when the unfamiliar species was

289   projected. Therefore, dogs were more likely to visually follow the dogs' video until it left the presentation area,

290   than the unfamiliar species video. The finding is consistent with the notion that familiarity drives attentional

291   responses for visual stimuli (Christie and Klein 1995). There is some direct evidence that this also applies to

292   dogs, in particular when presented with representations of dogs', such as face photographs (Racca et al. 2010) or

293   biological movement (Eatherington et al. 2019). Overall, the findings support the idea that dogs did at least

294   perceive the dog video as a familiar stimulus.

295   Evidence that dogs did recognise the dog-related stimuli as belonging to a dog, however, comes from the

296   analysis of attention patterns after the stimuli had disappeared. In this time interval, dogs spent less time

297   oriented towards the central part of the presentation area when a bark was followed by the appearance of a dog

298   video, than when any of such two stimuli was paired with an unfamiliar counterpart. In accordance with the

299   violation of expectancy paradigm, longer looking at the main projection area reflected a surprised reaction to the

300   pairing of an unfamiliar-species stimulus with a dog stimulus. Analogous interpretations of longer looking times

301   have been given in studies in dogs (Adachi et al. 2007), and other species including cats (Takagi et al. 2019),

302   horses (Lampe and Andre 2012; Nakamura et al. 2018), crows (Kondo et al. 2012), lions (Gilfillan et al. 2016).

303   Therefore, this result clearly indicates that dogs had perceived the appearance of the dog's video as an expected

304   consequence of the barking, implying they had appropriately recognized both stimuli as belonging to a dog.

305   After being presented with dog stimuli, dogs also spent longer time looking at the entrance region of the

306   presentation area, than when either dog stimulus was paired with an unfamiliar species stimulus. No such effect

307   was observed for attention to the exit region. Although the reason for this pattern of results is not immediately

308 clear, we believe this is a further indication that dogs retained the pair of dog stimuli as coherently representing

309 a dog; in this sense, dogs may have been interested in where the animal came from, especially since nothing

310 indicated the presence of such animal before its sudden appearance. The lack of differences in attention to the

311 exit region, on the other hand, could reflect a relatively low need to monitor an animal who was getting away

312 from the observer.

313 When both stimuli belonged to an unfamiliar species, the patterns of dogs' attention to the presentation area

314 were less clear cut than those observed when presented with dog stimuli. On the one hand, attention to the

315 central part of the presentation area when non-dog stimuli were paired was not different than that observed

316 when dog-stimuli were paired. The similarity in reaction may suggest dogs considered the appearance of the

317 unfamiliar individual as a plausible consequence of the unfamiliar vocalization, much as they considered the

318 appearance of the dog an unsurprising consequence of the bark. Unsurprised reactions to pairs of unfamiliar

319 stimuli in expectancy violation test have also been reported before (e.g. Adachi et al. 2007). As already

320 discussed for the pair of dog stimuli, the high amount of attention paid to the entrance region could indicate the

321 interest in where an unknown (but plausible) type of animal came from. On the other hand, dogs' attention to

322 the central part of the presentation area after non-dog stimuli pairs were presented was also not lower than when

323 a dog/non-dog stimuli pair was presented. A possible explanation is that dogs' attention patterns after being

324 exposed to the two unfamiliar stimuli was driven by the interest in such novel stimuli, rather than by a violated

325 expectation. Indeed, different studies showed neophilic reactions by dogs (e.g. Kaulfuß and Mills 2008; Racca

326 et al. 2010). Of particular relevance, as it deals with visual preference, the study by Racca and collaborators

327 (2010) showed that while dogs pay preferential attention to familiar rather than novel images of dogs, the

328 opposite is true for other classes of stimuli, including images of objects or of human faces. Along this reasoning,

329 hearing a novel auditory stimulus drove attention to the entrance region, and seeing a novel visual stimulus

330 drove attention to both the entrance and central region (the latter being predominantly occupied when the

331 stimulus became fully visible).

