Young children do not require perceptual-motor feedback to solve Aesop's Fable tasks

Rachael Miller [1] , Sarah A Jelbert [Corresp., 1] , Elsa Loissel [1] , Alex H Taylor [2] , Nicola S Clayton [1]

[1] Department of Psychology, University of Cambridge, Cambridge, United Kingdom

[2] School of Psychology, University of Auckland, Auckland, New Zealand

Corresponding Author: Sarah A Jelbert
Email address: saj48@cam.ac.uk

Aesop's Fable tasks – in which subjects drop objects into a water-filled tube to raise the water level and obtain out-of-reach floating rewards – have been used to test for causal understanding of water displacement in both young children and non-human animals. However, a number of alternative explanations for success on these tasks have yet to be ruled out. One hypothesis is that subjects may respond to perceptual-motor feedback: repeating those actions which bring the reward incrementally closer. Here, we devised a novel, forced-choice version of the Aesop's Fable task to assess whether or not subjects can solve water displacement tasks when visual feedback is removed. In the current experiment, 5-9 year old children were tested in six different conditions in which we either varied object properties (floating vs. sinking, hollow vs. solid, large vs. small objects and too large vs. small objects), the water level (high vs. low) and/or the tube size (narrow vs. wide). We found that children aged 8-9 years old were able to mentally simulate the result of their actions in most of the water displacement tasks, making correct choices from their first trial onwards. Children aged 5-7 years solved two conditions on their first trial (large vs. small objects and high- vs. low-water levels), and learnt to solve most of the remaining conditions over five trials. The developmental pattern shown here is comparable to previous studies using the standard Aesop's Fable task, where 8 year olds are typically successful from their first trial and 5-7 year olds learn to pass over five trials. Thus, our results indicate that children do not depend on visual feedback to solve these water displacement tasks. The forced-choice paradigm we describe could be used comparatively to test whether or not non-human animals require visual feedback to solve water displacement tasks.

3 Young children do not require perceptual-motor feedback to solve Aesop's Fable tasks

5 Rachael Miller [1†] & Sarah A. Jelbert*[†1], Elsa Loissel[1], Alex H. Taylor[2], Nicola S. Clayton[1]

8 Affiliations:

9 [1] Department of Psychology, University of Cambridge, Cambridge, UK

10 [2] School of Psychology, University of Auckland, Auckland, New Zealand

12 *Correspondence: Sarah Jelbert (saj48@cam.ac.uk)

13 [†]joint first authors

Abstract

Aesop's Fable tasks – in which subjects drop objects into a water-filled tube to raise the water level and obtain out-of-reach floating rewards – have been used to test for causal understanding of water displacement in both young children and non-human animals. However, a number of alternative explanations for success on these tasks have yet to be ruled out. One hypothesis is that subjects may respond to perceptual-motor feedback: repeating those actions which bring the reward incrementally closer. Here, we devised a novel, forced-choice version of the Aesop's Fable task to assess whether or not subjects can solve water displacement tasks when visual feedback is removed. In the current experiment, 5-9 year old children were tested in six different conditions in which we either varied object properties (floating vs. sinking, hollow vs. solid, large vs. small objects and too large vs. small objects), the water level (high vs. low) and/or the tube size (narrow vs. wide). We found that children aged 8-9 years old were able to mentally simulate the result of their actions in most of the water displacement tasks, making correct choices from their first trial onwards. Children aged 5-7 years solved two conditions on their first trial (large vs. small objects and high- vs. low-water levels), and learnt to solve most of the remaining conditions over five trials. The developmental pattern shown here is comparable to previous studies using the standard Aesop's Fable task, where 8 year olds are typically successful from their first trial and 5-7 year olds learn to pass over five trials. Thus, our results indicate that children do not depend on visual feedback to solve these water displacement tasks. The forced-choice paradigm we describe could be used comparatively to test whether or not non-human animals require visual feedback to solve water displacement tasks.

Introduction

37

38   Recently, a number of comparative, non-linguistic studies have been conducted to determine

39 what young children and non-human animals understand about elements of water displacement.

40 Researchers using the floating peanut task have demonstrated that children and some great apes

41 will spontaneously pour or spit water into a tube in order to bring a floating peanut within reach.

42 Among children, only a small proportion of 4 year olds, but up to 75% of 8 year olds, recognise

43 that they can use water as a tool to raise the level of floating rewards (Mendes, Hanus & Call,

44 2007; Hanus et al., 2011). A related line of research has used the Aesop's Fable paradigm to

45 assess whether or not children and non-human animals possess a causal understanding of water

46 displacement (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Clayton, 2014; Jelbert, Taylor

47 & Gray, 2015). These tasks are analogous to Aesop's famous tale in which a thirsty crow drops

48 stones into a pitcher of water to raise the water level until it is high enough for the bird to drink.

49 In Aesop's Fable tasks, subjects are typically presented with a choice of objects to drop into

50 tubes, or a choice of tubes to drop objects into, where one option is the most (or only) functional

51 choice to raise the water level and obtain a floating out-of-reach reward.

52   When presented with versions of this task, 4-7 year old children appear to learn, over the

53 course of 5 trials, which options will allow them to obtain the reward (Cheke, Loissel & Clayton,

54 2012). Across three conditions, 4-7 year olds could learn to drop stones into a tube containing

55 water, rather than one containing sawdust, and 5-7 year olds could learn to drop objects than

56 sank, rather than objects that floated on the water's surface. The majority of children aged 7 and

57 over, but few younger children, also learnt to pass a task including counter-intuitive causal cues,

58 where dropping a stone into one tube also raised the water level in a second tube via a concealed

59 connection (Cheke, Loissel & Clayton, 2012). More recently, 5-7 year old children failed to pass

60   a more difficult version of this task involving solid and hollow objects within 5 trials, but learnt

61   to do so within 20 trials (Miller et al., 2016).

