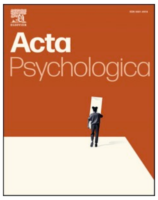
Neilson, Darcy, Cutting, Nicola ORCID logoORCID: https://orcid.org/0000-0002-3155-9566 and Tecwyn, Emma C ORCID logoORCID: https://orcid.org/0000-0002-2343-2282 (2026) Can Object Exploration or Explanation-Generation Facilitate Innovative Problem-Solving in 5–7-Year-Olds? Acta Psychologica, 264 (10639).

Acta Psychologica 264 (2026) 106396

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



Can object exploration or explanation-generation facilitate innovative problem-solving in 5–7-year-olds?☆

Darcy Neilson *, Nicola Cutting, Emma C. Tecwyn

School of Education, Language & Psychology, York St John University, United Kingdom of Great Britain and Northern Ireland

A R T I C L E I N F O

Keywords:
Cognitive development
Exploration
Explanation
Innovation
Problem-solving
Tool-use

A B S T R A C T

Children struggle with independently innovating solutions to physical problems until around 8-years-old. This pre-registered study tested whether encouraging 5–7-year-olds ($N = 83$) to (1) explore task materials or (2) generate explanations would support innovative problem-solving in two tasks that involved accessing an out-of-reach item (the hook task and water-displacement task). Children were assigned to one of three conditions (Explore, Explain, Baseline). In the pre-test phase children were prompted to: (1) explore task materials (the functional and non-functional items presented alongside each task), (2) explain how the materials could be used to problem solve, or (3) complete a filler task, respectively. During the test phase of the Explain condition the experimenter used verbal prompts (e.g., "Why did that happen when you did that?") to elicit explanations from children. These were matched with neutral "report" prompts ("*What happened when you did that*?") in the Explore and Baseline groups. Overall success rates were 48.2% (hook task) and 59% (water-displacement task), but there were no significant differences between conditions in success, whether children used the functional item first, or time to success. These findings indicate that neither the Explore nor the Explain manipulations used in the current study facilitated children's problem-solving performance. Exploratory analyses suggested that, regardless of condition, object manipulations, independent discoveries, and iterative tool refinement were associated with success on the hook task**.** Analyses of children's verbal utterances revealed an association between cognitive speech and success on the hook task. Taken together, these findings suggest that self-directed exploration and task engagement, rather than externally prompted efforts, supported problem-solving success.

1. Introduction

Innovating solutions to physical problems is crucial for navigating the world (Joh, Jaswal, & Keen, 2011), overcoming challenges (Chappell, Cutting, Apperly, & Beck, 2013), and performing tasks efficiently (Pauen & Bechtel-Kuehne, 2016). Tool-use is fundamental to problem-solving (Nielsen, 2012) and is often central to everyday tasks, such as using kitchen utensils for food preparation. Though proficient in tool *use* (selecting and using an appropriate tool), young children are remarkably unskilled at solitary tool *innovation* (generating a tool) in the context of generating novel solutions to physical problem-solving tasks (e.g., Beck, Apperly, Chappell, Guthrie, & Cutting, 2011; Cheke, Loissel, & Clayton, 2012).

Originally used to study tool innovation in New Caledonian crows (Taylor, Elliffe, Hunt, & Gray, 2010; Weir, Chappell, & Kacelnik, 2002), the 'hook task', where children must bend a pipe cleaner into a hook shape to retrieve an out-of-reach reward from a transparent vertical tube is the main paradigm for investigating tool innovation in children. Success rates in the hook task increase with age, from 8 to 20% in 4–5-year-olds (Beck et al., 2011; Cutting, Apperly, Chappell, & Beck, 2014) to 60–65% around ages 8–9 (Beck et al., 2011; Voigt, Pauen, & Bechtel-Kuehne, 2019; though Sheridan, Konopasky, Kirkwood, & Defeyter, 2016 report 44% success in children as young as 4–5-year-olds). This contrasts sharply with the 100% success rate observed in adult comparison groups (Beck et al., 2011). Despite extensive research, the factors constraining children's poor performance remain poorly understood (e.g., Beck, Williams, Cutting, Apperly, & Chappell, 2016; Chappell et al., 2013; Voigt et al., 2019).

Another task that has been used to investigate physical problem-solving processes in both non-human animals and children is the water-displacement task (e.g., Bird & Emery, 2009; Cheke et al., 2012; Loissel, Cheke, & Clayton, 2018; Miller, Jelbert, Loissel, Taylor, &

☆ This article is part of a Special issue entitled: 'Innovation and creativity' published in Acta Psychologica.

* Corresponding author at: York St John University, Lord Mayor's Walk, York, YO31 7EX, United Kingdom of Great Britain and Northern Ireland.
E-mail addresses: d.neilson@yorksj.ac.uk (D. Neilson), n.cutting@yorksj.ac.uk (N. Cutting), e.tecywn@yorksj.ac.uk (E.C. Tecwyn).

https://doi.org/10.1016/j.actpsy.2026.106396
Received 30 July 2025; Received in revised form 20 November 2025; Accepted 2 February 2026

Clayton, 2017; Taylor et al., 2011). This paradigm is based on Aesop's fable of the crow and the pitcher, where a thirsty crow uses stones as a tool to raise the level of water in a container to bring it within reach for drinking. In the experimental version, participants are typically presented with a transparent vertical tube part-filled with water with an out-of-reach reward floating on top. To solve the task, participants need to use functional objects (i.e., objects that sink) as tools to raise the water level enough to enable retrieval of the reward (e.g., Loissel et al., 2018).

When exclusively presented with heavy, sinking objects, children aged 5–7-years typically require five consecutive trials (where a trial ended after 2 min, or once all items had been inserted) to discover the solution of dropping these objects into the water to access the reward (Cheke et al., 2012). However, Miller et al. (2017) reported that when the task was made more challenging by presenting children with both solid objects that sank and hollow objects that floated, 5–7-year-olds required an average of 20 trials to succeed. Most 9-year-olds succeed on variations of the water-displacement task on their first trial, with Loissel et al. (2018) attributing this to their ability to mentally simulate the outcomes of their actions. Notably, success rates in water-displacement tasks resemble those observed in the hook task, where children begin to consistently succeed at around 8 years of age (Cheke et al., 2012).

In recent years, attention has shifted towards understanding the factors constraining children's tool innovation and innovative tool use (Rawlings & Legare, 2021), with work to date not having reached any clear conclusions (e.g., Beck et al., 2016; Cutting et al., 2014; Cutting, Apperly, & Beck, 2011). One cognitive process implicated in problem recognition and solution generation is causal reasoning (Rawlings & Legare, 2021): the ability to infer how actions influence outcomes (Sobel & Legare, 2014). Children's causal reasoning develops as they learn to identify and represent causal relationships in their environment to guide behaviour (Beller & Kuhnmünch, 2007; Bonawitz, van Schijndel, Friel, & Schulz, 2012; Gopnik, 2012). Although preschoolers and even toddlers can succeed in some observational causal learning tasks (e.g., Tecwyn, Mazumder, & Buchsbaum, 2023), children's causal reasoning skills continue developing throughout childhood (Goddu & Gopnik, 2024). It is not until much later in development (around 6 to 10 years) that children develop an appreciation for complex causal systems (e.g., understanding that a machine with more intricate internal mechanisms can perform a wider range of functions, Goddu and Gopnik (2024)). Since the hook and water-displacement task both require participants to differentiate between a functional (pipe cleaner; heavy ball) and a non-functional (string; light ball) item and understand how functional items can be used to bring about a desirable outcome, being able to solve these tasks likely requires causal reasoning skills.

Related to understanding of causality, another possible explanation for young children's difficulty with tool-innovation tasks may be the failure to discern the relevant affordances of the objects provided (Neldner, Mushin, & Nielsen, 2017). Children exhibit optimal performance on problem-solving tasks when the causal connection between a tool's form and functions are emphasised (Bechtel, Jeschonek, & Pauen, 2013; Fong, Puebla, & Nielsen, 2024; Gardiner, Bjorklund, Greif, & Gray, 2012; Goswami & Brown, 1990). Correspondingly, in Neldner et al.'s (2017) study, children aged 3–5-years were over 9 times more likely to innovate a hook tool when a hook shape was visible, compared to when it was not. Thus, children can identify a hook-shape as providing the necessary affordance to solve the task but encounter difficulty in innovating this tool shape autonomously (Neldner et al., 2017).

If causal reasoning and understanding of object affordances are important for being able to innovate solutions to physical problems, one way to facilitate this might be through encouraging children to explore the task objects. Haptic exploration helps children discover causally relevant object properties through perceptual feedback (Klatzky & Lederman, 1993; Bechtel et al., 2013). Early in development, children tend to explore more freely and unpredictably, often trying things at random (Meder, Wu, Schulz, & Ruggeri, 2021). As they mature, their actions become more systematic and goal-directed, at least in explore-exploit task paradigms (Meder et al., 2021). Gaining knowledge about the functional properties of the task materials via self-directed exploration could offer a means of self-scaffolding, aiding in children's causal understanding of how to use the resources independently to accomplish their goal (Bechtel et al., 2013; Meder et al., 2021). However, children seldom spontaneously explore task materials before attempting the hook task, possibly due to impulsivity (Chappell et al., 2013) or time constraints (Neldner et al., 2019), which may have hindered children's ability to recognise relevant object affordances in previous studies (Neldner et al., 2017).

Chappell et al. (2013) introduced a 10-s delay before the hook task and told children they could play with (i.e., explore) the materials, hypothesising that this would support affordance recognition. The delay period did not significantly impact success rates. Descriptive data collected in this study did, however, indicate that those in the delay/explore condition were more engaged with the materials (they showed more exploratory behaviours such as bending and combining materials) than the control group, suggesting they had a different learning experience by virtue of the opportunity to explore (Chappell et al., 2013). Older children (6–7-years) exhibited a greater number of tangible exploratory behaviours in the 10 s pre-task phase, including combining and bending materials, whereas younger children (4–5-years) spent notably less time devising a strategy before beginning the task. The authors concluded that the opportunity for exploration did not impact the likelihood of successful innovation, speculating that a lack of exploration of task materials cannot explain children's tool-innovation difficulties.

Various methodological constraints may have contributed to the lack of differences between the explore and baseline groups in Chappell et al. (2013). For example, the duration for exploration was very limited at 10 s. Neldner et al. (2017) argued that extending exploration time prior to testing could offer a better opportunity for material exploration, potentially increasing success rates. Additionally, although success in innovation tasks is predominately defined in binary terms (either the child successfully creates a tool or does not) focusing solely on this outcome can overlook the depth and complexity of the problem-solving strategies and the creative process which precedes it (Rawlings, 2022; Burdett & Ronfard, 2023). More fine-grained analyses of behaviour, such as Chappell et al.' (2013) descriptive analyses of interactions with task materials, are likely to offer more insight into innovative performance (Breyel & Pauen, 2022).

