

1 **ABSTRACT**

2 Accurate visual object recognition is essential to survival for a wide range of species across a wide
3 range of evolutionary histories and visual requirements. However this task is solved, it is a major
4 achievement because object recognition is far from simple. The appearance of an object can alter
5 almost completely as viewing condition change, not least under variations in lighting and orientation.
6 Determining the recognition limits of a species is important to understanding their visual ecology and
7 can help identify conditions under which recognition may fail. In this paper, we test whether a species
8 of fish is able to recognize objects from an unfamiliar object class (human faces) across changes in
9 viewing direction. Using operant conditioning, we trained archerfish (*Toxotes chatareus*) to
10 discriminate between two frontal views of standardised human faces and, critically, tested whether
11 they could continue to do so as the orientation in depth of the faces changed. All fish learned the
12 initial discrimination task and could also recognize rotated forms. These results represent the first
13 conclusive evidence that a species of fish is able to generalize recognition across views, speaking
14 against a strict image-matching process. The remarkable ability of these fish to generalize recognition
15 across views, speaks to the capacity of relatively simple brains to tackle the hard problem of view-
16 invariance, and provides insight into the mechanisms employed in more complex organisms such as
17 humans. Although we speculate that other fish species may demonstrate similar abilities, a visual
18 system capable of recognition across changes in viewpoint may be especially important to the unique
19 hunting strategy of archerfish.

20 **Keywords:** animal cognition, view-invariance, archerfish, visual system, human facial recognition

Object recognition is fundamental to many complex visual behaviours (e.g. food detection, mate selection, individual recognition) and yet it is a far from trivial task because changes in viewing conditions (e.g. lighting or viewing direction) can drastically alter the two-dimensional (2D) image cast in the eye of the observer by the three-dimensional (3D) object. The underlying mechanisms of recognition can constrain performance in terms of accuracy, flexibility, or speed, and have real consequences to the visual ecology of a species. Recognition which is robust allows for flexibility but may be slower than other systems and limited to animals with sufficiently sophisticated brain structures (Wallis & Bülthoff, 1999). Conversely, a view-dependent recognition system would require less processing power, allowing for faster recognition, but at the cost of reduced flexibility (Wallis & Bülthoff, 1999). The aim of this study is to test, for the first time, whether a species of fish is capable of view-invariant recognition of a complex object category. This should bring us closer to answering two things: how a fish might perceive visual stimuli in their environment, and the capabilities and limitations of the recognition system of a lower vertebrate with no visual cortex.

Experiments with primates (Logothetis & Sheinberg, 1996), rats (Alemi-Neissi et al., 2013; Rosselli et al., 2015; Tafazoli et al., 2012; Zoccolan, 2015; Zoccolan et al., 2009), sea lions (Stich et al., 2003), domestic chicks (Mascalzoni et al., 2012), and horses (Hanggi, 2010) have all demonstrated that some mammals can recognize objects from novel viewpoints. Conversely, some insects (Collett, 1992, 1995) appear to use ‘active vision’ in which they physically move until the image projected onto the retina matches a selection of less-flexible, stored views to the object. The results is that recognition is slower and more error prone.

Even in species that demonstrate view-invariant recognition, the underlying mechanism may vary and result in significant differences to their object recognition capabilities. Both pigeons and bees, for example, have been used to test how species lacking a neocortex recognize complex objects by testing if they can identify rotated human faces after experience with a single orientation (Dyer & Vuong, 2008; Jitsumori & Makino, 2004). When bees were trained with frontal views (0°) of two faces, they were unable to discriminate between the two when rotated by 30°, but given experience

48 with multiple views of the same faces (0° and 60°), the bees were able to recognize the rotated face
49 (30°) albeit with reduced accuracy (Dyer & Vuong, 2008). As a result, the authors concluded that
50 bees are most likely relying on a mechanism of image interpolation. Conversely, pigeons and chicks
51 demonstrated true spontaneous view-invariant recognition with a single exemplar, suggesting that
52 they are capable of using some form of image extrapolation (Jitsumori & Makino, 2004; Wood, 2013,
53 2015). One important lesson from this is that while bees are capable of performing many seemingly
54 sophisticated behaviours (e.g. Avarguès-Weber & Giurfa, 2013; Biernaskie et al., 2009; Howard et
55 al., 2017), it does not necessarily mean the underlying mechanisms or the functional limits are
56 identical to that of other species. Therefore it is worth exploring how a range of animals with different
57 life histories perform the same tasks in order to understand the evolution and application to complex
58 behaviours.

59 Contrary to common misconceptions about fish intelligence, studies of the visual ecology and
60 cognitive abilities of fish are providing increasing evidence that fish can have sophisticated visual
61 repertoires and possess impressive visual systems (e.g. Champ et al., 2014; Cheney et al., 2009;
62 Cheney & Marshall, 2009; Cheney et al., 2013; Newport et al., 2017; Newport et al., 2014, 2015;
63 Newport et al., 2013; Rosa Salva et al., 2014; Siebeck et al., 2009; Siebeck et al., 2010). Studies on
64 the underlying mechanisms of their visual system indicate that despite their relatively small
65 (Northcutt, 2002) and simple brain (i.e. no neocortex), fish demonstrate some sophisticated visual
66 abilities that are comparable to humans (e.g. Rischawy & Schuster, 2013; Schlegel & Schuster, 2008;
67 Schuster et al., 2004). We even know that fish can generalise recognition across some affine
68 transformations, including changes in size (Douglas et al., 1988; Frech et al., 2012; Schuster et al.,
69 2004), indicating some recognition flexibility. That said, there are very few studies investigating a
70 task as complicated as object recognition across changes in viewpoint.