62   Strikingly, performance on Aesop's Fable tasks by some bird species – primarily corvids –

63   has been shown to rival that of 5-7 year old children (Bird & Emery, 2009; Cheke, Bird &

64   Clayton, 2011; Taylor et al., 2011). Rooks (*Corvus frugilegus*), Eurasian jays (*Garrulus*

65   *glandarius*) and New Caledonian crows (*Corvus moneduloides*), for example, have all been

66   tested on various Aesop's Fable tasks. These experiments revealed that corvids will drop sinking

67   rather than floating objects into water-filled tubes, will drop large rather than small objects, and

68   drop solid objects (that displace a large amount of water) rather than hollow objects (that

69   displace only a small amount). They preferentially drop objects into tubes containing water,

70   rather than tubes containing sand, and drop objects into tubes with a high- rather than a low-

71   water level (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Taylor et al., 2011; Jelbert et

72   al., 2014; Logan et al., 2014). In most of these cases, birds do not solve the task on their very

73   first trial, but they do learn to solve the tasks over a small number of trials, rapidly learning to

74   exclusively select the most (or only) functional option. Thus, their behaviour is highly similar to

75   that of 5-7 year old children. Only by the age of 8 years do children reliably choose correct

76   options on their first trial, at which point children's performance clearly differs from that of

77   corvids (Cheke, Loissel & Clayton, 2012).

78   Although birds and 5-7 year old children show similar learning patterns on the Aesop's

79   Fable task, to date, it remains unclear whether their comparable performance is underpinned by

80   similar cognitive mechanisms (see Clayton, 2014; Jelbert, Taylor & Gray, 2015 for review).

81   Success on the Aesop's Fable tasks could be achieved through using a causal understanding of

82   water displacement and an ability to mentally simulate the effect that dropping objects will have

83   on the water level in each tube. However, the pattern of learning shown by birds and 5-7 year old

84   children could also be explained by other mechanisms. A common feature of the Aesop's Fable

85   tasks is that the reward incrementally moves closer to the subject's reach with each stone that is

86   dropped into the tube to raise the water level. Therefore, subjects could learn to solve these tasks

87   by repeating those actions which bring the reward incrementally closer – i.e. by responding to

88   perceptual-motor feedback (Taylor & Gray, 2009; Cheke, Bird & Clayton, 2011; Jelbert, Taylor

89   & Gray, 2015). This type of feedback is thought to underpin the seemingly 'insightful' behaviour

90   by which birds spontaneously learn to pull up strings to bring in attached rewards (New

91   Caledonian crows: Taylor et al., 2010; Taylor, Knaebe & Gray, 2012; common ravens, *Corvus*

92   *corax*: Heinrich & Bugnyar, 2005; California scrub jays, *Aphelocoma californica*: Hofmann,

93   Cheke & Clayton, 2016), and is a plausible explanation for the birds' behaviour on water

94   displacement tasks. For example, Cheke et al. (2011) found that 1 out of 2 Eurasian jays could

95   pass an arbitrary task where a reward was pushed incrementally towards the subject each time

96   they dropped a stone into an L-shaped apparatus. However, they failed to learn a task with the

97   same reward schedule where the subject was given a reward by the experimenter once they had

98   dropped a certain number of stones into one of two coloured tubes. This suggests that, in some

99   problem-solving situations, corvids potentially learn by attending to the position of a reward after

100   each action they make.

101   It is currently unclear whether responding to perceptual-motor feedback contributes to

102   children's performance on Aesop's Fable tasks. The increase in first trial success that occurs

103   between 7 and 8 years is roughly in line with performance on classic Piagetian conservation of

104   volume tasks, which are typically passed around age 7 (Piaget, 1930, 1974). However, Cheke

105   and colleagues found that performance on one conservation of volume task – in which water was

106 poured from a short, wide container into a thin, narrow container and children were asked

107 whether the amount of water was now more, less or the same – did not predict children's

108 performance on various Aesop's Fable tasks (Cheke, Loissel & Clayton, 2012). This suggests

109 that some children might pass these water displacement tasks by attending to covariation cues,

110 rather than by using an understanding of water displacement. With this issue in mind, in the

111 current study we devised a novel forced-choice version of the Aesop's Fable paradigm, capable

112 of determining whether subjects can solve Aesop's Fable tasks when opportunities for

113 perceptual-motor feedback are removed. This task was designed to be appropriate for use with

114 both human and non-human populations.

115   In our forced-choice paradigm, children were presented with two versions of a modified

116 water-tube apparatus, where a sliding barrier could be pushed to release a set of pre-positioned

117 objects, all at once, into a water-filled tube. On each trial, either the two water-tubes or the two

118 sets of objects varied, and children could choose one apparatus to interact with only. Tubes were

119 transparent on one side and opaque on the other. Thus, after the child had indicated their tube of

120 choice, but before they slid the barrier to release the objects, the tube could be rotated to the

121 opaque side, which denied the child visual access to the water level rising when the objects

122 dropped into the tube. Because of these design differences, here, subjects could not succeed by

123 observing the effect that dropping an object had on the water level. Thus, first trial success would

124 indicate that the participant can reason causally – mentally simulating which apparatus to select

125 to obtain the reward, before they received any kind of feedback from their actions.

126   Across six conditions, we either varied the properties of the objects (small vs. large objects,

127 too large vs. small objects, sinking vs. floating objects and hollow vs. solid objects) or the

128 properties of the tubes that were presented (narrow vs. wide tubes, and high vs. low water levels)

129   to assess what children understand about water displacement. Two conditions had been used

130   previously with young children (sinking vs. floating & solid vs. hollow objects: Cheke, Loissel

131   & Clayton, 2012; Miller et al., 2016) and four had not. Five of six tasks (all except the too large

132   vs. small objects condition) had been used in previous studies with corvids (Taylor et al., 2011;

133   Jelbert et al., 2014; Logan et al., 2014). Where possible, the conditions that we presented were

134   designed to counterbalance each other, so that the functional choice in one condition was non-

135   functional in another condition, and therefore a preference for one particular object or tube could

136   be ruled out (Jelbert, Taylor & Gray, 2015). For example, in the narrow vs. wide tube condition,

137   only the narrow tube was functional, whereas in the high- vs. low-water level condition, only the

138   wide tube was functional. Children received five trials in each condition, which allowed us to

139   also assess whether they could learn to solve these versions of the Aesop's Fable tasks over time,

140   despite not witnessing the water level rising. If children's success is dependent on perceptual-

141   motor feedback, we expected children to perform more poorly on the current forced-choice tasks

142   than in previous studies where visual feedback was freely available.

