Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis

The early development and challenges of tool innovation in preschool children

presented by
Sabine Breyel

year of submission
2022

Dean: Prof. Dr. Guido Sprenger
Advisor: Prof. Dr. Sabina Pauen

Danksagung

Zunächst möchte ich mich bei meiner Betreuerin Prof. Sabina Pauen bedanken. Vielen Dank, Sabina, dass du mir die Freiheit gegeben hast, eigene Ideen zu verfolgen und mich in deren Umsetzung immer unterstützt hast. Aus dem gemeinsamen Brainstormen, unseren Gesprächen und deinem Feedback habe ich so viel mitgenommen, das mich auch auf meinen weiteren Weg – vor allem beruflich, aber auch persönlich – begleiten wird. Vor allem bin ich dir sehr dankbar dafür, dass du mich immer wieder dazu angeregt hast, mental einen Schritt zurück zu treten, um die Dinge in einem größeren Zusammenhang zu sehen.

Vielen Dank an meine (ehemaligen) Kolleginnen aus der Abteilung Entwicklungspsychologie Miri, Christine, Lena, Cornelia, Milena, Jule, und ganz besonders Steffi. Gerade während der Corona-Pandemie hatte jeder eine ganze Menge eigene Probleme zu bewältigen und trotzdem hattet ihr für mich immer ein offenes Ohr und hilfreiche Ideen parat. Das weiß ich sehr zu schätzen und bin euch dafür sehr dankbar. Ich danke auch den PsTAs, die während meiner Zeit als Doktorandin für die Koordination der Labore zuständig waren, den Studierenden, die zeitweise im Projekt mitgearbeitet haben (Anna, Theresa, Sandra und Dorothea) und ich danke dir Christiane, nicht nur für die Unterstützung bei der Terminkoordination oder anderen organisatorischen Dingen, sondern vor allem dafür, dass du mir so oft dein Ohr geliehen und mich motiviert hast.

Ich bedanke mich ganz herzlich bei der Studienstiftung des deutschen Volkes für die finanzielle und ideelle Förderung während meiner Promotion.

Außerdem möchte ich mich bei meinen Freunden und meiner Familie bedanken. Vielen Dank vor allem an Wiebke, Katha, Lena, Natascha, Cornelius, Mama, Anne und David, dass ihr mich in unterschiedlichsten Phasen der Promotion immer unterstützt, motiviert, ausgehalten und inspiriert habt, dass ihr immer für mich da seid, mitfiebert und an mich glaubt.

Zu guter Letzt geht ein herzliches Dankeschön an die Familien, die mit ihren Kindern an meinen Studien teilgenommen haben und somit diese Arbeit überhaupt erst ermöglicht haben, und an die Kinder selbst, die mir in den letzten dreieinhalb Jahren so viele spannende, lustige und aufschlussreiche Momente geschenkt haben.

Table of contents

List of scientific publications for the publication-based thesis

I. Manuscript

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development*, *58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

II. Manuscript

Breyel, S., & Pauen, S. (2022a). *Private speech during problem-solving: Tool innovation challenges both preschoolers' cognitive and emotion regulation.* Manuscript submitted for publication in Journal of Cognition and Development.

III. Manuscript

Breyel, S., & Pauen, S. (2022b). *How prior experience in analogous tasks affects three-year-olds' tool making.* Manuscript submitted for publication in Cognitive Development.

1. Introduction

Every day we use a variety of tools – devices that enable us to efficiently achieve goals, mostly by overcoming the limitations of our own body. These goals might be simple like brushing teeth or writing something down but they might also be more complex like traveling large distances, building technical devices, or producing mRNA-based vaccines. The tools that are necessary for reaching most of these goals are so complex that it would be impossible for an individual to develop them from scratch. The mechanism that has nevertheless allowed humans to develop such sophisticated technologies is termed cumulative cultural evolution (Boyd & Richerson, 1985). New ideas are shared within a community and transmitted from one generation to the next, paving the way for the development of even more elaborate solutions. Thus, for the evolution of cumulative culture, both innovative ideas and faithful social learning are necessary (Legare & Nielsen, 2015).

The psychological processes that have been involved in the phylogenetic development of the human species might be closely linked to its ontogenetic development. Thus, to learn more about the evolution of culture, it is also important to learn more about the early development of innovation and social learning in childhood. Developmental research has shown that human children are very proficient social learners. This becomes evident in their tendency to copy with high fidelity (Flynn & Whiten, 2008; Horner et al., 2006), to over-imitate (for a review see Hoehl et al., 2019), and their ability to imitate rationally and flexibly (Gergely et al., 2002; Király et al., 2013; Legare et al., 2015). In contrast, children rarely prefer to learn individually or deviate from socially demonstrated behavior by exploring new ways to achieve a goal (Carr et al., 2015; Flynn et al., 2016; Whiten & Flynn, 2010). Even more so, generating new and useful ideas to solve a problem without any social information seems to pose a remarkably great challenge for children. Findings of a limited capacity to innovate have been reported both for tasks requiring innovative tool use, such as pouring water into a tube to get a buoyant toy (Hanus et al., 2011; Nielsen, 2013), and tasks requiring innovative tool making, such as constructing a hook to catch a prize (e.g., S. R. Beck et al., 2011; Cutting et al., 2011).

Despite a growing body of research on the development and cognitive underpinnings of tool-related innovation (see Rawlings & Legare, 2021, for an overview), the key question of why it is so difficult for children to come up with innovative solutions is still under debate. My thesis focuses on this question and aims to provide new insights about the challenges children face during tool innovation tasks by taking three different approaches. First, I describe the early development of innovation skills in the preschool age in more detail by examining three- to five-year-olds' skills in solving two versions of a tool-making task and suggest more differentiated ways to analyze children's solution approaches. Second, I analyze children's private speech while solving these problems to examine which processes on the cognitive, emotional, or motivational level might affect innovation performance. Third, in a training study, I examine whether young children benefit from prior experiences in tool making through analogical transfer.

2. What is (tool) innovation?

Innovation is a research topic shared by diverse disciplines such as anthropology, animal cognition, comparative research, economics, and psychology, resulting in a variety of conceptual understandings of this term. In the field of entrepreneurship, for example, innovation has been considered as the implementation of creative ideas to develop novel products or services that are successfully brought to market (e.g., Cropley & Cropley, 2015; Sarooghi et al., 2015). In the literature on social and cultural learning, innovation has been often viewed as a deviation from the socially demonstrated behavior by using another, potentially more efficient way to achieve a goal (e.g., Carr et al., 2015; Tennie et al., 2014; Whiten & Flynn, 2010). In research on animal behavior, innovation has been defined as "a process that results in new or modified learned behavior and that introduces novel behavioral variants into a population's repertoire" (Reader & Laland, 2003, p. 14) whereby this behavior serves as „a solution to a novel problem or a novel solution to an old problem" (Manrique et al., 2013, p. 195). Despite the diversity of concepts, there are three important commonalities. First, an innovation must be novel – either to the individual or to a broader community. Second, an innovation should be useful in the sense of solving a problem or achieving a specific goal. Third, innovation is linked to the practical implementation of products or the transmission of new and useful behaviors.

In an attempt to integrate these key points to a definition applicable to developmental research, Carr and colleagues (2016) proposed that "a behavioral innovation is a new, useful, and potentially transmitted learned behavior […] that is produced so as to successfully solve a novel problem or an existing problem in a novel manner" (Carr et al., 2016, p. 1515). They further suggested distinguishing between novel behavior that is based on a combination of individual learning and social information (innovation by modification) and such behavior that predominantly arises from individual learning (innovation by independent invention). Building on this definition, the term *tool innovation* more specifically refers to problem-solving that involves the new and useful application or combination of materials or objects. Some researchers narrow down this definition to problem-solving that requires tool construction*,* that is structurally modifying the original material to create a new object that serves as a tool (e.g., Cutting et al., 2011; Neldner et al., 2019). However, others also include problem-solving that requires uncommon or novel ways to use a material or resource as a tool in its current state (Ebel et al., 2019; Hanus et al., 2011; Nielsen, 2013). Furthermore, most studies refer to tool innovation as the form of novel behavior produced by an individual without direct social influences (resembling the term *innovation by independent invention* as suggested by Carr et al., 2016) and distinguish this from *tool manufacturing*, which refers to “the ability to make tools after instruction or observation” (Cutting et al., 2011, p. 497).

Drawing from this overview of definitions, in the scope of my dissertation I refer to tool innovation as a behavior that is generated by an individual without direct social information and that is implemented to solve a novel problem by using objects in a new and useful way or by modifying objects to construct a new tool. Since I focus on innovative problem-solving on the individual level, the transmission of behavior is not considered here.