71 Some of the earliest attempts to test recognition across more complex image transformations
72 were reported by Schuster and Amtsfeld (2002), who used 2D shapes to explore whether weakly
73 electric fish (*Gnathonemus petersii*) generalize between learned and novel stimuli based on similarity

74 of the retinal image. The authors concluded that a template matching recognition system explained
75 the generalization behaviour of this species. Schluessel et al. (2014) used 3D objects as stimuli for
76 experiments with Malawi cichlids, however the stimuli had features that were not affected by changes
77 in orientation (e.g. overall colour and size) making the results difficult to interpret. In addition, the
78 approach angle of the fish was not restricted, allowing the fish to view the objects from different
79 angles, before making a selection.

80 In this report, we adapt methods and a set of visual stimuli used in primate and human testing to
81 systematically investigate recognition generalization across rotation in depth in a species of fish. Our
82 test species is the archerfish (*Toxotes chatareus*), which are known for their ability to hunt aerial prey
83 by knocking them down with a jet of water spat from the mouth. Given the fish's proven ability to
84 perform visual discrimination tasks (Newport et al., 2016; Newport et al., 2013; Temple et al., 2013)
85 and its ecology, archerfish are a good candidate for testing. The stimulus set we chose are human
86 faces as they offer a number of points of interest as a test set. First, they are not biologically relevant
87 to the fish making it unlikely that the fish is endowed with an innate ability to discriminate them..
88 Second, recognizing faces is particularly taxing as their appearance changes dramatically as a
89 function of viewing angle, and a very large set of similar-looking distractors exist – namely other
90 faces (Wallis et al., 2008). Third, while other stimuli can be used to test view-invariance, a number
91 of previous studies in a range of taxa have used faces as a study set, permitting comparisons to be
92 drawn. If fish are capable of recognizing novel views of human faces it will provide the first evidence
93 of a remarkably robust recognition system in a fish, providing insight into their ecology and, in a
94 wider sense, into the mechanisms that might underlie sophisticated recognition systems – since this
95 fish does not possess a cortex.

96 **METHODS**

97 **General Procedure**

98 Six archerfish were used in this experiment. All fish were kept as described by Newport et al. (2013)
99 in accordance with the University of Queensland Animal Ethics Committee approval (AEC approval
100 number: SBMS/241/12) and all experimental protocols were approved by the same body.
101 Experiments followed similar procedures to those described by Newport et al. (2016). Briefly, stimuli
102 were displayed on a 15 inch (1024 x 768 pixels) LCD monitor (SyncMaster 153v, Samsung) with a
103 Plexiglas housing, suspended above the aquaria. The archerfish were presented with a two-alternative
104 forced-choice test (stimulus monitor coordinates: 0 -160, 0 160). Stimulus positions on the monitor
105 were balanced so that the rewarded stimulus (S+) appeared equally often on the left and right and was
106 further constrained to never appear in the same position on more than two consecutive trials.
107 Archerfish selected a stimulus by spitting a jet of water at the stimulus on the computer monitor. The
108 experimenter recorded the response of the fish as they could clearly see the jet of water and the
109 presented stimuli. The accuracy of the fish meant that the water jet hit within the area of only one
110 stimulus and therefore there was no ambiguity in the choice of the fish. The fish received a food pellet
111 as a reward (CichlidGold®, Kyorin Co.Ltd., Japan) which was dropped in by hand by the
112 experimenter when the correct stimulus was selected in training and reinforcement trials (frontal
113 views). Selection of the distractor stimulus (S-) resulted in no food reward and an immediate
114 termination of the trial.

115 **Stimuli**

116 The images comprised 2D renderings of scanned human faces from the 3D Head Models Database
117 created by researchers at the Max Planck Institute in Tübingen, Germany (Blanz & Vetter, 1999;
118 Troje & Bühlhoff, 1996; Vetter, 1998). All faces used in this experiment were of Caucasian women
119 and were 384 x 384 pixels when presented on the monitor. During training, the fish were presented
120 with frontal views of the faces. During testing they were presented with the same faces but rotated
121 about the vertical axis by 30°, 60° and 90° (Fig. 1. See Electronic Supplementary Material (ESM)
122 Fig. S1 for an animation of experimental procedure). Two separate experiments were run. In
123 Experiment 1, five individual fish (Fish A-E) were all trained to different pairs of faces in an attempt

124 to ensure that any observed positive result was not due to idiosyncrasies with the faces themselves.
125 In Experiment 2, the five fish as well as one naïve fish (Fish F) were trained to the same pair of faces
126 to determine if the fish would perform similarly to one another given the same stimuli. All fish
127 excluding B and F completed Experiment 1 first.

128 The faces used were chosen by the experimenter prior to experimentation. After
129 experimentation, the similarity of each image pair was analysed using the Structural Similarity index
130 (SSIM) (Wang et al., 2004). This method considers the luminance, contrast and structure of an image
131 and identical images will have an index of 1. As expected, images that match in orientation are more
132 similar than those of the same face but at different orientations (Fig. 2). In most cases, the frontal
133 view of a face is more similar to the rotated view of the same face than the opposing face (ESM,
134 Table S1).

135 **Pre-training**

136 During an initial task familiarization phase, the fish were trained to spit at the monitor and
137 discriminate between two stimuli. An orange circle (diameter: 2 cm, RGB value: 255, 127, 0) was
138 displayed in random positions on the monitor. Because of the inquisitive nature of archerfish, the fish
139 would spit at the circle voluntarily and each time they did so, they received a food pellet. The orange
140 circle was moved all over the monitor during this phase and often jiggled to attract the attention of
141 the fish. Trial times were not constrained in training phases as it is important to allow the fish time to
142 make a decision. Premature termination of trials during this phase would likely work against our
143 attempts to condition the fish to spit for a reward. Fish were considered ready for training once they
144 consistently made quick, accurate selection. Therefore, once the fish could spit at the circle, regardless
145 of where it was positioned within approximately 1 second of the circle being displayed, a purple circle
146 (diameter: 2 cm, RGB value: 127, 0, 127) was introduced as a distractor. Selection of the purple circle
147 was not rewarded and immediately terminated the trial.

