

2020-07-29

Piecing together the puzzle of pictorial representation: How jigsaw puzzles index metacognitive development

Doherty, MJ

<http://hdl.handle.net/10026.1/15569>

10.1111/cdev.13391

Child Development

Wiley

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

**Piecing together the puzzle of pictorial representation: How jigsaws index metacognitive
development**

Martin J. Doherty^a, Marina C. Wimmer^b, Cornelia Gollek^c, Charlotte Stone^b, & Elisabeth J.
Robinson^d

^aSchool of Psychology, University of East Anglia, Norwich, NR4 7TJ, UK

^bSchool of Psychology, Cognition Institute, University of Plymouth, Plymouth, UK PL4 8AA

^cSchool of Education, University of the West of Scotland, Ayr, KA8 0SX

^dDepartment of Psychology, University of Warwick, Coventry CV4 7AL

Piecing together the puzzle of pictorial representation: How jigsaw puzzles index metacognitive development

Abstract

Jigsaw puzzles are ubiquitous developmental toys in Western societies, used here to examine the development of metarepresentation. For jigsaw puzzles this entails understanding that individual pieces, when assembled, produce a picture. In Experiment 1, 3- to 5-year-olds ($N=117$) completed jigsaw puzzles that were normal, had no picture, or comprised non-interlocking rectangular pieces. Pictorial puzzle completion was associated with mental and graphical metarepresentational task performance. Guide pictures of completed pictorial puzzles were not useful. In Experiment 2, 3- to 4-year-olds ($N=52$) completed a simplified task, to choose the correct final piece. Guide-use associated with age and specifically graphical metarepresentation performance. We conclude that the pragmatically natural measure of jigsaw puzzle completion ability demonstrates general and pictorial metarepresentational development at 4 years.

Jigsaw puzzles are ubiquitous features of childhood in Western cultures. Playing with jigsaw puzzles is believed to benefit hand-eye coordination, spatial ability, social development, problem solving strategies, and specific mathematical skills (Fleer, 1990; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Young, Cartmill, Levine, & Goldin-Meadow, 2014). Surprisingly, there has been little research concerning the cognitive processes that underlie jigsaw puzzle completion.

Two existing studies examine jigsaw puzzle play in young children, relating it to the development of spatial abilities. Levine et al. (2012) observed children in their homes between the ages of 26 and 46 months. Frequency of play with jigsaw puzzles and similar puzzles was found to relate to the ability to mentally transform 2D shapes at 54 months, suggesting puzzle play promotes spatial skills. Young et al. (2014) provide partial evidence supporting this idea. They provided instruction during puzzle play with 4- to 5-year-old children. They varied the amount of spatial language (e.g., "... it will fit in one of the corners") and gesture (e.g., tracing the top corners in the frame). Spatial instruction produced substantial improvement, but only when combined with gesture. The improvement did not transfer to tests of mental transformation or spatial vocabulary.

Here we examine the development of jigsaw puzzle completion directly rather than as a predictor of other spatial skills. Our focus is another factor involved in jigsaw puzzle play, pictorial understanding. This was measured in 3- to 5-year-olds by i) comparing use of pictorial information versus shape information for completion, and ii) examining the tendency to consult and use information from a guide showing the final picture. Hereafter, for brevity we follow common usage by referring to jigsaw puzzles as 'jigsaws'.

Jigsaws are a particularly apt tool for measuring pictorial metarepresentation, defined here as understanding the relationship between the pictorial medium and the depicted scene. They form a natural split between seeing a picture as an object (when in pieces) and treating

it transparently in terms of its content when complete. Efficient completion of jigsaws requires understanding that the pieces combine to form a picture; the picture is not a mysterious side-effect of fitting the pieces appropriately. Understanding the representational nature of pictures should enhance jigsaw completion, because now content information as well as shape can be used.

Representation

Understanding representation is a key milestone in cognitive development. Metarepresentational ability, the understanding of the relation between a representational medium and its referent, has been argued to develop around four years (Doherty, 2009; Wimmer & Perner, 1983; Wellman, 1990; Wellman, Cross & Watson, 2001). This has been widely studied using the False Belief task. This demonstrates understanding that action is guided by mental states which represent external situations, and can misrepresent them (Wellman, 1990; Wimmer & Perner, 1983). Substantial research evidence has accumulated indicating comparable developments in understanding the representational nature of language occur around the same time. Children understand that objects can take alternative names, *bunny* and *rabbit*, or *rabbit* and *animal*, for example, around four years, and this understanding is associated with false belief understanding (Doherty & Perner, 1998; Perner, Stummer, Sprung, & Doherty, 2002). Similar developments have been shown in the ability to select objects as the referents of novel names (Gollek & Doherty, 2016; Karadaki & Doherty, 2017). These associations suggest that preschool children are developing a general concept of metarepresentation, applicable across multiple representational domains.

However, recent work on theory of mind has raised the possibility that younger children possess this understanding of mental representation much earlier than it can be demonstrated in standard tasks. This may be because of processing limitations (e.g., Scott & Baillargeon, 2017) or because children do not interpret the pragmatics of experimental

situations in the way the experimenter intends (Siegal & Beattie, 1991; Westra & Carruthers, 2017). If correct, these claims would challenge either the claim that a general concept of metarepresentation develops, or that it develops around the age of four years. A strong test of the general metarepresentational development hypothesis is to examine development in multiple domains. In addition to the mental and linguistic, also in pictorial metarepresentation.

When and how pictorial metarepresentational understanding develops is currently unclear. Young children are adept at interpreting pictures, but research suggests they do not understand pictures as representational vehicles before the age of four years (Allen & Armitage, 2017; Callaghan, 1999; Claxton, 2011; Doherty, 2008; Jolley, 2010; Liben, 2003; Simcock & DeLoache, 2006; Thomas, Jolley, Robinson, & Champion, 1999). They also become referentially confused between what the picture depicts and how the real referent currently appears (Beilin & Pearlman, 1991; Donnelly, Gjersoe, & Hood, 2013; Robinson, Nye, & Thomas, 1994; Wimmer, Robinson, Koenig, & Corder, 2014; Slaughter, 1998; Zaitchick, 1990).

Relatedly, children begin to develop an understanding of graphic metarepresentation such as false signs by the age of around four years (Parkin, 1994; Parkin & Perner, 1996). In the False Sign task a story is acted out in which a sign is supposed to indicate the location of a princess (Parkin, 1994). When she moves and the sign is not changed, the now-false sign misrepresents her location. To correctly answer the test question “where does the sign say the princess is?” children must answer on the basis of what the sign shows, rather than what is actually the case. This task can be considered a test of graphic metarepresentation: it assesses the understanding that one object (a sign) represents the location of another object. False Sign performance is typically associated with performance on the False Belief task (Sabbagh, Moses, & Shiverick, 2006; Leekam & Perner, 2008; Bowler, Briskman, Gurvidi, & Fornells-

Ambrojo, 2005, used a similar task based on signals of a model train set; see Leekam & Perner, 2008, for a summary) suggesting a link between graphic and mental metarepresentational development.

The finding of metarepresentational developments linked across domains is also supported by research on ambiguous figures, images with more than one interpretation. For example, the duck-rabbit (Jastrow, 1900) can be seen as either a duck or a rabbit (e.g., long elements in the picture forming the duck's bill or the rabbit's ears), which is understood by four years (Beck, Robinson, Ahmed, & Abid, 2011; Wimmer & Doherty, 2005; Wimmer & Doherty, 2011). Understanding how this can be requires a distinction between the picture and what it represents. In this sense ambiguous figures are analogous to homonyms, words with distinct unrelated meanings (e.g., bat, which can mean either sports equipment or a flying mammal). Doherty (2000) found the ability to identify homonyms was strongly related to alternative naming and false belief. The same association occurs with ambiguous figures (Wimmer & Doherty, 2011) with some qualifications. After having had the two interpretations pointed out to them, children can report both interpretations from the age of about four years. This ability is closely associated with false belief understanding (Doherty & Wimmer, 2005; Wimmer & Doherty, 2011), supporting the claim for the co-emergence of pictorial and mental metarepresentation. However, there are reasons for caution. Although they can acknowledge both interpretations of an ambiguous figure at four years, children do not appear to experience *reversal*, the characteristic switch from one interpretation to another, until roughly five years (Rock, Gopnik, & Hall, 1994; Gopnik & Rosati, 2001). Arguably, understanding of ambiguous figures is not yet mature (Beck et al., 2011). Additionally, ambiguous figures are peculiar, and plausibly children seldom experience them. It would be natural to question whether they are the best stimuli with which to measure typical development of pictorial metarepresentation. Other measures are required.

