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 that bring the reward incrementally closer. Here, we devised a novel, forced-choice version of the Aesop's Fable task to assess whether subjects can solve water displacement tasks when this type of feedback is removed. Subjects had to select only one set of objects, or one type of tube, into which all objects were dropped at once, and the effect the objects had on the water level was visually concealed. In the current experiment, fifty-five 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 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 solve most of the water displacement tasks on their first trial, without any opportunity for feedback, suggesting that they mentally simulated the results of their actions before making a choice. 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 perceptual-motor 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.

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

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

4 Affiliations:

5 [1] Department of Psychology, University of Cambridge, Cambridge, UK
6 [2] School of Psychology, University of Auckland, Auckland, New Zealand

7 *Correspondence: Sarah Jelbert (saj48@cam.ac.uk)
8 †Joint first authors

9 **Abstract**

10 Aesop's Fable tasks – in which subjects drop objects into a water-filled tube to raise the water

11 level and obtain out-of-reach floating rewards – have been used to test for causal understanding

12 of water displacement in both young children and non-human animals. However, a number of

13 alternative explanations for success on these tasks have yet to be ruled out. One hypothesis is that

14 subjects may respond to perceptual-motor feedback: repeating those actions that bring the reward

15 incrementally closer. Here, we devised a novel, forced-choice version of the Aesop's Fable task to

16 assess whether subjects can solve water displacement tasks when this type of feedback is

17 removed. Subjects had to select only one set of objects, or one type of tube, into which all objects

18 were dropped at once, and the effect the objects had on the water level was visually concealed. In

19 the current experiment, fifty-five 5-9 year old children were tested in six different conditions in

20 which we either varied object properties (floating vs. sinking, hollow vs. solid, large vs. small

21 and too large vs. small objects), the water level (high vs. low) and/or the tube size (narrow vs.

22 wide). We found that children aged 8-9 years old were able to solve most of the water

23 displacement tasks on their first trial, without any opportunity for feedback, suggesting that they

24 mentally simulated the results of their actions before making a choice. Children aged 5-7 years

25 solved two conditions on their first trial (large vs. small objects and high- vs. low-water levels),

26 and learnt to solve most of the remaining conditions over five trials. The developmental pattern

27 shown here is comparable to previous studies using the standard Aesop's Fable task, where 8 year

28 olds are typically successful from their first trial and 5-7 year olds learn to pass over five trials.

29 Thus, our results indicate that children do not depend on perceptual-motor feedback to solve

30 these water displacement tasks. The forced-choice paradigm we describe could be used

31 comparatively to test whether non-human animals require visual feedback to solve water

32 displacement tasks.

33 **Introduction**

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

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

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

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

38   (Mendes, Hanus & Call, 2007; Hanus et al., 2011). Among children, only a small proportion of 4

39   year olds, but up to 75% of 8 year olds, recognise that they can use water as a tool to raise the

40   level of floating rewards (Hanus et al., 2011). A related line of research has used the Aesop's

41   Fable paradigm to assess whether children and non-human animals possess a causal

42   understanding of water displacement (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Taylor

43   et al., 2011; Clayton, 2014; Jelbert, Taylor & Gray, 2015). These tasks are analogous to Aesop's

44   famous tale in which a thirsty crow drops stones into a pitcher of water to raise the water level

45   until it is high enough for the bird to drink except for the fact that the subjects in the experiments

46   are not thirsty but a reward is placed on top of the water. In Aesop's Fable tasks, subjects are

47   typically presented with a choice of objects to drop into tubes, or a choice of tubes to drop objects

48   into, where one option is the most (or only) functional choice to raise the water level and obtain a

49   floating out-of-reach reward.

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

51   of five trials, which options will allow them to obtain the reward (Cheke, Loissel & Clayton,

52   2012). Across three conditions, 4-7 year olds could learn to drop stones into a tube containing

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

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

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

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

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

58 a more difficult version of this task involving solid and hollow objects within five trials, but

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

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

61 also, to some degree, grackles (*Quiscalus mexicanus*) – has been shown to rival that of 5-7 year

62 old children (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Taylor et al., 2011; Logan,

63 2015, 2016). Rooks (*Corvus frugilegus*), Eurasian jays (*Garrulus glandarius*) and New

64 Caledonian crows (*Corvus moneduloides*), for example, have all been tested on various Aesop's

65 Fable tasks. These experiments revealed that corvids will drop sinking rather than floating objects

66 into water-filled tubes, will drop large rather than small objects, and drop solid objects (that

67 displace a large amount of water) rather than hollow objects (that displace only a small amount).