3. Measuring tool innovation

Developmental research on individual[1] innovative problem-solving has been inspired by paradigms from animal and comparative research (Bird & Emery, 2009a, 2009b; Hanus et al., 2011; Mendes et al., 2007; Weir et al., 2002). In general, the tasks can be described as reward-retrieval activities, since the goal is to obtain a favorable object from an apparatus (e.g., food or toys). In the following, I describe two tasks used most frequently in the literature and adapted in the current thesis to measure tool innovation in children.

---

[1] Since the current thesis focuses on individual innovative problem-solving, experimental approaches to study how innovations are transmitted in diffusion chains or groups of children (McGuigan et al., 2017; Reindl et al., 2017; Tennie et al., 2009; Whiten & Flynn, 2010) are not described here, but will be referred to in the general discussion.

The Hook Task was used to study tool innovation in human children for the first time by Beck and colleagues (2011). They tested whether three- to 11-year-old children could build a hook tool to retrieve a small basket with a handle containing a sticker from a narrow transparent tube. The children were given a straight pipe cleaner and a piece of string. The solution was to bend the pipe cleaner into a hook to pull up the basket. A surprising finding was that children up to school age had great difficulty solving this seemingly simple hook problem. For children between the ages of five and eight, the success rate ranged between 35% and 50%, and among the three- to five-year-olds not even 10% came up with the correct solution. Other studies confirmed these findings and reported similarly low performance for four- to eight-year-olds (S. R. Beck et al., 2014; Chappell et al., 2013; Cutting et al., 2011, 2014; Frick et al., 2017; Neldner et al., 2017). The fact that children from non-Western cultures also performed poorly in the task suggests that this might be a cross-cultural phenomenon (Nielsen et al., 2014).

In the Unbending Task, a sticker attached to a fluffy ball is placed in the middle of a transparent narrow tube that is open on both sides. As in the Hook Task, the materials available are a pipe cleaner and a piece of string. However, the pipe cleaner is bent in the middle and the solution is to pull it straight to push the ball through the tube toward an opening. Although success rates were slightly higher in this version (23%–33%) compared to the Hook version, again the majority of four- to five-year-olds had difficulty mastering the task (Chappell et al., 2013; Cutting et al., 2011).

4. Understanding the development of tool innovation

It seems surprising that creating simple tools is still very challenging for children until school age given the early and elaborate development of tool-using skills. Already preverbal infants have been shown to understand how the shape of artifacts is related to their functions (Träuble & Pauen, 2007, 2011). Beginning with the second year of life, toddlers use this knowledge to achieve goals with tools (e.g., pulling a toy into reach with a hook-shaped object) and even transfer their skills to different sets of objects (Brown, 1990; Rat-Fischer et al., 2012). When observing others' tool behavior, young children can also incorporate the actor's intentions into their own actions and omit unnecessary or inefficient actions to achieve a goal (Buttelmann et al., 2008; McGuigan & Whiten, 2009).

However, in contrast to the functional understanding of a tool or its adept use, innovating a tool requires at least two difficult cognitive steps (Chappell et al., 2013; Cutting et al., 2014). The first step is innovating the overall tool shape. In the example of the Hook task, the correct solution is a long tool with a hook. The second step is discovering how to achieve the solution with the given material. In the Hook task, the correct transformation is to bend one end of the pipe cleaner to form a hook. Since both the target state of the tool and the necessary transformation are unknown in this task, it is referred to as an ill-structured problem and deemed to be especially difficult for young children (Chappell et al., 2013; Cutting et al., 2014).

When children are presented with an integrated social demonstration of both pieces of information (i.e., solution and transformation) by an adult showing how to bend the pipe cleaner into a hook, the majority of three- to six-year-olds make the tool themselves afterward (S. R. Beck et al., 2011; Chappell et al., 2013; Cutting et al., 2011). Success rates in the youngest age groups were still variable even after this demonstration, however (Cutting et al., 2011; Gönül et al., 2018). Intriguingly, when solution and transformation information were presented separately, children below the age of five were hardly able to combine these clues in order to generate the correct solution (Cutting et al., 2014). For this purpose, it was first demonstrated that the pipe cleaner can be bent into a stable shape (i.e., a spiral), and then a ready-made pipe cleaner hook was presented. The information about the target form of the tool seemed to help younger children to infer the necessary transformation only if they used the ready-made tool themselves beforehand instead of just seeing it (Whalley et al., 2017). In sum, young children seem to have difficulties both with innovating the solution and the transformation to achieve this solution, but their own perceptual and motor experiences with the correct approach could help work out the necessary transformation. The question that remains largely unclear, however, is why the respective innovation step and the combination of information is so difficult.

One explanation would be that children up to a certain age do not yet possess the necessary cognitive abilities to achieve tool innovation. On a theoretical level, various cognitive abilities are discussed that could play an important role, including creativity, planning, hierarchical thinking, and executive functions (S. R. Beck et al., 2016; Carr et al., 2016; Gönül et al., 2018; Rawlings & Legare, 2021). However, empirical studies have not yet been able to uncover the corresponding correlations with such measures (S. R. Beck et al., 2016; Chappell et al., 2013; Gönül et al., 2018). An additional explanation could be that prior research methods have made it difficult for young children to develop and implement innovative ideas. In the following, I will briefly describe three methodological aspects that could affect tool innovation.

The first aspect is the selection of the task. The vast majority of research is limited to the Hook Task (S. R. Beck et al., 2011, 2014, 2016; Cutting et al., 2011; Frick et al., 2017; Gönül et al., 2018, 2019; Neldner et al., 2017; Whalley et al., 2017). The few studies including other task versions, such as the Unbending Task, provide initial evidence that innovation performance may depend on which basic solution method is required (i.e., a hook vs. a long straight tool). The studies report slightly higher success rates of four- to five-year-olds for the Unbending Task than the Hook Task (Chappell et al., 2013; Cutting, 2013). However, it has not yet been examined in detail whether these differences are statistically significant and whether they also apply to younger children.

The second aspect concerns the type of tool manufacture. In both the Hook and the Unbending Task, there is only one way to make a functional tool. The material must be bent to produce the correct shape. In addition to modifying the form of an object, which is called *Reshape*, three other manufacturing methods can be derived from behavioral observations of wild animals (B. B. Beck, 1980). When using the *Detach* method, an object is separated from another larger object or a solid surface to serve as a tool. When parts of an object are removed to render it a (more) functional tool, manufacturing is called *Subtract*. Finally, *Add* means putting several objects together to create a functional tool. Observations from both wild and captive animals indicate that Detach is the easiest manufacturing method, followed by Subtract, Add, and Reshape (Bania et al., 2009; B. B. Beck, 1980; Kacelnik et al., 2006; Shumaker et al., 2011; van Schaik et al., 2003). Whether this hierarchy could also apply in a similar form to children has not yet been clarified, but preliminary evidence indicates that Reshape is particularly difficult in the Hook task and Detach is particularly easy across different task versions (Cutting, 2013; Voigt et al., 2019). The findings on the difficulty level of Add and Subtract are inconsistent and seem to vary depending on the nature of the task, the availability of the specific materials, and the amount of time available for working out a solution (Cutting, 2013; Neldner et al., 2019; Voigt et al., 2019). For example, Voigt and colleagues (2019) showed that very few five-year-olds were able to make a tool within one minute regardless of the manufacturing method, but that success rates increased to about 70% after 10 minutes for Subtract, while remaining at 20% for Reshape.

This finding also highlights the third aspect that should be considered when studying tool innovation, namely the time that is granted to complete the task. With one exception (Voigt et al., 2019), previous studies using the Hook or Unbending task limited the working time to one (S. R. Beck et al., 2011, 2016; Chappell et al., 2013; Cutting et al., 2011; Gönül et al., 2018, 2019; Neldner et al., 2017) or two minutes (Nielsen et al., 2014; Whalley et al., 2017). The findings of Voigt et al. (2019) indicated that more time could be helpful for young children to come up with a solution to the innovation problem when different manufacturing options are available. However, which aspects of extended working time support the innovation process has not yet been investigated in detail.

5. Open questions and the current thesis

Previous findings suggest that there are several factors influencing children's performance in tool innovation tasks, such as prior experiences with the material and the solution, the type of task and manufacturing options, and working time. So far, however, we know little about how these factors change with age and what cognitive mechanisms are driving them. Therefore, the objective of the first study was to describe in more detail the cross-sectional development of preschoolers' innovation skills in two different tasks with different manufacturing options and extended time being available.