Pictures are made up of lines and patches of colour. Suitably arranged these go to form an image that can be considered to represent something. Understanding that marks on paper can be assembled to represent something is the foundation of drawing, painting, and writing, which children begin to do in the preschool period (Callaghan, Rochat, & Corbit, 2012; Callaghan, 1999). By definition this requires understanding of the relation between the representational medium and what it represents. Our claim is that this understanding will improve the efficiency with which children can complete jigsaws. Although a given jigsaw piece may appear a random set of lines and colour patches, suitably assembled these lines and patches form objects. Assembly requires connecting lines and colours between contiguous pieces, thus allowing better selection of candidate pieces. Without this understanding, the picture resulting from completing a jigsaw may seem a mysterious emergent property.

If this claim is correct, the development of pictorial metarepresentation should be evident in increased efficiency of jigsaw completion: children should be able to complete jigsaws faster and with less trial and error. The ecological validity of this measure is clear: children do play with jigsaws. They require no special instruction. The experimenter's intentions should be clear. Relative to all the measures above, pragmatic misinterpretation (Siegal & Beattie, 1991; Westra & Carruthers, 2017) is unlikely to affect jigsaw completion. Naturally spatial abilities, working memory, and various other factors will also be involved. We do not test these directly here. We test the hypothesis that in addition to these factors there is a relation between jigsaw completion ability and more traditional tests of metarepresentational ability, that is false belief (mental) and false sign (graphical) performances.

Using a guide

Jigsaws typically are supplied with a copy of the completed image, often on the box lid. This can aid completion if the participant utilises the correspondence between the guide

and the final jigsaw in selecting and arranging pieces. The relation between guide and completed picture is one of geometrical correspondence.

Relevant research about geometrical correspondence in pictures and related physical media suggests that some ability to use the relation between public representations and their referents emerges around the age of two-and-a-half or three years (DeLoache & Burns, 1994). A particularly well-known demonstration uses a scale model of a room (DeLoache, 1987). The model is laid out in the same way as the room, with model furniture corresponding to the real furniture in the room. Shown where an object is hidden in one of the spaces, children's task is to find it in the corresponding space. Children can do this from roughly three years, and in some studies younger (e.g., DeLoache, Kolstad, & Anderson, 1991).

This suggests a new skill to detect correspondences at this age. However, when identical pieces of furniture are used (two identical chairs placed in different locations for example) 3-year-olds appear to choose between identical hiding places at chance (Blades & Cooke, 1994). Children may simply be utilising 'element-to-element' correspondence (Newcombe & Huttenlocher, 2000): for example, shown the small object hidden in the model room behind a small chair, children know to look behind the chair in the large room, regardless of where it is. When there are two chairs in each space, for example one by the door and one next to the bed, they choose at random. Using the additional geometrical information to distinguish between identical chairs appears to develop around four years. If so, we would not expect children to be able to utilise the correspondence relation between jigsaw guide and jigsaw before the age of four years.

The present study

To sum up: For jigsaw puzzle completion metarepresentational understanding should help understand that pictorial elements can combine to produce an image, providing

additional strategies to fitting pieces together by trial and error. The additional presence of a box lid or equivalent guide to the final picture may also help children who understand geometrical correspondence. We anticipate that children who understand that the pictorial guide is *supposed* to be related to the final jigsaw will be particularly able to exploit this relation.

To assess the possible involvement of metarepresentation in jigsaw completion and guide use, we compare with performance on the False Belief and False Sign tasks. The False Belief task is a widely used measure of mental metarepresentation, and as discussed has been shown to have substantial associations with measures of linguistic metarepresentation and, in the form of ambiguous figures, pictorial metarepresentation. The False Sign task (Parkin, 1994; Parkin & Perner, 1996) can be considered a test of graphic metarepresentation. It may be a particularly appropriate comparator task for use of the guide, in the sense that both are physical representations that indicate something about how the world should be: where the princess should be, or how the finished jigsaw should appear.

Experiment 1

Experiment 1 examined what information (shape, pictorial content, or both) preschool children use to complete jigsaws. We varied whether the jigsaw bore a picture and whether the pieces were characteristically irregular shapes, which allow solution based solely on fitting pieces together, or equal-sized rectangles, which do not. Half of participants had a guide picture of the completed jigsaw. We hypothesised that children who show evidence of metarepresentational understanding, indexed by False Belief and False Sign tasks, will complete jigsaws with pictorial content more efficiently, and will be more able to utilise a guide. We modelled the False Sign task closely on Parkin's (1994) version, since it was brief and closely analogous to the unexpected transfer False Belief task.

Jigsaw completion ability can be measured in a number of ways. We originally planned completion time to be the primary measure. Extensive piloting with 68 3- to 5-year-olds showed that all could complete a normal 6-piece jigsaw within 3 minutes. We set this as the maximum time for each jigsaw to avoid fatigue effects. However, as many younger children proved unable to complete the harder shape-cue or picture-cue jigsaws within this time, we use two additional measures: dichotomous ability to complete the jigsaw within the time limit, and the number of times children attempted to join jigsaw pieces.

Method

Participants

Participants were 117 children (74 girls), 29 3-year-olds ($M = 3;8$, $SD = 4$ months), 39 4-year-olds ($M = 4;7$, $SD = 3$ months), and 49 5-year-olds ($M = 5;7$, $SD = 5$ months) of white ethnic origin predominantly from lower middle-class backgrounds from _____. All children spoke German as a first language, as did the experimenter. For both experiments, inclusion criteria were informed parental consent and child assent immediately prior to testing. The exclusion criterion would have been teacher or parental indication of a special-needs diagnosis, but this did not prove the case in either experiment. The stopping criterion was that all available children had been tested.

Materials and Procedure

Children were seen in a quiet familiar room for approximately 15 minutes.

Jigsaws. There were three types of 6-piece jigsaws measuring 22 cm x 22 cm (Figure 1). 1) The *normal jigsaw* had typically-shaped interlocking pieces, which could only connect to the appropriate adjacent pieces. The picture's contents were chosen so that they could not readily be divided into separate objects when in pieces. Otherwise the task could have been considered as arranging a set of objects rather than creating a picture. 2) The *shape-cue* jigsaw had the same shaped pieces as the normal jigsaw but no picture. Each piece was a

different colour. 3) The *picture-cue* jigsaw had the same picture as the normal jigsaw but consisted of equally sized rectangles (11 cm x 7.3 cm) so that shape did not aid completion.

Figure 1 about here

Each child was shown the assembled jigsaw and asked: “Do you think you could put this back together if we mess it up?” The jigsaw was then disassembled and the child allowed three minutes to assemble it. Jigsaw presentation order was counterbalanced between participants. Half ($N = 59$) had a guide picture placed without comment above the pieces showing the finished jigsaw; half ($N = 58$) had no guide.

The dependent variables were the ability to complete each jigsaw within three minutes, the time taken, and the number of *unnecessary* attempts to connect one side of one piece to another piece, providing an index of trial and error. Immediate correct assembly thus scored zero attempts, and lower scores indicate more efficient completion.

False Sign task. A short story was acted out with two Playpeople dolls, a cardboard castle (a façade with a picture of a castle 16cm high x 14 cm wide) and forest (a façade with picture of woods 16cm high x 14 cm wide), a road leading to both, and a cardboard signpost, (a blue arrow five cm long, four cm wide and on a five cm pole that allowed it to rotate). Children were first introduced to the sign. The experimenter pointed it toward the child, and asked where it was pointing. The child was then asked to point the sign at the experimenter. All children were able to answer the question and point the sign.

The test phase closely followed Parkin’s (1994) procedure: “Look, here’s a castle and here’s a forest. And this is the princess. Now, sometimes the princess likes to play in the

castle, and sometimes she likes to play in the woods. And look, here's a signpost. This shows where the princess is. So, today the princess is playing in the castle [experimenter placed princess out of sight behind the castle]. And look, the signpost points to the castle. Now, here comes the prince to visit the princess. He comes and he sees the sign. Then he thinks "I will fetch the princess a present", and off he goes again. But what's happening here? The princess is bored of playing in the castle, so she goes to play in the woods [experimenter moved princess to the woods, out of sight]. Now here comes the prince again." Children were asked the following questions:

False sign test question: Where does the sign show the princess is?

Reality question: Where is the princess really?

Memory question: Can you remember where the princess was in the beginning?

Children passed the false sign task if they answered all three questions correctly.

False Belief task. A short story was acted out with two Playpeople dolls, a marble, and an opaque jar and box. Sally placed a marble in the box and left. Tony then moved the marble to the jar. Sally returned and children were asked the following questions:

False belief test question: "Where will Sally look first for her marble?"

Reality question: "Where is the marble really?"

Memory question: "Where did Sally put the marble in the beginning?"

Children passed the False Belief task if they answered all three questions correctly.

Results

Bonferroni confidence interval adjustments and post-hoc Bonferroni analysis were used for the analyses of variance (ANOVAs) throughout.