148 In this stage, the two stimuli only appeared in the two positions described above. Each session with
149 the target and distractor consisted of 30 trials and the fish had to select the target at a statistically
150 significant frequency of $\geq 73\%$ (binomial: $p < 0.01$, $n = 30$ trials) in one session in order to move on
151 to the main experiment.

152 **Training**

153 The archerfish were first trained to discriminate between two frontal views of faces. For Fish A in
154 Experiment 1 and Fish F in Experiment 2 only (this represents the first instance of training for each
155 individual), two additional steps were added to the training procedure: 1) selection of S- was
156 associated with no penalty in the first session but in all following sessions it resulted in a 30 second
157 delay before the next trial began, 2) six trials were designated 'prep-trials' in which the fish were
158 neither rewarded nor penalized regardless of their stimulus choice. The prep-trials were used to
159 prepare the fish for trials with the testing images, which would garner no response (i.e. no reward or
160 delay).

161 All training sessions consisted of 30 trials, divided into six blocks of 5 trials. Each block had
162 one randomly distributed prep-trial. All other fish were given no prep-trials or time penalties for
163 selection of S- as this seemed to have little impact on performance but considerably slowed the
164 training session and had the potential to affect fish motivation. Sessions were run until the fish had
165 achieved a statistically significant S+ selection frequency of $\geq 73\%$ in two consecutive sessions
166 (binomial: $p < 0.01$, $n = 30$ trials).

167 **Testing**

168 Once the fish could consistently select S+, 'test trials' were introduced. Fifteen of the 30 trials showed
169 the frontal views of the two trained faces and were used as rewarded reinforcement trials. The
170 remaining 15 trials showed three novel orientations (30° , 60° , 90°) of the same two faces (five trials
171 for each rotation). The first two trials of every session were always reinforcement trials. In the
172 remaining trials, presentations of reinforcement and test trials were pseudorandom with the constraint

173 that neither was presented in more than two consecutive sessions. Test trials were never rewarded or
174 penalized in order to prevent the fish from learning the correct response. Five sessions of 30 trials
175 were run so that a sample size of 25 trials was gathered for each rotation and fish.

176 Response times were recorded for each trial and the trial time was calculated based on the
177 difference in time between when stimuli were first presented on the monitor and when the
178 experimenter recorded a response by pressing a key on a keyboard. These times were recorded
179 automatically by the stimulus presentation software. The mean trial response time (RT) for each
180 testing condition was calculated for each fish trained to different faces (Experiment 1) and the same
181 faces (Experiment 2). The difference in mean response time of the treatment (90° - 30° , 90° - 60° , 60° -
182 30°) was analysed for each individual using a permutations test ($n = 10,000$). Linear regression was
183 carried out on the results of all fish to investigate the relationship between accuracy and mean
184 response time.

185 **RESULTS**

186 **Training**

187 All six fish completed the pre-training with both the orange and purple circle within 1-3 sessions. All
188 fish passed the training stage and learned to discriminate the frontal views of two faces within 2-19
189 sessions (Experiment 1) and 4-21 sessions (Experiment 2).

190 **Testing**

191 *Stimulus selection frequency*

192 The results of all fish were then analysed using a generalized linear mixed model (GLMM) with a
193 binomial distribution (log-link function) to test whether the selection frequencies of each stimulus
194 type were significantly different from that expected due to chance (using the glmer function in the
195 lme4 package; Bates et al., 2015). The individual S+ selection frequencies of each fish were analyzed
196 using a binomial test.

197 **Experiment 1**

198 When trained to different face pairs, fish could identify the learned face given the frontal view
199 (GLMM: $z > 9.086$, $N_{\text{observations}} = 20$, $N_{\text{fish}} = 5$, $p < 0.001$), 30° rotations (GLMM: $z > 3.533$, $N_{\text{observations}}$
200 $= 20$, $N_{\text{fish}} = 5$, $p < 0.001$), 60° rotations (GLMM: $z > 3.380$, $N_{\text{observations}} = 20$, $N_{\text{fish}} = 5$, $p < 0.001$) but
201 not 90° rotations (GLMM: $z > 1.504$, $N_{\text{observations}} = 20$, $N_{\text{fish}} = 5$, $p = 0.133$) (Fig. 3). All individual
202 fish continued to select S+ in the reinforcement trials at a high degree of accuracy ($\geq 73\%$) (see ESM
203 Table S2 for all individual results and statistical analyses). When the faces were rotated by 30°, two
204 fish (Fish B and C) selected S+ at a frequency different from chance and two fish (Fish A and B)
205 continued to select S+ when the faces were rotated by 60°. Only one individual (Fish B) selected S+
206 at a higher frequency than predicted by chance when exposed to 90° rotations.

207 **Experiment 2**

208 When the fish were trained to the same face pair, fish could identify the learned face given the frontal
209 view (GLMM: $z > 7.477$, $N_{\text{observations}} = 24$, $N_{\text{fish}} = 6$, $p < 0.001$), 30° rotations (GLMM: $z > 6.846$,
210 $N_{\text{observations}} = 24$, $N_{\text{fish}} = 6$, $p < 0.001$), 60° rotations (GLMM: $z > 6.402$, $N_{\text{observations}} = 24$, $N_{\text{fish}} = 6$, $p <$
211 0.001) and 90° rotations (GLMM: $z > 3.006$, $N_{\text{observations}} = 24$, $N_{\text{fish}} = 6$, $p < 0.05$) (Fig. 3). All fish
212 selected S+ at a statistically significant frequency ($\geq 72\%$) when presented with the frontal view, 30°
213 and 60° rotations (Fig. 3B; see ESM Table S3 for all individual results and statistical analyses). When
214 presented with 90° rotations, three fish (Fish C, D, and E) continued to preferentially select S+ but
215 the remaining three fish (Fish A, B and F) did not.