Existing research has mainly focused on whether children manage to get the target object out of the apparatus as a measure of tool innovation. In a few cases, there were attempts to differentiate between the successful construction of the tool and the successful solution of the task (see Neldner et al., 2017). Nevertheless, how the manufacturing process and more generally the innovation process works at the individual level has not been examined so far. A second aim of the first study was therefore to analyze in more detail the type of solution strategies children chose when different manufacturing options were simultaneously present, and to describe the reasons for failure. Moreover, I analyzed children's speech as a new methodological approach in the second study to find out more about how they work on innovation tasks and on which levels of regulation (i.e., cognitive, emotional, motivational) they are challenged during the process of problem-solving.

As the literature review has shown, there is some empirical evidence on how children can solve tool-making tasks after a (partial) demonstration. Especially for younger children, it is not yet clear which information can help them to work out a solution on their own. However, an important aspect seems to be their own active experience with the tool and the respective task (Whalley et al., 2017). Therefore, the third study examined whether young children also benefit from prior experience in analogous tasks by transferring it to the context of the tool innovation task.

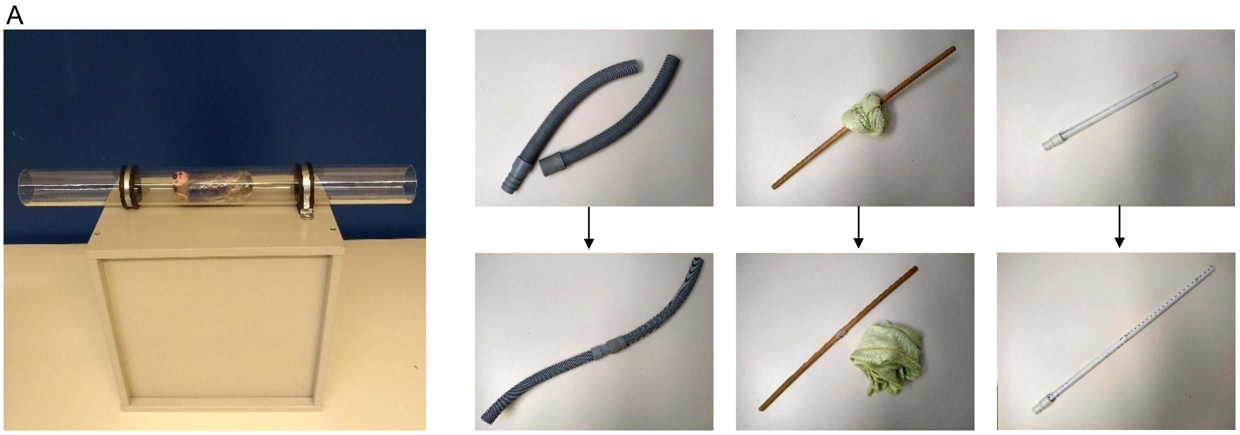
5.1 Study 1 – Age- and task-related differences in preschoolers' tool innovation

Previous studies suggest that the ability to solve innovative problems by using tools develops rather late in childhood (e.g., S. R. Beck et al., 2011; Frick et al., 2017). However, this conclusion is predominantly based on work examining children's specific skill of bending a pipe cleaner into a functional form to get a reward from a tube within one minute. Current evidence indicates that this particular study design might have underestimated children's capacity for tool innovation (Voigt et al., 2019). The authors found that 70−90% of five-year-olds were able to solve innovation tasks when more time and other manufacturing options than only Reshape were available. In the first study of my dissertation, I used this new task design to examine for the first time how children below the age of five perform in tool innovation tasks with multiple solution options and an extended time frame of 10 minutes (for a detailed description of the procedure see Breyel & Pauen, 2021). For both task versions (i.e., horizontal tube, vertical tube), six items were presented. Three of those were distractors and the other three were potentially functional to get a toy out of the respective tube if modified by Adding, Subtracting, or Reshaping (see Figure 1).

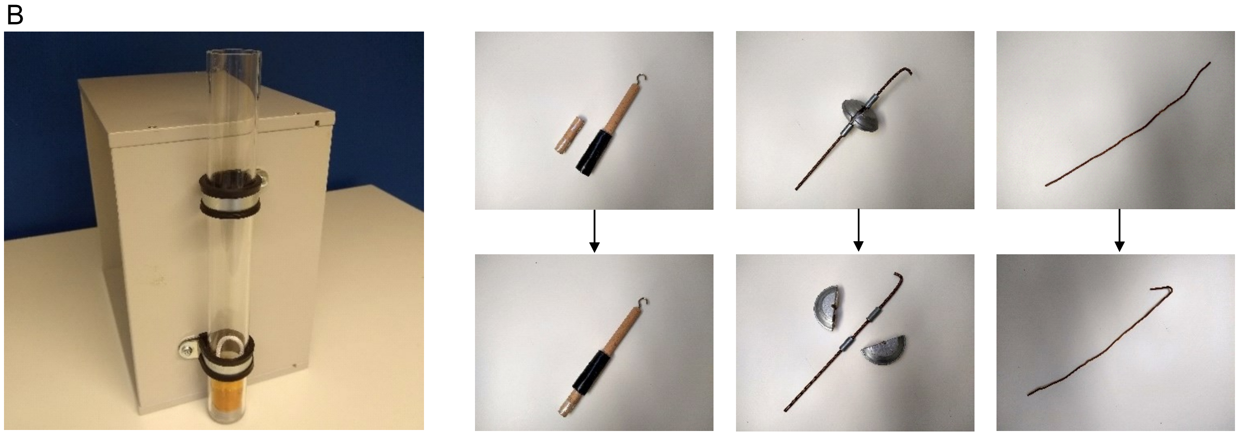
Another aim of this study was to analyze in more detail *how* children solve the two versions of the task to generate new ideas about which cognitive skills are particularly required for tool innovation and to what extent different tasks demand these skills. Therefore, I categorized the solution approaches into two types depending on whether or not constructing the tool was separated in time and/or space from applying it. If a functional tool was built before being applied to the problem, the solution was termed *First-order Innovation* (including the intended manufacturing options Reshape, Subtract, and Add). If materials were combined functionally during an ongoing attempt without Adding, Subtracting, or Reshaping the items, this was called S*econd-order Innovation* (e.g., inserting two or more items into the horizontal tube to push the target out).

Figure 1

Modifications for Relevant Items in the Horizontal (A) and Vertical Tube Task (B)



Horizontal Tube Task Add-Horizontal Subtract-Horizontal Reshape-Horizontal



Vertical Tube Task Add-Vertical Subtract-Vertical Reshape-Vertical

Regarding age effects, we expected success rates to increase and latency to success to decrease from three- to five-year-olds in both tasks. Concerning task effects, we expected children to be more successful and faster in finding a solution for the Horizontal Tube Task (HTT) than the Vertical Tube Task (VTT). We did not specify hypotheses on age- and task-related effects on the type of innovation (i.e., First-order vs. Second-order) since this was the first approach to differentiate the quality of innovative solution attempts.

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development, 58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

Confirming our hypotheses, success rates increased significantly with age for the VTT, from 7% in three-year-olds to 88% in five-year-olds, and overall the HTT was solved faster than the VTT. In contrast to our prediction, the HTT was solved at a very high level by all age groups. The exploratory analysis revealed that the repertoire of solution strategies differed between task versions. Most children solved the HTT using Second-order Innovation (i.e., combining items without manufacturing a tool first), whereas First-order Innovation with the Add and Subtract method predominated in the VTT.

The current study replicates the results from Voigt et al. (2019) by showing that five-year-olds competently solve the two innovation tasks when different solution options and enough time were available. However, the results suggest that the different task demands play a more important role for younger children who struggled to solve the VTT but performed very well in the HTT. One reason for the higher difficulty of the VTT might be that the necessary action sequence appears to be more complex than in the HTT. First, firm contact with the object must be established, but then the motion's direction must be reversed to pull up the object. Furthermore, not only the length and circumference must be considered as dimensions of the tool as in the case of the horizontal tube, but also the hook shape must be taken into account in the vertical tube. Overall, not only demands on manual skills but also the cognitive load appears to be higher in the VTT than in the HTT. This seems to be especially challenging for the younger children and might be related to the fact that relevant skills such as planning, executive functions, and attentional control are still developing between the ages of three and five (Garon et al., 2008; Jongbloed-Pereboom et al., 2013; Kaller et al., 2008; Zelazo et al., 2003).