False Belief and False Sign

Performances on these tasks for the two age groups are shown in Table 1.

Performance on the memory control questions were good, and all children passed the reality

control question for both tasks. Performances on the test questions improved with age for both False Belief, $\chi^2 = 27.98$, $df = 2$, $p < .001$, and False Sign tasks, $\chi^2 = 39.70$, $df = 2$, $p < .001$. Performance increased between 3- and 4-year-olds (both tasks, $p = .001$) and 4- and 5-year-olds (false belief: $p = .04$; false sign: $p < .001$) (Mann-Whitney test). The two performances were strongly associated; $r = .57$, $p < .001$, even when age was controlled for, $r_{\text{partial}} = .43$, $p < .001$. Given their close values and association, for the remainder of the analysis we combine the two variables as a single more-stable measure of metarepresentational ability.

Table 1 about here

Jigsaws

Completion. There were considerable differences between completion of jigsaw types within three minutes, $\chi^2 = 22.24$, $p < .001$ (Friedman-test) (Table 2). Four- and 5-year-old children completed virtually all jigsaws within the time (95% of participants or greater). Three-year-olds performed less well. Completion rates of the normal and shape-cue jigsaws were equally good (Binomial, $p = 0.625$), and both were completed more often than the picture-cue jigsaw (McNemar, $ps \leq .001$).

There were no differences in completion rate between the guide and no-guide conditions for any jigsaw, either in the whole sample (all $ps > .32$, Mann-Whitney test) or within the 3-year-olds, the group that made most errors (all $ps > .50$).

We examined completion ability and metarepresentational performance in the 3-year-old group only, since older children's performance approached ceiling. There was no association between metarepresentation and the Normal or Shape-cue jigsaws ($r = 0.11$ and r

= -0.37, respectively) but a substantial correlation with the Picture-cue jigsaw, $r = 0.45$, $p = 0.015$, which did not change after age was partialled out, $r = 0.45$, $p = 0.017$; 15 of the 17 children who did not complete the Picture-cue jigsaw failed both metarepresentational tasks, compared to five of the 12 who completed it.

Table 2 about here

Completion time. Twenty children did not complete at least one jigsaw in three minutes, 17 of them in the 3-year-old group. Analysing completion time can thus only be done for part of the sample, one disproportionately successful at the measure of metarepresentation. It is nevertheless of interest to analyse for the remaining children the time it took to complete each jigsaw type. Twelve 3-year-olds completed all three jigsaws, six in each of the guide- and no-guide conditions. For the following analysis we pooled the 3- and 4-year-old participants, to create two roughly equal groups: 49 3- to 4-year-olds, *mean age* = 52 months, *SD* = 5.4 months, and 48 5-year-olds, *mean age* = 67 months, *SD* = 5.3 months.

Figure 2 about here

Figure 2 shows the completion time for the three jigsaws in Guide and No Guide conditions. A three (jigsaw type: normal vs. shape-cue vs. picture-cue; within participants) x two (guide picture: no guide vs. guide; between participants) x two (age group: 3- to 4-year-olds vs. 5-year-olds; between participants) repeated measures ANOVA on completion times showed a main effect of jigsaw type, $F(2, 186) = 7.75$, $p < .001$, $\eta_p^2 = .077$. Children completed the normal jigsaw more quickly ($M = 37.6$ seconds) than both the shape-cue ($M =$

50.3 seconds, $p < .002$) and picture-cue jigsaw ($M = 49.8$ seconds, $p < .001$) which did not differ. Completion time decreased with age, $F(1, 93) = 22.04$, $p < .001$, $\eta_p^2 = .19$, where 3- to 4-year-olds ($M = 57.2$) required more time than 5-year-olds ($M = 34.4$, $p < .001$). There was no significant jigsaw type x age group interaction, $F(2, 186) = 2.01$, $p = .14$, $\eta_p^2 = .02$.

As shown in Figure 2, children in each age group completed more quickly when a guide picture was present ($M = 39.8$) than without ($M = 52.1$), $F(1, 93) = 6.47$, $p = .013$, $\eta_p^2 = .07$. There was no age x guide interaction, $F(1, 93) = 0.97$, $p = .33$, $\eta_p^2 = .01$. However, guide presence interacted with jigsaw type, $F(2, 186) = 6.14$, $p < .003$, $\eta_p^2 = .06$. Completion was faster with a guide in the shape-cue jigsaw: 37.2 seconds without guide and 63.7 seconds with guide, $p < .001$. There was no significant difference for either the normal (38.5 second attempts without and 36.7 seconds with guide) or the picture-cue jigsaw (54.0 seconds without and 39.8 seconds with guide).

Number of attempts. For children who completed all jigsaws, number of attempts and completion time were very strongly related: partialling out age, the correlations between completion time and number of attempts were for the normal jigsaw $r = 0.77$, shape-cue $r = 0.85$, and picture-cue, $r = 0.69$, all $ps < 0.001$. Thus we are reasonably confident that the two measures are comparable, and using number of attempts allows the following analysis to retain all participants.

Figure 3 about here

Figure 3 shows the number of attempts for the three jigsaws in Guide and No Guide conditions. A three (jigsaw type: normal vs. shape-cue vs. picture-cue; within participants) x two (guide picture: no guide vs. guide; between participants) x three (age group: 3- vs. 4- vs. 5-year-olds; between participants) repeated measures ANOVA on the number of attempts

made within the time period showed a main effect of jigsaw type, $F(2, 222) = 19.39, p < .001, \eta_p^2 = .15$. Children required fewer attempts for the normal jigsaw ($M = 2.48$) than both the shape-cue ($M = 4.19, p < .001$) and picture-cue jigsaw ($M = 4.91, p < .001$) which did not differ ($p = .15$). The number of attempts decreased with age, $F(2, 111) = 38.27, p < .001, \eta_p^2 = .41$. Adjacent age group comparisons show that 3-year-olds ($M = 7.06$) required more attempts than 4-year-olds ($M = 3.59, p < .001$) who in turn required more attempts than 5-year-olds ($M = 2.12, p = .02$). These two main effects were qualified by a jigsaw type x age group interaction, $F(4, 222) = 2.67, p = .03, \eta_p^2 = .05$; 3-year-olds required fewer attempts in both normal ($p < .001$) and shape-cue jigsaw ($p = .02$) than in the picture-cue jigsaw, where the first two did not differ. Four-year-olds required fewer attempts in the normal jigsaw than both the shape-cue ($p < .001$) and picture-cue jigsaw ($p = .001$), that did not differ. In contrast, 5-year-olds' attempts did not differ across all three jigsaw types (all $ps > .17$).

As shown in Figure 3, children in each age group required fewer attempts when a guide picture was present ($M = 3.49$) than without ($M = 5.03$), $F(1, 111) = 11.34, p = .001, \eta_p^2 = .09$. There was no age x guide interaction, $F(2, 111) = 1.18, p = .31, \eta_p^2 = .02$. However, guide presence interacted with jigsaw type, $F(2, 222) = 9.81, p < .001, \eta_p^2 = .08$. Fewer attempts occurred with a guide in the shape-cue jigsaw: 5.97 mean attempts without guide and 2.27 mean attempts with guide, $p < .001$. There was no difference for either the normal (2.74 attempts without and 2.22 with guide) or the picture-cue jigsaw (5.06 attempts without and 4.76 attempts with guide).

Table 3 about here

Correlational analyses

Completion time data exclude most of the 3-year-olds, and thus present disproportionate success on the metarepresentational measures. Thus, we restrict correlational analysis to the number of attempts measure. Table 3 shows correlations between age, metarepresentational performance, and jigsaw attempts. Number of attempts on all three jigsaws are associated with both age and metarepresentational performance, and with each other. Results for the shape-cue attempts should be interpreted cautiously given the disproportionate effect of guide condition on this jigsaw. We therefore partial out age and condition (guide or no guide; Table 3, below the diagonal). When this was done metarepresentational performance remained associated with the normal and picture-cue jigsaw, but not the shape-cue.

Discussion Experiment 1

Jigsaw puzzle completion ability improves markedly between the ages of 3- and 5-years. Four- and 5-year-olds could complete jigsaws in which the only cues available were based on either shape or pictorial content. Three-year-olds performed less well, although most could complete the normal and shape-cue six-piece jigsaws within three minutes.

There was a substantial correlation between 3-year-olds' ability to complete the picture-only jigsaw and their performance on the metarepresentational tasks. This relatively blunt measure supports the hypothesis that the use of pictorial information when completing jigsaws draws on metarepresentational understanding.