Another possible route to enhancing causal understanding, and therefore potentially improving innovation of solutions in physical problem-solving tasks, may be via encouraging children to generate explanations regarding why their actions produce particular outcomes (Legare & Clegg, 2014). Research by Legare and Lombrozo (2014) suggests that children's explanations are especially effective in promoting causal learning. Asking 3–6-year-olds to explain the workings of a novel toy led to better causal learning compared to simply observing the toy (Legare & Lombrozo, 2014). Walker, Bonawitz, and Lombrozo (2016) offer further support for this idea, demonstrating that asking children to explain their observations, e.g., "*Why did that happen* when you did that?", can help them discover underlying causal structures. This suggests an association between generating explanations and causal learning; however, it does not explain the direction of this relationship. For example, a child with a stronger initial understanding of causality might articulate their explanations more clearly and use this information more productively. The benefits of explaining likely depend on their content, quality, and the underlying knowledge they represent, rather than the specific type of instruction given (Walker et al., 2016; Legare, 2014).

In a follow-up study, Legare and Lombrozo (2014) examined the type of verbal responses children gave to examine whether improvements in causal learning were due to global language use or a specific benefit of explanations. This time, children were either asked to explain or simply describe how the toy worked. Across both conditions, children who provided explanatory verbal responses outperformed those who only described the toy, on understanding the toy's mechanical functions and in reconstructing it. These results suggest that the advantages of explaining are distinct and support learning in ways that general language use does not. Interestingly, children were equally likely to generate causal explanations in the explain and describe conditions, possibly because the instruction to describe or explain were too subtle for children to differentiate between (Legare & Lombrozo, 2014). In more recent studies, explain-prompts were instead matched with report-prompts since they both (1) direct children's attention towards available evidence (Walker & Gopnik, 2017; Legare & Lombrozo, 2014), (2) create a pedagogical context for learning (Brockbank & Walker, 2022) and (3) ensure equivalent time engagement across trials (Walker et al., 2016). The current study employs the use of explain and report prompts, consistent with these suggestions.

Encouraging children to explain how they might approach solving a task prior to them actually attempting it might also facilitate innovative problem-solving via the generation of an appropriate plan. It is possible that young children may have the capacity to generate the correct idea (i.e., form a hook in the hook task) but struggle to devise an appropriate plan for using the materials effectively to achieve this (Bjorklund & Gardiner, 2010; Neldner et al., 2017), or struggle to mentally simulate the outcomes of their actions (Loissel et al., 2018). Generating an explanation ahead of starting a task might help children foresee potential shortcomings in their idea, reducing the likelihood of ineffective strategy perseveration (Kaller, Rahm, Spreer, Mader, & Unterrainer, 2008; Neldner et al., 2017).

Planning is a sophisticated cognitive process operating in tandem with numerous supporting executive functions, including cognitive flexibility (McCormack & Atance, 2011), self-monitoring (Veenman, 2016), and goal-orientation (Keen, 2011). Whereas most 4–5-year-olds struggled with a physical problem-solving task that required formulation of a plan prior to acting, children aged 9–10-years were able to plan multiple steps ahead (Tecwyn, Thorpe, & Chappell, 2014). If a limited ability to anticipate consequences of actions is a factor contributing to poor innovative problem-solving, prompting younger children to explain how they might solve a problem could improve their planning and decision-making skills. Though there is evidence that asking 3-year-olds to visualise a solution facilitates their spatial problem-solving (Joh et al., 2011), to our knowledge, whether asking children to generate explanations before and during innovative problem-solving tasks might facilitate performance has not yet been examined.

Studying whether children's verbal explanations help them to innovate solutions could expand current understanding of how private speech and planning skills contribute towards successful problem-solving. Children spontaneously and frequently talk to themselves whilst engaging with problem-solving activities (Winsler, 2009; Lidstone, Meins, & Fernyhough, 2010). This self-directed speech has repeatedly been shown to aid in planning and performance (Berk, 2014; Fernyhough & Fradley, 2005). Indeed, Lidstone et al. (2010) found that suppressing self-directed speech in a sample of 7–10-year-olds of negatively impacted performance on the Tower of London task, suggesting planning during middle-childhood is largely verbally mediated.

Breyel and Pauen (2022) were the first to examine private speech during tool-innovation tasks, studying the verbal utterances made by 3–5-year-olds during two variations of the hook task. Successful children generated more cognitive speech (orienting to planning, strategizing, and idea-generation) than non-successful children. Because tool-innovation tasks require open-ended problem-solving, planning, and creativity to solve, these findings suggest that cognitive speech may play a key role in supporting such behaviours. Meta-cognitive speech, associated with goal, task progress and error-monitoring, was also more frequent in successful children, attributed to increased persistence and motivation to overcome difficulties (Breyel & Pauen, 2022).

Taken together, current research suggests that young children's difficulties with innovative problem-solving may stem from limited causal reasoning skills (Rawlings, 2022), failure to recognise object affordances (Neldner et al., 2017), or underdeveloped planning skills (Tecwyn et al., 2014). Recent findings also indicate that private speech (especially cognitive and metacognitive talk) may support success by fostering strategy consideration, causal reasoning, and persistence (Breyel & Pauen, 2022). Thus, prompting children to explore or explain may offer self-scaffolding opportunities that enhance performance, potentially via improved causal understanding of task-relevant features (Bechtel et al., 2013; Legare & Lombrozo, 2014; Meder et al., 2021).

1.1. *The current study*

This study aimed to examine whether prompting 5- to 7-year-old children to explore task materials or generate explanations could improve innovative problem-solving. Children completed two tasks (the hook and water-displacement tasks), solving both of which involves exploiting the functional properties of objects to use them as tools. We chose 5- to 7-year-olds because they demonstrate relatively poor performance in these tasks (e.g., Beck et al., 2011; Chappell et al., 2013; Cheke et al., 2012), therefore increasing the chances of seeing an effect if either of our manipulations is effective. Children were allocated to one of three conditions: Explore, Explain or Baseline, differing in (a) the pre-task activities and (b) the verbal prompts used by the experimenter during testing (explain vs report, adapted from Walker et al., 2016).

If object exploration enhances innovative problem-solving, children in the Explore condition should be more likely to succeed, in comparison to children in the Baseline condition. If encouraging children to generate explanations enhances innovative problem-solving, children in the Explain condition should be more likely to succeed, in comparison to children in the Baseline condition. Independent exploration has been shown to improve children's understanding of object affordances (Bonawitz et al., 2011; Loissel et al., 2018; Neldner et al., 2017), aiding their causal understanding (Legare, 2012; Schulz & Bonawitz, 2007). Therefore, we expect children in the Explore condition to select the functional item (pipe cleaner or heavy ball) for the first insertion into the tube more often and solve the tasks more quickly than those in the Baseline condition. No specific predictions were made for the Explain condition regarding the first item inserted or time to success. We expected success to increase with age on both tasks, consistent with prior research showing innovative problem-solving performance improves over this age range (e.g., Hanus, Mendes, Tennie, & Call, 2011; Cutting et al., 2011; Miller et al., 2017; Loissel et al., 2018; Rawlings et al., 2022).

Current research relies overwhelmingly on binary performance metrics (e.g., success, first object inserted), limiting our understanding of innovative problem-solving processes to a simplistic pass or fail categorisation (Breyel & Pauen, 2022; Burdett & Ronfard, 2023; Rawlings, 2022). This study aimed to overcome this limitation and explore the richness, depth and complexity of children's behaviours during innovative problem-solving tasks by also examining children's behaviours and verbalisations throughout the tasks.

2. Method

This study's design, hypotheses and analysis plan were pre-registered; see https://doi.org/10.17605/OSF.IO/FHY5T. Ethical approval for this study was granted by the [anonymised for review] Ethics Committee [ETH2425-0120].

2.1. *Participants*

Eighty-three children (45 male; 38 female), aged 5 years 4 months to 7 years 9 months (M = 6 years 6 months, SD = 8 months), were recruited from three primary schools across, West Yorkshire, UK. The target sample size of 90 children (30 per condition) was based on previous similar research (20–30 children per cell, e.g., Cutting et al., 2014; Miller et al., 2017). Our final sample reflects what was achievable given the economic and time constraints of the project. Gatekeeper consent was obtained from school headteachers prior to administering a digital information sheet and consent form to caregivers. Descriptive statistics for gender and age by condition, as provided by parents upon form completion, are presented in Table 1.

The sample was predominantly White British (91.6%), with a small number of children from other ethnic backgrounds. Most parents had completed higher education or vocational equivalents (61.4%), and the largest proportion of households reported annual incomes between £50,001 and £75,000 (see Table S1 for full profile).

2.2. *Materials*

Children were presented with two physical problem-solving tasks: a hook task and water-displacement task. Both tasks consisted of a transparent vertical tube containing an out-of-reach token which could be exchanged for a reward (sticker) upon retrieval. Children were provided with objects to assist with retrieval, which varied according to whether they were functional (could be used effectively to retrieve the token) or not (ineffective for retrieving the token). The materials and apparatus used for each task are described in detail below.

2.2.1. *Hook task*

A transparent 12 × 4.5 cm acrylic tube was affixed to a 24 × 32 × 1.2 cm polyethylene board using strong adhesive glue-gel (Fig. 1a). A foam sponge cut into a circular shape with a slightly smaller diameter than the tube was attached to the bottom of a miniature 2.7 × 4.5 cm bucket with a metal-wire handle and dropped inside the tube. Children were also supplied with a red 30 cm pipe cleaner (*functional*-item) (Fig. 1c) and a red 30 cm piece of polypropylene rope string cord (*non-functional*-item) (Fig. 1d). A wooden 1.5 × 1.5 cm teddy-shaped token was placed inside the bucket as an incentive for retrieval during the test trials. To succeed on this task, participants bend the pipe-cleaner into a hook and use the hook to fish out the bucket containing the token.

2.2.2. *Water-displacement task*

The dimensions and construction of the transparent tube was the same as for the hook task (Fig. 1b). Children were provided with 15 (3 g) marbles (*functional*-item) (Fig. 1c) and 15 (0.22 g) cork balls (*non-functional*-item), which were each 1.4 cm in diameter (Fig. 1d). Each item was coated in one layer of nonporous wax-paint undercoat and two layers of water-resistant yellow paint to make them appear visually indistinguishable, as per Cheke et al. (2012). Balls were presented intermixed together in a transparent bowl during testing. The vertical tube was filled with 200 ml of water with a wooden teddy-token floating on the surface approximately 7 cm from the top of the tube (out of reach). To succeed, children needed to insert the heavy balls to raise the water level to bring the token within reach.