216 Initially, Fish B was given an unlimited time to make a choice (up to almost 17 minutes in one case).
217 After 15 trials, that individual was thought to have stopped responding to 90° rotation trials as there
218 was a substantial increase in response time. In the final 10 trials, if the fish had not made a choice
219 within at least five minutes then the trial was terminated. As a result, only 15 trials were included in
220 the analysis of stimulus selection frequency. Because fish choices cannot be measured in terminated
221 trials, in all other tests with other subjects, the experimenter waited until the fish made a selection
222 regardless of how long it took.

223 *Response time*

224 **Experiment 1**

225 When trained to different faces, the mean response time varied by individual fish and testing condition
226 (Fig. 4A). For three fish, there was a significant difference in the mean time of 90° rotations compared
227 to 30° (Fish A: $p < 0.001$, Fish B: $p < 0.001$, Fish C: $p = 0.032$) and 60° rotations (Fish A: $p < 0.001$,
228 Fish B: $p < 0.001$, Fish C: $p = 0.006$). The mean response time of Fish D and E was not significantly
229 different for 90°- 30° rotations (Fish D: $p = 0.695$, Fish E: $p = 0.989$) or 90°- 60° rotation (Fish D: p
230 $= 0.188$, Fish E: $p = 0.868$). There was no significant difference in mean response time between 60°
231 and 30° rotations for any fish (Fish A: $p = 0.279$, Fish B: $p = 0.501$, Fish C: $p = 0.734$, Fish D: $p =$
232 0.922 , Fish E: $p = 0.969$). See ESM Table S4 for detailed individual results and Fig. S2 for graphical
233 results of the statistical analysis. For Fish A, the response time for one control trial was not recorded
234 due to experimenter error. For Fish E, the response time for one 90° rotation trial was excluded from
235 the analysis as the fish did not look at the computer monitor for approximately two minutes after the
236 stimuli were displayed. Once the fish looked at the monitor it rapidly made a choice but this time was
237 not recorded. There was no significant group relationship between accuracy and mean response time
238 ($R^2 = 0.005$, $F[1, 13] = 0.07$, $p = 0.802$).

239 **Experiment 2**

240 Similarly to Experiment 1, when trained to the same two faces, the mean response time varied by
241 individual fish and testing condition (Fig. 4B). When all fish were trained to the same two faces, four
242 fish had a mean response time for 90° rotations that was significantly different from that of 30° (Fish
243 A: $p < 0.001$, Fish B: $p < 0.001$, Fish C: $p < 0.001$, Fish F: $p < 0.001$) and 60° rotations (Fish A: $p <$
244 0.001 , Fish B: $p < 0.001$, Fish C: $p < 0.001$, Fish F: $p < 0.001$). There was no significant difference
245 in mean response time for two fish between 90° and 30° (Fish D: $p = 0.471$, Fish E: $p = 0.912$) or 60°
246 rotations (Fish D: $p = 0.444$, Fish E: $p = 0.959$). As in Experiment 1, there was no significant
247 difference in mean response time between 60° and 30° rotations for any fish (Fish A: $p = 0.397$, Fish
248 B: $p = 0.482$, Fish C: $p = 0.286$, Fish D: $p = 0.532$, Fish E: $p = 0.359$, Fish F: $p = 0.251$). See ESM

249 Table S5 for individual results and Fig. S3 for graphical results of the statistical analysis. As
250 previously described, 15 trials were terminated before a choice was made for Fish B. The time per
251 trial was then set to a maximum of five minutes and for the sake of analysis these trials were all given
252 a standard time of 300 seconds. There was a significant group relationship between accuracy and
253 mean response time ($R^2 = 0.571$, $F[1, 16] = 21.3$, $p < 0.001$).

254 **DISCUSSION**

255 We tested whether archerfish could continue to discriminate between two learned human faces when
256 the faces were rotated about the vertical axis by 30°, 60° and 90°. The degree of rotation at which
257 fish could still recognize the learned face was variable per individual and experiment, but ranged
258 from 30° to 90°. The fact that all subjects continued to recognize the learned faces when rotated
259 (either in Experiment 1 or 2), demonstrates that archerfish exhibit some degree of spontaneous view
260 generalization. However, accuracy and decision speed were affected by the degree of rotation
261 indicating that the archerfish is not perfectly view-invariant but rather view-dependent. That said, it
262 is worth noting that humans too, show a limited ability to generalize face recognition across depth
263 rotations, despite their high biological relevance (Patterson & Baddeley, 1977; Wallis et al., 2009).
264 The degree of rotation in depth that can be tolerated by archerfish may be greater given more
265 dissimilar stimuli as is the case with rats (Rosselli et al., 2015) or stimuli that they are more familiar
266 with. However, some fish species regularly face similarly difficult recognition tasks including the
267 Ambon damselfish (*Pomacentrus amboinensis*) which distinguishes individual conspecifics based on
268 intricate facial patterns (Siebeck et al., 2010). For opportunistic feeders like the archerfish, accurately
269 identifying unpalatable prey from a wide range of possible items may be computationally difficult
270 but crucial to survival.