5.2 Study 2 – Private speech as a window into innovative problem-solving

Based on the existing literature, the question of which processes lead to success or failure in a tool innovation task remains largely open. Furthermore, the role of motivation and emotion regulation in the development of tool innovation has hardly been addressed so far. As a complement to previous research methods, I proposed to use children's language as a source of information in Study 2 to find out more about which challenges they face at different levels of regulation (i.e., cognitive, emotional, motivational) when solving tool innovation tasks.

That children often talk to themselves in cognitively challenging situations has been documented in various tasks, such as planning, categorization, and executive functions (Al-

Namlah et al., 2012; Alarcón-Rubio et al., 2014; Aro et al., 2015; Benigno et al., 2011; Chiu & Alexander, 2000; Thibodeaux et al., 2019; Winsler et al., 2007). This phenomenon of *private speech* is seen as an intermediate step between externalized social speech and silent inner speech and is supposed to help children regulate their thoughts, emotions, and behavior (Vygotsky, 1962). Most previous research on the relation between private speech and cognitive performance focused on the level of speech internalization, ranging from task-irrelevant and task-relevant overt speech to whispers or silent lip movements as manifestations of partially internalized speech (Berk, 1986; Diaz, 1992; Winsler, 2009). However, there are approaches to additionally classify the content of private speech and its emotional valence more precisely in order to determine correlations with performance indicators (Chiu & Alexander, 2000; Krafft & Berk, 1998; Sawyer, 2017). For the present study, an integrated coding scheme was chosen that differentiates utterances into five broad categories of cognitive, metacognitive, motivational, playful, and partially internalized speech, and also includes a valence dimension differentiating positive and negative speech content (Sawyer, 2017). Each speech category is further divided into several subcategories (for more details see manuscript of Study 2).

For the speech analysis, the video recordings of children who participated in Study 1 were transcribed and the speech was separated into utterances and classified as private speech according to common rules in the field (Winsler, 2009). The number of utterances per minute and the proportion of utterances in the total private speech were calculated for each private speech category. It was expected that both cognitive and metacognitive speech would be positively related to performance in terms of success and latency to success. These categories include statements about identifying problems and planning one's actions, evaluating those actions, and reflecting on progress in the task (Sawyer, 2017) and have also been associated with better performance in other challenging tasks (Chiu & Alexander, 2000; Manning et al., 1994). In addition, very similar skills (e.g., planning, problem recognition, and attentional control) are also considered essential for successfully generating an innovative solution (e.g., Rawlings & Legare, 2021). No specific hypotheses could be derived from previous research regarding the relationship of the remaining speech categories and possible differences related to age and type of innovation task. Thus, these aspects were analyzed in an exploratory manner.

Breyel, S., & Pauen, S. (2022a). *Private speech during problem-solving: Tool innovation challenges both preschoolers' cognitive and emotion regulation.* Manuscript submitted for publication in Journal of Cognition and Development.

As hypothesized, successful children produced more cognitive and metacognitive speech per minute than unsuccessful children. Success was also associated with a higher percentage of cognitive speech in total private speech. Regarding latencies, a higher density of cognitive speech was related to faster success. The effects were small and did not allow for further analyses on the level of speech subcategories. Although the results should therefore be interpreted cautiously, they provide first hints that successful tool innovation is reflected in children's spontaneous private speech about their thoughts, paving the way for more research on potentially facilitating effects of talking out loud.

The exploratory analyses further revealed that longer latency to success correlated positively with the amount of negative speech and that more negative speech was produced in the VTT than in the HTT, which is probably an effect of task difficulty. These correlations raise the question of how much time children should be given to solve corresponding tasks. Voigt and colleagues (2019) showed that success rates improved with time when different solutions were available. However, the current study points to the risk that frustration and the need to regulate it also increase with time if children do not find a solution immediately.

Overall, the findings of this study show that additionally analyzing language has great potential for future research on tool innovation and may improve our understanding of the challenges children face at different levels of regulation and how to support them in mastering these challenges.

5.3 Study 3 – How prior experience in analogous tasks affects tool making

The ability to transfer knowledge about tool functions emerges early in toddlerhood. For example, Brown (1990) demonstrated that even 18-month-olds transferred functional tool knowledge to similar sets of materials without being distracted by irrelevant features such as color. Other studies have shown that young preschool children can solve multiple analogous tool use problems if their attention is focused on the similarity between the problems (Brown et al., 1986; Brown & Kane, 1988; Crisafi & Brown, 1986). Given these findings, it seems surprising that three- to four-year-olds did not significantly improve in tool making across several trials when slightly different tool materials and apparatuses were presented (S. R. Beck et al., 2014). Thus, compared to children above four years of age (S. R. Beck et al., 2014; Gerson et al., 2018), younger children's transfer skills in the domain of tool making seem to be limited.

To examine more closely why young children struggle to transfer tool-making knowledge and/or skills to structurally similar tasks, the third study of my thesis tested the effect of tool-making training on subsequent performance in the Vertical Tube Task (VTT). Differing from previous work (S. R. Beck et al., 2014; Gerson et al., 2018), a broader range of indicators to measure performance was used instead of focusing on successful retrieval of the target object only. A group of three-year-olds received training consisting of a demonstration of how to make a wire hook (Reshape exercise) and two practice tasks in which a toy had to be retrieved with a wire hook from apparatuses which were analogous to the test task. A control group was presented with the test task without any prior exercises. We expected the trained children to perform better in the VTT and to use the Reshape method more often than untrained children. Next to success (i.e., removing the target from the tube), latency to success, the first approach to the task, durations of different attempt categories (e.g., attempts with modified items vs unmodified items), the manufacturing method, and the type of innovation were analyzed as indicators of performance.

Breyel, S., & Pauen, S. (2022b). *How prior experience in analogous tasks affects three-year-olds' tool making*. Manuscript submitted for publication in Cognitive Development.

Contrary to our hypothesis, success rates did not differ significantly between trained and untrained children. On the one hand, the rate was relatively low in the training group (i.e., 56%) considering that children had practiced on two similar tasks. On the other hand, the rate was relatively high in the control group (i.e., 36%) compared to the poor performance of three-year-olds found in Study 1 (i.e., 7%). However, the effect of the training was still evident in children's behavior: Trained children more often used the exercised tool-making method as an attempt to solve the task, while children in the control group mainly used material in the original form, failing to recognize that modifications were necessary. Thus, the majority of trained children transferred the right approach to solve the problems, but most of them failed because they did not manage to adjust the tool to the respective apparatus. This discrepancy between having the proper idea and failing to implement it was also evident across the training tasks. While the proportion of spontaneous hook building increased significantly from about 40% to almost 90% from the first to the second task, individual success rates remained at about 40%. These findings illustrate that executing tool-making ideas and adjusting a tool to the particularities of the respective apparatus represent major challenges for three-year-olds. This is consistent with current evidence that even older children have great difficulty with adequately adjusting the shape of an oversized hook tool (Cutting et al., 2019). One explanation might be that children predominantly focus on functional aspects when using tools – such as the overall hook shape – and fail to take into account the appropriate sizes and proportions simultaneously (Casler et al., 2011; Casler & Kelemen, 2007).

Consistent with the results of Study 1 for the VTT, First-order Innovation was the dominant solution type in Study 3. However, an interesting difference between conditions occurred: While most children in the Control Condition first tested the material in its original form and only then made the necessary modification (*First-order Step-by-Step Tool Innovation*), trained children applied their tool significantly more often directly in its modified form (*First-order Direct Tool Innovation*). Such a close look at behavioral patterns and solution approaches could be an avenue for examining individual differences in planning behavior during tool innovation in future work.

6. General discussion

A prominent conclusion from existing empirical evidence on tool innovation is that young children are poor innovators (Carr et al., 2016; Nielsen, 2013; Nielsen et al., 2014; Reindl et al., 2016). However, my work shows that a more differentiated view of preschool children's performance reveals potentials that may have been underestimated in previous studies. On the one hand, there are large differences between versions of tube innovation tasks, affecting the likelihood of success, especially in younger children, and determining the type of solution strategies (Study 1). On the other hand, my findings suggest that not only cognitive skills play a role in successful tool innovation, but also the regulation of emotions (Study 2) and motor-coordinative skills (Study 3) should be considered as factors influencing young children's performance.