332 One question arising from our results it whether dogs showed a different response to the pairing of the bark and

333 dog video merely because they were familiar with both stimuli, without implying classification of the stimuli as

334 belonging to a dog. The literature provides some indications that this may not be the case. For instance, Gergely

335 and collaborators (2019) showed that dogs exposed to a conspecific vocalization pay more attention to pictures

336 of dogs than of humans, a species dogs were highly familiar with. Moreover, a recent functional neuroimaging

337 study revealed greater activation of visual cortical areas in dogs, when exposed to videos of conspecific faces

than when exposed to human faces, suggesting the existence of species-specific processing mechanisms (Bunford et al. 2020). Taken together, these findings suggest dogs do possess the ability to visually discriminate dogs from another familiar species. Whether such ability would be the result of exposure alone or it is aided by a predisposition it is impossible to state by results of the present or of other studies in dogs. Findings in humans indicate that experience builds on top of predispositions in determining one's ability to identify motion features as belonging to a conspecific (reviewed by Hirai and Senju 2020). A thorough understanding of if and how the same factors impact on dogs' ability to recognize other animals would require further experiments, which are currently ongoing at our laboratory.

Few other studies have attempted to demonstrate dogs' ability to recognize the species of other conspecifics in figurative representations, providing suggestive though not conclusive evidence (Autier-Dérian et al. 2013; Gergely et al. 2019). The present findings differ in important ways from all previous attempts. First, in all other studies the stimuli depicted animal heads, whereas our stimuli represented lateral views of the animal's whole body. Our finding imply that a detailed frontal view of the head is not a necessary stimulus for dogs to the recognize a conspecific, at least if motion information is available. Indeed, a crucial difference between the present and earlier studies was that we presented videos rather than still images, allowing us to incorporate information about movement. Our own laboratory showed dogs are attracted by the motion of a laterally walking dog (Eatherington et al. 2019) and studies in other species highlight how motion cues alone can be used for the recognition of conspecifics (Jitsumori et al. 1999; Nunes et al. 2020). Thus, the presence of motion information in our experiment may have played a role in allowing dogs to appropriately identify the conspecific's video. The abovementioned studies indicate that morphology, independently from motion, can also be individually sufficient to the aims of recognition (Jitsumori et al. 1999; Nunes et al. 2020). However, these studies only depicted heads, a stimulus that is rich in features useful to the aims of recognition, even to the level of the individual. Our findings indicate that even more limited morphological details provided by a lateral, whole body view, paired with motion information may be sufficient for dogs to recognize a conspecific. Finally, the area of dogs' visual cognition is not new to the application of cross-modal and expectancy violation paradigms; for instance, similar paradigms have been successfully used to demonstrate dogs' recognition of humans' identity or sex (Adachi et al. 2007; Ratcliffe et al. 2014), or expectations about conspecifics' body size (Taylor et al. 2011). However, to the best of our knowledge, this method had never been used in dogs with videos and some methodological considerations seem useful at this stage. First, while videos were projected, dogs spent most of their time oriented towards the presentation area, indicating the stimuli were able to attract

368 dogs' attention (at least from a behavioural standpoint), a crucial and often problematic aspect of research on

369 visual cognition. Second, even after the stimulus disappeared, dogs remained oriented towards the presentation

370 area for a significant portion of the allowed 30s - suggesting maintenance of interest in what had been projected.

371 Third, the analysis of dogs' orientation across subsequent presentations suggests limited habituation through the

372 first two trials, but a significant decrement starting from the third trial. Overall, these results indicate the method

373 is suitable to study dogs' spontaneous cross-modal processing of auditory and animated visual stimuli, and that

374 dogs can be presented with up to two presentations before their attention starts to decline.

375

376 CONCLUSIONS

377 This study provides the first evidence that dogs recognize videos of dogs as actually representing dogs. These

378 findings will hopefully be a starting point towards the more extensive use of videos in dog behavioural and

379 cognitive research. At the same time, several questions arise from our results; for instance, our stimuli depicted

380 a laterally walking dog, but it would be important to assess whether recognition extends to other dynamic

381 behaviours. A related question is whether motion information alone would be sufficient for dogs to recognize

382 dogs in videos or if, in fact, other figurative information (e.g. shape, color, etc.) is needed to the aims of

383 recognition. Finally, as some of our findings suggest a role of experience or familiarity with the class of stimuli,

384 more studies are needed to determine how exposure impacts on dogs' ability to recognize conspecifics or other

385 species in videos.