2.3. *Design*

The study was divided into two phases: a 1-min pre-test which varied between the three conditions (exploration / explanation / baseline) and a 5-min test phase per task in which children attempted to retrieve the token using the materials provided (Fig. 2). Children were semi-randomly allocated to one of the three conditions for both problem-solving tasks, with the aim of achieving an approximately even balance of age and gender across conditions.

Table 1
Gender distribution and average age across conditions.

Condition	Gender		Age (months)		Number of participants
	Male: n (%)	Female: n (%)	M	SD	
Explore	17 (60.7)	11 (39.3)	76.71	7.98	28
Explain	13 (48.1)	14 (51.9)	77.22	8.91	27
Baseline	15 (53.6)	13 (46.4)	77.96	8.04	28
Total	**45 (54.2)**	**38 (45.8)**	**77.3**	**8.23**	**83**

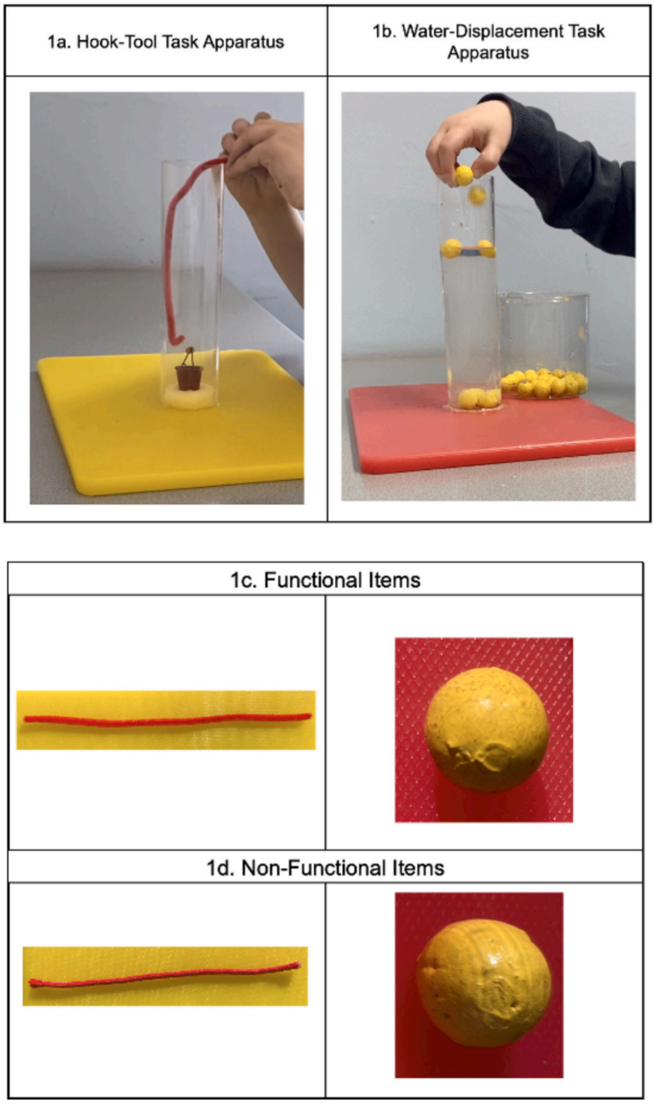


Fig. 1. Task apparatus for hook and water-displacement problem-solving tasks with examples of visually matched functional (pipe cleaner; marbles) and non-functional items (string; cork ball) provided in each task.

2.4. *Procedure*

Children were tested individually in a quiet area of their school. All children provided verbal assent to be video recorded during testing, with parental consent confirmed prior to data collection. In all conditions for both tasks, children were shown the task apparatus and received an initial description explaining, "*If you can get the token out of this tube, you can swap it with me for a sticker!*".

2.4.1. *Pre-test phase*

In all conditions, the puzzles were in view but out of reach for the child. The conditions varied as follows:

2.4.1.1. Exploration condition. The child was presented with the task materials available to help them solve the corresponding puzzle (pipe

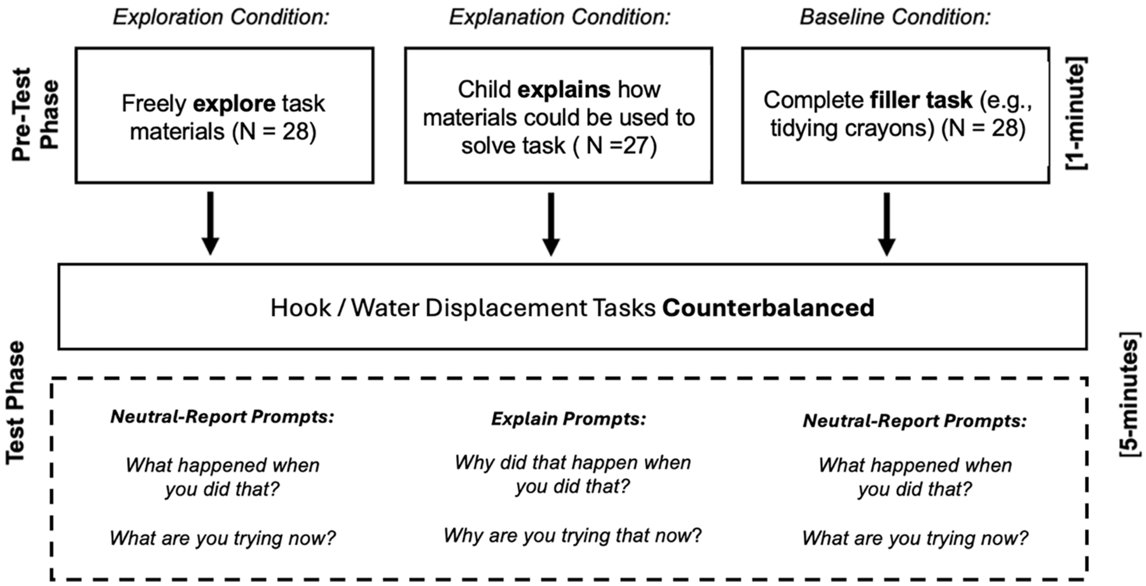


Fig. 2. An illustrative diagram depicting the study design and task structure.

cleaner/string alongside the hook task and marbles/cork balls with the water-displacement task) and told, *“You can have a go in a minute, but first can you play with these things here to find out how you could use them to get the teddy?”* Children were given 1-min to freely explore the objects in any way they desired. If the child only played with one type of object (i.e., only functional or only non-functional), the researcher asked, *“Have you explored all the objects?”*

2.4.1.2. Explanation condition. The experimenter held up the corresponding task materials and explained to the child, “You can have a go in a minute, but first can you think about and tell me how you could use these things here to get the teddy?”. Children were given 1-min to discuss their ideas with the researcher. To encourage open-ended idea generation and avoid influencing subsequent behaviour (e.g., favouring one idea over another), the researcher consistently responded to each idea by saying, “That's a good idea! Is there anything else you'd like to try?”

2.4.1.3. Baseline condition. Children completed a filler task whereby they were introduced the task and told, “*You can have a go in a minute, but first please can you help me tidy these things up?*” Before their first task, children were asked to tidy up crayons. Before their second task, children were asked to help tidy up craft items. This ensured the pre-test phase was equal duration across conditions as well as matched for the active involvement of the child.

2.4.2. 5-Minute test phase

Following the pre-test phase, the puzzle was moved within the child's reach, and they were told “*“It's your turn now. You can use these things here in any way that you like to help you get the teddy!”*. The testing duration was five minutes, aligning with Neldner et al.'s (2017) observations on the restrictive nature of time constraints in exploration. This extension from the traditional one-minute timeframe also allowed participants sufficient time to develop and articulate their explanations. The verbal prompts used by the experimenter during the test phase varied across conditions as follows:

2.4.2.1. Explanation condition. Two *explain*-prompts were used to support a child in explaining their behaviours ‘live’ as they worked through the tasks during the *explanation* condition. The number of prompts was not experimentally controlled, as prompt delivery was contingent on each child's engagement with the apparatus. Taken from Walker et al. (2016), the first prompt, *“Why did that happen when you did that?”,* was utilised each time the child attempted to solve the task (e.g., inserted an object (either functional or non-functional) into the tube).Thus, prompt frequency was not an independent manipulation but instead varied reactively in response to the child's behaviour. However, restricting child-researcher dyads to a meticulously controlled script constrains our understanding of the authentic nature of social interactions and their influence on social learning mechanisms (Rawlings, 2022). Therefore, a second *explain*-prompt, “*Why are you trying that now*?”, was devised to increase the ecological validity of the experimental procedure. This prompt was used when a child manipulated the materials in ways not immediately relating token retrieval (i.e., no object was directly inserted into the tube), but which did demonstrate task engagement, such as weighing the heavy and light balls in their hands or curling the end of the string and noticing it did not hold its shape.

2.4.2.2. Exploration and baseline condition. Two neutral ‘report’ prompts were employed during the test phase of the *exploration* and *baseline* conditions, aiming to encourage children to describe rather than explain their behaviours. The experimenter asked, “*What happened when you did that?*” after the child tried to solve the problem (e.g., inserted the pipe cleaner or dropped a ball into the tube). If the child was engaging with the materials in a way that was related to the task but not actually attempting to solve it (such bending the pipe cleaner into different shapes or sorting the balls into categories on the table), the experimenter instead asked, “*What are you trying now?*”**.** These prompts were designed to match with the level of verbal-input and experimenter interaction in the *explanation* condition, consistent with Walker et al. (2016).

2.5. Data coding and analyses

2.5.1. Pre-registered confirmatory analyses

Task success, measured as a binary dependent variable (yes/no), was contingent on successful item-retrieval, coded once the token had exited the top of the tube. Whether the child first inserted a functional (pipe cleaner; sinking ball) or non-functional (string; floating ball) item into the tube was recorded as *first insertion*. For successful trials, the duration between the start of the test phase (once the instructions had been delivered) and the retrieval of the token-item was measured as *latency to success* (in seconds). A series of regression models for each task were used to address the main hypotheses. First, a binary logistic regression with success (yes/no) as the outcome variable and condition (explanation/exploration/baseline), age (in months), and a condition x age interaction as predictors. A second binary logistic regression examined first object insertion (functional or non-functional) as the outcome, with the same predictors. Finally, a linear regression model with success (in seconds) as a continuous outcome variable, again with the same predictors.

2.5.2. *Exploratory analyses*

A McNemar test was used to test whether children who succeeded at one task were also likely to succeed at the other task.