68 They preferentially drop objects into tubes containing water, rather than tubes containing sand,

69 and drop objects into tubes with a high- rather than a low-water level (rooks: Bird & Emery,

70 2009; Eurasian jays: Cheke, Bird & Clayton, 2011; New Caledonian crows: Taylor et al., 2011;

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

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

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

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

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

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

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

78 task, to date, it remains unclear whether their comparable performance is underpinned by similar

79 cognitive mechanisms (see Clayton, 2014; Jelbert, Taylor & Gray, 2015 for review). Success on

80 the Aesop's Fable tasks could be achieved through using a causal understanding of water

81 displacement and an ability to mentally simulate the effect that dropping objects will have on the

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

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

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

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

86   by repeating those actions that bring the reward incrementally closer – i.e., by responding to

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

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

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

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

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

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

93   displacement tasks. For example, Cheke et al. (2011) found that one out of two Eurasian jays

94   could pass an arbitrary task where a reward was pushed incrementally towards the subject each

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

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

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

98   some problem-solving situations, corvids potentially learn by attending to the position of a

99   reward after each action they make.

100   It is currently unclear whether responding to perceptual-motor feedback contributes to children's

101   performance on Aesop's Fable tasks. The increase in first trial success that occurs between 7 and

102   8 years is roughly in line with performance on classic Piagetian conservation of volume tasks,

103   which are typically passed around age 7 (Piaget, 1930, 1974). However, Cheke and colleagues

104   found that performance on one conservation of volume task – in which water was poured from a

105   short, wide container into a thin, narrow container and children were asked whether the amount

106   of water was now more, less or the same – did not predict children's performance on various

107   Aesop's Fable tasks (Cheke, Loissel & Clayton, 2012). This suggests that some children might

108   pass these water displacement tasks by attending to covariation cues, rather than by using an

109   understanding of water displacement. With this issue in mind, in the current study we devised a

110   novel forced-choice version of the Aesop's Fable paradigm, capable of determining whether

111   subjects can solve Aesop's Fable tasks when opportunities for perceptual-motor feedback are

112   removed. This task was designed to be appropriate for use with both human and non-human

113   populations.

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

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

116   all at once, into a water-filled tube. On each trial, either the two water-tubes or the two sets of

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

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

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

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

121   dropped into the tube. Because of these design differences, (in that one set of objects were

122   dropped in a tube, all at once, and the effect on the water level was concealed) here, subjects

123   could not succeed by observing the effect that dropping each object had on the water level, as

124   may have been the case in previous studies.

125   Subjects received five trials in each condition, which allowed us to address two questions. The

126   first was whether participants could solve these tasks on their very first trial. First trial success

127   would indicate that the participant could reason causally – likely mentally simulating the effects

128   that objects would have on the water level – before they received any kind of feedback from their

129   actions. The second question was whether children could learn to solve the tasks over a small

130   number of trials, despite not witnessing the water level rising in response to objects being

131   dropped into the tube during the trials. At the end of each trial, children received feedback on the

132   overall success of their actions (whether the reward could now be reached), but they did not

133   receive the specific type of perceptual-motor feedback of observing the reward move

134   incrementally closer, that has been suggested as an explanation for success on these types of

135   water displacement task (Taylor & Gray, 2009; Cheke, Bird & Clayton, 2011; Jelbert, Taylor &

136   Gray, 2015).

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

138   large vs. small objects, sinking vs. floating objects and hollow vs. solid objects) or the properties

139   of the tubes that were presented (narrow vs. wide tubes, and high vs. low water levels) to assess

140   what children understand about water displacement. Two conditions had been used previously

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

142   2012; Miller et al., 2016) and four had not. Five of six tasks (all except the too large vs. small

143   objects condition) had been used in previous studies with corvids (Taylor et al., 2011; Jelbert et

144   al., 2014; Logan et al., 2014). Where possible, the conditions that we presented were designed to

145   counterbalance each other, so that the functional choice in one condition was non-functional in

146   another condition, and therefore a preference for one particular object or tube could be ruled out

147   (Jelbert, Taylor & Gray, 2015). For example, in the narrow vs. wide tube condition, only the

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

149   tube was functional. If children's success is indeed dependent on perceptual-motor feedback, we

150   expected children to perform more poorly on the current forced-choice tasks than they did in

151   previous studies where perceptual-motor feedback was freely available.

152   **Methods**

153   **Subjects**
154   Subjects were 55 children aged between 5 and 9 years old: 10 5-year olds (Mean: 5.4 years;

155   Range: 5.0-5.9 years), 13 6-year olds (M: 6.4; R: 6.0-6.9), 11 7-year olds (M: 7.6; R: 7.2-7.9), 11

156   8-year olds (M: 8.4; R: 8.0-8.9) and 10 9-year olds (M: 9.4, R:9.1-9.9), of which 27 were male

157 and 28 were female. This sample size was chosen to ensure we had a minimum of 10 children per

158 age group, and that we included a similar number of participants to the previous Aesop's Fable

159 study with children (Cheke, Loissel & Clayton, 2012). All children completed both testing

160 sessions. Children were recruited and tested at five primary schools in Cambridgeshire, serving

161 predominantly white, middle-class communities, between February and May 2016.