386

387 REFERENCES

388

389 Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. Anim

390   Cogn 10:17–21. https://doi.org/10.1007/s10071-006-0025-8

391 Albuquerque N, Guo K, Wilkinson A, Savalli C, Otto E, Mills DS (2016) Dogs recognize dog and human

392   emotions. Biol Lett 12:1-5. https://doi:10.1098/rsbl.2015.0883.

393 Autier-Dérian D, Deputte BL, Chalvet-Monfray K, Coulon M, Mounier L (2013) Visual discrimination of

394   species in dogs (*Canis familiaris*). Anim Cogn 16:637–651. https://doi.org/10.1007/s10071-013-0600-8

395 Bálint A, Faragó T, Meike Z, Lenkei R, Miklósi Á, Pongrácz P. (2015) "Do not choose as I do!"-Dogs avoid the

396   food that is indicated by another dog's gaze in a two-object choice task. Appl Anim Behav Sci 170: 44-53

397 Bunford N, Hernández-Pérez R, Farkas EB, Cuaya LV, Szabó D, Szabó ÁG, Gácsi M, Miklósi A, Andics A.

(2020). Comparative brain imaging reveals analogous and divergent patterns of species-and face-sensitivity in humans and dogs. J Neurosci JN-RM-2800-19. https://doi.org/10.1523/JNEUROSCI.2800-19.2020

Christie J, Klein R (1995) Familiarity and attention: Does what we know affect what we notice? Mem Cognit 23:547–550. https://doi.org/10.3758/BF03197256

Eatherington CJ, Marinelli L, Lõoke M, Battaglini L, Mongillo P (2019) Local Dot Motion, Not Global Configuration, Determines Dogs' Preference for Point-Light Displays. Animals 9:661. https://doi.org/10.3390/ani9090661

Eatherington CJ, Mongillo P, Lõoke M, Marinelli L (2020) Dogs (*Canis familiaris*) recognise our faces in photographs: implications for existing and future research. Anim Cogn 23:711–719. https://doi.org/10.1007/s10071-020-01382-3

Faragó T, Pongrácz P, Miklósi À, Huber L, Virányi Z, Range F (2010) Dogs' expectation about signalers' body size by virtue of their growls. PLoS One 5: e15175. https://doi.org/10.1371/journal.pone.0015175

Fox MW (1971) Integrative development of brain and behavior in the dog. University of Chicago Press, Chicago.

Gergely A, Petró E, Oláh K, Topál J (2019) Auditory–visual matching of conspecifics and non-conspecifics by dogs and human infants. Animals 9:17. https://doi.org/10.3390/ani9010017

Gherardi F, Aquiloni L, Tricarico E (2012) Revisiting social recognition systems in invertebrates. Anim Cogn 15:745–762. https://doi.org/10.1007/s10071-012-0513-y

Gilfillan G, Vitale J, McNutt JW, McComb K (2016) Cross-modal individual recognition in wild African lions. Biol Lett 12:20160323. https://doi.org/10.1098/rsbl.2016.0323

Hepper PG (1994) Long-term retention of kinship recognition established during infancy in the domestic dog. Behav Processes 33:3–14. https://doi.org/10.1016/0376-6357(94)90056-6

Herzog M, Hopf S (1986) Recognition of Visual Pattern Components in Squirrel Monkeys. Eur Arch Psychiatr Neurol Sci 236:10–16. https://doi.org/10.1007/BF00641051

Hillstrom AP, Yantis S (1994) Visual motion and attentional capture. Percept Psychophys 55:399–411. https://doi.org/10.3758/BF03205298

Hirai M, Senju A. (2020). The two-process theory of biological motion processing. Neurosci Biobehav Rev 111:114-124.

Holmes WG, Sherman PW (1983) Kin recognition in animals. Am Sci 7:46–55.