271 Although all fish were able to complete the task, the results were not entirely consistent. Of
272 the six fish tested, two were unable to accurately select S+ when rotated by 90° in either Experiment
273 1 or 2 (Fish A and F). Both of these individuals were exposed to a slightly different training procedure

274 whereby a 30 second penalty was given for incorrect choices and 20% of trials were unrewarded or
275 delayed regardless of whether the fish was correct or not (Fish A: Experiment 1; Fish F: Experiment
276 2; see Methods section for more detail). In theory this should have better prepared the fish for
277 unrewarded testing trials and should not have had a particular impact on only 90° rotation trials but
278 it is possible that this training method had some effect on the choices of the fish that we are unaware
279 of. For some animals, different training schedules, such as overtraining, can lead to variation in
280 discrimination strategies (e.g. Mandler & Hooper, 1967), but that has not yet been tested in fish. In
281 this case, these fish did not experience more training than others (Fish A: 8 sessions; Fish F: 4
282 sessions) and the training procedure was only changed in one experiment per fish.

283 We also observed that the mean response time was significantly higher for some fish when
284 presented with faces rotated by 90° compared to 30° and 60° (Experiment 1: Fish A, B, C; Experiment
285 2: Fish A, B, C, F. See ESM Tables S4 and S5 for individual *p*-values). In Experiment 2, the two fish
286 that did not show an increased response time also showed no decrease in selection accuracy when
287 exposed to 90° rotations (no speed/accuracy trade-off). As test trials were never rewarded, it is
288 possible the fish were inadvertently trained to treat profile views as different to standard, rewarded
289 frontal views. If that were the case, you might expect to see fish hesitating to make a choice until
290 finally simply choosing at random. While some fish could indeed have been doing this, at least two
291 fish (Experiment 1: Fish B; Experiment 2: Fish C) were able to maintain a significant S+ selection
292 accuracy but simply took much longer to make a choice. Therefore, it seems likely that the fish were
293 still motivated to make a choice but it simply took them longer. Overall, there was a significant
294 relationship between accuracy and response time in Experiment 2 ($p < 0.001$), where testing
295 conditions with higher accuracy also had shorter response times. However we found no significant
296 relationship in Experiment 1 ($p = 0.8$). The observed increase in response time suggests there may be
297 a cost in performance as the appearance of the faces becomes increasingly different from the
298 exemplar. In a laboratory setting where decision time does not impact the quality of the food reward,
299 it makes sense to take the time to make a correct decision, but recognition requiring extended decision

300 time is unlikely to be practicable for archerfish in natural conditions where interspecific competition
301 for food is intense and catching prey quickly is imperative to avoid theft (Davis & Dill, 2012).

302 So where do these data fit into the wider debate on face and object recognition? There is
303 accumulating evidence that in primates at least, faces are processed by a specialized cortical system
304 distinct from other objects. Proponents of this view point to a range of perceptual effects and modes
305 of generalization which are unique to faces (for a review see Wallis, 2013). The idea has also been
306 supported by electrophysiological studies, most notably by Tsao and colleagues (Tsao et al., 2008).
307 They combined functional imaging and single cell recording to identify face selective ‘patches’ of
308 neurons dotted throughout the primate visual hierarchy, from early visual areas through to anterior
309 regions of the temporal lobe. At the same time, in the human behavioral literature, theorists have
310 argued that the underlying representation is different too. They have developed a model of face
311 representation using a norm-based model (built around specific whole-face prototypes) rather than
312 the feature analyzers common to biological models of object recognition (Riesenhuber & Poggio,
313 2000; Wallis, 2013; Wallis & Bühlhoff, 1999). However, their model is not universally accepted.
314 Computational models exist which can explain how many of the unusual properties of face
315 recognition can be encapsulated in a single, self-organizing, hierarchical model of vision exactly like
316 the currently accepted model of object recognition. The key is to regard faces as a highly over-learned
317 category of objects in which the supporting features are driven, through learning, to high intrinsic
318 complexity/dimensionality (Wallis, 2013). Interestingly, the most recent data from Tsao’s lab speaks
319 in favor of a feature-based representation, rather than a norm-based one (Chang & Tsao, 2017). We
320 would argue that the data reported here also supports the idea that recognition of faces is possible
321 using a more generic system suited to all forms of object recognition. The striking fact is not merely
322 that a fish does not require a specialized area of cortex to do the task, but that they do not appear to
323 need any cortex at all! That is not to say that the performance of these fish is on a par with human
324 observers, but all things considered, the performance is remarkable. Humans have many more
325 neurons at their disposal, they have much more experience with human faces (years vs hours), they

326 probably possess cortical regions dedicated to the task, and yet still they struggle with depth rotation
327 and struggle to cope with rotation in the picture plane.

328 The feature-based model of recognition (Bülthoff & Edelman, 1992; Ullman, 2007; Wallis &
329 Bülthoff, 1999) finds echoes in the Common Elements Model proposed by (Soto & Wasserman,
330 2012; Soto & Wasserman, 2016) to describe the performance of pigeons. The observed decrease in
331 accuracy and increase in response time from our current study is like-wise consistent with a view-
332 based/feature-based model and suggests that fish form representations of objects, including faces, as
333 a collection of features. The object features or elements encoded can dictate how robust recognition
334 is when viewed under new conditions and may provide an explanation for the observed variation in
335 fish generalization performance. During training, fish were required to discriminate between two
336 faces, a task that can be achieved by identifying relatively few indicative features. For example, the
337 two faces could be discriminated by only differences in overall luminance or the presence or absence
338 of a specific pixel. It is for this reason that Newport et al. (2016) suggested using a large number of
339 faces to test whether fish could discriminate human faces. A similar procedure was not used in this
340 experiment as it would have required a vast number of unrewarded testing trials to collect enough
341 responses for each face to be statistically useful. In addition, this would have required a significant
342 training and testing period which we felt would have made the experiment so long that this species
343 would likely have lost motivation to participate. Unlike some animal models which can perform
344 hundreds of trials per session (e.g. rats: Zoccolan et al., 2009), we have found that the size of
345 archerfish used in the current experiments demonstrate the most consistent performance when given
346 only 30 trials per session.