162 **Apparatus**

163 In all experimental trials, subjects were presented with two water-filled Perspex tubes each

164 containing a magnetic floating token, which could be retrieved using a magnetic 'fishing rod'

165 when the water level reached 60mm from the top of the tube. A removable slanting Perspex tube

166 containing objects (referred to as the 'object-tube') was placed on top of each tube (diameter

167 5cm, Figure 1). Subjects could slide a barrier at the base of the object-tube to release the objects,

168 all at once, into the water-filled tube. Different tubes and different objects were used in each

169 condition, and details of these are provided in the experimental procedures below.

170 Tokens were small pieces of cork attached to a small magnet (~5-10mm²), which would float on

171 the water's surface. The fishing rod comprised 60mm of string with a small magnet attached to

172 one end and a sheet of clear plastic (~90x60mm) at the other, which prevented the rod from being

173 inserted fully into the tube. Using a fishing rod ensured that the token was accessible at the same

174 level, across all conditions and for all children. To maintain motivation, we used a sticker reward

175 trail where subjects would receive a sticker after every few correct choices. Each time a subject

176 retrieved a token, they could move a playing piece one step on this trail and stickers were placed

177 intermittently across the trail (approximately every 3rd step) for the child to obtain.

178 **Pre-Training**

179 Subjects received two training steps, first to learn to slide the barriers attached to the object-tube,

180 and second to learn that only one option was rewarded, and that only one choice was permitted.

181 In the first training step, subjects were presented with a Perspex collapsible platform apparatus

182 (as per Bird & Emery 2009), and observed the experimenter placing the slanting object-tube on

183   top of the apparatus. The experimenter demonstrated inserting a training object (a plastic, light

184   blue oblong, 25x15x10mm) into the object-tube. Then, the experimenter pushed the barrier to

185   allow the object to drop down and collapse the platform inside the apparatus to release a token.

186   We used this opportunity to explain that the tokens were equivalent to one step along the sticker

187   reward trail and, if they reached a sticker, then the sticker belonged to the subject. The subjects

188   were then asked to drop the blue training object into the collapsible platform apparatus

189   themselves, and show that they were able to push the barrier to release the object.

190   In the second training step, subjects received five trials of a pebbles vs. water condition

191   (analogous to earlier sawdust vs. water conditions with birds). We presented the subject with two

192   medium vertical Perspex tubes (as used in the main tests) with the object-tube attachment in

193   position and demonstrated inserting three of the blue oblong training objects into the object-

194   tubes. One vertical tube was filled with small blue pebbles, and the other tube was filled with

195   water to the same height. At the start of the test the level of both the water and the pebbles was

196   too low for the child to reach the token with the magnetic fishing rod. Releasing objects into the

197   water-filled tube only would raise the water level sufficiently to bring the token within reach.

198   Here, we used small blue pebbles in place of sawdust in order to avoid any potential allergies to

199   this substrate in the children. Subjects completed five trials, or three correct consecutive trials,

200   where they could choose only one tube by pushing the barrier, which released all of the objects

201   into the vertical tube. They could then use the fishing rod to attempt to obtain the token. This step

202   allowed us to ensure the subject understood that they could only make one choice of tube and that

203   there was only one correct choice (in this case the water-filled tube). They could also practice

204   using the fishing rod to obtain the token.

205   **Test Conditions**
206   Immediately after training, subjects began the experimental trials. They received 30 trials in total

207   (six conditions, five trials per condition), generally completing 18 trials in session one (which

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

209   Condition order was pseudorandomised (see details in Test Procedure), with trials from each

210   condition spanning both testing days. In each condition, a different set of tubes or objects were

211   used (see Supplementary Movie for example trials).

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

213   = 5cm, height = 15cm) were presented, with the water set to the same level. One tube was

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

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

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

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

218   and four small objects were presented, equivalent to those used in Condition 1. Here, two

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

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

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

222   choose the four small objects which would raise the water level sufficiently to bring the token

223   within reach.

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

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

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

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

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

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

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

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

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

233   the trial.

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

235   as used in Condition 1. One tube was presented with three grey hollow objects (metal cubes:

236     $20mm^3$), and one tube with three grey solid objects of the same size and shape (clay cubes:

237     $20mm^3$). Hollow objects consisted of a wire frame without solid sides (see Figure 1) and

238   therefore displaced only a small amount of water in the tube, but solid objects would raise the

239   water sufficiently to bring the token within reach.