143

144

145   **Methods**

146   **Subjects**

147   Subjects were 55 children aged between 5 and 9 years old: 10 5-year olds, 13 6-year olds, 11

148   7-year olds, 11 8-year olds and 10 9-year olds, of which 27 were male and 28 were female.

149   Children were recruited and tested at five primary schools in Cambridgeshire between February

150   and May 2016.

151

Apparatus

In all experimental trials, subjects were presented with two water-filled Perspex tubes each containing a magnetic floating token, which could be retrieved using a magnetic 'fishing rod' when the water level reached 60mm from the top of the tube. A removable slanting Perspex tube containing objects (referred to as the 'object-tube') was placed on top of each tube (diameter 5cm, Figure 1). Subjects could slide a barrier at the base of the object-tube to release the objects into the water-filled tube. Different tubes and different objects were used in each condition, and details of these are provided in the experimental procedures below.

Tokens were small pieces of cork attached to a small magnet ($\sim$5-10mm²), which would float on the water's surface. The fishing rod comprised 60mm of string with a small magnet attached to one end and a sheet of clear plastic ($\sim$90x60mm) at the other, which prevented the rod from being inserted fully into the tube. Using a fishing rod ensured that the token was accessible when it reached a standardized distance across all conditions and for all children. To maintain motivation, we used a 'token track' reward trail, where subjects could move a figure along the trail for each token they retrieved during training and experimental trials. Stickers were placed intermittently across the trial (approximately every 3rd step) for the child to obtain.

Pre-Training

Subjects received two training steps, first to learn to slide the barriers attached to the object-tube, and second to learn that only one option was rewarded, and that only one choice was permitted. In the first training step, subjects were presented with a Perspex collapsible platform apparatus (as per Bird & Emery 2009), and observed the experimenter placing the slanting object-tube on top of the apparatus. The experimenter demonstrated inserting a training object (a

175 plastic, light blue oblong, 25x15x10mm) into the object-tube. Then, the experimenter pushed the

176 barrier to allow the object to drop down and collapse the platform inside the apparatus to release

177 a token. We used this opportunity to explain that the tokens were equivalent to one step along the

178 token track and if they reached a sticker, then the sticker belonged to the subject. The subject

179 was then asked to drop the blue training object into the collapsible platform apparatus by

180 themselves and obtain a sticker, and show that they were able to push the barrier to release the

181 object.

182   In the second training step, subjects received five trials of a sawdust vs. water condition. We

183 presented the subject with 2 vertical Perspex tubes with the object-tube attachment in position

184 and demonstrated inserting 3 training objects into the object-tubes. One vertical tube was filled

185 with small blue stones, and the other tube was filled with water to the same height. We used

186 small blue stones in place of sawdust in order to avoid any potential allergies to this substrate in

187 the children. Subjects completed 5 trials, or 3 correct consecutive trials, where they could choose

188 only one tube by pushing the barrier, which released all of the objects into the vertical tube. They

189 could then use the fishing rod to attempt to obtain the token. This step allowed us to ensure the

190 subject understood that they could only make one choice of tube and that there was only one

191 correct choice (in this case the water-filled tube). They could also practice using the fishing rod

192 to obtain the token.

193

194 **Test Conditions**

195   Immediately after training, subjects began the experimental trials. They received 30 trials in

196 total (6 conditions, 5 trials per condition), generally completing 18 trials in session 1 (which

197 lasted 30-45 minutes), and 12 trials in a second session the following day (lasting 20-30

198   minutes). In each condition, a different set of tubes or objects were used (see Supplementary

199   Movie for example trials).

200   **Condition 1: Large vs. Small objects.** Two identical medium sized water-filled tubes

201   (diameter = 5cm, height = 15cm) were presented, with the water set to the same level. One tube

202   was presented with a single large object (grey clay sphere: 40mm diameter), and one tube with a

203   single small object (grey clay sphere: 13mm diameter). When released, the large object raised the

204   water level sufficiently to bring the token within reach, but the small object did not.

205   **Condition 2: Too large vs. Small objects in narrow tubes**. In the second condition, 4 large

206   and 4 small objects were presented, equivalent to those used in Condition 1. Here, two identical

207   *narrow* water-filled tubes were used (diameter = 3.5cm, height = 15cm). The water level was

208   equivalent in both. Because narrower tubes were used, the large object was now too large to fit

209   inside the water-filled tube, and could not displace any water. The subject should instead choose

210   the 4 small objects which would raise the water level sufficiently to bring the token within reach.

211   **Condition 3: Floating vs. Sinking objects.** Here, two identical medium sized tubes were

212   presented, as used in Condition 1. One set of heavy, sinking objects (clay spheres, 20mm

213   diameter), and one set of light, floating objects were presented (polystyrene spheres, 20mm

214   diameter). Heavy objects would sink and displace the water in the tube, whereas light objects

215   floated on the surface of the water and were therefore non-functional. To make the objects

216   visually distinct, one set was painted white and one painted black, with the colours

217   counterbalanced across children. Unlike the other conditions, the relevant property here - weight

218   - was not directly detectable through observation; therefore, in this condition the child was given

219   the opportunity to handle each set of objects and place them into the object-tubes at the start of

220   the trial.

221   **Condition 4: Hollow vs. Solid objects.** Here, two identical medium sized tubes were

222   presented, as used in Condition 1. One tube was presented with 3 grey hollow objects (metal

223   cubes: 20mm^3), and one tube with 3 grey solid objects of the same size and shape (clay squares:

224   20mm^3). Hollow objects displaced only a small amount of water in the tube, but solid objects

225   would raise the water sufficiently to bring the token within reach.

226   **Condition 5: Wide vs. Narrow tubes.** In this condition, the properties of the tubes were

227   varied. One wide (diameter = 7cm, height = 15cm) and one narrow (diameter = 3.5cm, height =

228   15cm) water-filled tube were presented. The water level was equal for both tubes, and each were

229   presented with an identical set of three medium sized grey objects in place (clay spheres: 20mm

230   diameter). If the subject chose to release objects into the narrow tube, the water level would rise

231   by enough to bring the floating token within reach, but not in the wide tube.