Completion time and the number of attempts to connect pieces provide graded measures of efficiency. Completion time for children who completed all jigsaws showed that the normal jigsaw was completed faster than either the shape-cue or picture-cue jigsaws, and these other jigsaws took almost the same time. Completion time was faster when a guide was present, but only significantly so for the shape-cue jigsaw. This jigsaw took the longest without a guide, and roughly the same time as the normal jigsaw with a guide. The pattern of

findings was comparable when analysing the number of attempts made. Again, presence of a guide was effective for all age groups, and this effect was largely restricted to the shape-cue jigsaw.

The number of attempts measure was significantly associated with metarepresentational understanding for all jigsaw types; only the jigsaws with pictorial information remained correlated with overall metarepresentational performance when age and guide presence was partialled out. This supports our hypothesis that children who show evidence of metarepresentational understanding will complete jigsaws with pictorial content more efficiently

The effectiveness of the guide suggests children were using the spatial correspondence. However, they were *not* doing so for jigsaws with pictorial content. This challenges our claim that metarepresentational understanding of pictures should allow children to exploit this correspondence relationship.

However, other factors may make this difficult: the visual simplicity of the shape-cue jigsaw plausibly makes the correspondence with the guide easier to utilise than for the pictorial jigsaws: the bottom right corner of the jigsaw is blue, and one jigsaw piece is entirely blue, whereas in the other jigsaws the bottom right piece has part of a sheep in it, as do four of the other five pieces. Thus the question remains whether children consulted the guide for the jigsaws with pictorial content, but were not good at utilizing the information, or whether they only consulted it for the shape-cue jigsaw. The design of Experiment 1 does not allow us to distinguish these two possibilities. This is the aim of Experiment 2.

Experiment 2

To examine whether children also consult the guide for other jigsaw types, and whether they can use the information in principle, we reduced jigsaw task demands to a

minimum. Children had to select the correct final piece of a partially-completed four-piece jigsaw, avoiding a plausible and implausible distractor. Being certain of the correct choice required consulting a guide picture.

There were three jigsaw-guide conditions, selected so the match could be: only made on the basis of colour; only on the basis of pictorial content (removing the colour from the guide); or on the basis of both. Thus conditions were: i) colour (a jigsaw with each piece a different colour, and corresponding guide); ii) outline (e.g., a coloured car jigsaw, and a line-drawing car on the guide); and iii) normal (e.g., coloured tree jigsaw and coloured tree on guide).

Performance was compared to the same measures of metarepresentational development as in Experiment 1. Verbal mental age was measured, to examine the specificity of any relationships.

Method

Participants

Participants were 52 children (30 girls), 27 3-year-olds ($M = 3;6$, $SD = 3$ months, verbal mental age (VMA) $M = 3;10$, $SD = 11$ months) and 25 4-year-olds ($M = 4;5$, $SD = 3$ months, VMA $M = 5;2$, $SD = 16$ months) from predominantly white middle-class nurseries (two university preschools and one private nursery) in _____ and _____. All children spoke English as a first language. One 3- and three 4-year-olds' verbal mental age could not be measured due to test scheduling constraints. Inclusion, exclusion, and stopping criteria were as for Experiment 1.

Design

The jigsaw, False Sign and False Belief tasks were administered in counterbalanced order, after which the British Picture Vocabulary Scale 2nd Edition (Dunn, Dunn, Whetton, & Burley, 2007) was administered according to the manual.

Materials

There were two trials each of three different jigsaw types. See Figure 4. All were four-piece jigsaws with one piece missing. Each jigsaw had a guide. There were three candidate pieces shaped to fit in the empty position. One completed the picture shown in the guide; one formed a satisfactory picture but differed from that shown in the guide; one was unsuitable. The unsuitable pieces were, for the Normal and Outline jigsaws part of one of the other pictures but rotated clockwise through 90 degrees, and for the Colour jigsaws an uncoloured (i.e., white) piece. The guides for the Normal and Colour jigsaws were identical to the final jigsaw image. The guide for the Outline jigsaw was a monochrome outline otherwise identical to the colour jigsaw image. Each picture jigsaw featured in either the normal guide or outline guide trials, randomised between children.

Figure 4 about here

Procedure

The experimenter placed the partially completed jigsaw in front of the child with the guide above it, and said “make this jigsaw [points at the jigsaw] look exactly like this picture [points at the guide]”. Three pieces that all fitted in shape were then placed below the jigsaw in random order: the correct piece, an alternative piece that differed from the guide but made a valid picture, and a piece with clearly incongruous content. The experimenter watched the child, noting whether the child looked back at the guide after the pieces had been laid out before choosing

The dependent variables were whether the correct piece was selected and whether children looked back at the guide after the pieces had been laid out before choosing. Children first had one jigsaw of each type, then False Belief and False Sign tasks administered as in

Experiment 1, followed by the remaining jigsaws, one of each type. The order of jigsaw type and False Belief and False Sign tasks was fully counterbalanced between children.

Results

Jigsaws

Both 3- and 4-year-olds performed significantly above chance on the six trials, $p = 1/3$, $t(23) = 3.54$, $p = .002$, $t(27) = 10.26$, $p < .001$, respectively. Number of times children chose the correct jigsaw piece for each jigsaw type is shown in Figure 5. There were no differences for the three jigsaw types, either overall (Friedman $\chi^2 = 2.75$, $p = .25$) or within each age group (3 years: Friedman $\chi^2 = 1.20$, $p = .55$; 4 years: Friedman $\chi^2 = 3.96$, $p = .138$). Figure 6 shows that similarly, there were no differences between the number of looks-to-guide for the three jigsaw types, either overall (Friedman $\chi^2 = 0.98$, $p = .67$) or within each age group (three years: Friedman $\chi^2 = 0.78$, $p = .68$; four years: Friedman $\chi^2 = 1.54$, $p = .46$). Correct choices for each type were correlated (all $r_s > .32$, all $p_s < .03$) as were looks for each type (all $r_s > .44$, all $p_s \leq .001$). The scores for correct choices and for looks-to-guide were therefore summed. Total Jigsaw-correct-choice improved between three and four years (Mann-Whitney $U = 193$, $p = .007$). Total Looks-to-guide increased between the two age groups, falling short of conventional significance (Mann-Whitney $U = 235$, $p = .056$).

Figures 5 and 6 about here

False Belief and False Sign tasks

Performances on each task improved between three and four years: False Belief, $M = .37$ and $M = .80$ respectively, Mann-Whitney $U = 193.00$, $p = .002$; False Sign, $M = .22$ and $M = .72$, respectively, Mann-Whitney $U = 166$, $p < .001$. The two tasks did not differ in difficulty,

Wilcoxon $Z = 1.34$, $p = 0.18$. However, the two tasks were not strongly correlated in this experiment, $r = .25$, $p = .078$. We report them combined as in Experiment 1, as well as separately.

Correlational analyses

Table 4 examines relations between age, verbal mental age, jigsaw mean accuracy, jigsaw number of looks at guide, and performances on the false belief and false sign tasks. Jigsaw accuracy was related to whether the guide was consulted, and remained so after partialling out age and verbal mental age. Jigsaw accuracy and guide use were related to metarepresentational performance, the relation with guide use remaining strong after age and verbal mental age had been partialled out. As can be seen from Table 4, these relationships are strongest for False Sign performance.

Table 4 about here

Discussion

The jigsaw task in Experiment 2 reduced the demands to a minimum, requiring children to simply select the matching piece from three alternatives. To be certain of being correct children had to consult the guide because two of the pieces would form a valid picture. Under these simplified conditions guide use did not differ between the different types of jigsaw, and increased between three and four years.

Guide use was significantly associated with metarepresentational performance, beyond common effects of chronological and verbal mental age. This finding should be interpreted with caution, as the tendency to consult the guide was weakly related to False Belief performance. Despite very similar levels of performance the False Belief and False

Sign tasks were not strongly related in this experiment; we do not know why. Nevertheless, we conclude that understanding of the usefulness of the guide increases between three and four years, and is related to at least one measure of metarepresentational development.

General Discussion

In two experiments we employed jigsaw puzzle completion to examine children's understanding of pictorial metarepresentation. We find that the use of pictorial information to aid jigsaw completion and the tendency to utilise a guide picture develop in the preschool period. Both abilities are associated with mental and graphical measures of metarepresentation. We conclude this is confirmatory evidence of development of a general metarepresentational understanding around the age of four years. Jigsaws are a particularly apposite test of this ability: children do play with jigsaws outside of experiments, and, unlike many other measures of metarepresentation, performance is unlikely to be affected by misunderstanding of the task demands or pragmatic misinterpretation.