2.5.3. *Behavioural coding*

A series of ethograms were developed and programmed into Behavioural Observation Research Interactive Software (BORIS) (Friard & Gamba, 2016) to facilitate an in-depth assessment of children's behaviours during innovative problem-solving tasks, which is lacking in most previous research (Breyel & Pauen, 2022; but see e.g., Burdett & Ronfard, 2023). Ethograms were developed using an initial set of 18 video-recordings, with the first author (DN) coding commonalities in children's actions and verbalisations for each task (see Table S2a-S2b for full ethograms and Table 2 for descriptions of variables used in exploratory analyses). Detailed behavioural observations were carried out using each ethogram, to record children's behaviours and verbalisations during the pre-test (excluding baseline) and test phase for each task. Due to technical issues, 66 of the videos were coded manually, external to BORIS software but following the same process.

To control for differences in overall duration of the Test Phase, which varied between participants, the time each child spent on each behaviour was converted into a percentage of their total Test Phase duration before running the analyses.

Exploratory analyses were planned to address the following questions:

- Explore Condition Pre-Test Phase:
  - How long did children spend on different behaviours?
  - Which behaviours (if any) were associated with task success?
- Test Phase:
  - Did behaviours differ between conditions?
  - Did behaviours differ between successful and unsuccessful individuals?
  - Was amount of prompting related to success?

2.5.4. *Verbal utterances*

Strings of speech separated by at least 2 s were coded as an ‘utterance’, consistent with Winsler et al. (2005) and Breyel and Pauen (2022). A novel coding scheme was created based on previous work (Brockbank & Walker, 2022; Breyel & Pauen, 2022; Sawyer, 2017; Liquin & Lombrozo, 2020). Each utterance was designated one of six discrete codes relative to its semantic content: *(1) physical-properties, (2) cognitive, (3) meta-cognitive, (4) question, (5) causal-explanation, or (6) non-relevant*, further defined in Table 3.

The *physical-properties* coding category captures statements about the physical attributes of materials. Brockbank and Walker's (2022) subcategories of cognitive speech were collapsed to form the present *cognitive* category, which captures the idea generation stage of innovation as well as the development and execution of an action plan. Meta-cognitive utterances represent children's capacity to evaluate reflect on their own performance (Breyel & Pauen, 2022). The *question* category expands on Brockbank and Walker's (2022) ‘thinking’ category to include Liquin and Lombrozo's (2020) definition wherein questions represent explanation-seeking curiosity. A novel category was introduced into the current coding scheme encapsulating explanations which orient to causal mechanisms, i.e., how objects could be used (cause) to achieve a specific outcome (effect).

Total number of utterances (overall per child and frequency per category) were analysed and compared between conditions and between successful vs non-successful children.

Table 2
Key behaviours coded with illustrative examples.

Behavioural Code	Description	Example from current study
Interactions with Materials	Total time (s) the child spent interacting with all task materials (including functional and non-functional items, and tube apparatus).	Participant handled the heavy and light balls for a total of 92 s during the trial.
Tool Generation (hook task only)	Time (s) spent actively attempting to create a tool.	Participant spent 34 s bending the pipe cleaner to form a functional hook.
Tool Modification (hook task only)	Time (s) spent adjusting, reshaping, and refining an existing tool (i.e., hook).	Participant spent 13 s adjusting their pipe cleaner hook.
Contour following (Klatzky & Lederman, 1993)	Tracing the shape of the materials (exploring objects without changing their shape or form).	Holding the pipe cleaner between thumb and forefinger while squeezing and tracing along its length or rolling balls in finger tips.
Form manipulation (hook task only)	Interactions with an object that result in a change to its original state (e.g., manipulate its shape) but would not function as a tool.	Curling pipe cleaner into a spiral shape or looping it to form closed shapes (e.g., a circle or heart).
Random object handling	Unfocussed, non-strategic and seemingly distracted interactions with objects without a clear goal.	Winding the pipe cleaner in a circular motion or passing balls from one hand to another whilst talking to the experimenter.
Visual exploration	Looking at or inspecting materials without touching or manipulating them.	Close visual examination of the string or tube apparatus while seated, without physically touching them.
Twisted object inside tube (hook task only)	Rotating or turning the inserted object after making prolonged contact with the target (bucket).	Rotated the string inside the tube, trying to attach onto the bucket handle.
Wiggled object inside tube (hook task only)	Repetitively wiggling the PC into bucket in quick succession (resembled a stabbing motion) making no contact/intermittent contact with the target.	Rapidly poked the pipe cleaner into the tube in short thrusts that failed to hook or lift the bucket.
Weight exploration (WD task only) (See: Unsupported Holding, Klatzky & Lederman, 1993)	Lifting an object off the table or holding suspended in the air	Participant picked up a heavy and light ball and held each in their palm, weighing them up respectively.
Hardness exploration (WD task only) (Klatzky & Lederman, 1993)	Applying opposing forces with hands or fingers to compress or squeeze an object, often using thumb and forefinger or both hands.	Pinching a light ball between fingers or knocking hard ball against the table.

Note. WD = Water Displacement.

3. Results

Preliminary analyses revealed no differences in age ($F(2, 80) = 0.160$, $p = .852$) or gender distribution ($\chi^2(2) = 0.882$, $p = .644$) across conditions. Table 4 presents descriptive statistics for our key measures (success, first item insertion and latency to success). A subset of videos (20%, 6 randomly selected from each of the 3 conditions) were independently coded by trained coders to assess inter-rater reliability. Cohen's kappa was calculated for all categorical variables (success and first item insertion), with $\kappa = 1.00$ for both measures on both tasks,

Table 3
Coding scheme used to categorise verbal responses.

Coding Category	Definition	Examples from current study
1. Physical description	Utterances relating to task material characteristics, such as weight, frequency, pliability, size and length.	'This one is softer.' 'Oh some are floaty ones!'
2. Cognitive	Statements of intention to use/ manipulate materials, describing specific instructions related to ongoing tasks or outlining behavioural action plans (Brockbank & Walker, 2022).	'Maybe if I try this end instead it could work' 'I'm gonna try and put the ball under it and just lift the teddy up'
3. Meta-cognitive	Recognizing solutions, monitoring errors, or reflecting upon task progress (Sawyer, 2017).	'I did that a bit just too fast!' 'I didn't even see that!' 'I'll just come up with ideas if it fails'
4. Question	General reasoning about the function or about properties of task materials strictly in question format (Breyel & Pauen, 2022), seeking to discover why something is the case (Liquin & Lombrozo, 2020).	'Can I take these out?' 'What is this one?'
5. Causal Explanation	Reference to causal relationships between variables; how one event influences another – e.g., relating to sinking objects raising water level/ pipe cleaner hooking onto bucket.	'I need the hard balls to make the teddy come up higher.' 'If this end hooks onto the thing and can pull it up to my hand.'
6. Non-relevant	Arbitrary statements unrelated to the task.	'I've got this game at home.'

indicating perfect agreement. Intraclass correlation coefficients (ICCs) were calculated for latency as a continuous variable, yielding values of approximately 0.999 for both tasks, reflecting excellent agreement between raters.

A series of regression analyses are presented, examining (1) success rates, (2) frequency of functional vs non-functional first item insertion and (3) latency to success across conditions in each task. Thereafter, exploratory analyses are presented, first investigating children's actions during the 'Explore' pre-task phase. The frequency and duration of actions in the test phase (controlling for time) are then described in greater depth. Next, actions of successful children across tasks and conditions are investigated. To conclude, the verbal analyses, comparing the type and frequency of utterance across conditions and successful children, are presented for both pre-task and test phases of each task. Results for all exploratory analyses are provided in full in the Supplementary Materials.

3.1. *Preregistered confirmatory analyses*

Anonymised data for the main analyses are available in the project's files section at the following OSF link: https://osf.io/2temj/overview?view_only=5498a1e4c01843a18c2d65e52aaae2bb. Success rates for each task are presented in Table 4. Success rates were very similar across conditions in the hook task, with 40 (48.2%) of the 83 participants succeeding overall. A total of 49 (59.0%) children succeeded on the water-displacement task. Which category of item (functional or non-functional) was inserted first and average duration to success (for successful children) per condition for both tasks are also presented in Table 4.

3.1.1. *Successful performance: were children more likely to succeed in the explore or explain condition than in the baseline condition?*

Binary logistic regression was used to assess the effects of condition (with the explore and explain condition compared with the baseline group respectively) and age (in months) on the likelihood of a child

Table 4
Summary of success, first item inserted, and latency to success (for successful children) by task and condition.

Condition	Success	
	Hook Task: n (%)	Water-Displacement Task: n (%)
Explain	13 (48.1)	16 (59.3)
Explore	13 (46.4)	18 (64.3)
Baseline	14 (50.0)	15 (53.6)
Overall	40 (48.2)	49 (59.0)

Condition	First Item Inserted							
	Hook Task: n (%)				Water-Displacement Task: n (%)			
	Functional	Non-Functional	Both	None	Functional	Non-Functional	Both	None
Explain	13 (48.1)	13 (48.1)	0	1 (3.7)	9 (33.0)	17 (63.0)	0	1 (3.7)
Explore	21 (75.0)	4 (14.3)	3 (10.7)	0	15 (53.6)	11 (39.3)	0	2 (7.1)
Baseline	15 (53.6)	10 (35.7)	3 (10.7)	0	12 (42.9)	14 (50.0)	1 (3.6)	0
Overall	49 (59.0)	27 (32.5)	6 (7.2)	1 (1.2)	36 (43.4)	42 (50.6)	1 (1.2)	4 (4.8)

Condition	Latency (s) to Success (Successful Children Only)	
	Hook Task: M (SD)	Water-Displacement Task: M (SD)
Explain	108.07 (76.70)	114.38 (96.32)
Explore	140.62 (83.43)	132.67 (90.81)
Baseline	86.21 (77.24)	112.88 (94.32)
Overall	110.93 (80.26)	120.87 (91.96)

$N = 83$.

solving the puzzle within the 5-min time frame for each task. For the hook task, analyses revealed that the model accounted for only 2.2% of variance in performance and was not statistically significant, $\chi^2(5) = 1.353$, $p = .929$. Neither condition nor age, nor an interaction term between the two, emerged as significant predictors of successful performance (lowest $p = .361$, see Table S3). Similarly, the model for the water-displacement task was non-significant, $\chi^2(5) = 2.228$, $p = .817$, accounting for 3.6% of variance in performance. Consistent with the previous task, neither condition nor age, nor their interaction, emerged as significant predictors in this model, (lowest $p = .544$, see Table S3).

3.1.2. First item insertion

Were Children More Likely to Insert the Functional Item first following prompts to explore?