232   **Condition 6: High vs. Low water levels in wide and narrow tubes.** This condition was

233   identical to Condition 5 except that the water levels in each tube varied. Here, the wide tube was

234   presented with a higher initial water level than in Condition 5, meaning that, now, objects

235   released into the wide tube would bring the floating token within reach. The narrow tube was

236   presented with a very low water level, and was therefore non-functional.

237

238   **Test Procedure**

239   On each experimental trial, two water-filled tubes were presented, each containing a floating

240   out-of-reach token. One side of the tube was transparent, and one side was opaque. Tubes were

241   initially presented with the transparent side facing the child, with the water level and token both

242   visible. The object-tubes were pre-attached, and the subject observed as the experimenter

243   inserted the objects into the object-tube (with the exception of Condition 3: floating vs. sinking

244 objects, where the child handled the objects and inserted them by themselves). In each condition,

245 the water level was set so that for the correct choice, dropping the objects into the tube would

246 raise the level sufficiently to allow the token to be removed, but not for the incorrect choice. The

247 subject was asked to make their choice by pointing at the specific tube. The experimenter

248 removed the tube that wasn't chosen. They then rotated the chosen water-filled tube so that the

249 opaque side faced the subject, and indicated that the subject could now slide the barrier to release

250 the objects into the water-filled tube. Hence, visual access was blocked after the choice was

251 made, but before the resulting action of dropping the objects into the tube. The experimenter

252 removed the object-tube, and the subject was then able to attempt to fish the token from the tube

253 of choice using the fishing rod.

254   Trials of each condition were presented in a pseudorandomised order. The trials were

255 intermixed within each block of 6 conditions, so the correct choice in the following trial (e.g.

256 small objects) was not the same as in the previous trial (e.g. large objects). The positions of the

257 correct choice (left, right) were counterbalanced within and between all test blocks so that the

258 correct choice was on the left three times and the right three times within each block of 6 trials,

259 though not on the same side more than two times in a row within a block. The experimenter was

260 RM or EL, with RM, EL or SAJ assisting during testing by re-setting tubes (emptying out the

261 water and objects, replacing the water at correct level ready for next trial) as required. The

262 experimenter followed a set script and procedure with each subject.

263   Conditions 1, 2 and 5, 6 were selected so that for each variation in size of tube or object,

264 both options would be correct in some trials, though not in others, depending on the context of

265 the trial. For example, the narrow tube would be correct vs. the wide one when the water level

266   was equal, whereas the wide tube would be correct vs. the narrow one when the water level was

267   unequal.

268

Data Analysis

270   We recorded the choice per trial for each subject as 'correct' or 'incorrect'. All test sessions

271   were coded live as well as being video-recorded unless parental consent requested otherwise.

272   The full data set is available on FigShare:10.6084/m9.figshare.3899787.

273   We used General Linear Mixed Models (GLMM) in SPSS (v.21) to assess which factors

274   influenced accuracy in the children. Accuracy was a binary variable of whether the subject

275   correctly solved the trial (1) or not (0), and was entered as a dependent variable in the models.

276   We ran two models as we had two measures of interest: (1) accuracy on the first trial and (2)

277   accuracy across all five trials. Condition (1-6), age (5-9 years), gender (male/female) and trial

278   number (1-5, model 2 only) were entered as main effects, and subject ID was entered as a

279   random effect. Data were fitted using a binomial distribution with a logit link function. A

280   stepwise backward elimination analysis removed the least significant variable in each step until

281   the final model was obtained. Each dropped variable was then re-added separately to the final

282   model to check that it remained significant ($p > 0.05$).

283

Ethics Statement

285   The study was conducted under the European Research Council Executive Agency Ethics

286   Team (application: 339993-CAUSCOG-ERR) and University of Cambridge Psychology

287   Research Ethics Committee (pre.2013.109). Informed written consent was obtained from parents

288   prior to participation of the child. The parents of the child identified in the supplementary movie

289   gave their informed written consent for this information to be published.

290

291   **Results**

292   On trial 1, we found significant effects of condition and age on accuracy (correct vs.

293   incorrect choice; Table 1). Accuracy on the first trial increased with age, and was significantly

294   lower in condition 2 (too large vs. small objects) and condition 5 (wide vs. narrow tubes) than in

295   the other conditions (Table 1). Across all 5 trials, we also found significant effects of condition,

296   age and trial number on accuracy (correct vs. incorrect choices; Table 2), with accuracy

297   increasing with both age (Figure 2) and trial number. Accuracy across all 5 trials was

298   significantly lower only in condition 5 (wide vs. narrow tubes) compared to the other conditions

299   (Figure 3; Table 2). There was no significant effect of gender on accuracy on trial 1 or across all

300   trials (Table 2; Table 3).

301   We further explored correct choices within each condition per age group (5-7, 8-9 years old;

302   Table 3). In condition 1 (large vs. small objects) and condition 6 (high vs low water level),

303   children made significantly more correct choices on *trial 1* in both the 5-7 and 8-9 years old age

304   groups. In condition 2 (too large vs. small condition), children significantly made the correct

305   choices *across all trials* in both age 5-7 years and 8-9 years, but not on trial 1. In condition 5

306   (wide vs. narrow tubes), only children aged 8-9 years significantly made correct choices *across*

307   *all trials*, and not from trial 1.

308   For the two conditions that have been previously tested in children using the standard

309   Aesop's Fable task: floating vs. sinking and hollow vs. solid objects, we found that children

310   made significantly more correct choices *across all 5 trials* from age 5-7, and from *trial 1* from

311   age 8-9 (Table 3). These results are similar to previous findings testing the sinking vs. floating

312   condition in a standard Aesop's Fable paradigm (Cheke, Loissel & Clayton, 2012; Miller et al.,

313   2016), but the children performed better than in previous tests using solid vs. hollow objects

314   (Miller et al., 2016), which found that children struggled to select the correct option over five

315   trials, but learnt to do so over twenty trials.

316

317   **Discussion**

318   In the current study, we developed a forced-choice Aesop's Fable paradigm, comprised of

319   six different conditions, which could not be solved by responding only to perceptual-motor

320   feedback. Subjects were not able to observe the water level rising when objects were dropped

321   into the water-filled tubes, and had to select only one set of objects or one type of tube, into

322   which all objects were dropped at once. To solve these tasks on their very first trial, children

323   needed to mentally simulate the effect that dropping objects would have on the water level of the

324   tubes. Over five trials, children received feedback about the success of their actions (whether or

325   not the token could now be reached using the fishing rod), but at no point did they observe the

326   water level rising.