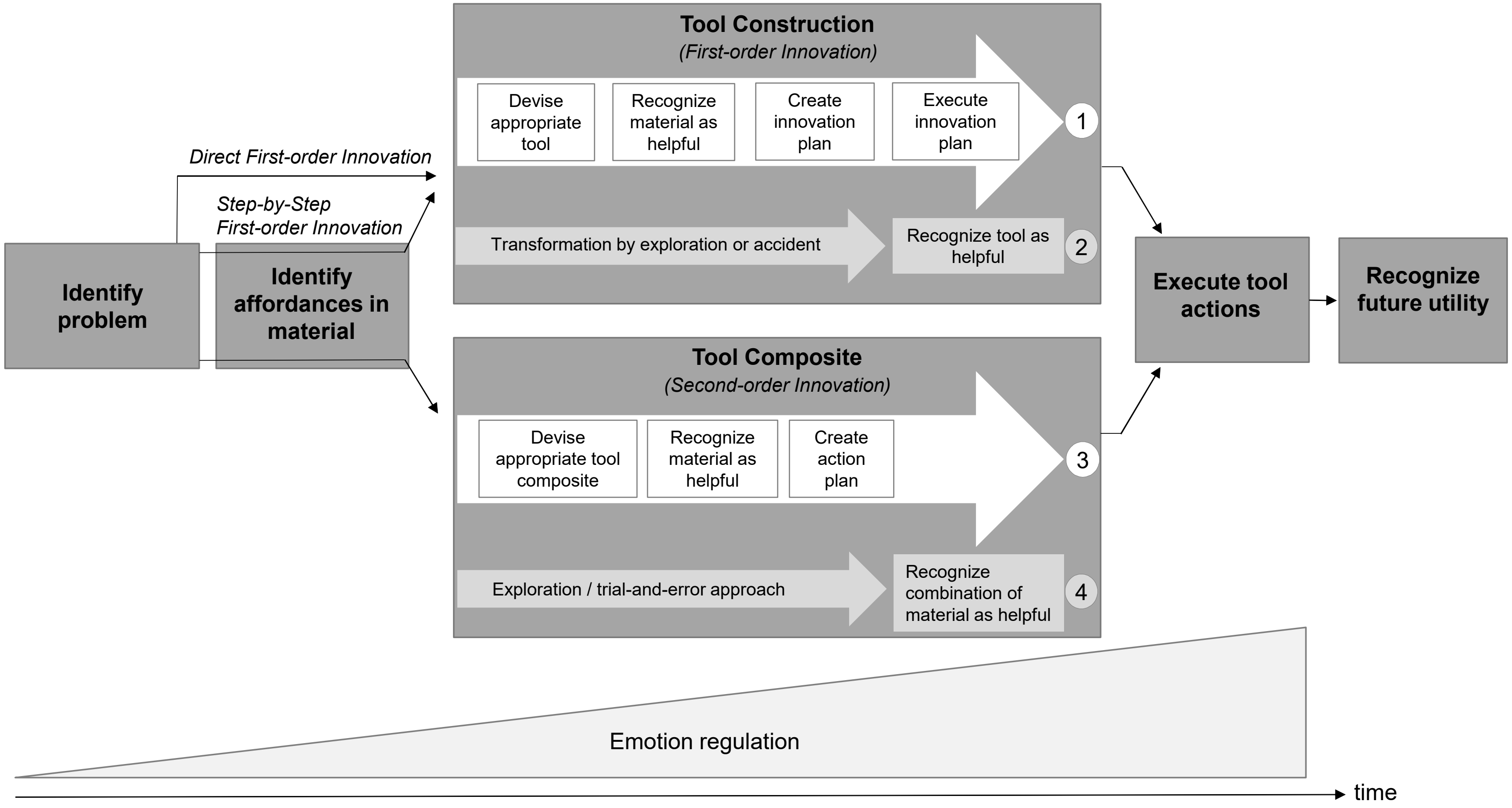
6.1 A pathway model to success in tool innovation tasks

Recently, Neldner (2020) suggested that children need to master eight steps on the way to successful tool innovation. After identifying the problem that the target object is out of reach (step 1), the affordances of the materials available have to be identified (step 2) to conclude that these materials are not useful in the current form. To find a solution, a mental representation of an appropriate tool has to be devised (step 3), and the child has to recognize the material as helpful to accomplish this tool representation (step 4). The transformation of the material into the desired form must be planned (step 5) before this innovation plan (step 6) and the action plan for the tool can be executed (step 7) to successfully retrieve the target. Step 8 entails remembering this solution and recognizing its utility for future situations.

Based on my findings, I propose a more differentiated view of this process. In the following, I describe the different paths of my adapted and extended version (see Figure 2) and outline approaches for future research.

Figure 2

Adapted Pathway Model to Success in Tool Innovation Tasks



Note. This model is based on the tool innovation process model by Neldner (2020, p.160). For both Tool Construction and Tool Composite, one route refers to solution generation by deliberate planning (routes 1 and 3, respectively) whereas another route illustrates solution generation via trial and error or exploration (routes 2 and 4, respectively).

Identification Phase: the role of exploring the material’s affordances

My findings support Neldner's (2020) assumption that identifying the problem is no major challenge for most children. This is evidenced both by the fact that few three- to five-year-olds needed a specific prompt that they should use the materials to solve the task, and those children who did need a prompt quickly moved on to exploring the materials (Study 1). For the next step of identifying the material’s affordances, however, children seemed to display different strategies. Exploring the material’s functions in its current state before transforming it (Step-by-Step First-order Innovation) was the predominant behavior of children in the control group without prior task experiences in Study 3. However, about one-third of the children seemed to take a shortcut and constructed the tool without trying the item in its original state (Direct First-order Innovation). This may reflect individual differences in the capacity for mental imagery, i.e., the ability to internally simulate external events or perceptual experiences (Borst, 2013). Children who are better at imagining objects or actions might recognize the nonfunctional aspect of the items’ current state quickly by simulating the action of inserting the item into the tube.

In the Training Condition of Study 3, the proportion of Direct First-order Innovation was significantly higher than in the Control Condition. The trained children had previously worked on two analogous tasks, which may have led them to directly recognize the usefulness of the material for the test task (see the last step of the model, Figure 2). They either applied this knowledge directly by transforming the material they already knew (Reshape) or were able to transfer the knowledge to create an innovation plan for new material (Add or Subtract). Further studies need to clarify whether the two forms of First-order Innovation (i.e., Direct vs. Step-by-Step) can be replicated in other versions of innovation tasks and which cognitive mechanisms might facilitate them.

Generating a solution: tool construction vs. tool composite

The upper part of the model refers to First-order Innovation – the construction of a novel tool. In the present work, different materials were presented simultaneously, each of which could be transformed into a functional tool by one of three methods (i.e., Reshape, Add, Subtract). How often each solution method was selected varied strongly. Without prior practice (i.e., excluding the Training Condition in Study 3), only three of the 114 children tested ever came up with the idea of bending the wire into a hook in the VTT, and none of these children had success with this Reshape method. In contrast, Add (about 47%) and Subtract (about 32%) occurred much more frequently in the VTT collapsed across studies. It seems reasonable to conclude that Add and Subtract are simpler transformation methods than Reshape, which is also largely consistent with the findings of a hierarchy of tool manufacturing in the animal kingdom (Bania et al., 2009; Shumaker et al., 2011; van Schaik et al., 2009). However, it is important to discuss that the Add and Subtract tool, which I used for my studies, differed from the Reshape tool not only in terms of the necessary transformation method. For the Reshape tool, children had to generate the idea of a hook shape themselves since it was not present in the original form of the straight wire. In contrast, the hooks were already present in the Add and Subtract items, and the necessary transformation needed to be performed on another part of the object. For example for the Add tool, a short hook was extended by attaching a wooden stick. In a study using the Hook Task, the hook shape for the Add tool was created by assembling two wooden dowels at a specific angle and the success rate was only 11%, much lower than in my studies (Cutting, 2013). These observations suggest that it is easier to mentally devise the appropriate tool and/or to develop the tool construction plan if the key feature of the tool (i.e., a hook in the Hook Task) is already visible (see also Neldner et al., 2017). To better understand how the tool construction process is affected by these aspects and to examine the relative difficulty of transformation methods in children, future studies using the HTT and VTT should either vary the transformation method while keeping the affordances of the tools constant or vary the material's affordances within the same transformation method.