347 While our results demonstrate that fish can recognize a rotated object, generalization to novel
348 views was relatively poor for many fish in Experiment 1. In most cases, the fish were more accurate
349 when all trained to the same faces in Experiment 2. One possible explanation is that there was
350 something intrinsically easier about discriminating the faces in Experiment 2. The face pair was
351 selected prior to any experiments including those described by Newport et al. (2016) and the authors

352 did not select the stimuli based on what they thought might be easier for the fish. Instead, the images
353 chosen were because they were both high quality (i.e. no artifacts from the image rendering process)
354 and had a similar, though not identical, face height and image brightness. An analysis of the structural
355 similarity of the two faces did not indicate that the faces used in Experiment 2 are more dissimilar
356 than other face pairs used, which might make the task easier (Fig. 2). Alternatively, the previous
357 experience of the fish may explain their improved performance. Excluding Fish B and F, all subjects
358 had previously participated in an experiment following this procedure. Their previous experience may
359 have led the fish to attend to cues that were more useful for identifying a rotating object. However,
360 further experiments exploring the image elements used by the fish for face identification are required
361 to confirm this hypothesis.

362 The ecological benefit of object recognition, despite changes in orientation based on
363 experience with a single learned view, is that unfamiliar objects may be identifiable under a broader
364 range of conditions. This may be particularly important for fish as, unlike many terrestrial animals,
365 they can freely move vertically and horizontally with little additional energy investment. As a result,
366 the number of potential viewing angles under which a fish must recognize an object may be greatly
367 increased. For archerfish that must track a prey item and hit it with a jet of water, identification of
368 moving objects would very likely require a system that can cope with continual changes in object
369 orientation. However, our results indicate that recognition accuracy and speed can be negatively
370 impacted as objects are increasingly rotated. Therefore, it is worth exploring in future studies whether
371 archerfish can also make use of other mechanisms, such as active vision, to improve accuracy and
372 decision speed.

373 **REFERENCES**

- 374 Alemi-Neissi, A., Rosselli, F. B., & Zoccolan, D. (2013). Multifeatural shape processing in rats
375 engaged in invariant visual object recognition. *The Journal of Neuroscience*, 33(14), 5939-
376 5956. doi: 10.1523/jneurosci.3629-12.2013
- 377 Avarguès-Weber, A., & Giurfa, M. (2013). Conceptual learning by miniature brains. *Proceedings of*
378 *the Royal Society B: Biological Sciences*, 280(1772). doi: 10.1098/rspb.2013.1907
- 379 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
380 Using lme4. [sparse matrix methods; linear mixed models; penalized least squares; Cholesky
381 decomposition]. 2015, 67(1), 48. doi: 10.18637/jss.v067.i01
- 382 Biernaskie, J., xa, M, Walker, S., xa, C, . . . Editor: Michael, C. W. (2009). Bumblebees Learn to
383 Forage like Bayesians. *The American Naturalist*, 174(3), 413-423. doi: 10.1086/603629
- 384 Blanz, V., & Vetter, T. (1999). *A morphable model for the synthesis of 3D faces*. Paper presented at
385 the Proceedings of the 26th annual conference on Computer graphics and interactive
386 techniques.
- 387 Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view
388 interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*,
389 89(1), 60-64.
- 390 Champ, C., Wallis, G., Vorobyev, M., Siebeck, U., & Marshall, J. (2014). Visual acuity in a species
391 of coral reef fish: (*Rhinecanthus aculeatus*). *Brain, Behavior and Evolution*, 83(1), 31-42.
392 doi: doi.org/10.1159/000356977
- 393 Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, 169(6),
394 1013-1028.e1014. doi: 10.1016/j.cell.2017.05.011
- 395 Cheney, K. L., Grutter, A. S., Blomberg, S. P., & Marshall, N. J. (2009). Blue and Yellow Signal
396 Cleaning Behavior in Coral Reef Fishes. *Current Biology*, 19(15), 1283-1287. doi:
397 10.1016/j.cub.2009.06.028

398 Cheney, K. L., & Marshall, N. J. (2009). Mimicry in coral reef fish: how accurate is this deception
399 in terms of color and luminance? *Behavioral Ecology*, 20(3), 459-468. doi:
400 10.1093/beheco/arp017

401 Cheney, K. L., Newport, C., McClure, E. C., & Marshall, N. J. (2013). Colour vision and response
402 bias in a coral reef fish. *Journal of Experimental Biology*, 216(15), 2967-2973. doi:
403 10.1242/jeb.087932

404 Collett, T. S. (1992). Landmark Learning and Guidance in Insects. *Philosophical Transactions:*
405 *Biological Sciences*, 337(1281), 295-303. doi: 10.2307/57061

406 Collett, T. S. (1995). Making learning easy: the acquisition of visual information during the
407 orientation flights of social wasps. *Journal of Comparative Physiology A*, 177(6), 737-747.
408 doi: 10.1007/BF00187632

409 Davis, B. D., & Dill, L. M. (2012). Intraspecific kleptoparasitism and counter-tactics in the
410 archerfish (*Toxotes chatareus*). *Behaviour*, 149, 1367-1394. doi: 10.1163/1568539X-
411 00003026

412 Douglas, R. H., Eva, J., & Guttridge, N. (1988). Size constancy in goldfish (*Carassius auratus*).
413 *Behavioural Brain Research*, 30(1), 37-42. doi: 10.1016/0166-4328(88)90006-X