327   We found that there was a significant effect of both age and condition on accuracy, and that

328   accuracy increased significantly across trials. Notably, children's performance in the current

329   paradigm followed a similar developmental pattern to that found on previous versions of Aesop's

330   Fable tasks, where responding to perceptual-motor feedback had been a possible strategy for

331   success (Cheke, Loissel & Clayton, 2012; Miller et al., 2016). We found that children aged 8-9

332   years passed the majority of conditions on their first trial (4 of the 6 conditions), while 5-7 year

333   olds passed two conditions on their first trial, and learnt to solve three of the four remaining

334   conditions over five trials. This is comparable to standard Aesop's Fable tasks, which 8 year old

335   children typically pass on their first trial and 5-7 year olds can learn to pass over five trials. The

336   finding that children's performance was not impaired in the current study, relative to previous

337   standard Aesop's Fable tasks, indicates that children do not require visual feedback of the water

338   level rising to solve these types of water displacement tasks.

339   Our results also highlight that the six conditions we presented were not equally easy for

340   children to solve, and therefore may have each tapped slightly different cognitive processes.

341   Younger children, aged 5-7 years, were able to pass the large vs. small condition and the high-

342   vs. low-water level conditions on their first trial. These two conditions have not previously been

343   used with children, though corvids have also consistently solved versions of these tasks over a

344   small number of trials (Bird & Emery, 2009; Taylor et al., 2011; Jelbert et al., 2014; Logan et al.,

345   2014). Given that younger children passed these two conditions only, it is possible they may

346   have been solved using simpler mechanisms than the other variations of the task. For example,

347   young children may have selected the tube with a high-water level simply because the token was

348   already closest to the top of this tube, not because they imagined the effect that dropping objects

349   would have on the future water level. Equally, young children may have had a general preference

350   for the larger objects. There is some support for this as 5-7 year old children also selected the

351   large objects more often than the small objects on the too large vs. small condition, where the

352   large object could not fit into the narrow tubes and was therefore non-functional (though this

353   trend was non-significant with a Bonferroni correction). Based on this pattern of performance,

354   the evidence that children aged 5-7 mentally simulated the effects of dropping objects into the

355   tubes is equivocal.

356   Older children, aged 8-9 years, were able to solve both these tasks, and additionally were

357   able to pass the floating vs. sinking objects condition and the hollow vs. solid object condition on

358   their first trials. Younger children learnt to solve these two conditions within five trials. Here, the

359   developmental pattern on the floating vs. sinking condition is entirely in line with previous

360   research (Cheke, Loissel & Clayton, 2012), while children's performance on the hollow vs. solid

361   objects task was actually better than that observed in a previous study (Miller et al., 2016). Miller

362   and colleagues found that 5-7 year old children learnt to select solid over hollow objects over the

363   course of 20 trials, and though 8-10 year olds solved the task within 5 trials, they did not do so

364   from their very first trial. One explanation for the children's superior performance on the present

365   task is that children in the current study were allowed to select one type of object only, which

366   may have simplified their decision-making process. In line with this, in the earlier explorative

367   task making some mistakes would not typically prevent the subject from obtaining the reward;

368   thus, there was no penalty for testing out both of the presented options in early trials. Another

369   possibility, which cannot be ruled out here, is that children's performances could have been

370   scaffolded by their experience in the other concurrent test conditions. For example, when

371   obtaining rewards using large or small clay spheres, children may have gained information that

372   influenced their choices of solid over hollow cubes in this particular task. The information used

373   could be simple, such as generalising the appearance of successful objects, or more complex,

374   such as drawing inferences about the mechanics of water displacement from observing successes

375   and failures in other contexts. While this is unlikely to account for first trial successes, given that

376   trial orders were randomised, the opportunity to learn from other conditions over multiple trials

377   may have contributed to children's ability to quickly acquire the correct option in the solid vs

378   hollow task.

379   Performance in the remaining two conditions – wide vs. narrow tubes and too large vs. small

380   objects – was significantly lower than in the other conditions, across all ages. In the too large vs.

381   small condition children were less accurate on their first trial, but not over five trials. As

382   discussed, there was a non-significant trend for 5-7 year old children to prefer the large object on

383   their first trial, while 8-9 year old children chose at chance. This suggests that the children

384   initially failed to recognise that the large object would not be able to fit into the narrow tube.

385   However, once they had experienced this surprising event, they rapidly learnt to avoid choosing

386   large objects on subsequent trials of this condition. Given that the difference in the sizes of the

387   narrow tube and the large objects was quite subtle, it may be the case that children would solve

388   this condition from the first trial if the disparity had been greater. In the narrow vs. wide

389   condition performance remained significantly poorer than in other conditions both on the first

390   trial of the task and over five trials. This condition has not been previously used with children;

391   however, when tested with corvids, the majority have also struggled on this task (Jelbert et al.,

392   2014), though some passed when the number of objects was restricted (Logan et al., 2014). The

393   poor performance observed here strongly suggests that children between 5 and 9 years are not

394   yet able to accurately simulate the different effects that objects will have on the water level of

395   differently sized tubes. Thus, although by the age of 7 children can recognise that water volume

396   is conserved when it passes between two containers of different sizes (Piaget, 1930, 1974), they

397   do not yet appear to possess a full, intuitive understanding of the behaviour of liquids in different

398   containers.

399   Overall, the results reported here suggest that young children's success on Aesop's Fable

400   tasks cannot be attributed to learning from perceptual-motor feedback. Performance on this task,

401   which restricted access to visual feedback, was equivalent to (or better than) performance on

402 previous versions of the Aesop's Fable task. The pattern of results we found, in which the

403 majority of tasks were solved on the first trial by children over the age of 8, is consistent with a

404 number of studies suggesting that children only reliably solve various innovative tool-use

405 problems at around 7-8 years of age (Beck et al., 2011, 2016; Hanus et al., 2011; Nielsen, 2013).

406 Younger children performed above chance on two of the Aesop's Fable tasks on their first trial;

407 though, as discussed, it is possible they achieved this without mentally simulating the effects of

408 their actions.