428 Ishikawa Y, Mills D, Willmott A, Mullineaux D, Guo K (2018) Sociability modifies dogs' sensitivity to

429   biological motion of different social relevance. Anim Cogn 21:245–252. https://doi.org/10.1007/s10071-

430   018-1160-8

431 Jitsumori M, Natori M, Okuyama K (1999) Recognition of moving video images of conspecifics by pigeons:

432   Effects of individuals, static and dynamic motion cues, and movement. Anim Learn Behav 27:303–315.

433   https://doi.org/10.3758/BF03199729

434 Kaulfuß P, Mills DS (2008) Neophilia in domestic dogs (*Canis familiaris*) and its implication for studies of dog

435   cognition. Anim Cogn 11:553–556. https://doi.org/10.1007/s10071-007-0128-x

436 Kondo N, Izawa EI, Watanabe S (2012) Crows cross-modally recognize group members but not non-group

437   members. Proc R Soc B Biol Sci 279:1937–1942. https://doi.org/10.1098/rspb.2011.2419

438 Lampe JF, Andre J (2012) Cross-modal recognition of human individuals in domestic horses (*Equus caballus*).

439   Anim Cogn 15:623–630. https://doi.org/10.1007/s10071-012-0490-1

440 Mateo JM (2004) Recognition systems and biological organization: The perception component of social

441   recognition. Ann Zool Fennici 41:729-745.

442 Nakamura K, Takimoto-Inose A, Hasegawa T (2018) Cross-modal perception of human emotion in domestic

443   horses (*Equus caballus*). Sci Rep 8:1–9. https://doi.org/10.1038/s41598-018-26892-6

444 Nunes AR, Carreira L, Anbalagan S, Blechman J, Levkowitz G, Oliveira RF (2020) Perceptual mechanisms of

445   social affiliation in zebrafish. Sci Rep 10:1–14. https://doi.org/10.1038/s41598-020-60154-8

446 Plimpton EH, Swartz KB, Rosenblum LA (1981) Responses of juvenile bonnet macaques to social stimuli

447   presented through color videotapes. Dev Psychobiol 14:109–115. https://doi.org/10.1002/dev.420140204

448 Pongracz P, Miklosi A, Doka A, Csanyi V (2003) Successful Application of Video-Projected Human Images for

449   Signalling to Dogs. Ethology 109:809–821. https://doi.org/10.1046/j.0179-1613.2003.00923.x

450 Pratt J, Radulescu PV, Guo RM, Abrams RA (2010) It's Alive!: Animate Motion Captures Visual Attention.

451   Psychol Sci 21:1724-1730. https://doi.org/10.1177/0956797610387440

452 Racca A, Amadei E, Ligout S, Guo K, Meints K, Mills D (2010) Discrimination of human and dog faces and

453   inversion responses in domestic dogs (*Canis familiaris*). Anim Cogn 13:525–533.

454   https://doi.org/10.1007/s10071-009-0303-3

455 Range F, Aust U, Steurer M, Huber L (2008) Visual categorization of natural stimuli by domestic dogs. Anim

456   Cogn 11:339–347. https://doi.org/10.1007/s10071-007-0123-2

457 Ratcliffe VF, McComb K, Reby D (2014) Cross-modal discrimination of human gender by domestic dogs.

458   Anim Behav 91:127–135. https://doi.org/10.1016/j.anbehav.2014.03.009

459   Siniscalchi M, Lusito R, Vallortigara G, Quaranta A (2013) Seeing left- or right-asymmetric tail wagging

460   produces different emotional responses in dogs. Curr Biol 23:2279–2282.

461   https://doi.org/10.1016/j.cub.2013.09.027

462   Takagi S, Arahori M, Chijiiwa H, Saito A, Kuroshima H, Fujita K (2019) Cats match voice and face: cross-

463   modal representation of humans in cats (*Felis catus*). Anim Cogn 22:901–906.

464   https://doi.org/10.1007/s10071-019-01265-2

465   Taylor AM, Reby D, McComb K (2011) Cross modal perception of body size in domestic dogs (*Canis*

466   *familiaris*). PLoS One 6:2–7. https://doi.org/10.1371/journal.pone.0017069

467

468