Our specific focus was on two aspects of jigsaw completion that relate to the tendency to use pictorial information to aid completion: development of use of pictorial versus shape information, and the tendency to consult and use information from a guide showing the final picture. In Experiment 1 we examined the ability to use pictorial information in jigsaw completion. Younger children were specifically poor at using pictorial cues to aid completion. Use of pictorial information was strongly and specifically associated with tasks measuring mental and graphical metarepresentation. These findings suggest that the use of pictorial information in jigsaw completion reflects the development of a general concept of metarepresentation, which is evident in the mental, graphical and pictorial domain. This adds to the body of literature showing linked metarepresentational developments across the linguistic, mental, and pictorial domains (Diaz & Farrar, 2018; Doherty & Perner, 1998; Doherty & Wimmer 2005; Gollek, & Doherty, 2016; Perner, Stummer, Sprung, & Doherty,

2002; Perner & Roessler, 2012; Wimmer & Doherty, 2011). Specifically, we suggest that children begin to understand that pictorial elements, such as lines and patches of colour, can be arranged to produce a picture. This understanding is a critical stage in the ability to draw and paint.

More broadly, this finding also has theoretical relevance to suggestions that young children only fail theory of mind tasks because of the pragmatic demands of the tasks (e.g., Westra & Carruthers, 2017). These claims imply that performance should not associate with representational tasks that do not make the same pragmatic demands. A case could be made that previous non-mental metarepresentational tasks share many of the pragmatic demands of theory of mind tasks. They were after all designed with direct comparison to the False Belief task in mind (e.g., Doherty & Perner, 1998; Farrar, Ashwell, & Maag, 2005; Wimmer & Doherty, 2011). However, it is hard to see how the jigsaw tasks employed here would be subject to pragmatic misinterpretation. Children were simply asked to complete jigsaws, an activity we believe that most of our participants have experience of. Granted we did not measure degree of familiarity with jigsaws, but potentially poor performance as a result of lack of familiarity would differ in underlying nature to poor performance as a result of pragmatic misinterpretation. Thus it is hard to see why it should relate to pragmatic difficulties hypothesised to impair false belief performance.

Together, the experiments showed that children could use the spatial correspondence between a guide picture and the jigsaw, as long as task complexity was kept to a minimum. The effectiveness of the guide for the shape-cue jigsaw in Experiment 1 was unrelated to age. This suggests that even three-year-olds can use spatial correspondence in principle, in turn suggesting it is a precursor to understanding representation rather than a simultaneous development. A key theoretical distinction between understanding simple correspondence and understanding correspondence in a representational relation is understanding that the

relation should or is meant to obtain (Blades & Cooke, 1994; Perner, 1991). The shift to understanding the correspondence relation as representational should increase the probability of attending to and utilising this relation. This was confirmed in Experiment 2 by children's increasing tendency to consult the guide before selecting the jigsaw piece, and the substantial and significant association of this tendency to metarepresentational performance.

In other words, the ability to detect and use correspondence information in our jigsaw task is present from at least the age of 3 years and does not require metarepresentational understanding. However, metarepresentational understanding allows children to understand why the relationship exists, and therefore increases the probability of their attending to and utilising it.

Limitations

This conclusion must be made with caution because in Experiment 2 the two chosen measures of metarepresentational understanding were not themselves closely related. Guide consultation was primarily related to False Sign performance. The False Sign task had been hypothesised to be particularly apt comparator for guide use, since both indicate how a state of affairs should be. However, although these might be grounds to predict a stronger relation with the False Sign task, we still predicted a relation with the False Belief task. Since this did not obtain, we make no more detailed speculation.

It is possible that our choice of False Sign procedure influenced the reliability of the association between the two tasks. We modelled our task on Parkin's (1994) procedure. Although simple and brief, other versions of the task provide more plausible reasons for the sign to indicate a location. Arguably signs do not usually function to indicate the location of people. Another version used by Parkin (1994) involved the sign indicating the location of an ice cream van, which is a more obviously useful function. Bowler et al. (2005) went a step further by having instead a false signal to guide an automatic train, which closely matches the

real use of signals. Thus, children may have performed less well on our version than in previous studies because the informing function of the sign was less natural or obvious. Nevertheless, if this were so it is not clear why it should only have been so in one experiment, nor why performance was not measurably different from False Belief performance. The issue remains to be resolved in future work.

A further limitation is one of scope. We did not examine spatial abilities, other than indirectly measuring children's ability and inclination to use geometrical correspondence. We are confident that the development of jigsaw completion ability is related to spatial development more generally. Levine et al.'s (2012) observational study of 2- to 3-year-olds naturalistic play demonstrates subsequent associations with at least one spatial skill. Young et al. (2014) show that training with spatial language can enhance jigsaw completion skill, if accompanied by gesture. This places jigsaws in the context of other work examining the rapid development of spatial skills from preschool onwards (e.g., Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017; Schmitt, Korucu, Napoli, Bryant, & Purpura, 2018). Verdine et al. (2017) demonstrate that spatial skills measured at 3 years predict mathematical skills up to two years later, and argue for spatial skills' foundational importance in STEM education. As Kuhl, Lim, Guerriero, & van Damme (2019) suggest, play with spatial toys, including jigsaw puzzles, is likely to enhance these skills.

A full examination of the development of jigsaw completion ability would include the many spatial skills such play may enhance. The present study serves to suggest that jigsaw play also involves fundamental understanding of the nature of pictures, a finding which is unique to this study. Whether play with jigsaws enhances this understanding remains to be tested, as does whether pictorial understanding interacts with spatial understanding at this age. Young et al.'s (2014) training study's findings are promising in this regard. Their control condition involved pointing out non-spatial *pictorial* information, e.g., "This piece

has some light blue colors, so it will go in the sky.” This also led to substantial improvement, regardless of gesture.

Other issues where pictorial and spatial understanding interact include the understanding of maps. Maps have clear spatial functions, and typically correspond geometrically to the space they represent. Mature understanding of them also requires understanding of the symbols on them and the fact that they are graphic representations (Blades & Spencer, 1987; Liben & Downs, 1989). This understanding may be protracted, particularly if one considers understanding of the intentions of the cartographer (Myers & Liben, 2012). Mental imagery, may also in some contexts plausibly involve both the understanding of the mental image as a picture (e.g., Estes, 1998) and as a spatial object, as occurs in mental rotation (Levine et al., 2012; Wimmer, Maras, Robinson, Doherty, & Pugeault, 2015; Wimmer, Maras, Robinson, & Thomas, 2016).

Summary and Conclusion

In two experiments we examine the role of pictorial metarepresentation in jigsaw puzzle completion. In Experiment 1 we compared completion ability of jigsaws varying the presence of pictorial information with pieces of characteristic interlocking shapes or rectangles. Normal jigsaws with both information were completed faster than either shape-cue or picture-cue jigsaws. Most 3-year-olds completed the normal and shape-cue jigsaws but could not complete the picture-cue jigsaw within the 3 minutes allowed. Ability to do so was substantially associated with metarepresentational ability, measured by the False Belief and False Sign tasks. For the entire sample, the number of attempts made to connect jigsaw pieces was associated with metarepresentational ability, for the jigsaws with pictorial information. This suggests improvements in jigsaw skill is part of a general development of metarepresentational ability at around 4 years.

We also examined the effectiveness of providing a guide to the completed jigsaw. The ability to use a guide requires the ability to utilize its correspondence with the completed jigsaw. In Experiment 1, children were able to do this at all ages, but only for the shape-cue jigsaw. We suggest that this was because the simplicity of the shape-cue jigsaw allowed the effective use of the information. This was examined in Experiment 2 by minimising the demands of the jigsaw: children were provided with a nearly-completed jigsaw and asked to select the appropriate final piece. The primary measure was whether they consulted the guide before doing so. The tendency to do this did not differ for jigsaws for which the guide provided information about colour, pictorial content, or both. The tendency to use the guide increased from 3- to 4-years, and was associated with metarepresentational performance. We suggest that older children understand more about the intended relationship between the guide and the completed jigsaw, specifically that the two *should* correspond. This understanding is a precursor to understanding representation, and arguably an immediate precursor. This conclusion is limited by the fact that only the False Sign task was associated with guide-use.