409   The perceptual-motor feedback hypothesis was first suggested as an alternative explanation

410 for the impressive performance of corvids that appear to demonstrate causal reasoning on the

411 Aesop's Fable tasks (Taylor & Gray, 2009; Cheke, Bird & Clayton, 2011; Jelbert, Taylor &

412 Gray, 2015). To date, the Aesop's Fable task has been used to assess the cognitive abilities of

413 various species of corvid, including rooks, New Caledonian crows, Eurasian jays and California

414 scrub jays (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Taylor et al., 2011; Jelbert et al.,

415 2014; Logan et al., 2014, 2016; Miller et al., 2016), as well as grackles, *Quiscalus mexicanus,*

416 another innovative species of bird (Logan, 2015, 2016). A number of great apes have also been

417 tested on the comparable floating-peanut task (Mendes, Hanus & Call, 2007; Hanus et al., 2011).

418 Recently, Miller and colleagues demonstrated that in New Caledonian crows, but not in human

419 children, performance on object-choice tasks can be influenced by pre-existing preferences for

420 certain types of objects, casting some doubt on the suggestion that success on Aesop's Fable

421 tasks reflects causal understanding – at least when considering their selection of objects (Miller

422 et al., 2016). To date, it is unclear whether or not the opportunity to receive perceptual-motor

423 feedback accounts for non-human animals' ability to rapidly solve various water displacement

424 tasks. Perceptual-motor feedback has been suggested to underpin spontaneous string pulling

425 behaviour performed by birds (Taylor et al., 2010; Taylor, Knaebe & Gray, 2012; see also Jacobs

426 & Osvath, 2015; Hofmann, Cheke & Clayton, 2016), as well as performance on certain problem-

427 solving tasks by great apes (Volter & Call, 2012). The methodology that we describe here could

428 be adopted for use with non-human animals to test whether or not their success depends on

429 visual feedback, with the present study allowing for comparison with young children. Use of this

430 paradigm would help us to understand the learning mechanisms that might underpin the

431 remarkable performance of certain species on water displacement tasks.

432

433 **Acknowledgements**

434   We would like to thank the staff, parents and children at Great Abington Primary School,

435 Holywell C of E Primary School, Sutton C of E Primary School, Stretham Community Primary

436 School and Spinney Primary School in Cambridgeshire for their participation in this study.

437 Thank you to Ian Millar for help in apparatus construction.

438

439

440

References

Beck SR., Apperly IA., Chappell J., Guthrie C., Cutting N. 2011. Making tools isn't child's play. *Cognition* 119:301–306. DOI: 10.1016/j.cognition.2011.01.003.

Beck SR., Williams C., Cutting N., Apperly IA., Chappell J. 2016. Individual differences in children's innovative problem-solving are not predicted by divergent thinking or executive functions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20150190. DOI: 10.1098/rstb.2015.0190.

Bird CD., Emery NJ. 2009. Rooks use stones to raise the water level to reach a floating worm. *Current Biology* 19:1410–1414. DOI: 10.1016/j.cub.2009.07.033.

Cheke LG., Bird CD., Clayton NS. 2011. Tool-use and instrumental learning in the Eurasian jay (Garrulus glandarius). *Animal cognition* 14:441–455. DOI: 10.1007/s10071-011-0379-4.

Cheke LG., Loissel E., Clayton NS. 2012. How do children solve Aesop's Fable? *PloS one* 7:e40574. DOI: 10.1371/journal.pone.0040574.

Clayton NS. 2014. Ways of thinking: From crows to children and back again. *Quarterly Journal of Experimental Psychology* 218:1–33. DOI: 10.1080/17470218.2014.943673.

Hanus D., Mendes N., Tennie C., Call J. 2011. Comparing the performances of apes (Gorilla gorilla, Pan troglodytes, Pongo pygmaeus) and human children (Homo sapiens) in the floating peanut task. *PloS one* 6:e19555. DOI: 10.1371/journal.pone.0019555.

Heinrich B., Bugnyar T. 2005. Testing problem solving in ravens: string-pulling to reach food. *Ethology* 111:962–976.

Hofmann MM., Cheke LG., Clayton NS. 2016. Western scrub-jays (Aphelocoma californica) solve multiple-string problems by the spatial relation of string and reward. *Animal Cognition* 19:1103–1114. DOI: 10.1007/s10071-016-1018-x.

Jacobs IF., Osvath M. 2015. The string-pulling paradigm in comparative psychology. *Journal of Comparative Psychology* 129:89–120.

Jelbert SA., Taylor AH., Cheke LG., Clayton NS., Gray RD. 2014. Using the Aesop's Fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PLoS ONE* 9:e92895. DOI: 10.1371/journal.pone.0092895.

Jelbert SA., Taylor AH., Gray RD. 2015. Investigating animal cognition with the Aesop's Fable paradigm: Current understanding and future directions. *Communicative and Integrative Biology* 8:1–6. DOI: 10.1080/19420889.2015.1035846.

Logan CJ. 2015. Innovation frequency does not indicate behavioral flexibility in great-tailed grackles. *Cognition* 30:31. DOI: 10.1007/s13398-014-0173-7.2.

Logan CJ. 2016. Behavioral flexibility and problem solving in an invasive bird. *PeerJ* 4:e1975. DOI: 10.7717/peerj.1975.

476   Logan CJ., Harvey BD., Schlinger BA., Rensel M. 2016. Western scrub-jays do not appear to
477   attend to functionality in Aesop's Fable experiments. *PeerJ* 4:e1707. DOI:
478   10.7717/peerj.1707.

479   Logan CJ., Jelbert SA., Breen AJ., Gray RD., Taylor AH. 2014. Modifications to the Aesop's
480   Fable paradigm change New Caledonian crow performances. *PLoS ONE* 9:e103049. DOI:
481   10.1371/journal.pone.0103049.

482   Mendes N., Hanus D., Call J. 2007. Raising the level: orangutans use water as a tool. *Biology
483   letters* 3:453–455. DOI: 10.1098/rsbl.2007.0198.