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

241   One wide (diameter = 7cm, height = 15cm) and one narrow (diameter = 3.5cm, height = 15cm)

242   water-filled tube were presented. The water level was equal for both tubes, and each were

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

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

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

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

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

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

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

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

251   **Test Procedure**
252   On each experimental trial, two water-filled tubes were presented, each containing a floating out-

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

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

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

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

257   where the children handled the objects and inserted them by themselves). In each condition, the

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

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

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

261   removed the tube that was not chosen. They then rotated the chosen water-filled tube so that the

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

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

264   made, but before the resulting action of dropping the objects into the tube. The opaque side of the

265   tube concealed the water level, but not the bottom of the tube (Figure 1) to ensure that the child

266   could see that the objects had rolled into the tube, but the objects' effects on the water level

267   remained obscured. The experimenter then removed the object-tube, and the subject was able to

268   attempt to fish the token from the tube of choice using the fishing rod.

269   Trials of each condition were presented in a pseudorandomised order across children. Each block

270   of six trials contained one trial from each condition, with the provisions that the correct choice in

271   any one trial (e.g., small objects) was not the same as in the previous trial, and that the correct

272   side (left or right) was counterbalanced, ensuring that the correct choice was on the left three

273   times and the right three times within each block of six trials, though not on the same side more

274   than twice in a row within a block. The experimenter was RM or EL, with RM, EL or SAJ

275   assisting during testing by re-setting tubes (emptying out the water and objects, replacing the

276   water at correct level ready for next trial) as required. The experimenter followed a set script and

277   procedure with each subject.

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

279   option is correct on some trials but not on others, depending on the context of the trial. For

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

281   whereas the wide tube would be correct vs. the narrow one when the water level was unequal.

282 **Data Analysis**

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

284 coded live as well as being video-recorded unless parental consent requested otherwise. 10% of

285 trials were coded from video and compared to the live coding, finding 100% agreement with the

286 data. The full data set is available on FigShare: https://figshare.com/s/d1cf8b3b925c58495fd2.

287 We conducted Generalized Linear Mixed Models (GLMM: Baayen, 2008) using R (version

288 2.15.0; R Core Team, 2014) to assess which factors influenced success rate in the children.

289 Success was a binary variable indicating whether the subject correctly solved the trial (1) or not

290 (0), and was entered as a dependent variable in the models. We ran two models as we had two

291 measures of interest: model 1) success on the first trial and model 2) success across all five trials.

292 We included the random effect of subject ID, fixed effects of age in years (continuous: ages 5-9

293 in individual years), condition (1-6), gender (male/female), trial number (1-5; model 2 only) and

294 the interaction between age and condition, and age and trial number (model 2 only). We used

295 likelihood ratio tests to compare the full model (all predictor variables, random effects and

296 control variables) firstly with a null model, and then with reduced models to test each of the

297 effects of interest (Forstmeier & Schielzeth, 2011). The null model consisted of random effects,

298 control variables and no predictor variables. The reduced model comprised of all effects present

299 in the full model, except the effect of interest (Göckeritz, Schmidt & Tomasello, 2014). We then

300 ran further analyses using exact two-tailed Binomial tests to assess success rate in each condition

301 across two age groups (5-7 years & 8-9 years). We selected these two age groups to allow for

302 easy comparison to previous Aesop's Fable studies (e.g., Cheke, Loissel & Clayton, 2012; Miller

303 et al., 2016), whilst also minimising the possibility of Type II errors due to small samples size.

304 Further analyses assessing success rate across all subjects (Table S1), and per age in years (Table

305 S2) are presented in the supplementary materials.

Ethics Statement

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

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

Ethics Committee (pre.2013.109). Informed written consent was obtained from parents prior to

participation of the child. The parents of the child identified in the supplementary movie gave

their informed written consent for this information to be published.

Results

On trial 1, the full model differed significantly from the null model ($X^2 = 19.39$, $df = 3$, $p = 0.0002$). We found a significant main effect of age ($X^2 = 8.92$, $df = 1$, $p = 0.002$: Table 1) on

success rate (correct vs. incorrect choice), with success on the first trial increasing with age.

Across all five trials, we also found that the full model differed significantly from the null model

($X^2 = 74.62$, $df = 5$, $p < 0.001$). We found significant main effects of age ($X^2 = 16.42$, $df = 1$, $p < 0.001$: Table 1) and condition ($X^2 = 7.17$, $df = 1$, $p = 0.007$), and a significant interaction effect

of age and condition ($X^2 = 8.5$, $df = 1$, $p = 0.004$) on success rate (correct vs. incorrect choices per

trial). Success rate increased with age (Figure 2), and success across all five trials was

significantly poorer in the wide vs. narrow tubes condition, compared with the other conditions.

Considering all subjects together, and over all five trials, children chose the correct option

significantly more often that chance in all conditions except narrow vs. wide tubes (Table S1).