This and other questions about jigsaw development remain for future research. We maintain that jigsaws are a natural and useful tool to study the development of both children's spatial abilities and their understanding of pictorial representation

References

- Allen, M. L., & Armitage, E. (2017). How children learn to navigate the symbolic world of pictures: the importance of the artist's mind and differentiating picture modalities. In *Advances in Child Development and Behavior, Vol. 52*, pp. 153-184.
- Beck, S. R., Robinson, A. N., Ahmed, S., & Abid, R. (2011). Children's understanding that ambiguous figures have multiple interpretations. *European Journal of Developmental Psychology, 8*, 403-422.
- Beilin, H., & Pearlman, E. G. (1991). Children's iconic realism: Object versus property realism. *Advances in Child Development and Behavior, 23*, 73-111.
- Blades, M., & Cooke, Z. (1994). Young children's ability to understand a model as a spatial representation. *The Journal of Genetic Psychology, 155*(2), 201-218.
- Blades, M., & Spencer, C. (1987). The use of maps by 4–6-year-old children in a large-scale maze. *British Journal of Developmental Psychology, 5*(1), 19-24.
- Bowler, D. M., Briskman, J., Gurvidi, N., & Fornells-Ambrojo, M. (2005). Understanding the mind or predicting signal-dependent action? Performance of children with and without autism on analogues of the false-belief task. *Journal of Cognition and Development, 6*(2), 259-283.
- Callaghan, T. C. (1999). Early understanding and production of graphic symbols. *Child Development, 70*, 1314-1324.
- Callaghan, T. C., Rochat, P., & Corbit, J. (2012). Young children's knowledge of the representational function of pictorial symbols: Development across the preschool years in three cultures. *Journal of Cognition and Development, 13*(3), 320-353.
- Claxton, L. J. (2011). An investigation of preschoolers' misattributions of the properties of two-dimensional images: understanding the relationship between a symbol and its referent. *Infant and Child Development, 20*(3), 301-312.

- DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, 238(4833), 1556-1557.
- DeLoache, J. S. & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition*, 52, 83-110.
- DeLoache, J. S., Kolstad, V., & Anderson, K. N. (1991). Physical similarity and young children's understanding of scale models. *Child Development*, 62(1), 111-126.
- Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: a longitudinal study. *Developmental Science*, 21(4), e12594.
- Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, 27(2), 367-392.
- Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press.
- Doherty, M. J. & Wimmer, M. C. (2005). Children's understanding of ambiguous figures: Which cognitive developments are necessary to experience reversal? *Cognitive Development*, 20, 407-421.
- Donnelly, K. E., Gjersoe, N. L., & Hood, B. (2013). When pictures lie: Children's misunderstanding of photographs. *Cognition*, 129, 51-62.
- Dunn, L. M., Dunn, L., Whetton, C., & Burley, J. (1997). British picture vocabulary scale II. *Windsor, UK: NFER-Nelson*.
- Farrar, M. J., Ashwell, S., & Maag, L. (2005). The emergence of phonological awareness: Connections to language and theory of mind development. *First language*, 25(2), 157-172.
- Fleer, M. (1990). The value of jigsaw puzzles in early childhood education. *Early Child Development and Care*, 60, 73-88.
- Gollek, C., & Doherty, M. J. (2016). Metacognitive developments in word learning: Mutual exclusivity and theory of mind. *Journal of Experimental Child Psychology*, 148, 51-69.

- Gopnik, A., & Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science*, 4(2), 175-183.
- Jastrow, J. (1900). *Fact and fable in psychology*. Houghton, Mifflin and Co.
- Jolley, R. P. (2010). *Children and pictures: Drawing and understanding*. John Wiley & Sons.
- Karadaki, T., & Doherty, M.J. (2017) Preschool metacognitive developments allow learning and use of co-referential words. Poster presented at the IX.Dubrovnik Conference on Cognitive Science, Dubrovnik, Croatia, May 2017.
- Kuhl, P. K., Lim, S. S., Guerriero, S., & van Damme, D. (2019), *Developing Minds in the Digital Age: Towards a Science of Learning for 21st Century Education*, Educational Research and Innovation, OECD Publishing.
- Leekam, S., Perner, J., Healey, L., & Sewell, C. (2008). False signs and the non-specificity of theory of mind: Evidence that preschoolers have general difficulties in understanding representations. *British Journal of Developmental Psychology*, 26, 485-497.
- Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: a predictor of preschoolers' spatial transformation skill. *Developmental psychology*, 48, 530-542.
- Liben, L. S. (2003). Beyond point and shoot: Children's developing understanding of photographs as spatial and expressive representations. *Advances in Child Development and Behavior*, 31, 1-43.
- Liben, L. S., & Downs, R. M. (1989). Understanding maps as symbols: The development of map concepts in children. In *Advances in child development and behavior* (Vol. 22, pp. 145-201). JAI.
- Nye, R., & Thomas, G. V. (1994). Children's conceptions of the relationship between pictures and their referents. *Cognitive Development*, 9, 165-191.
- Newcombe, N. S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. MIT Press.

- Parkin, L. J. (1994). Children's understanding of misrepresentation. Unpublished D.Phil Thesis, University of Sussex.
- Parkin, L. J., & Perner, J. (1996). Wrong directions in children's theory of mind: What it means to understand belief as representation. *Unpublished manuscript, University of Sussex.*
- Perner, J. (1991). *Understanding the Representational Mind*. The MIT Press.
- Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *The Quarterly Journal of Experimental Psychology, 61*(1), 76-89.
- Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*, 519-525.
- Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development, 17*, 1451-1472.
- Robinson, E. J., Nye, R., & Thomas, G. V. (1994). Children's conceptions of the relationship between pictures and their referents. *Cognitive Development, 9*(2), 165-191.
- Rock, I., Gopnik, A., & Hall, S. (1994). Do young children reverse ambiguous figures?. *Perception, 23*(6), 635-644.
- Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Development, 77*(4), 1034-1049.
- Schmitt, S. A., Korucu, I., Napoli, A. R., Bryant, L. M., & Purpura, D. J. (2018). Using block play to enhance preschool children's mathematics and executive functioning: A randomized controlled trial. *Early Childhood Research Quarterly, 44*, 181-191.

- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21(4), 237-249.
- Siegel, M. & Beattie, K. (1991). Where to look for children's understanding of false beliefs. *Cognition*, 38, 1-12.
- Simcock, G., & DeLoache, J. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology*, 42(6), 1352.
- Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, 69, 321-332
- Thomas, G. V., Jolley, R. P., Robinson, E. J., & Champion, H. (1999). Realist errors in children's responses to pictures and words as representations. *Journal of Experimental Child Psychology*, 74(1), 1-20.
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monographs of the society for research in child development*, 82(1), 7-30
- Wellman, H.M., (1990). *The Child's Theory of Mind*. Cambridge MA: Bradford.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72, 655-684.
- Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165-176.
- Wimmer, M. C., & Doherty, M. J. (2011). The Development of Ambiguous Figure Perception. *Monographs of the Society for Research in Child Development*, 76(1), 1-130.
- Wimmer M.C., Maras K.L., Robinson E.J., Doherty M.J., & Pugeault, N. (2015) How visuo-spatial mental imagery develops: Image generation and maintenance. *PLOS ONE* 10, (11) e0142566-e0142566.

- Wimmer, M. C., Maras, K. L., Robinson, E. J., & Thomas, C. (2016). The format of children's mental images: Evidence from mental scanning. *Cognition*, *154*, 49-54.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.
- Wimmer, M. C., Robinson, E. J., Koenig, L., & Corder, E. (2014). Getting the Picture: Iconicity Does Not Affect Representation-Referent Confusion. *PloS one*, *9*, e107910.
- Young, C., Cartmill, E., Levine, S., & Goldin-Meadow, S. (2014, January). Gesture and speech input are interlocking pieces: The development of children's Jigsaw Puzzle Assembly ability. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 36, No. 36.

Table 1: Mean proportion correct responses on the False Belief, False Sign, and respective memory control questions in Experiment 1.

	3-year-olds (N = 29)	4-year-olds (N = 39)	5-year-olds (N = 40)
False Belief	.21	.62	.82
False Sign	.24	.64	.94
FB Memory	.93	1.00	1.00
FS Memory	.98	1.00	.98

Table 2: Mean proportion of completion within 3 minutes as a function of age group, jigsaw type, and guide use in Experiment 1.

		3-year-olds (N = 29)	4-year-olds (N = 39)	5-year-olds (N = 40)
Normal jigsaw	No guide	.93	1.00	1.00
	Guide	.80	1.00	1.00
Shape-only	No guide	.71	.95	1.00
	Guide	.87	1.00	1.00
Picture-only	No guide	.43	1.00	.96
	Guide	.40	.95	1.00

Table 3. Correlations between age, metarepresentational performance, and jigsaw attempts in Experiment 1. Partial correlations controlling for age are shown below the diagonal.

	Metarep ⁿ	Normal	Shape	Picture
Age	.55***	-.44***	-.35***	-.46***
Metarepresentation	—	-.46***	-0.29***	-.42***
Normal	-.29*	—	.41***	.45***
Shape	-.12	.29*	—	.30***
Picture	-.22*	.32*	.17	—

Note. *** = $p < .001$; * = $p < .05$

Table 4: Correlations of performance and partial correlations below the diagonal controlling for age and verbal mental age in Experiment 2.

	Verbal mental age	Jigsaw accuracy	Jigsaw guide look	False Belief	False Sign	Metarep ⁿ
Age	.50***	.37**	.26†	.43**	.50***	.61***
VMA	---	.35*	.00	.53***	.28†	.50***
Accuracy		---	.47***	.22	.45***	.43**
guide look			---	.18	.40**	.37**
False Belief				---	.25†	
False Sign					---	
Metarep ⁿ						

Note. *** = $p < .001$; ** = $p < .01$; * = $p < .05$; † = $p < .10$

Figure 1. Jigsaws used in Experiment 1, normal jigsaw (left), shape-only (middle), and picture only (right).