A binary logistic regression was executed for each task to assess the effects of condition (*explain / explore / baseline*) and age (in months) on the likelihood of a given child inserting the functional item into the tube first. Cases where the child inserted both objects simultaneously (hooks: 6 cases; water-displacement: 1 case) or inserted no objects (hooks: 1 case; water-displacement: 4 cases) were excluded from these analyses. This took the total sample size to 76 for analyses for the hook task and 78 for the water-displacement task.

The model using data from the hook task was statistically significant, $\chi^2(5) = 12.982$, $p = .024$, indicating that it had predictive power in determining whether a child would insert a functional or non-functional item on their first attempt. The logistic regression model explained 21.6% of the variance in performance (Nagelkerke R^2) and correctly classified 68.4% of cases overall. It accurately identified 75.5% of cases where the functional object was inserted first (37 out of 49) and 55.6% of cases where a non-functional object was used first (15 out of 27). Age was the only significant predictor of successful first object insertion ($B = 0.116$, $SE = 0.059$, $p = .048$). The odds of selecting the functional item first increased by 12.3% for each month of age. No other predictors were significant (lowest $p = .171$, see Table S4).

The model using data from the water-displacement was not statistically significant, χ^2 (5) = 6.407, $p = .269$ and explained 10.5% of variance in performance, with no predictors emerging as significant, (lowest $p = .079$, see Table S4).

3.1.3. Latency to success: were successful children quicker to succeed following prompts to explore?

To assess the relationship between latency to success (in seconds) with condition (explain / explore / baseline) and age (in months) for each task a series of linear regression analyses were run, with latency as a continuous outcome variable. Initial analyses indicated a violation of the linearity assumption, assessed through scatter plots and a plot of standardised residual against the predicted values produced for each task. Given the violation, a log transformation was applied to the latency to success variable for both tasks. The linearity assumption was now met, with the transformed variable entered into all subsequent analyses.

A significant regression was not found for the hook task data ($F(3,34) = 0.188$, $p = .904$), with the model explaining 7.1% of variability (Adjusted R^2) and neither condition, age, nor their interactions, emerged as significant predictors (lowest $p = .184$, see Table S5). Similarly, the regression model was not significant for data from the water-displacement task ($F(3,43) = 0.203$, $p = .893$), with this model accounting for 5.5% variance (Adjusted R^2) in duration to success. Again, neither condition, age (in months), nor their interaction significantly predicted latency to success for this task (lowest $p = .56$, see Table S5).

3.2. Exploratory analyses

3.2.1. Did the same children succeed in both tasks?

Overall, 25% of the sample did not solve either task, while only 32% of the children successfully completed both tasks. An exact McNemar's test determined this difference to be non-statistically significant, $p = .176$.

Video recorded footage was inaccessible for one participant due to technical error. Therefore, for all subsequent exploratory analyses of behaviours and verbalisations, the maximum total sample size was 82 participants. Due to non-normal data distribution and unmatched group sizes, non-parametric tests were used to examine group differences.

3.2.2. Explore condition pre-test phase: how long did children spend on different actions? Which actions were associated with task success?

To examine if any actions carried out in the 1-min pre-task phase facilitated performance in the test phase, the duration of time spent interacting with the task materials was compared between successful and unsuccessful children for each task (Table S6).

3.2.2.1. Hook pre-test phase. Successful children on the hook task spent significantly more time interacting with task materials ($M = 43.88$ s, $SD = 10.17$) than unsuccessful children ($M = 27.65$ s, $SD = 9.62$) ($U = 156.0$, $p = .001$). To probe further into what behaviours occurred during this period, the frequency of different behaviours were calculated and compared between groups. No differences were statistically significant (average frequency of behaviours are presented in S7).

3.2.2.2. Water-displacement pre-test phase. Total time interacting with objects did not differ significantly between successful and unsuccessful children on the water-displacement task ($U = 61.0$, $p = .195$). Since no significant differences were found in time spent on any specific behaviours, further analyses were not conducted.

3.3. Test phase

3.3.1. Did behaviours differ between conditions?

When comparing task related behaviours (e.g., type of object manipulation, visual exploration, proportion of functional item insertions, and interactions with items as tools) across conditions, no significant differences were found between conditions (Table S8).

3.3.2. Did behaviours differ between successful and unsuccessful individuals?

To investigate if any behaviours were related to success, time spent on individual behaviours during the test phase was compared between successful and unsuccessful participants for each task, collapsed across conditions. A summary of the average percentage of time spent on different task related behaviours in both tasks is presented in Table S8.

3.3.2.1. Hook task. Both successful and unsuccessful children created tools (95% hooks, 5% other functional tool (Table S9)); however, successful children devoted a significantly greater percentage of task time to this activity ($M = 4.87\%$, $SD = 5.58$) than their unsuccessful peers ($M = 0.69\%$, $SD = 1.34$; $U = 292.0$, $p < .001$), indicating that greater time investment in tool creation was associated with task success. Similarly, successful children spent a significantly greater percentage of time modifying their hook tool ($M = 6.77\%$, $SD = 7.41$) than unsuccessful children ($M = 0.56\%$, $SD = 1.78$; $U = 380.0$, $p < .001$). Where unsuccessful children made a hook, the design often remained unsuitable for success (e.g., too large, meaning it could not fit into the tube).

Successful children also spent a greater proportion of their time twisting the objects (slowly and deliberately rotating the pipe cleaner or string when attempting to attach it onto the bucket) inside the tube ($M = 11.34\%$, $SD = 16.76$) compared to unsuccessful children ($M = 5.87\%$, $SD = 7.42$; $U = 609.50$, $p = .033$). In contrast, unsuccessful children spent a significantly greater proportion of their task time wiggling the objects (rapidly shaking the materials inside the tube without necessarily contacting the bucket) in the tube ($M = 38.35\%$, $SD = 20.40$) than successful children ($M = 13.87\%$, $SD = 12.59$; $U = 262.0$, $p < .001$).

3.3.2.2. Water-displacement task. Overall, the data suggest that there were no statistically significant differences in the time spent on various behaviours between successful and unsuccessful children on the water-displacement task. Full results are available in Supplementary Materials.

3.3.3. Was amount of prompting related to success?

We examined whether prompting frequency was related to task success. To standardise this measure, the total number of prompts was divided by total time spent on the task for each child, computing a "prompts per minute" metric. Descriptive statistics are presented in Table 5. Because prompt rates were not normally distributed in either task (Shapiro–Wilk $p < .05$), Mann–Whitney U tests were conducted to compare whether number of prompts per minute varied between successful and unsuccessful children, on both tasks.

In the hook task, results indicated that successful children received a significantly higher rate of prompts per minute (Mdn = 1.75) than unsuccessful children (Mdn = 1.40), U = 578, p = .009, $r_{r}b$ = 0.31. To probe this further, a binary logistic regression was performed with success (0 = fail, 1 = success) as the dependent variable and prompts per minute as the predictor. The model was statistically significant, $\chi^2(1) = 6.20$, p = .013, explaining approximately 12% of the variance in success (Nagelkerke $R^2 = 0.12$). Higher prompting frequency predicted greater likelihood of success (B = 0.71, SE = 0.29, p = .013, OR = 2.04).

Results using data from the water displacement task also indicated a significant difference in prompts per minute between successful (Mdn = 2.31) and unsuccessful (Mdn = 1.70) children, U = 549, p = .012, $r_{r}b$ = 0.33. A follow up binary logistic regression (outcome = success (0 = fail, 1 = success), predictor = prompts per minute). Again, prompt frequency significantly predicted success (B = 0.65, p = .010, OR = 1.91, Nagelkerke $R^2 = 0.119$). Including condition in the models did not change the pattern of results for either task.

A Spearman's rank correlation (used due to non-normality) was used to assess whether the number of prompts was associated with iterative behaviour, operationalised as the percentage of time spent modifying the hook tool. Results indicated no significant relationship between prompts per minute and time spent modifying the tool ($r_s(80) = 0.07$, p = .52).

3.4. Verbal analyses

The total number of utterances was calculated and categorised for analysis using the coding scheme in Table 3. To control for total task time, utterances were converted to percentages by dividing each category's count by the total number of utterances per child for all subsequent analyses.

3.4.1. Pre-test phase for the explain condition

No significant differences were found in either the type or quantity of speech between successful and unsuccessful children in the hook task (Table S10). In the water-displacement task, successful children produced significantly more cognitive speech (M = 27.30, SD = 17.68) than unsuccessful children (M = 17.47, SD = 16.63), U = 518.50, p = .003. No other comparisons were significant.

Table 5

Mean prompt frequency (prompts per minute) for successful and unsuccessful children in both tasks.

	Minimum	Maximum	M	SD
Hook Task				
Successful children	0.00	4.78	1.95	1.03
Unsuccessful children	0.20	2.80	1.41	0.56
Water Displacement Task				
Successful children	0.20	4.49	2.31	1.08
Unsuccessful children	0.23	3.80	1.70	0.86

3.4.2. Test phase

In the hook task, total number of utterances did not differ significantly across conditions (*Explore:* M = 13.21, SD = 8.67; *Explain:* M = 13.81, SD = 10.22; *Baseline:* M = 14.54, SD = 12.44), H = 0.032, p = .984, nor between successful (M = 13.74, SD = 9.19) and unsuccessful children (M = 13.95, SD = 11.58), U = 814.500, p = .823. Successful children produced a significantly greater proportion of cognitive speech (M = 27.30, SD = 17.68) than unsuccessful children (M = 17.47, SD = 16.63), U = 518.500, p = .003. No other differences in utterance types across conditions or success levels were significant (Table S11).

In the water-displacement task, successful children produced significantly fewer utterances overall (M = 11.94, SD = 9.09) than unsuccessful children (M = 16.79, SD = 11.65), U = 585.000, p = .040. However, total number of utterances did not differ across conditions (*Explore:* M = 13.64, SD = 11.06; *Explain:* M = 14.15, SD = 9.86; *Baseline:* M = 14.15, SD = 10.80), H = 0.215, p = .643. No significant differences were found in speech category proportions in the water-displacement task across conditions or between successful and unsuccessful group (Table S11).

3.5. Bayesian analyses

Although not part of our preregistered analysis plan, to assess whether the null findings reflected a genuine absence of condition effects or limited statistical power, we conducted Bayesian logistic regressions predicting task success from condition for both the hook and water displacement tasks. We focused specifically on task success because this was the primary outcome variable that our experimental manipulations (the 1-min exploration or explanation period) were designed to influence; as such, it provides the most direct test of whether these brief interventions altered children's problem-solving performance. Models were compared to an intercept-only null model using bridge sampling to compute Bayes factors (BF_{01}) (Note: We report BF_{01} rather than BF_{10} because our primary question was whether the data supported the absence of an effect; BF_{01} directly quantifies evidence for the null over the alternative (Quintana & Williams, 2018). For interpretability, log-odds estimates were converted to odds ratios, with 95% credible intervals.