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

Table 2). In condition 1 (large vs. small objects) and condition 6 (high vs low water level),

children made significantly more correct than incorrect choices on *trial one* in both the 5-7 and 8-

9 years old age groups. In condition 2 (too large vs. small condition), children significantly made

the correct choices *across all trials* at both age 5-7 years and 8-9 years, but not on trial 1. In

329   condition 5 (wide vs. narrow tubes), only children aged 8-9 years significantly made correct

330   choices *across all trials*, and not from trial 1.

331   For the two conditions that have been previously tested in children using the standard Aesop's

332   Fable task: floating vs. sinking and hollow vs. solid objects, we found that children made

333   significantly more correct choices *across all five trials* in the age 5-7 year group, and from *trial*

334   *one* in the 8-9 year group (Table 2). These results are similar to previous findings testing the

335   sinking vs. floating condition in a standard Aesop's Fable paradigm (Cheke, Loissel & Clayton,

336   2012; Miller et al., 2016), but the children performed better than in previous tests using solid vs.

337   hollow objects (Miller et al., 2016), which found that children struggled to select the correct

338   option over five trials, but learnt to do so over twenty trials.

339   Discussion

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

341   different conditions, which could not be solved by responding only to perceptual-motor feedback.

342   Subjects were not able to observe the water level rising when objects were dropped into the

343   water-filled tubes, and had to select only one set of objects or one type of tube, into which all

344   objects were dropped at once. To solve these tasks on their very first trial, children likely needed

345   to mentally simulate the effect that dropping objects would have on the water level of the tubes,

346   or, at a minimum, children needed to use prior semantic knowledge of the relevant properties of

347   objects or tubes, to choose correctly on their first trial. Over five trials, here, children received

348   feedback about the success of their actions at the end of each trial (whether the token could now

349   be reached using the fishing rod), but at no point did they observe the water level rising. This

350   contrasts with previous studies in which subjects were able to observe the change in the water

351   level, within a trial, each time an object was dropped into a tube.

352   We found that there was a significant effect of age on success (correct vs. incorrect choices per

353   trial) in trial one and across all trials, with success rate increasing significantly with age. We also

354   found a significant effect of condition on success across all trials, and an interaction effect of age

355   and condition across trials.. Notably, children's performance in the current paradigm followed a

356   similar developmental pattern to that found on previous versions of Aesop's Fable tasks, where

357   responding to perceptual-motor feedback had been a possible strategy for success (Cheke, Loissel

358   & Clayton, 2012; Miller et al., 2016). We found that children aged 8-9 years passed the majority

359   of conditions on their first trial (four of the six conditions), although 5-7 year olds passed two

360   conditions on their first trial, and learnt to solve three of the four remaining conditions over five

361   trials. This is comparable to standard Aesop's Fable tasks, which 8 year old children typically

362   pass on their first trial and 5-7 year olds can learn to pass over five trials. The finding that

363   children's performance was not impaired in the current study, relative to previous standard

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

365   rising to solve these types of water displacement tasks.

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

367   to solve, and therefore may have each tapped slightly different cognitive processes. Younger

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

369   water level conditions on their first trial. These two conditions have not previously been used

370   with children, though corvids have also consistently solved versions of these tasks over a small

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

372   Given that younger children passed these two conditions only, it is possible they may have been

373   solved using simpler mechanisms than the other variations of the task. For example, young

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

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

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

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

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

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

380 non-significant with a Bonferroni correction). Based on this pattern of performance, the evidence

381 that children aged 5-7 mentally simulated the effects of dropping objects into the tubes is

382 equivocal.

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

384 pass the floating vs. sinking objects condition and the hollow vs. solid object condition on their

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

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

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

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

389 standard version of the Aesop's Fable task, in which children chose to insert solid and hollow

390 objects into a water-filled tube one at a time, Miller and colleagues found that 5-7 year old

391 children learnt to select solid over hollow objects over the course of 20 trials, and though 8-10

392 year olds solved the task within five trials, they did not do so from their very first trial. One

393 explanation for the children's superior performance on the present task is that children in the

394 current study were allowed to select one type of object only, which may have simplified their

395 decision-making process. In line with this, in the earlier explorative task making some mistakes

396 would not typically prevent the subject from obtaining the reward; thus, there was no penalty for

397 testing out both of the presented options in early trials. Another possibility, which cannot be ruled

398 out here, is that children's performances could have been scaffolded by their experience in the

399 other concurrent test conditions. For example, when obtaining rewards using large or small clay

400 spheres, children may have gained information that influenced their choices of solid over hollow

401 cubes in this particular task. The information used could be simple, such as generalising the

402 appearance of successful objects, or more complex, such as drawing inferences about the

403 mechanics of water displacement from observing successes and failures in other contexts. While

404   this is unlikely to account for first trial successes, given that trial orders were randomised, the

405   opportunity to learn from other conditions over multiple trials may have contributed to children's

406   ability to quickly acquire the correct option in the solid vs hollow task.