414 Dyer, A. G., & Vuong, Q. C. (2008). Insect brains use image interpolation mechanisms to recognise
415 rotated objects. *PLoS ONE*, 3(12), e4086. doi: 10.1371/journal.pone.0004086

416 Frech, B., Vogtsberger, M., & Neumeyer, C. (2012). Visual discrimination of objects differing in
417 spatial depth by goldfish. *Journal of Comparative Physiology A*, 198(1), 53-60. doi:
418 10.1007/s00359-011-0685-y

419 Hanggi, E. B. (2010). Rotated object recognition in four domestic horses (*Equus caballus*). *Journal*
420 *of Equine Veterinary Science*, 30(4), 175-186. doi: 10.1016/j.jevs.2010.02.003

421 Howard, S. R., Avarguès-Weber, A., Garcia, J., & Dyer, A. G. (2017). Free-flying honeybees
422 extrapolate relational size rules to sort successively visited artificial flowers in a realistic
423 foraging situation. *Animal Cognition*, 20(4), 627-638. doi: 10.1007/s10071-017-1086-6

424 Jitsumori, M., & Makino, H. (2004). Recognition of static and dynamic images of depth-rotated
 425 human faces by pigeons. *Learning & Behavior*, 32(2), 145-156.

426 Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of*
 427 *Neuroscience*, 19, 577-621. doi: 10.1146/annurev.ne.19.030196.003045

428 Mandler, J. M., & Hooper, W. R. (1967). Overtraining and goal approach strategies in
 429 discrimination reversal. *Quarterly Journal of Experimental Psychology*, 19(2), 142-149. doi:
 430 10.1080/14640746708400083

431 Mascalcioni, E., Osorio, D., Regolin, L., & Vallortigara, G. (2012). Symmetry perception by poultry
 432 chicks and its implications for three-dimensional object recognition. *Proceedings of the*
 433 *Royal Society B: Biological Sciences*, 279(1730), 841-846. doi: 10.1098/rspb.2011.1486

434 Newport, C., Green, N., McClure, E., Osorio, D., Vorobyev, M., & Cheney, K. L. (2017). Fish use
 435 colour to learn compound visual signals. *Animal Behaviour*, 125, 93-100. doi:
 436 10.1016/j.anbehav.2017.01.003

437 Newport, C., Wallis, G., Reshitnyk, Y., & Siebeck, U. E. (2016). Discrimination of human faces by
 438 archerfish (*Toxotes chatareus*). [Article]. *Scientific Reports*, 6, 27523. doi:
 439 10.1038/srep27523

440 Newport, C., Wallis, G., & Siebeck, U. E. (2014). Concept learning and the use of three common
 441 psychophysical paradigms in the archerfish (*Toxotes chatareus*). *Frontiers in Neuroscience*
 442 (*Neural Circuits*), 8, 1-13. doi: 10.3389/fncir.2014.00039

443 Newport, C., Wallis, G., & Siebeck, U. E. (2015). Same/different abstract concept learning by
 444 archerfish (*Toxotes chatareus*). *PLOS ONE*, 10(11), e0143401. doi:
 445 10.1371/journal.pone.0143401

446 Newport, C., Wallis, G., Temple, S. E., & Siebeck, U. E. (2013). Complex, context-dependent
 447 decision strategies of archerfish, *Toxotes chatareus*. *Animal Behaviour*, 86(6), 1265-1274.
 448 doi: 10.1016/j.anbehav.2013.09.031

449 Northcutt, R. G. (2002). Understanding Vertebrate Brain Evolution. *Integrative and Comparative*
450 *Biology*, 42(4), 743-756. doi: 10.1093/icb/42.4.743

451 Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *J Exp Psychol Hum Learn*,
452 3(4), 406-417. doi: doi.org/10.1037/0278-7393.3.4.406

453 Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nat Neurosci*.

454 Rischawy, I., & Schuster, S. (2013). Visual search in hunting archerfish shares all hallmarks of
455 human performance. *Journal of Experimental Biology*, 216(16), 3096-3103. doi:
456 10.1242/jeb.087734

457 Rosa Salva, O., Sovrano, V. A., & Vallortigara, G. (2014). What can fish brains tell us about visual
458 perception? [Review]. *Frontiers in Neural Circuits*, 8. doi: 10.3389/fncir.2014.00119

459 Rosselli, F. B., Alemi, A., Ansuini, A., & Zoccolan, D. (2015). Object similarity affects the
460 perceptual strategy underlying invariant visual object recognition in rats. *Frontiers in Neural*
461 *Circuits*, 9, 10. doi: 10.3389/fncir.2015.00010

462 Schlegel, T., & Schuster, S. (2008). Small circuits for large tasks: High-speed decision-making in
463 archerfish. *Science*, 319(5859), 104-106. doi: 10.1126/science.1149265

464 Schluessel, V., Kraniotakes, H., & Bleckmann, H. (2014). Visual discrimination of rotated 3D
465 objects in Malawi cichlids (*Pseudotropheus* sp.): a first indication for form constancy in
466 fishes. *Animal Cognition*, 17(2), 359-371. doi: 10.1007/s10071-013-0667-2

467 Schuster, S., & Amtsfeld, S. (2002). Template-matching describes visual pattern-recognition tasks
468 in the weakly electric fish *Gnathonemus petersii*. *Journal of Experimental Biology*, 205(4),
469 549-557.