The lower part of the model refers to Second-order Innovation – the use of two or more items without construction a new tool. This solution path has not been included in Neldner's model (2020). In Study 1, I observed that a majority of children solved the HTT by combining multiple materials in their original form without building a tool per se. I labeled this strategy Second-order Innovation to express that the new and useful combination of materials can also be considered as innovative problem-solving (see also Carr et al., 2016). In animal research, similar behavior has been termed *Tool Composite Use* (Shumaker et al., 2011). As a form of *Associative Tool Use*, a Tool Composite occurs when two or more objects are used simultaneously to achieve a goal (often but not necessarily in a different mode), but no modification of the individual objects is required. For a comprehensive understanding of the development of innovative problem-solving with tools, I think it is important to examine this strategy in more detail because it could be an intermediate step between simple tool use and tool making (see also Reindl et al., 2022).

There is much research showing that children are capable of successfully using a single tool in its existing form to achieve goals at a very early age (e.g., Barrett et al., 2007; Brown, 1990; Rat-Fischer et al., 2012; Reindl et al., 2016). In contrast, there is little empirical research on children's ability to engage in various forms of Associative Tool Use. These studies suggest that children below the age of four especially struggle with coordinating two sequential tool actions (i.e., pull a stick into reach with a rake; pull a toy into reach with this stick) even though they mastered both tool actions individually very well (Metevier, 2006; Reindl et al., 2022). Unlike sequential tool use, the objects are used simultaneously to achieve one shared goal in a Tool Composite (Shumaker et al., 2011). Thus, the spatial and temporal separation of action planning and execution is smaller, possibly decreasing the demands on working memory. This could explain why even three-year-olds in Study 1 achieved such high success rates with the Second-order Innovation strategy, even though working memory as one of the components of executive functions shows tremendous development during later preschool age (Garon et al., 2008; Zelazo et al., 2003).

Whether a Tool Composite is an effective solution depends on the selection and quantity of materials as well as the apparatus of a tool innovation task. Presumably, the large selection of items in the current version of the HTT compared to other studies with fewer items (Cutting, 2013; Cutting et al., 2011; Neldner et al., 2019) has also led to more frequent use of the Second-order strategy. In contrast, it occurred very rarely in the VTT, although there were as many objects available as in the HTT. Thus, not only the number of objects might be important, but also whether single objects have the necessary properties for a useful combination. In the VTT, a necessary property is the presence of a hook. Since only two items had a pre-made hook, the possibility for Second-order Innovation was probably limited. When planning future studies, it should therefore be considered that the task design strongly determines which range of innovation strategies can be investigated.

Generating a solution: conscious planning vs. trial and error

Neldner's process model (2020) implies that tool innovation is the result of deliberate planning. To generate a solution, an appropriate tool needs to be devised mentally and the material needs to be recognized as helpful before a plan to transform the material (innovation plan) can be created and executed. This process of deliberate Tool Construction is depicted in the first route of my adapted model (see Figure 2). Analogous cognitive steps can be involved in deliberate Tool Composite Use (see route 3 in Figure 2). For example in the HTT, children might devise an appropriate Tool Composite (i.e., two or more short items will serve as a long tool together), notice that they can achieve it by combining the given material, and create an action plan how to combine which items.

However, in accordance with other theoretical considerations on childhood innovation (see Carr et al., 2016), Neldner (2020) also describes an alternative process in which tool manufacturing occurs by trial and error or accident (see route 2 in Figure 2). For example, the non-functional part could be separated from the Subtract-item because it falls on the ground, or the Reshape-item is bent while playing around. If a tool is constructed this way, its potential to solve the task still has to be recognized. The behavior descriptions of unsuccessful children in Study 1 and Study 3 suggest that some children indeed failed to realize the potential of their tools, especially in the VTT. Some of them did not use the correctly transformed tool at all on the tube, and others used it only in the wrong direction (i.e., hook upwards). Similarly, children might not deliberately plan to use a Tool Composite, but rather insert items into the tube by trial and error (see route 4 in Figure 2). However, they also have to recognize that a combination of items is required to solve the task.

Since it is difficult to interpret children's intentions from solely observing their behavior, an intention of the speech analysis was to gain additional insights into the cognitive processes described in the different pathways. However, the quantitative analysis of private speech in Study 2 revealed only small differences in the amount of cognitive and metacognitive utterances between successful and unsuccessful children, which did not methodologically allow for a closer look at the subcategory level. On a qualitative level, there was some evidence that children were aware of their tool actions, reflected by utterances from the speech subcategories of planning, strategic thinking, and reflecting on the task progress. These examples occurred both for Second-order Innovation (e.g., a three-year-old child has inserted one part of the horizontal Add-item and says, „I have to push it [*first part of Add-item*] with the other tube [*second part of Add-item*], then the mouse comes out!") and First-order Innovation (e.g., four-year-old: "I will make a longer stick."; three-year-old: "I have to bend it [*the wire*] a little bit smaller to make it go around the corner [*inside the tube*]").

To investigate the cognitive mechanisms and phases of the innovation process, it could be helpful for future studies to evaluate speech transcripts of a larger sample of children in a broader age range with the help of qualitative content analysis. The procedure of inductive category development (Mayring, 2010) might be a promising approach. In contrast to applying pre-formulated speech categories from the literature, as has been the approach of Study 2, the inductive content analysis creates categories based on the current transcript material. With this procedure, speech categories could be tailored more specifically to the tool innovation process, which could help identify prominent obstacles in innovative problem-solving and the most relevant (meta)cognitive skills to overcome them. Furthermore, it might be useful to include social speech into such an analysis since it reveals insights into children's need for co-regulation in the problem-solving process.

Execute tool actions: coordinating motor skills and perceptual feedback

Beyond innovating a tool, Study 3 highlighted that applying the tool can be a major challenge for young children. Even though all three-year-olds recognized that bending the wire would produce the overall necessary hook shape, only a minority managed to adjust the hook shape to make it fit into the tube and retrieve the target. The bending action requires fine motor skills including manual dexterity, finger strength, and bilateral hand coordination to produce the correct tool shape. However, the fact that children were able to form a correct hook at the beginning of the training phase in Study 3 (in some cases after repeated practice) shows that fine motor skills alone cannot explain the difficulties. Other challenges might be to integrate visual and perceptual feedback of the motor actions (Gardiner et al., 2012) and to account for the correct size and proportions between tool and apparatus (Casler et al., 2011; Casler & Kelemen, 2007). The Reshape material in the VTT is special in this respect because its transformation from a straight to a hook shape is continuous and the ideal stopping point of this movement must be determined beforehand, or the hook must be subsequently adapted to the size of the apparatus. In contrast, transformations for the Add and Subtract tools were dichotomous (e.g., non-functional part of Subtract is removed vs. not removed) and did not require much adjustment, except for making sure that the two parts of the Add tool were firmly plugged into each other. For future studies, one might consider designing a material for Reshape that can be formed into a hook by a simpler, dichotomous action as well (e.g., by using a hinge) to keep the demands for tool construction as similar as possible between the transformation methods.

Beyond cognition and motor skills: the role of emotion regulation

My empirical work suggests that, in addition to cognitive processes and their motor-perceptual implementation, regulation at the emotional level should be considered when examining the innovation process. In Study 2, I found that time to success was positively correlated with the amount of verbal negative emotion expression and negative task evaluations. This correlation may imply that more negative emotions led to later success or that later success left more room for developing and/or expressing negative emotions.

The level of emotion regulation has not been examined at all in previous work on tool innovation. One reason might be that the short working times of one to two minutes in these studies did not allow much opportunity for negative emotions to arise, demanding less regulation. However, as is discussed in Study 2, the likelihood of experiencing negative feelings might increase with more time being available to work on the task. Similarly to an ego depletion effect[2], the higher demands for regulation at the emotional level could in turn limit the availability of cognitive resources like attentional control, cognitive flexibility, working memory, or inhibition – skills that are considered important for generating innovative solutions (Gönül et al., 2018; Rawlings & Legare, 2021). In line with this idea, when four- to six-year-olds regulated their feelings of disappointment by suppressing or disguising them, performance in subsequent cognitive tasks was impaired (Oeri & Roebers, 2020). Thus, it might be especially important to have a closer look at the strategies children use to regulate their emotions. For example, distraction has been found as the most effective emotion

---

[2] Ego depletion is considered as a short-term state in which self-control performance is impaired after repeatedly demanding it in previous tasks, assuming that self-control is a limited resource (Baumeister et al., 2007).

regulation strategy of preschoolers to reduce anger in a frustrating task (Day & Smith, 2013). In the case of an innovation task, on the one hand, it seems rather unlikely that children can generate solutions when they distract themselves from the problem. On the other hand, process theories on creativity suggest that a phase of incubation, i.e., not thinking consciously about a creative solution and paving the way for unconscious processes, is part of generating new and useful ideas or products (see Kozbelt et al., 2010 for an overview). Thus, another possibility is that distracting attention away from the problem for some time might not hinder but even support children's innovation process. Future research should describe and analyze the activities children perform during innovation tasks in more detail (e.g., actively working in the task, distracting, asking someone for help, etc.) to examine regulation strategies and their effects on success.