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

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

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

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

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

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

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

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

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

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

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

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

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

420   2014), though some passed when the number of objects was restricted (Logan et al., 2014). The

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

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

423   differently sized tubes in this context. Thus, although by the age of 7, children can recognise that

424   water volume is conserved when it passes between two containers of different sizes (Piaget,

425   1930, 1974), they do not appear to possess a full, intuitive understanding of the behaviour of

426   liquids in different containers.

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

428   cannot be attributed to learning from perceptual-motor feedback because they did not observe the

429   reward moving incrementally closer after each object drop. Furthermore there was no evidence

430   that the children found this task more difficult than other versions of water displacement tasks:

431   indeed, performance on this task, which restricted access to visual feedback, was equivalent to (or

432   better than) performance on previous versions of the Aesop's Fable task. In evaluating these

433   findings it is important to note that children did receive feedback on the overall success of their

434   choices at the end of each trial. They may also have gained some information from other visible

435   features, such as estimating the final water level from peering into the tube (though this was

436   difficult to judge from above) or observing the light objects floating, and the too-large objects

437   becoming stuck in the tube. However, they did not observe the reward moving incrementally

438   closer after each object drop. Therefore the perceptual-motor feedback explanation cannot

439   account for children's success on Aesop's Fable tasks (Taylor & Gray, 2009; Cheke, Bird &

440   Clayton, 2011; Jelbert, Taylor & Gray, 2015). The pattern of results we found, in which the

441   majority of tasks were solved on the first trial by children over the age of 8, is also consistent

442   with a number of studies suggesting that children reliably solve various innovative tool-use

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

444   2013). This adds to the growing body of evidence that spontaneously recognising the causal

445   relations involved in tool use tasks can be remarkably difficult for young children.

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

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

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

449   2015). To date, the Aesop's Fable task has been used to assess the cognitive abilities of various

450   species of corvid, including rooks, New Caledonian crows, Eurasian jays and California scrub

451   jays (Bird & Emery, 2009; Cheke, Bird & Clayton, 2011; Taylor et al., 2011; Jelbert et al., 2014;

452   Logan et al., 2014, 2016; Miller et al., 2016), as well as grackles, another behaviourally flexible

453   species of bird (Logan, 2015, 2016). A number of great apes have also been tested on the

454   comparable floating-peanut task (Mendes, Hanus & Call, 2007; Hanus et al., 2011). Recently,

455   Miller and colleagues demonstrated that in New Caledonian crows, but not in human children,

456   performance on object-choice tasks can be influenced by pre-existing preferences for certain

457   types of objects, casting some doubt on the suggestion that birds' success on Aesop's Fable tasks

458   reflects causal understanding – at least when considering their selection of objects (Miller et al.,

459   2016). Perceptual-motor feedback has been suggested to underpin spontaneous string pulling

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

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

462   solving tasks by great apes (Völter & Call, 2012). However, to date, it is unclear whether the

463   opportunity to receive perceptual-motor feedback accounts for non-human animals' ability to

464   rapidly solve various water displacement tasks. The methodology that we describe here could be

465   adopted for use with non-human animals to test whether their success depends on perceptual-

466   motor feedback, with the present study allowing for comparison with young children. Use of this

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

468   remarkable performance of certain species on water displacement tasks.

469 **Acknowledgements**

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

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

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

473 you to Ian Millar for help in apparatus construction, and to Markus Boeckle for assistance with

474 the statistical analysis. We are also very grateful to Jennifer Vonk (Editor), Eva Reindl and an

475 anonymous reviewer for constructive comments on the manuscript.

References

Baayen RH. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. DOI: 10.1017/CBO9780511801686.

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.

Forstmeier W., Schielzeth H. 2011. Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology* 65:47–55. DOI: 10.1007/s00265-010-1038-5.

Göckeritz S., Schmidt MFH., Tomasello M. 2014. Young children's creation and transmission of social norms. *Cognitive Development* 30:81–95. DOI: 10.1016/j.cogdev.2014.01.003.

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*

513   *Biology* 8:1–6. DOI: 10.1080/19420889.2015.1035846.

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

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

518   Logan CJ., Harvey BD., Schlinger BA., Rensel M. 2016. Western scrub-jays do not appear to
519   attend to functionality in Aesop's Fable experiments. *PeerJ* 4:e1707. DOI:
520   10.7717/peerj.1707.

521   Logan CJ., Jelbert SA., Breen AJ., Gray RD., Taylor AH. 2014. Modifications to the Aesop's
522   Fable paradigm change New Caledonian crow performances. *PLoS ONE* 9:e103049. DOI:
523   10.1371/journal.pone.0103049.