Across both tasks, posterior distributions for condition effects were centred near zero, with odds ratios close to 1 (Hook task: Explain vs Baseline OR = 1.08 [95% CI 0.42–2.70], Explore vs Baseline OR = 0.89 [0.36–2.21]; Marble task: Explain vs Baseline OR = 1.10, Explore vs Baseline OR = 0.91). These wide credible intervals, all including 1, indicate substantial uncertainty around the direction and magnitude of any potential effects (Kruschke & Liddell, 2018).

Bayes factors favoured the null model for both tasks (Hook BF_{01} = 4.66; Marble BF_{01} = 4.32), providing moderate evidence that condition did not meaningfully influence task success (Wagenmakers et al., 2018). However, this level of evidence is also consistent with the presence of small effects that the current sample (N = 83) was underpowered to detect. (Etz & Vandekerckhove, 2016; Kruschke & Liddell, 2018). Thus, while the Bayesian analyses reinforce the absence of clear condition differences in the present data, they do not rule out the possibility of subtle effects that may emerge in larger samples or under more robust manipulations.

4. Discussion

This study examined whether prompting 5- to 7-year-olds to explore task materials or generate explanations would improve their performance at solving problems requiring tool innovation or innovative tool use. Children were assigned to one of three conditions (Explore, Explain, Baseline), and completed two five-minute problem-solving tasks (the hook task and the water-displacement task). Research prompts administered during testing were designed to encourage explanations (in the Explain condition) or descriptions (report prompts, used in Explore and

Baseline groups). Contrary to our predictions, neither encouraging children to explore nor explain had a significant effect on success rates, latency to success, or insertion of the functional item first in either of the tasks we used to assess innovative problem-solving. However, exploratory analyses of behavioural data and verbalisations revealed interesting differences in levels of exploration and the type of explanations made by successful children, regardless of condition. Although exploratory, these data offer potential insight into how young children may approach, and succeed in, physical problem-solving tasks. This discussion first considers the null findings of the pre-registered confirmatory analyses, then turns to the exploratory findings to further contextualise these results and identify promising directions for future research. The absence of significant differences between conditions in success rates, latency to solution, and functional item selection suggests that the explore and explain interventions were not effective in facilitating performance overall. The explore pre-task activities were intentionally open-ended, corresponding with evidence that unrestricted exploration facilitates independent discoveries (Bonawitz et al., 2011; Meder et al., 2021; Sobel et al., 2018), and reduces fixation on adult-demonstrated behaviours (e.g., Cutting et al., 2012). However, this lack of structure may have inadvertently limited performance. Flexible, unconfined exploration *can* lead to the incidental discovery of the correct solution to a problem, but this is not guaranteed (Meder et al., 2021). The absence of group differences does not mean explanation or exploration are irrelevant to innovation; our brief and lightly structured interventions may have simply been insufficient to reveal their effects. This corresponds with the methodological constraint highlighted by Rawlings (2022) whereby children may incorrectly interpret task instructions due to a lack of dynamic conversation and scaffolding from adult input. Indeed, tuning into social cues from adults (or other perceived ‘knowledgeable experts’) is a learning mechanism children rely on in pedagogical learning contexts (Bonawitz et al., 2011). The lack of interactive support may have limited the effectiveness of the explanation prompts (despite it having been effective in other contexts; Walker & Gopnik, 2017). Future research should explore how much and when adult support during exploratory learning is optimal (Newman & DeCaro, 2018), particularly within innovative problem-solving contexts.

In addition, exclusively presenting tools or materials with opaque affordances may have worsened children's performance through augmenting the ill-structured nature of the tasks (Chappell et al., 2013; Neldner et al., 2017). In ill-structured problems, information essential for solving the problem is not immediately apparent (Cutting, Apperly, Chappell, & Beck, 2019; Goel & Grafman, 2000; Wood, 1983). Instead, the individual must internally compute how to use or transform the available stimuli to reach the end-goal state (e.g., fashion the pipe cleaner into a hook; Beck et al., 2011; or isolate the heavy balls; Cheke et al., 2012) required to solve the task. Without further guidance, there is potential that, even after gathering this information, children still struggle to apply discovered affordances to goal-directed actions without adult scaffolding (consistent with Cutting, Beck, & Apperly, 2013; Neldner et al., 2017). This supports the criticism that perhaps the pre-task activities were too open-ended and underscores the limitations of entirely unstructured learning environments (Kirschner, Sweller, & Clark, 2006; Likourezos & Kalyuga, 2017). The ill-structured nature of the tasks may have also contributed to age-related differences in functional material selection. Our exploration pre-task was designed to nudge children to discover the object's functional affordances. Yet, children in the explore condition were not significantly more likely to select the functional items first in either task, suggesting exploration did not automatically translate into a deeper understanding of object functional properties. We did, however, observe that older children were significantly more likely to use the functional pipe cleaner first in the hook task. These findings align with previous research demonstrating developmental improvements in problem-solving and innovation. Cutting et al. (2014) showed that older children (5–6 years) are better than younger children (4–5 years) at coordinating relevant information for success in the hook task (e.g., the need for a hook or the functional properties of pipe cleaners) when this information is explicitly highlighted. However, even older children may struggle generating such information independently. In our study, children were required to identify and integrate this information on their own, forming what Jonassen, Yacci, and Beissner (2013) and Cutting et al. (2014) refer to as “structural knowledge”: a flexible understanding of how different elements interrelate to solve a problem. Although older children may better recognise the pipe cleaner's affordances, they still appear to have difficulty applying this knowledge without further guidance, as reflected in the absence of age-related improvements in success rates in the current study.

In light of these null findings, it is important to consider the potential influence of sample size. We used a convenience stopping rule driven by time and resource constraints rather than a pre-determined target based on an a priori power calculation. Our post hoc Bayesian analyses provided moderate evidence for the absence of an effect, though small effects remain plausible due to the sample size and the subtlety of our manipulations (a brief 1-min explore/explain pre-task and slight variations in prompt phrasing) (Etz & Vandekerckhove, 2016; Kruschke & Liddell, 2018).These considerations suggest that larger-scale replications may be needed to determine whether effects of prompting emerge under different conditions.

As implemented, the prompts did not meaningfully influence task success. To better understand why, we examined the variation in prompt frequency across participants. Our data suggest that children who received more prompts per minute were roughly twice as likely to succeed on the tasks. However, because prompting was contingent on children's engagement, this association may simply reflect that more engaged children made more attempts and were therefore exposed to more prompts, rather than indicating a causal effect of prompts leading to success. Determining whether prompting merely accompanies children's natural engagement or actively supports their problem solving remains an important direction for future research.

Since our current explore or explain manipulations did not affect performance, it may be more informative to investigate how the behaviours of successful children differed from those who did not succeed. Conceptualising success as “pass or fail” obliges children to converge on a single solution to be considered ‘innovative’ (Beck et al., 2016). However, Burdett and Ronfard (2023) instead highlight the creative processes underlying innovation. Rather than focussing solely on success rates, we sought to capture the creative process by coding and analysing behaviours and utterances in detail, allowing for a deeper examination of behaviours contributing to innovation.

Exploratory analyses of children's behaviour suggest that tactile exploration is associated with success in the hook task. While overall success rates between conditions did not differ significantly, children who succeeded spent significantly more time interacting with the task materials during the main task phase. These interactions were not random; they were systematic and iterative, aligning with existing literature which links feedback-driven exploration with improved problem-solving (Brockbank & Walker, 2022; Burdett & Ronfard, 2023; Gopnik, 2012). More specifically, in our study, successful participants dedicated a significantly larger proportion of their time to creating and subsequently refining (manipulating the materials in ways that optimised their functional properties; Burdett & Ronfard, 2023) hook tools. These iterative modifications co-occurred with success, converging with prior work emphasising exploration's role in children's tool innovation (Burdett & Ronfard, 2023).

Perhaps encouraging specific forms of exploration (i.e., iterative refinements in design) could facilitate better performance on the hook task. It may be that the focus of children's exploration efforts was not directed appropriately. In our sample of 5- to 7-year-olds, children may have been insensitive to the distinction between a request to “explain” and one to “describe” due to their age and cognitive maturity. Indeed, children often struggle to detect subtle linguistic differences (Legare &

Lombrozo, 2014). Future work could examine whether prompts specifically targeting iteration (e.g., "*What might you try differently next time?"*) elicit more purposeful task engagement than static prompts such as "What happened?" or "Why?". Researchers might also manipulate the timing of prompts, presenting explanation prompts before, during, or after task interaction, to determine when they are most effective in supporting planning versus reflection (Newman & DeCaro, 2018; Wise & O'Neill, 2009). Using open-ended or multi-solution tasks may amplify the influence of prompts, as these formats better capture creative and divergent problem-solving (Kittredge et al., 2015). Exploring a broader range of innovation and problem-solving tasks, and considering developmental differences in prompt interpretation, could provide valuable insights for future research.

Another pattern observed in successful children in the hook task was the greater use of goal-directed exploration strategies, rather than random trial-and-error (Meder et al., 2021). Across development, children begin to rely less on randomness and more on strategic approaches to exploration (Meder et al., 2021; Sobel & Letourneau, 2018). In context of the hook task, twisting objects inside the tube (attempting to "hook" onto the bucket) may reflect a goal-driven exploration strategy aimed at retrieving the target. We observed that successful children carried out this action nearly twice as often as those who were unsuccessful. In contrast, unsuccessful children spent a significantly greater proportion of their task time wiggling the string and pipe cleaner aimlessly inside the tube, lacking focus and control. These findings support prior work suggesting that persistent, focused, and repeated engagement with tool properties is critical to innovation (Meder et al., 2021; Neldner et al., 2019) and that strategic exploration and form manipulations may be a more reliable predictor of success than age alone. Still, it is important to note that these associations do not establish causality.

It is possible that a third factor, such as general intelligence, underlies both the tendency to engage in goal-directed exploration and success in the task. Indeed, Beck et al. (2016) found that children's receptive vocabulary (measured via the BPVS) predicted success on a hook innovation task and suggested that this could reflect a more general cognitive ability, possibly general intelligence. Thus, one plausible interpretation is that general intelligence might be driving both strategic behaviour and innovation success. Future work should explicitly measure and control for general cognitive ability (e.g., IQ or vocabulary) to disentangle these possibilities.