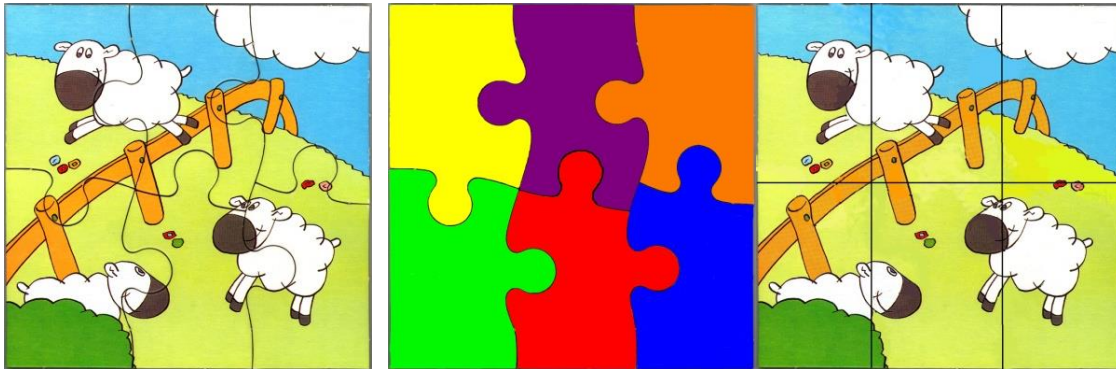
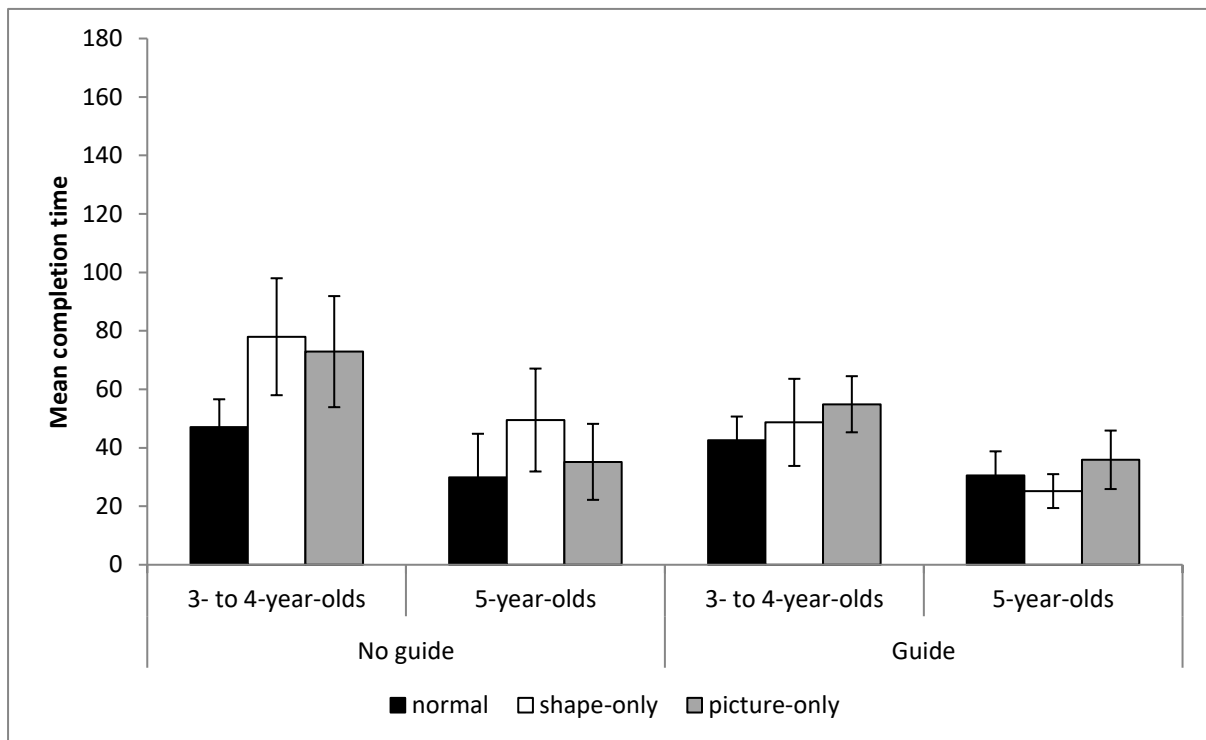
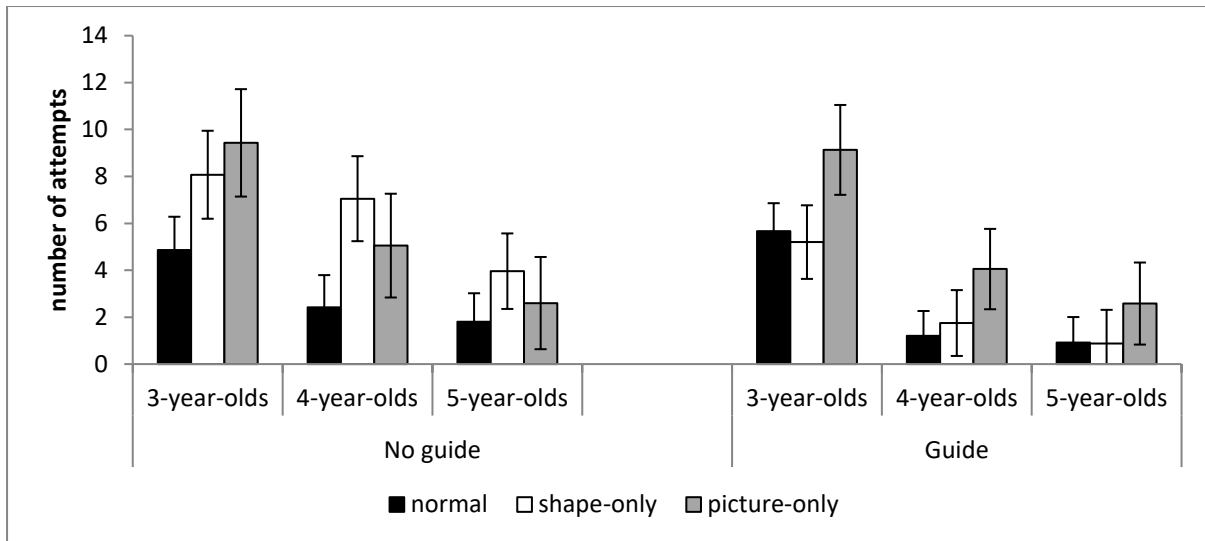


Figure 2. Mean time in seconds taken to complete each jigsaw for each age group in the guide and no guide condition of Experiment 1.



Note: error bars show 95% CI.

Figure 3. Mean number of attempts for each jigsaw for each age group in the guide and no guide condition of Experiment 1.



Note: error bars show 95% CI.

Figure 4. An example set of jigsaws in Experiment 2: from top to bottom, two normal jigsaws, 2 color jigsaws, 2 outline jigsaws.