484   Miller R., Jelbert SA., Taylor AH., Cheke LG., Gray RD., Loissel E., Clayton NS. 2016.
485   Performance in object-choice Aesop's Fable tasks are influenced by object biases in New
486   Caledonian crows but not in human children. *PLoS ONE* 11:e0168056. DOI:
487   10.1371/journal.pone.0168056.

488   Nielsen M. 2013. Young children's imitative and innovative behaviour on the floating object
489   task. *Infant and Child Development* 22:44–52. DOI: 10.1002/Icd.1765.

490   Piaget J. 1930. *The child's conception of physical causality*. London: Routledge and Kegan Paul.

491   Piaget J. 1974. *Understanding causality*. New York: Norton.

492   Taylor AH., Elliffe DM., Hunt GR., Emery NJ., Clayton NS., Gray RD. 2011. New Caledonian
493   crows learn the functional properties of novel tool types. *PloS one* 6:e26887. DOI:
494   10.1371/journal.pone.0026887.

495   Taylor AH., Gray RD. 2009. Animal cognition: Aesop's Fable flies from fiction to fact. *Current
496   Biology* 19:R731–R732. DOI: 10.1016/j.cub.2009.07.055.

497   Taylor AH., Knaebe B., Gray RD. 2012. An end to insight? New Caledonian crows can
498   spontaneously solve problems without planning their actions. *Proceedings of the Royal
499   Society B: Biological Sciences* 279:4977–4981. DOI: 10.1098/rspb.2012.1998.

500   Taylor AH., Medina FS., Holzhaider JC., Hearne LJ., Hunt GR., Gray RD. 2010. An
501   investigation into the cognition behind spontaneous string pulling in New Caledonian
502   crows. *PloS one* 5:e9345. DOI: 10.1371/Journal.Pone.0009345.

503   Volter CJ., Call J. 2012. Problem solving in great apes (Pan paniscus, Pan troglodytes, Gorilla
504   gorilla, and Pongo abelii): the effect of visual feedback. *Animal cognition* 15:923–936.
505   DOI: 10.1007/s10071-012-0519-5.

506

507   **Tables & Figures**

508

509   **Table 1.** Generalized Linear Mixed Model on factors affecting accuracy in trial 1 (model 1).

Fixed term	Full Model			Final Model		
	F	df1, df2	p	F	df1, df2	P
Condition	8.106	5, 318	**<0.001**	7.923	5, 323	**<0.001**
Age	4.442	4, 318	**0.002**	4.482	4, 324	**0.002**
Gender	0.988	1, 318	0.321	1.014	1, 323	0.315

510

511

512   **Table 2.** Generalized Linear Mixed Model on factors affecting accuracy across all 5 trials

513   (model 2).

Fixed term	Full Model			Final Model		
	F	df1, df2	p	F	df1, df2	P
Condition	21.491	4, 1635	**<0.001**	20.831	5, 1644	**<0.001**
Age	9.286	4, 1635	**<0.001**	9.453	4, 1645	**<0.001**
Gender	0.045	4, 1635	0.832	0.040	1, 1644	0.842
TrialNumber	10.420	4, 1635	**<0.001**	9.557	4, 1645	**<0.001**

514

515

516

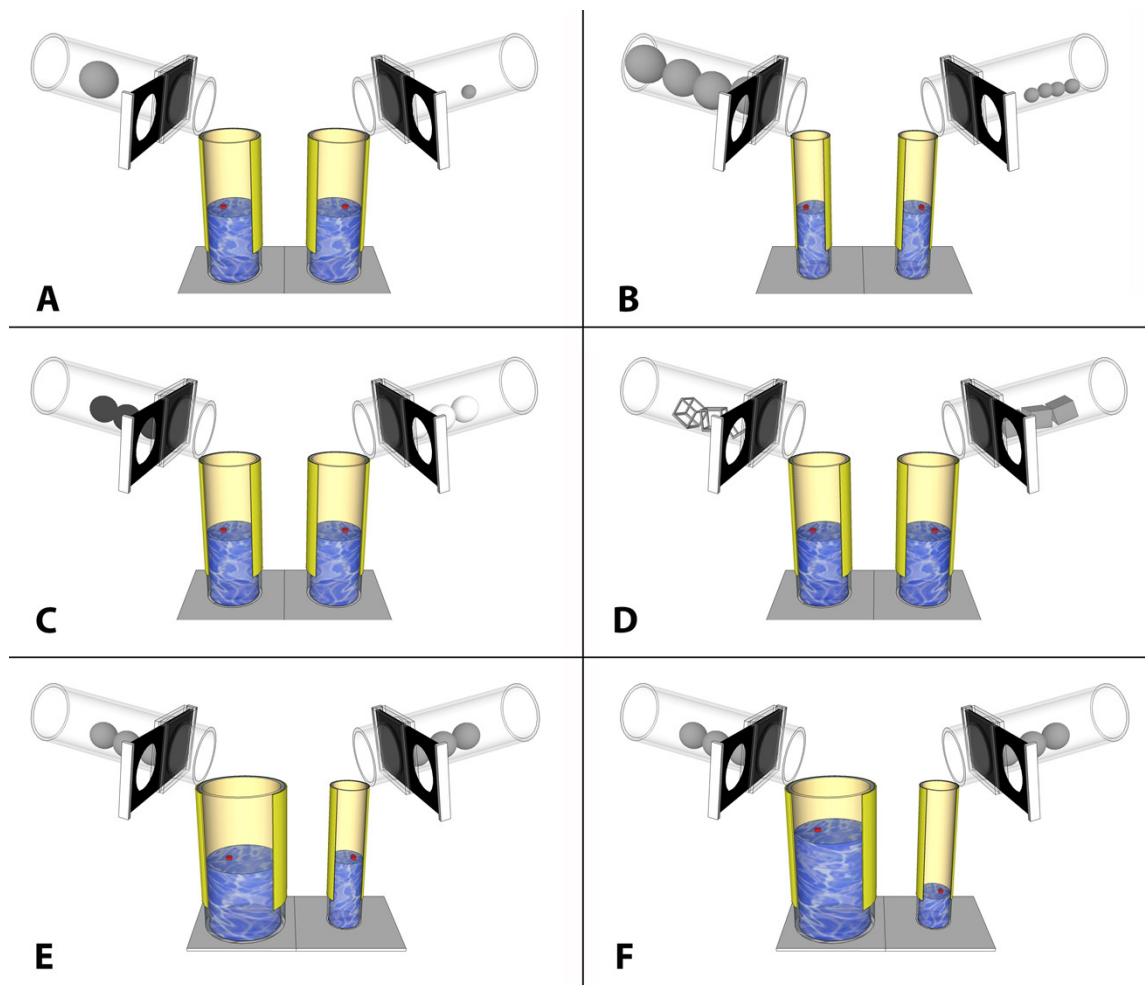
517 **Table 3.** Correct choices (%) in each condition by each age group: 5-7 years old (n=34) and 8-9

518 years old (n=21). P-values ('p') are calculated from exact two-tailed binomial tests. Significant

519 p-values are highlighted in bold. NS = not significant with a Bonferroni correction.