There are several possible explanations for why higher levels of tactile exploration were associated with better performance in the current study. Perhaps, through their targeted exploration efforts, children developed a better understanding of what constitutes an appropriate tool. This may reflect a greater understanding of how the object's affordances can be optimised to meet task demands (Neldner et al., 2019). Alternatively, it is possible that some children brought preexisting knowledge about hook functions to the task, allowing them to identify relevant properties more efficiently within their explorations. Those with a weaker conceptual understanding may have experienced increased difficulty uniting the necessary pieces of information for tool construction, and to solve the task (Cutting et al., 2019; Neldner et al., 2019).

These findings align with the idea that causal understanding is not the only path to successful problem-solving. Loissel et al. (2018) describe how causal learning and associative learning are often confounded in innovative problem-solving tasks; a tool which is causally functional is also rewarded. Objects may be selected due to paired association with the reward (e.g., elevating the floating token closer within reach) rather than through engaging causal reasoning (choosing the object based on their utilitarian features). In the water-displacement task, haptic exploration co-occurred with children noticing weight differences, but this does not necessarily reflect a true grasp of the materials' causal properties. While 8-year-olds in previous studies quickly identified and exclusively used the heavy, functional balls (Cheke et al., 2012), children in the current study rarely did so, indicating they did not possess a sophisticated understanding of the objects' causal affordances.

Many children in our study instead seemed to adopt a trial-and-error approach, inserting all the balls before then "reverse engineering" the solution by removing floating ones that blocked access to the reward. If children are biased towards co-variation between action and outcome (rewards), they may have judged the light balls to be ineffectual simply because they obstructed visibility of the reward, rather than due to their buoyancy. Thus, no conceptual/causal understanding of "weight" was necessarily applied. This supports Taylor et al.'s (2010) hypothesis that children can use perceptual-motor feedback to enhance their performance - repeatedly performing actions that bring them closer to the reward rather than engaging a causal understanding of object affordances (Cheke et al., 2012; Miller et al., 2017).

In the hook task, successful children produced significantly more cognitive speech than unsuccessful peers, aligning with previous analyses of children's utterances in different problem-solving contexts (Hook task: Breyel & Pauen, 2022; Fishing Tasks: Chiu & Alexander, 2000; Sawyer, 2017). Cognitive speech is thought to help children to articulate their ideas, assess their progress and adjust their strategies as needed (Breyel & Pauen, 2022; Sawyer, 2017). This externalization can make it easier for them to organise their thoughts and identify a clear plan of action (Russo, 2019). Doebel and Munakata (2022) similarly found that children who used more self-directed speech demonstrated better executive control in tasks demanding flexibility or planning. The finding that successful children used more cognitive speech adds to our understanding of private speech in problem-solving contexts (e.g., Fernyhough & Borghi, 2023) by exemplifying how a distinct component of private speech (relating to internal cognitions) featured more prominently in children who produced innovative solutions.

In conventional pedagogical settings, cognitive speech may serve as an advantageous resource for children to receive immediate feedback from others on the worthiness of their ideas and approaches (Mercer, 2002). For example, if a child answers a teacher's question regarding how to complete a task, children may be accustomed to receiving feedback on the response they offer (Liquin & Lombrozo, 2020). Such feedback will ordinarily be used to refine their strategies and improve their problem-solving approach in real-time (Newman & DeCaro, 2018). Therefore, they may have approached the current tasks with the expectation that the researcher would offer useful guidance and feedback (Bonawitz et al., 2011). Perhaps verbalising plans out loud was habitual and reflective of children's attempts to accrue social support in solving the problem. It would be interesting to compare whether children who have not become habituated to classic Western pedagogical teaching methods use a comparable amount of cognitive speech.

In evaluating their ability to complete the task, many children sought requests for help (e.g., "I cannot do this; I need some help"). This may reflect the normative pedagogical relationship between a naïve student and a proficient instructor whereby children expect 'experts' to intervene and support their performance (Bonawitz et al., 2011). On one hand, seeking social support can mitigate the costs associated with risky trial-and-error learning, therefore serving as a potential shortcut to success (Legare & Nielsen, 2015). Seeking help may also represent individual preferences in learning strategy (Rawlings, Flynn, and Kendal (2017), uncertainty about the boundaries of the task (e.g., *"What am I allowed to do?"* (Sheridan et al., 2016)) or a lack of belief in one's own competency. A detailed examination of trait differences, such as self-efficacy, may shed light upon how individual trait differences influence performance (Rawlings et al., 2022). However, using private speech as the only marker of this would offer little construct validity, necessitating cross-validation using supplementary measures (Russo, 2019; Wang & Sun, 2024).

Children's verbal reasoning skills were not controlled for in this study. Beck et al. (2016) specifically concluded that verbal reasoning has a supportive role in problem-solving, but its effectiveness is relative to the child's stage of cognitive development. Research indicates that children with higher verbal intelligence often use more sophisticated private speech (Berk, 2014), which can aid their cognitive processes and problem-solving skills (Lidstone et al., 2010). To better understand the relationship between cognitive development, language skills, and private speech in children's problem-solving, future research should include assessments of explicit cognitive abilities, such as language and/or meta-cognitive skills (Breyel & Pauen, 2022).

Both tasks in the current study required children to recognise a single correct solution and in doing so inherently favour convergent thinking skills over idea generation (Breyel & Pauen, 2022; Burdett & Ronfard, 2023). This may lessen the effectiveness of pre-task exploration or explanation – which appear to motivate divergent thinking skills instead (i.e., generating multiple ideas and fostering creativity; Legare & Lombrozo, 2014). Fixation on a convergent approach to the tasks may have hindered performance if children failed to converge on the correct behaviour. A task that allows for multiple solutions may be better suited to test the efficacy of pre-task manipulations.

5. Conclusion

Neither pre-task nor test-phase exploration or explanation prompts were associated with differences in children's likelihood of successful problem-solving. In the context of our short-duration tasks and manipulations, these interventions alone did not appear sufficient to support innovation. This aligns with existing concerns about the limitations of unstructured learning environments for tasks that demand a single correct solution (Beck et al., 2016; Cutting et al., 2019).

Although the different conditions did not yield differences in performance, exploratory analyses revealed some differences between successful and unsuccessful children's behaviours that should be explored in future work. These include potential benefits of iterative, trial-and-error adjustments of materials and the use of purposeful, outcome-oriented strategies during self-directed interactions with the task materials. It is possible that the frequency and type of individual children's interactions with task materials, rather than the nature of any instructional prompt they receive, may be more strongly associated with innovative problem-solving. Future research should further investigate how to support and scaffold effective exploratory behaviours within diverse problem-solving contexts.

CRediT authorship contribution statement

Darcy Neilson: Writing – review & editing, Writing – original draft, Resources, Project administration, Investigation, Formal analysis, Data curation. **Nicola Cutting:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Emma C. Tecwyn:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Funding

This project was supported by York St John University School of Education, Language and Psychology Quality Related funding awarded to ECT and NC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express our sincere gratitude to the participating schools for their cooperation with participant recruitment and data collection which made this research possible. We are also deeply thankful to our research assistants, Naomh Weston, Tom Lawton, and Halyna Moldovanova, whose diligent efforts in video coding contributed to the analysis and overall completeness of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2026.106396.

Data availability

Anonymised data for the main analyses are available in the project’s files section at the following OSF link: https://osf.io/2temj/overview?view_only=5498a1e4c01843a18c2d65e52aaae2bb

References

Bechtel, S., Jeschonek, S., & Pauen, S. (2013). How 24-month-olds form and transfer knowledge about tools: The role of perceptual, functional, causal, and feedback information. *Journal of Experimental Child Psychology, 115*(1), 163–179. https://doi.org/10.1016/j.jecp.2012.12.004

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn’t child’s play. *Cognition, 119*(2), 301–306. https://doi.org/10.1016/j.cognition.2011.01.003

Beck, S. R., Williams, C., Cutting, N., Apperly, I. A., & 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*(1690), 20150190. https://doi.org/10.1098/rstb.2015.0190

Beller, S., & Kuhnmünch, G. (2007). What causal conditional reasoning tells us about people's understanding of causality. *Thinking & Reasoning, 13*(4), 426–460. https://doi.org/10.1080/13546780701319122

Berk, L. E. (2014). Children’s private speech: An overview of theory and the status of research. In *Private Speech* (pp. 17–53).

Bird, C. D., & Emery, N. J. (2009). Rooks use stones to raise the water level to reach a floating worm. *Current Biology, 19*(16), 1410–1414. https://doi.org/10.1016/j.cub.2009.07.033

Bjorklund, D. F., & Gardiner, A. K. (2010). *Object play and tool use: Developmental and evolutionary perspectives*. https://doi.org/10.1093/oxfordhb/9780195393002.013.0013

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bonawitz, E., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Breyel, S., & Pauen, S. (2022). Private speech during problem-solving: Tool innovation challenges both preschoolers’ cognitive and emotion regulation. *Journal of Cognition and Development, 24*(3), 354–374. https://doi.org/10.1080/15248372.2022.2144319

Brockbank, E., & Walker, C. M. (2022). Explanation impacts hypothesis generation, but not evaluation, during learning. *Cognition, 225*, Article 105100. https://doi.org/10.1016/j.cognition.2022.105100

Burdett, E. R., & Ronfard, S. (2023). Tinkering to innovation: How children refine tools over multiple attempts. *Developmental Psychology*. https://doi.org/10.1037/dev0001512

Chappell, J., Cutting, N., Apperly, I. A., & Beck, S. R. (2013). The development of tool manufacture in humans: What helps young children make innovative tools? *Philosophical Transactions of the Royal Society, B: Biological Sciences, 368*(1630), 20120409. https://doi.org/10.1098/rstb.2012.0409

Cheke, L. G., Loissel, E., & Clayton, N. S. (2012). How do children solve Aesop’s fable? *PLoS One, 7*(7), Article e40574. https://doi.org/10.1371/journal.pone.0040574

Chiu, S., & Alexander, P. A. (2000). The motivational function of preschoolers’ private speech. *Discourse Processes, 30*(2), 133–152. https://doi.org/10.1207/S15326950DP3002_03

Cutting, N., Apperly, I. A., & Beck, S. R. (2011). Why do children lack the flexibility to innovate tools? *Journal of Experimental Child Psychology, 109*(4), 497–511. https://doi.org/10.1016/j.jecp.2011.02.012

Cutting, N., Apperly, I. A., Chappell, J., & Beck, S. R. (2014). The puzzling difficulty of tool innovation: Why can’t children piece their knowledge together? *Journal of Experimental Child Psychology, 125*, 110–117. https://doi.org/10.1016/j.jecp.2013.11.010

Cutting, N., Apperly, I. A., Chappell, J., & Beck, S. R. (2019). Is tool modification more difficult than innovation? *Cognitive Development, 52*, Article 100811. https://doi.org/10.1016/j.cogdev.2019.100811

Cutting, N., Beck, S. R., & Apperly, I. A. (2013). *Is there a complexity hierarchy in human children’s tool making?*. Unpublished manuscript. UK: University of Birmingham.