470 Schuster, S., Rossel, S., Schmidtman, A., Jäger, I., & Poralla, J. (2004). Archer fish learn to
471 compensate for complex optical distortions to determine the absolute size of their aerial
472 prey. *Current Biology*, 14(17), 1565-1568. doi: 10.1016/j.cub.2004.08.050

473 Siebeck, U. E., Litherland, L., & Wallis, G. M. (2009). Shape learning and discrimination in reef
474 fish. *Journal of Experimental Biology*, 212(13), 2113-2119. doi: 10.1242/jeb.028936

475 Siebeck, U. E., Parker, A. N., Sprenger, D., Mathger, L. M., & Wallis, G. (2010). A species of reef
 476 fish that uses ultraviolet patterns for covert face recognition. [Article]. *Current Biology*,
 477 20(5), 407-410. doi: 10.1016/j.cub.2009.12.047

478 Soto, F., & Wasserman, E. (2012). Visual object categorization in birds and primates: Integrating
 479 behavioral, neurobiological, and computational evidence within a “general process”
 480 framework. *Cognitive, Affective, & Behavioral Neuroscience*, 12(1), 220-240. doi:
 481 10.3758/s13415-011-0070-x

482 Soto, F. A., & Wasserman, E. A. (2016). Promoting rotational-invariance in object recognition
 483 despite experience with only a single view. *Behavioural Processes*. doi:
 484 10.1016/j.beproc.2015.11.005

485 Stich, K. P., Dehnhardt, G., & Mauck, B. (2003). Mental Rotation of Perspective Stimuli in a
 486 California Sea Lion (*Zalophus californianus*). *Brain, Behavior and Evolution*, 61(2), 102-
 487 112. doi: 10.1159/000069355

488 Tafazoli, S., Di Filippo, A., & Zoccolan, D. (2012). Transformation-Tolerant Object Recognition in
 489 Rats Revealed by Visual Priming. *The Journal of Neuroscience*, 32(1), 21-34. doi:
 490 10.1523/jneurosci.3932-11.2012

491 Temple, S. E., Manietta, D., & Collin, S. P. (2013). A comparison of behavioural (Landolt C) and
 492 anatomical estimates of visual acuity in archerfish (*Toxotes chatareus*). *Vision Research*,
 493 83(0), 1-8. doi: 10.1016/j.visres.2013.02.014

494 Troje, N. F., & Bühlhoff, H. H. (1996). Face recognition under varying poses: The role of texture
 495 and shape. *Vision Research*, 36(12), 1761-1771. doi: 10.1016/0042-6989(95)00230-8

496 Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. A. (2008). Patches of face-selective cortex
 497 in the macaque frontal lobe. *Nature Neuroscience*, 11, 877. doi: 10.1038/nn.2158

498 Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in*
 499 *Cognitive Sciences*, 11(2), 58-64. doi: 10.1016/j.tics.2006.11.009

500 Vetter, T. (1998). Synthesis of novel views from a single face image. *International Journal of*
501 *Computer Vision*, 28(2), 103-116. doi: 10.1023/a:1008058932445

502 Wallis, G. (2013). Towards a unified model of face and object recognition in the human visual
503 system. [Original Research]. *Frontiers in Psychology*, 4, 1-25. doi:
504 10.3389/fpsyg.2013.00497

505 Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bülthoff, H. (2009). Learning illumination-
506 and orientation-invariant representations of objects through temporal association. *Journal of*
507 *Vision*, 9(7), 6-6. doi: 10.1167/9.7.6

508 Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*,
509 3(1), 22-31. doi: 10.1016/s1364-6613(98)01261-3

510 Wallis, G., Siebeck, U. E., Swann, K., Blanz, V., & Bülthoff, H. H. (2008). The prototype effect
511 revisited: Evidence for an abstract feature model of face recognition. *Journal of Vision*, 8(3).
512 doi: 10.1167/8.3.20

513 Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from
514 error visibility to structural similarity. *IEEE Trans Image Process*, 13(4), 600-612.

515 Wood, J. N. (2013). Newborn chickens generate invariant object representations at the onset of
516 visual object experience. *Proceedings of the National Academy of Sciences*, 110(34), 14000-
517 14005. doi: 10.1073/pnas.1308246110

518 Wood, J. N. (2015). Characterizing the information content of a newly hatched chick's first visual
519 object representation. *Developmental Science*, 18(2), 194-205. doi: 10.1111/desc.12198

520 Zoccolan, D. (2015). Invariant visual object recognition and shape processing in rats. *Behavioural*
521 *Brain Research*, 285(Supplement C), 10-33. doi: doi.org/10.1016/j.bbr.2014.12.053

522 Zoccolan, D., Oertelt, N., DiCarlo, J. J., & Cox, D. D. (2009). A rodent model for the study of
523 invariant visual object recognition. *Proceedings of the National Academy of Sciences*,
524 106(21), 8748-8753. doi: 10.1073/pnas.0811583106

526 **FIGURE CAPTIONS**

527

528 **Figure 1:** Experimental stimuli. (A) Two examples of faces and the rotated views used for testing.
529 (B) Frontal views of all stimuli pairs and corresponding reward condition for all experiments. All
530 face images were provided by the Max-Planck Institute for Biological Cybernetics in Tübingen,
531 Germany.

532 **Figure 2:** Analysis of stimuli similarity. Structural Similarity index (SSIM) calculated for face pairs
533 presented to each fish. An index=1 indicates images are identical.

534 **Figure 3:** The mean and individual stimulus selection frequency results are shown for Experiment 1
535 and Experiment 2. Error bars represent 90% confidence intervals. The red line at 50% represents the
536 expected S+ selection frequency if the fish were choosing at random.

537 **Figure 4:** Relative mean response time (RT) in seconds for individual fish for Experiment 1 and 2.
538 Bars represent the 95% confidence interval. Response times in the figure have been transformed to
539 the percent of the total time taken per fish for all trials. The data is presented this way as there are
540 large differences in the actual response times of the fish. All analyses were conducted on actual
541 response times.