Beyond current success: Recognize the future utility of solutions

The last part of the process model refers to remembering the successful solution and recognizing its potential for future applications (see Figure 2). Several authors consider this aspect important for the definition of innovations – either in terms of the fact that the innovative behavior is learned as evidenced in repeated occurrence, or in terms of sharing the behavior with others (Carr et al., 2016; Reader & Laland, 2003; von Hippel & Suddendorf, 2018). To investigate the future utility of innovations, multiple trials have to be implemented in individual problem-solving, or two or more individuals have to be examined to focus on the transmission of innovative behaviors or ideas. As discussed in Study 3, when applying multiple individual-level trials to examine tool-making transfer, the task design in terms of sizes, proportions, and idiosyncratic features of the materials becomes crucial. Moreover, by including multiple trials that are separated by a certain time interval, future work could investigate children's ability to recall solutions depending on the pathway through which this solution emerged. In the light of findings indicating an insight memory advantage (e.g., Danek et al., 2013), this might be of special interest in terms of differentiating between the conscious, insightful vs. unconscious, trial-and-error pathway.

The transmission of behaviors at a group level can be studied using diffusion chain or open diffusion designs even in young preschool children (Dean et al., 2012; McGuigan et al., 2017; Reindl et al., 2017; Tennie et al., 2009; Whiten & Flynn, 2010). A diffusion chain design consists of multiple sequences of dyadic interactions involving an active participant and an observing participant. Each observer subsequently becomes an active participant and is observed by the next participant in the chain. In an open diffusion design, all participants are present at the same time and can observe or work on the task freely. These methods could be used to investigate how children transmit tool innovation skills (for a first approach see McGuigan et al., 2017) and to what extent they can further modify tool innovations based on what they have learned (Carr et al., 2016; Cutting et al., 2019).

A future perspective on measuring tool innovation

Several starting points for future research have already been outlined in the description of the pathway model. Finally, I sketch out one additional idea that could enrich the study of tool innovation. Sheridan and colleagues (2016) reported much higher innovation rates in their study in a children’s museum compared to previous laboratory studies. This finding gives rise to the question of whether innovation is affected by context variables and might be facilitated by an environment that is associated with positive experiences of playful exploration and learning. Building on this idea, I propose investigating innovative problem-solving in the context of Live Escape Games outside the laboratory.

The goal of a Live Escape Game is to get out of a room or to accomplish another quest of a fictional scenario (e.g., finding a cure for a disease) within a certain time by solving several puzzles. Thus, there are many opportunities to integrate different problem-solving tasks into a scenario suitable for children, requiring simple tool use, associative tool use, or tool innovation. By constructing so-called meta-puzzles, in which several partial solutions must be combined to generate the overarching solution, the ability to recognize future utilities of single solutions can also be tested. Moreover, the context of Live Escape Games allows investigating problem-solving both at the individual and group level. Examining problem-solving in small teams could also extend the current focus on technical innovations (i.e., making a physical tool) by taking social innovations into account, i.e., developing new and useful solutions to problems through collaboration or other social skills (see von Hippel & Suddendorf, 2018). Since both outstanding tool-related skills and elaborate social cognition are considered essential for the successful cultural evolution of humans (e.g., Herrmann et al., 2007; Tomasello, 2020), it would be paramount to extend the study of tool innovation by also examining the ontogenetic development of social innovation.

7. Conclusion

Innovation is an essential part of the cumulative cultural evolution that has enabled humans to develop increasingly complex technologies and a uniquely elaborate culture through the social transmission of new and useful ideas or artifacts. To improve our understanding of the psychological processes underlying the development of innovation, I conducted three studies to investigate the ability of children between three and five years of age to solve new problems by innovatively using tools.

My work provides an important contribution to the field, as it demonstrates that the type of measurement strongly influences the extent to which young children can show their innovative potential, suggesting that previous studies have underestimated children’s capacities to innovate. In particular, the apparatus design (i.e., horizontal vs. vertical tube), the number of available materials and concluding solution options (i.e., First-order vs. Second-order Innovation strategies), and working time are important parameters, the interaction of which needs to be further investigated in future studies.

Based on the model put forward by Neldner (2020) and the observed solution strategies in my studies, I developed an adapted pathway model of the tool innovation process. My findings suggest that a detailed analysis of behavioral and linguistic outcomes could provide new insights into the differences and commonalities of the cognitive mechanisms involved in the different pathways to tool innovation success. However, another important conclusion of my work is that – especially in young children – not only do cognitive skills determine success, but also interactions with other domains such as motor skills, perception, and emotion regulation should be taken into account. This holistic view may allow for a deeper understanding of the emergence and developmental trajectory of innovative problem-solving in childhood.

8. References

Al-Namlah, A. S., Meins, E., & Fernyhough, C. (2012). Self-regulatory private speech relates to children's recall and organization of autobiographical memories. *Early Childhood Research Quarterly*, *27*(3), 441–446. https://doi.org/10.1016/j.ecresq.2012.02.005

Alarcón-Rubio, D., Sánchez-Medina, J. A., & Prieto-García, J. R. (2014). Executive function and verbal self-regulation in childhood: Developmental linkages between partially internalized private speech and cognitive flexibility. *Early Childhood Research Quarterly*, *29*(2), 95–105. https://doi.org/10.1016/j.ecresq.2013.11.002

Aro, T., Poikkeus, A. M., Laakso, M. L., Tolvanen, A., & Ahonen, T. (2015). Associations between private speech, behavioral self-regulation, and cognitive abilities. *International Journal of Behavioral Development*, *39*(6), 508–518. https://doi.org/10.1177/0165025414556094

Bania, A. E., Harris, S., Kinsley, H. R., & Boysen, S. T. (2009). Constructive and deconstructive tool modification by chimpanzees (Pan troglodytes). *Animal Cognition*, *12*(1), 85–95. https://doi.org/10.1007/s10071-008-0173-0

Barrett, T. M., Davis, E. F., & Needham, A. (2007). Learning about tools in infancy. *Developmental Psychology*, *43*(2), 352–368. https://doi.org/10.1037/0012-1649.43.2.352

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Beck, B. B. (1980). *Animal tool behavior*. Garland STPM Publishing.

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., Cutting, N., Apperly, I. A., Demery, Z., Iliffe, L., Rishi, S., & Chappell, J. (2014). Is tool-making knowledge robust over time and across problems? *Frontiers in Psychology*, *5*(1395), 1–10. https://doi.org/10.3389/fpsyg.2014.01395

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 of London. Series B, Biological Sciences*, *371*(1690), 20150190. https://doi.org/10.1098/rstb.2015.0190

Benigno, J. P., Byrd, D. L., McNamara, J. P., Berg, W. K., & Farrar, M. J. (2011). Talking through transitions: Microgenetic changes in preschoolers' private speech and executive functioning. *Child Language Teaching and Therapy*, *27*(3), 269–285. https://doi.org/10.1177/0265659010394385

Berk, L. E. (1986). Relationship of Elementary School Children's Private Speech to Behavioral Accompaniment to Task, Attention, and Task Performance. *Developmental Psychology*, *22*(5), 671–680. https://doi.org/10.1037/0012-1649.22.5.671

Bird, C. D., & Emery, N. J. (2009a). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370–10375. https://doi.org/10.1073/pnas.0901008106

Bird, C. D., & Emery, N. J. (2009b). 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

Borst, G. (2013). Neural underpinnings of object mental imagery, spatial imagery, and motor imagery. In K. N. Ochsner & S. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience* (Vol. 2, pp. 74–87). Oxford University Press.