Doebel, S., & Munakata, Y. (2022). Unravelling the nature of children’s self-directed speech: Correlates of five-and six-year-olds’ overt and partially covert speech on three tasks. *Collabra: Psychology, 8*(1), Article 57543. https://doi.org/10.1525/collabra.57543

Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology. *PLoS One, 11*(2), Article e0149794.

Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.08.014

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: Relations with task difficulty and task performance. *Cognitive Development, 20*(1), 103–120. https://doi.org/10.1016/j.cogdev.2004.11.002

Fong, F. T., Puebla, G., & Nielsen, M. (2024). The role of conventionality and design in children's function judgments about malfunctioning artifacts. *Journal of Experimental Child Psychology, 240*, Article 105835. https://doi.org/10.1016/j.jecp.2023.105835

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution, 7*(11), 1325–1330.

Gardiner, A. K., Bjorklund, D. F., Greif, M. L., & Gray, S. K. (2012). Choosing and using tools: Prior experience and task difficulty influence preschoolers' tool-use strategies. *Cognitive Development, 27*(3), 240–254. https://doi.org/10.1016/j.cogdev.2012.05.001

Goddu, M. K., & Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology, 3*(5), 319–339. https://doi.org/10.1038/s44159-024-00300-5

Goel, V., & Grafman, J. (2000). Role of the right prefrontal cortex in ill-structured planning. *Cognitive Neuropsychology, 17*(5), 415–436. https://doi.org/10.1080/026432900410775

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*(6102), 1623–1627. https://doi.org/10.1126/science.1223416

Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition, 35*(1), 69–95. https://doi.org/10.1016/0010-0277(90)90037-K

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*(6), Article e19555.

Joh, A. S., Jaswal, V. K., & Keen, R. (2011). Imagining a way out of the gravity bias: Preschoolers can visualize the solution to a spatial problem. *Child Development, 82*(3), 744–750. https://doi.org/10.1111/j.1467-8624.2011.01584.x

Jonassen, D. H., Yacci, M., & Beissner, K. (2013). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Routledge.

Kaller, C. P., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain and Cognition, 67*(3), 360–370. https://doi.org/10.1016/j.bandc.2008.02.003

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology, 62*(1), 1–21. https://doi.org/10.1146/annurev.psych.031809.130730

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75–86.

Kittredge, A. K., Klahr, D., & Fisher, A. V. (2015). *Instruction of discovery: Pedagogy's effect on exploration*. Manuscript in preparation.

Klatzky, R. L., & Lederman, S. J. (1993). Toward a computational model of constraint-driven exploration and haptic object identification. *Perception, 22*(5), 597–621. https://doi.org/10.1068/p220597

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review, 25*(1), 155–177.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development, 83*(1), 173–185. https://doi.org/10.1111/j.1467-8624.2011.01691.x

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives, 8*(2), 101–106. https://doi.org/10.1111/cdep.12070

Legare, C. H., & Clegg, J. M. (2014). The development of children's causal explanations. In *The Routledge International Handbook of Young Children's Thinking and Understanding*. Routledge Handbooks Online.

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology, 126*, 198–212. https://doi.org/10.1016/j.jecp.2014.03.001

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences, 19*(11), 688–699. https://doi.org/10.1016/j.tics.2015.08.005

Lidstone, J. S., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology, 107*(4), 438–451. https://doi.org/10.1016/j.jecp.2010.06.002

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science, 45*, 195–219. https://doi.org/10.1007/s11251-016-9399-4

Liquin, E. G., & Lombrozo, T. (2020). A functional approach to explanation-seeking curiosity. *Cognitive Psychology, 119*, Article 101276. https://doi.org/10.1016/j.cogpsych.2020.101276

Loissel, E., Cheke, L. G., & Clayton, N. S. (2018). Exploring the relative contributions of reward-history and functionality information to children's acquisition of the Aesop's fable task. *PLoS One, 13*(2), Article e0193264. https://doi.org/10.1371/journal.pone.0193264

McCormack, T., & Atance, C. M. (2011). Planning in young children: A review and synthesis. *Developmental Review, 31*(1), 1–31. https://doi.org/10.1016/j.dr.2011.02.002

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science, 24*(4), Article e13095. https://doi.org/10.1111/desc.13095

Mercer, N. (2002). *Words and minds: How we use language to think together*. Routledge.

Miller, R., Jelbert, S. A., Loissel, E., Taylor, A. H., & Clayton, N. S. (2017). Young children do not require perceptual-motor feedback to solve Aesop's fable tasks. *PeerJ, 5*, Article e3484. https://doi.org/10.7717/peerj.3484

Neldner, K., Mushin, I., & Nielsen, M. (2017). Young children's tool innovation across culture: Affordance visibility matters. *Cognition, 168*, 335–343. https://doi.org/10.1016/j.cognition.2017.07.015

Neldner, K., Redshaw, J., Murphy, S., Tomaselli, K., Davis, J., Dixson, B., & Nielsen, M. (2019). Creation across culture: Children's tool innovation is influenced by cultural and developmental factors. *Developmental Psychology, 55*(4), 877–889. https://doi.org/10.1037/dev0000672

Newman, P. M., & DeCaro, M. S. (2018). How much support is optimal during exploratory learning?. In *, Vol. 40. Proceedings of the Annual Meeting of the Cognitive Science Society*. https://escholarship.org/uc/item/17r1k961.

Nielsen, M. (2012). Imitation, pretend play, and childhood: Essential elements in the evolution of human culture? *Journal of Comparative Psychology, 126*(2), 170–181. https://doi.org/10.1037/a0025168

Pauen, S., & Bechtel-Kuehne, S. (2016). How toddlers acquire and transfer tool knowledge: Developmental changes and the role of executive functions. *Child Development, 87*(4), 1233–1249. https://doi.org/10.1111/cdev.12532

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 18*(1), 178.

Rawlings, B., Flynn, E., & Kendal, R. (2017). To copy or to innovate? The role of personality and social networks in children's learning strategies. *Child Development Perspectives, 11*(1), 39–44. https://doi.org/10.1111/cdep.12206

Rawlings, B., & Legare, C. H. (2021). Toddlers, tools, and tech: The cognitive ontogenesis of innovation. *Trends in Cognitive Sciences, 25*(1), 81–92. https://doi.org/10.1016/j.tics.2020.10.006

Rawlings, B. S. (2022). After a decade of tool innovation, what comes next? *Child Development Perspectives, 16*(2), 118–124. https://doi.org/10.1111/cdep.12451

Rawlings, B. S., Flynn, E. G., & Kendal, R. L. (2022). Personality predicts innovation and social learning in children: Implications for cultural evolution. *Developmental Science, 25*(1), Article e13153. https://doi.org/10.1111/desc.13153

Russo, A. (2019). Thinking out loud or speaking in loud: A review on inner speech. *Journal of Psychosocial Systems, 3*(2), 53–65.

Sawyer, J. (2017). I think I can: Preschoolers' private speech and motivation in playful versus non-playful contexts. *Early Childhood Research Quarterly, 38*, 84–96. https://doi.org/10.1016/j.ecresq.2016.09.004

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Sheridan, K. M., Konopasky, A. W., Kirkwood, S., & Defeyter, M. A. (2016). The effects of environment and ownership on children's innovation of tools and tool material selection. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*(1690), 20150191. https://doi.org/10.1098/rstb.2015.0191

Sobel, D. M., & Legare, C. H. (2014). Causal learning in children. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(4), 413–427. https://doi.org/10.1002/wcs.1291

Sobel, D. M., & Letourneau, S. M. (2018). Curiosity, exploration, and children's understanding of learning. In *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms* (pp. 57–74). https://doi.org/10.1007/978-3-319-77182-3_4

Taylor, A. H., Elliffe, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural innovation in new Caledonian crows. *Proceedings of the Royal Society B: Biological Sciences, 277*(1694), 2637–2643. https://doi.org/10.1098/2Frspb.2010.0285

Taylor, A. H., Elliffe, D. M., Hunt, G. R., Emery, N. J., Clayton, N. S., & Gray, R. D. (2011). New Caledonian crows learn the functional properties of novel tool types. *PLoS One, 6*(12), Article e26887. https://doi.org/10.1371/journal.pone.0026887

Tecwyn, E. C., Mazumder, P., & Buchsbaum, D. (2023). One-and two-year-olds grasp that causes must precede their effects. *Developmental Psychology, 59*(8), 1519.

Tecwyn, E. C., Thorpe, S. K., & Chappell, J. (2014). Development of planning in 4-to 10-year-old children: Reducing inhibitory demands does not improve performance. *Journal of Experimental Child Psychology, 125*, 85–101. https://doi.org/10.1016/j.jecp.2014.02.006

Veenman, M. V. (2016). Learning to self-monitor and self-regulate. In *Handbook of research on learning and instruction* (pp. 249–273). Routledge.

Voigt, B., Pauen, S., & Bechtel-Kuehne, S. (2019). Getting the mouse out of the box: Tool innovation in preschoolers. *Journal of Experimental Child Psychology, 184*, 65–81. https://doi.org/10.1016/j.jecp.2019.03.005

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57.

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2016). Effects of explaining on children's preference for simpler hypotheses. *Psychonomic Bulletin & Review, 24*, 1538–1547. https://doi.org/10.3758/s13423-016-1144-0

Walker, C. M., & Gopnik, A. (2017). Discriminating relational and perceptual judgments: Evidence from human toddlers. *Cognition, 166*, 23–27. https://doi.org/10.1016/j.cognition.2017.05.013

Wang, Y., & Sun, P. P. (2024). Development and validation of scales for speaking self-efficacy: Constructs, sources, and relations. *PLoS One, 19*(1), Article e0297517. https://doi.org/10.1371/journal.pone.0297517

Weir, A. A., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science, 297*(5583), 981. https://doi.org/10.1126/science.1073433

Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3–41). Cambridge University Press. https://doi.org/10.1017/CBO9780511581533.003.

Winsler, A., Fernyhough, C., McClaren, E. M., & Way, E. (2005). *Private speech coding manual*. Unpublished manuscript. Fairfax, VA, USA: George Mason University. Available at: http://classweb.gmu.edu/awinsler/Resources/PsCodingManual.pdf.

Wise, A. F., & O'Neill, K. (2009). Beyond more versus less: A reframing of the debate on instructional guidance. In S. Tobias, & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 82–105). Routledge/Taylor & Francis Group.

Wood, P. K. (1983). Inquiring systems and problem structure: Implications for cognitive development. *Human Development, 26*(5), 249–265. https://doi.org/10.1159/000272887