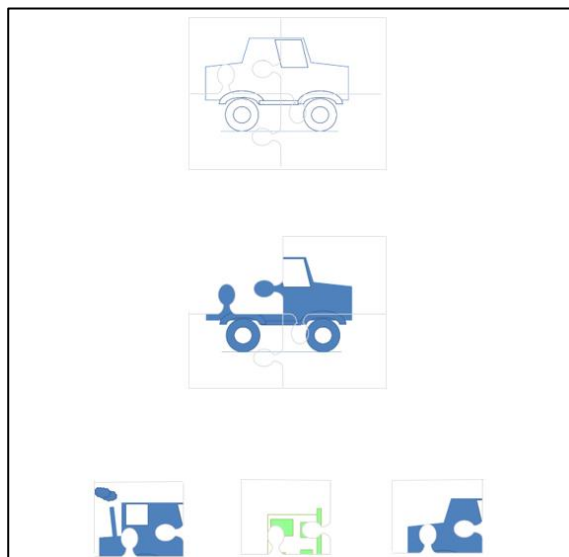
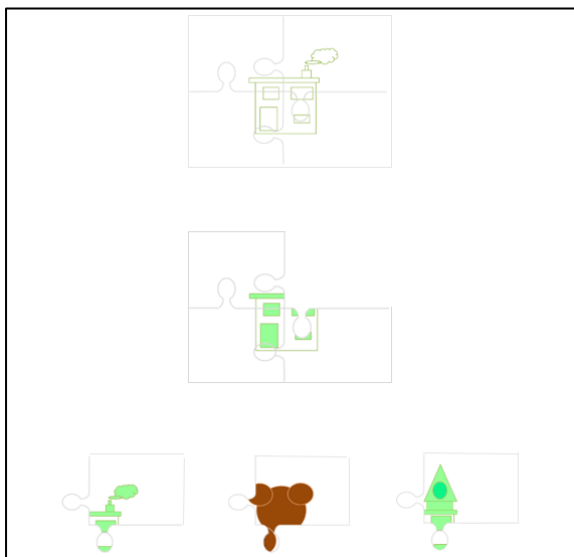
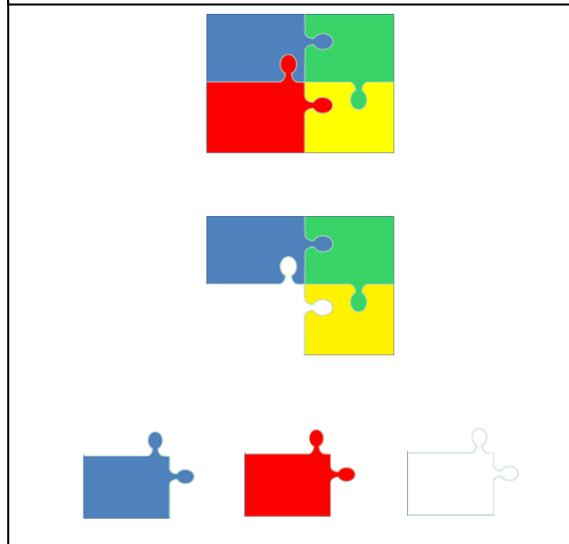
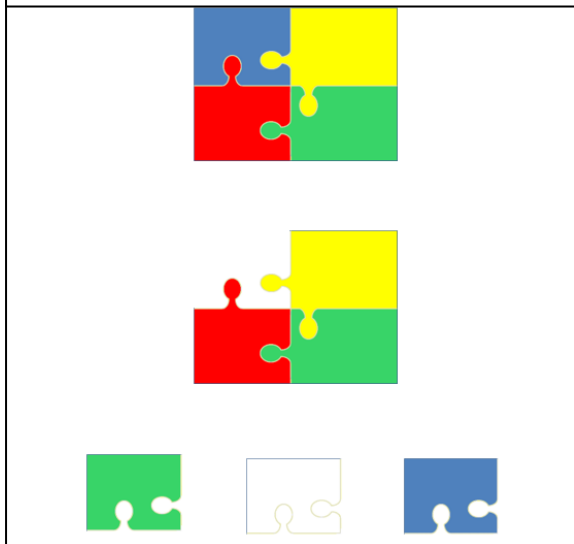
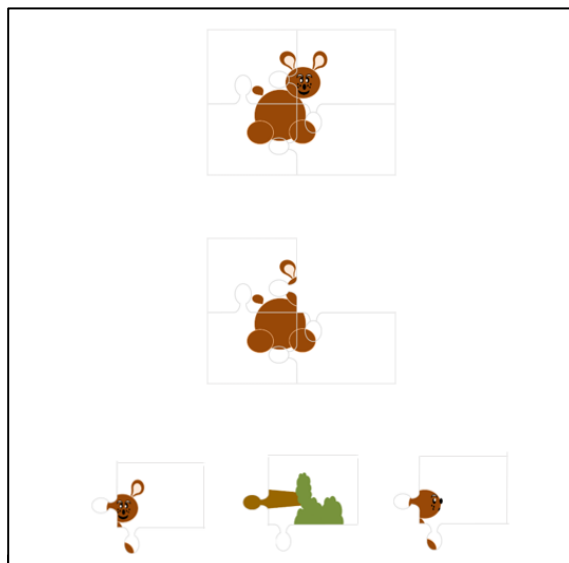
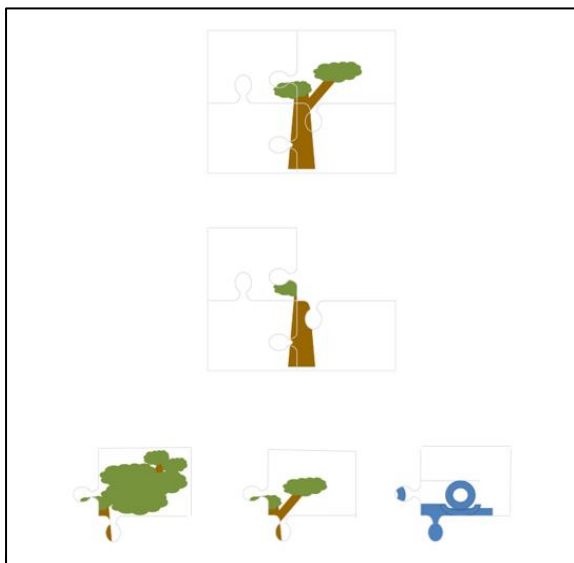
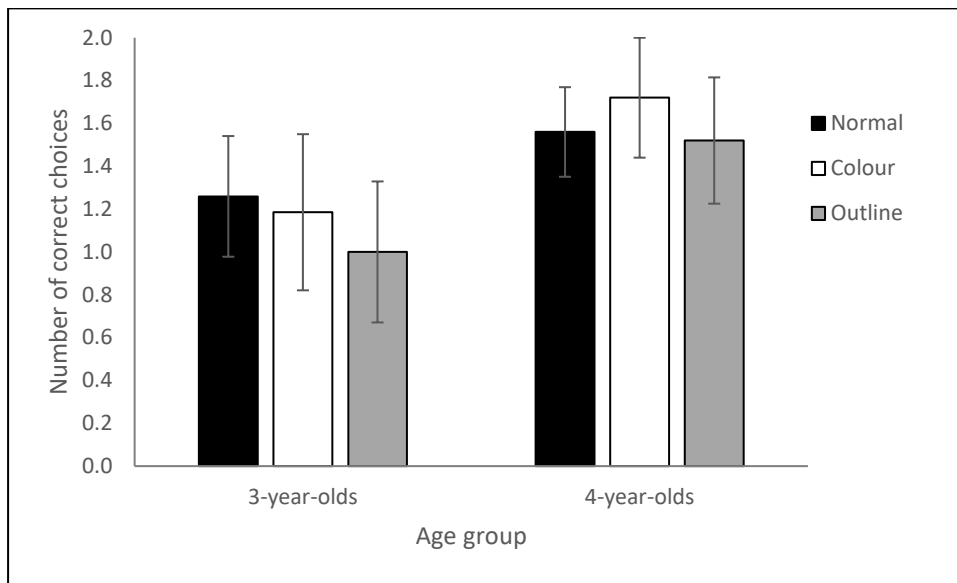
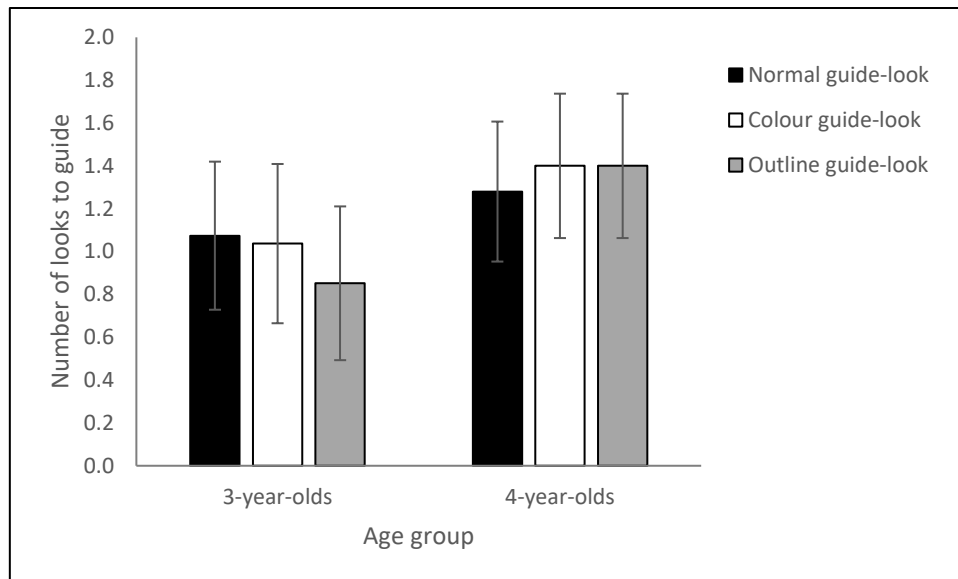


Figure 5. Mean choices of correct jigsaw piece for the three jigsaw types in Experiment 2.



Note: error bars show 95% CI.

Figure 6. Mean looks to guide before choosing a jigsaw piece for the three jigsaw types in Experiment 2.



Note: error bars show 95% CI.