Age Group	Large vs. small		Too large vs. small		Floating vs. sinking		Hollow vs. solid		Wide vs. narrow		High vs. low	
	%	p	%	p	%	p	%	P	%	p	%	P
Trial 1												
5-7	76	**0.003**	29	0.024 NS	56	0.608	53	0.864	38	0.229	79	**0.001**
8-9	95	**<0.001**	48	>0.999	90	**<0.001**	86	**0.002**	62	0.383	81	**0.007**
Across all trials												
5-7	62	**0.002**	71	**<0.001**	75	**<0.001**	69	**<0.001**	41	0.017 NS	85	**<0.001**
8-9	91	**<0.001**	80	**<0.001**	94	**<0.001**	86	**<0.001**	65	**0.003**	87	**<0.001**

522

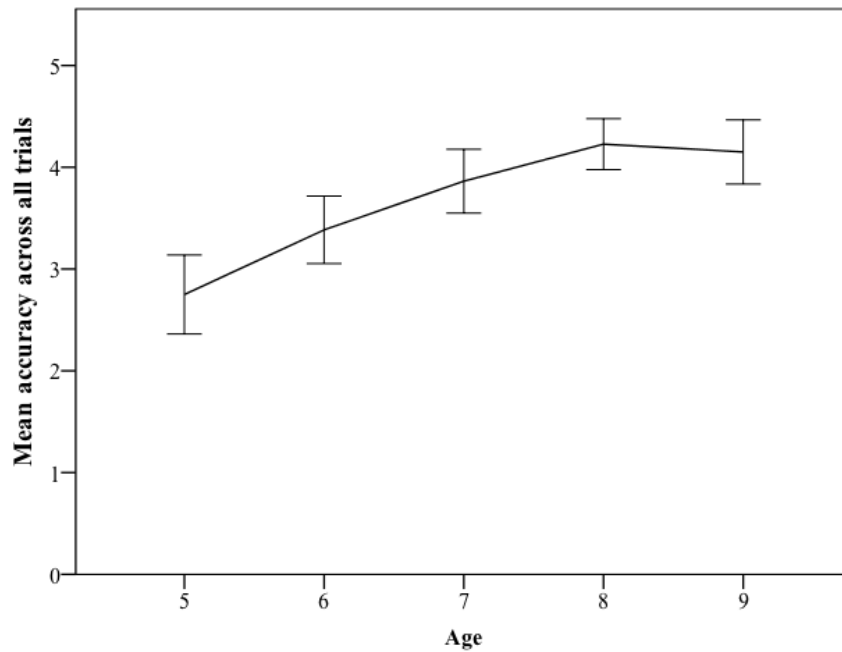
523

524 **Figure 1: Diagram of each experimental condition.** Children were first presented with two

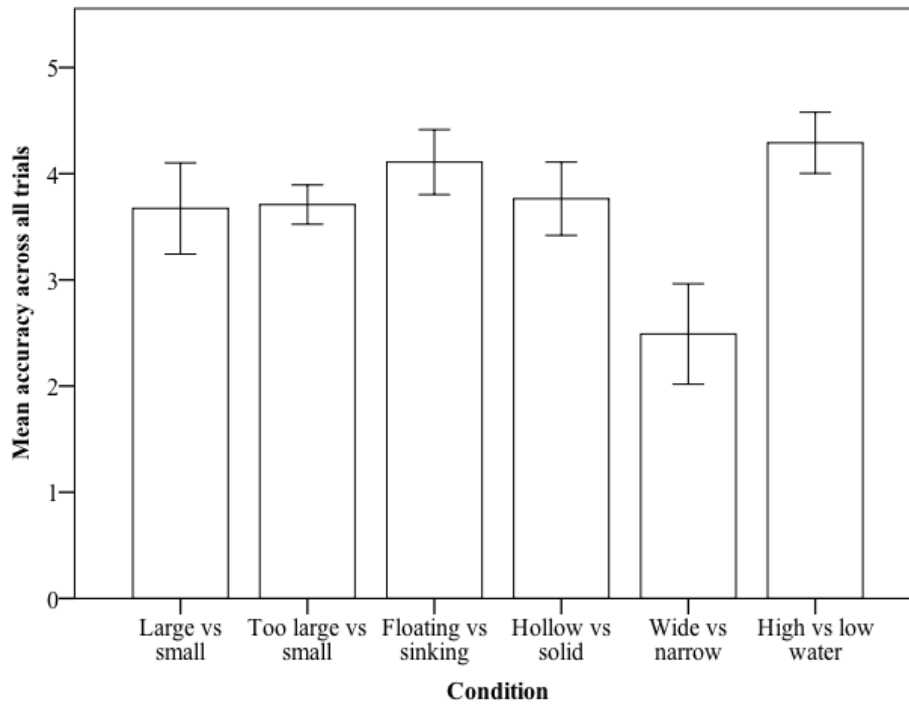
525 tubes in the orientations displayed here. After selecting one tube, the experimenter removed the

526 non-chosen tube, and rotated the chosen water-filled tube 180° to obscure the child's view of the

527 water level with the yellow covering. The child could then slide the barrier to release objects into

528 the tube. A: large vs. small, B: too large vs. small, C: floating vs. sinking, D: hollow vs. solid, E:

529 wide vs. narrow tubes, F: high vs. low water levels in wide vs. narrow tubes.

530

531   **Figure 2: Mean number of correct choices across all 5 trials for all conditions by age**

532

533   **Figure 3: Mean number of correct choices across all 5 trials by condition**