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

526   Miller R., Jelbert SA., Taylor AH., Cheke LG., Gray RD., Loissel E., Clayton NS. 2016.
527   Performance in object-choice Aesop's Fable tasks are influenced by object biases in New
528   Caledonian crows but not in human children. *PLoS ONE* 11:e0168056. DOI:
529   10.1371/journal.pone.0168056.

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

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

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

534   R Core Team. 2014. R: A language and environment for statistical computing. *R Foundation for*
535   *Statistical Computing: Vienna, Austria*:2014. DOI: 10.1017/CBO9781107415324.004.

536   Taylor AH., Elliffe DM., Hunt GR., Emery NJ., Clayton NS., Gray RD. 2011. New Caledonian
537   crows learn the functional properties of novel tool types. *PloS one* 6:e26887. DOI:
538   10.1371/journal.pone.0026887.

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

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

544   Taylor AH., Medina FS., Holzhaider JC., Hearne LJ., Hunt GR., Gray RD. 2010. An investigation
545   into the cognition behind spontaneous string pulling in New Caledonian crows. *PloS one*
546   5:e9345. DOI: 10.1371/Journal.Pone.0009345.

547   Völter CJ., Call J. 2012. Problem solving in great apes (Pan paniscus, Pan troglodytes, Gorilla
548   gorilla, and Pongo abelii): the effect of visual feedback. *Animal cognition* 15:923–936. DOI:
549   10.1007/s10071-012-0519-5.

550 **Tables & Figures**

551 **Tables**

552 **Table 1.** Generalized Linear Mixed Models on factors affecting the number of correct trials in

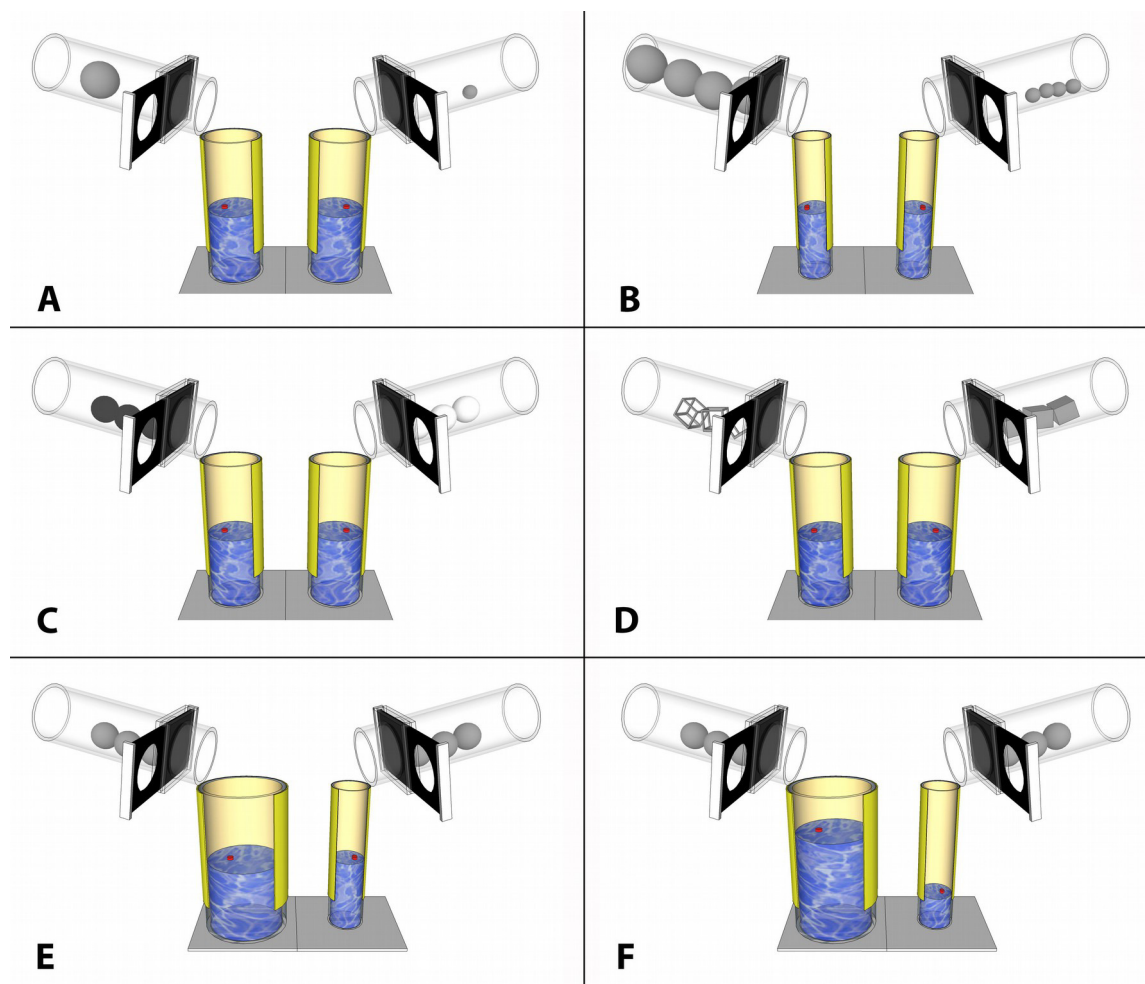
553 trial 1 (model 1) and across all trials (model 2).

Fixed term	Full Model 1: Trial 1			Full Model 2: All Trials		
	Estimate	*z-value*	*p-value*	**Estimate**	*z-value*	*p-value*
Age in years	**0.6072**	**2.876**	**0.004**	**0.5995**	**4.008**	**<0.001**
Condition	0.435	1.192	0.233	**0.4754**	**2.673**	**0.008**
Age*Condition	-0.0581	-1.1	0.271	**-0.0767**	**-2.909**	**0.004**
Gender	0.4348	1.129	0.259	0.0163	0.082	0.934
Trial Number	/	/	/	0.0085	0.040	0.968
Age*Trial Number	/	/	/	0.0377	1.167	0.243

554   **Table 2.** Correct choices (%) in each condition by each age group: 5-7 years old (n=34) and 8-9

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

556   values are highlighted in bold. NS = not significant with a Bonferroni correction.

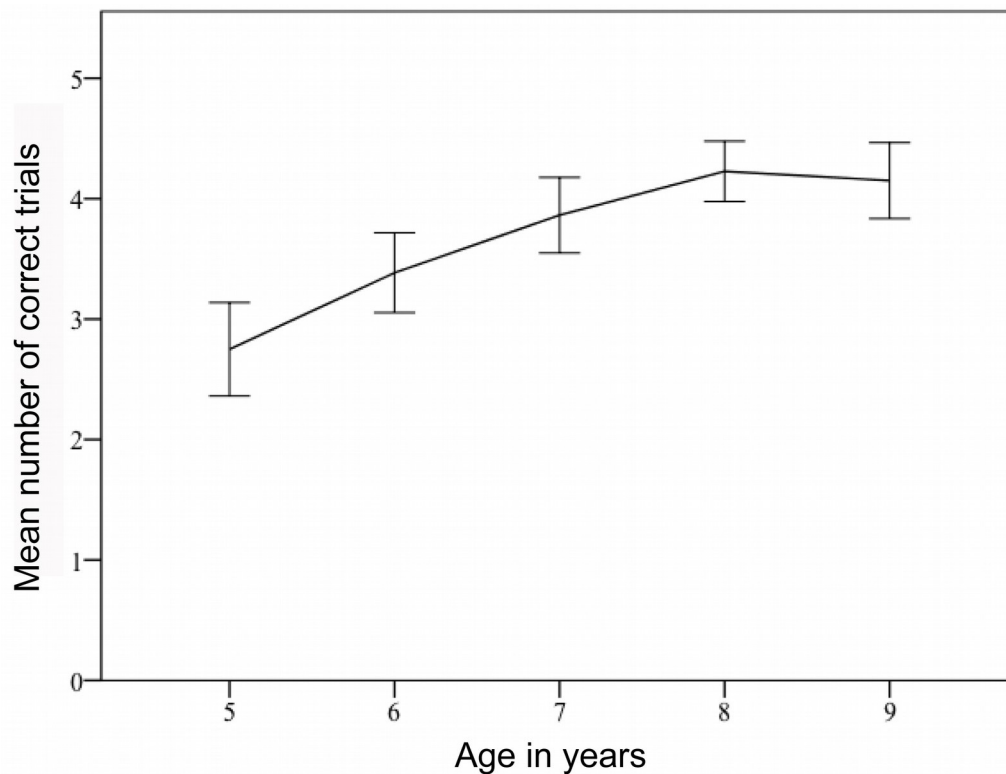
Age Group	Large vs. small		Too large vs. small		Floating vs. sinking		Hollow vs. solid		Wide vs. narrow		High vs. low	
	%	*p*	%	*p*	%	*p*	%	*p*	%	*p*	%	*p*
Trial 1												
5-7	76	**0.003**	29	0.024 NS	56	0.608	53	0.864	38	0.229	79	**0.001**
8-9	95	**<0.001**	48	>0.999	90	**<0.001**	86	**0.002**	62	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**

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

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

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

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

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

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

563 **Figure 2: Mean number of correct trials across all 5 trials for all conditions by age (in**

564 **years).** Success rate increased significantly with age. Error bars indicate standard errors.