Boyd, R., & Richerson, J. P. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development*, *58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

Brown, A. L. (1990). Domain-specific principles affecting learning and transfer in children. *Cognitive Science*, *14*(1990), 107–133. https://doi.org/10.1207/s15516709cog1401_6

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523. https://doi.org/10.1016/0010-0285(88)90014-X

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development*, *1*(2), 103–121. https://doi.org/10.1016/S0885-2014(86)80014-4

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2008). Rational tool use and tool choice in human infants and great apes. *Child Development*, *79*(3), 609–626. https://doi.org/10.1111/j.1467-8624.2008.01146.x

Carr, K., Kendal, R. L., & Flynn, E. G. (2015). Imitate or innovate? Children's innovation is influenced by the efficacy of observed behaviour. *Cognition*, *142*, 322–332. https://doi.org/10.1016/j.cognition.2015.05.005

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What Is Innovation, How Does It Develop, and Who Does It? *Child Development*, *87*(5), 1505–1519. https://doi.org/10.1111/cdev.12549

Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental Psychology*, *47*(3), 857–866. https://doi.org/10.1037/a0021174

Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition*, *103*(1), 120–130. https://doi.org/10.1016/j.cognition.2006.02.006

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 of London. Series B, Biological Sciences*, *368*(1630), 20120409. https://doi.org/10.1098/rstb.2012.0409

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

Crisafi, M. A., & Brown, A. L. (1986). Analogical Transfer in Very Young Children : Combining Two Separately Learned Solutions to Reach a Goal. *Child Development*, *57*(4), 953–968.

Cropley, D. H., & Cropley, A. J. (2015). *The psychology of innovation in organizations*. Cambridge University Press.

Cutting, N. (2013). *Children's Tool Making: From innovation to manufacture* [PhD Thesis, University of Birmingham]. http://etheses.bham.ac.uk/3969/1/CUTTING12PhD.pdf

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*(1), 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*. https://doi.org/10.1016/j.cogdev.2019.100811

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2013). Aha! experiences leave a mark: Facilitated recall of insight solutions. *Psychological Research*, *77*(5), 659–669. https://doi.org/10.1007/s00426-012-0454-8

Dean, A. L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture. *Science*, *335*(6072), 1114–1118. https://doi.org/10.1126/science.1213969

Diaz, R. M. (1992). Methodological concerns in the study of private speech. In R. M. Diaz & L. E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 55–81). Erlbaum.

Ebel, S. J., Hanus, D., & Call, J. (2019). How prior experience and task presentation modulate innovation in 6-year-old-children. *Journal of Experimental Child Psychology*, *180*, 87–103. https://doi.org/10.1016/j.jecp.2018.12.004

Flynn, E. G., Turner, C., & Giraldeau, L.-A. (2016). Selectivity in social and asocial learning: investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(20150189). https://doi.org/10.1098/rstb.2015.0189

Flynn, E. G., & Whiten, A. (2008). Cultural transmission of tool use in young children: A diffusion chain study. *Social Development*, *17*(3), 699–718. https://doi.org/10.1111/j.1467-9507.2007.00453.x

Frick, A., Clément, F., & Gruber, T. (2017). Evidence for a sex effect during overimitation: boys copy irrelevant modelled actions more than girls across cultures. *Royal Society Open Science*, *4*(12), 170367. https://doi.org/10.1098/rsos.170367

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*. https://doi.org/10.1016/j.cogdev.2012.05.001

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755–756. https://doi.org/10.1038/415755a

Gerson, S. A., Burdett, E. R. R., & Beck, S. R. (2018). Preschoolers use analogy to facilitate innovative problem-solving. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1711–1716). Cognitive Science Society. https://cogsci.mindmodeling.org/2018/papers/0329/0329.pdf

Gönül, G., Hohenberger, A., Corballis, M., & Henderson, A. M. E. (2019). Joint and individual tool making in preschoolers: From social to cognitive processes. *Social Development*, *28*(4), 1037–1053. https://doi.org/10.1111/sode.12373

Gönül, G., Takmaz, E. K., Hohenberger, A., & Corballis, M. (2018). The cognitive ontogeny of tool making in children: The role of inhibition and hierarchical structuring. *Journal of Experimental Child Psychology*, *173*, 222–238. https://doi.org/10.1016/j.jecp.2018.03.017

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). https://doi.org/10.1371/journal.pone.0019555

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*, 1360–1366. https://doi.org/10.1126/science.1146282

Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). ‘Over-imitation’: A review and appraisal of a decade of research. *Developmental Review*, *51*, 90–108. https://doi.org/10.1016/j.dr.2018.12.002

Horner, V., Whiten, A., Flynn, E. G., & de Waal, F. B. M. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13878–13883. https://doi.org/10.1073/pnas.0606015103

Jongbloed-Pereboom, M., Nijhuis-van der Sanden, M. W. G., Saraber-Schiphorst, N., Crajé, C., & Steenbergen, B. (2013). Anticipatory action planning increases from 3 to 10 years of age in typically developing children. *Journal of Experimental Child Psychology*, *114*(2), 295–305. https://doi.org/10.1016/j.jecp.2012.08.008

Kacelnik, A., Chappell, J., Kenward, B., & Weir, A. A. S. (2006). Cognitive adaptations for tool-related behaviour in New Caledonian Crows. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: experimental explorations of animal intelligence* (pp. 515–528). Oxford University Press.

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

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology*, *116*(2), 471–486. https://doi.org/10.1016/j.jecp.2012.12.003

Kozbelt, A., Beghetto, R. A., & Runco, M. A. (2010). Theories of Creativity. In J. C. Kaufman & J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 20–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.004

Krafft, K. C., & Berk, L. E. (1998). Private speech in two preschools: Significance of open-ended activities and make-believe play for verbal self-regulation. *Early Childhood Research Quarterly*, *13*(4), 637–658. https://doi.org/10.1016/S0885-2006(99)80065-9

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

Legare, C. H., Wen, N. J., Herrmann, P. A., & Whitehouse, H. (2015). Imitative flexibility and the development of cultural learning. *Cognition*, *142*, 351–361. https://doi.org/10.1016/j.cognition.2015.05.020

Manning, B. H., White, C. S., & Daugherty, M. (1994). Young children's private speech as a precursor to metacognitive strategy use during task engagement. *Discourse Processes*, *17*(2), 191–211. https://doi.org/10.1080/01638539409544866

Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal Behaviour*, *85*(1), 195–202. https://doi.org/10.1016/j.anbehav.2012.10.026

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11th ed.). Beltz.

McGuigan, N., Burdett, E., Burgess, V., Dean, L., Lucas, A., Vale, G., & Whiten, A. (2017). Innovation and social transmission in experimental micro-societies : exploring the scope of cumulative culture in young children. *Philosophical Transactions of the Royal Society B*, *372*, 20160425. https://doi.org/10.1098/rstb.2016.0425

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23- and 30-month-olds. *Journal of Experimental Child Psychology*, *104*(4), 367–381. https://doi.org/10.1016/j.jecp.2009.07.001

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: Orangutans use water as a tool. *Biology Letters*, *3*(5), 453–455. https://doi.org/10.1098/rsbl.2007.0198

Metevier, C. M. (2006). *Tool-using in rhesus monkeys and 36-month-old children: acquisition, comprehension, and individual differences*. (Publication No. 3215775) [Doctoral dissertation, University of Massachusetts Amherst]. ProQuest Dissertations Publishing.

Neldner, K. (2020). *The development and evolution of tool innovation in human children (Homo sapiens) and chimpanzees (Pan troglodytes): a cross-cultural and comparative investigation* [PhD Thesis, The University of Queensland]. https://doi.org/10.14264/uql.2020.949

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/https://doi.org/10.1037/dev0000672

Nielsen, M. (2013). Young Children's Imitative and Innovative Behaviour on the Floating Object Task. *Infant and Child Development*, *22*, 44–52. https://doi.org/10.1002/icd.1765

Nielsen, M., Tomaselli, K., Mushin, I., & Whiten, A. (2014). Exploring tool innovation: A comparison of Western and Bushman children. *Journal of Experimental Child Psychology*, *126*. https://doi.org/10.1016/j.jecp.2014.05.008

Oeri, N., & Roebers, C. M. (2020). Regulating disappointment can impair cognitive performance in kindergarten children: Individual differences in ego depletion. *Journal of Experimental Child Psychology*, *190*, 104728. https://doi.org/10.1016/j.jecp.2019.104728

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2012). The emergence of tool use during the second year of life. *Journal of Experimental Child Psychology*, *113*(3), 440–446. https://doi.org/10.1016/j.jecp.2012.06.001

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

Reader, S. M., & Laland, K. N. (2003). *Animal innovation* (Vol. 10). Oxford University Press.

Reindl, E., Apperly, I. A., Beck, S. R., & Tennie, C. (2017). Young children copy cumulative technological design in the absence of action information. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-01715-2

Reindl, E., Beck, S. R., Apperly, I. A., & Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1825), 20152402. https://doi.org/10.1098/rspb.2015.2402

Reindl, E., Tennie, C., Apperly, I. A., Lugosi, Z., & Beck, S. R. (2022). Young children spontaneously invent three different types of associative tool use behaviour. *Evolutionary Human Sciences*, 1–48. https://doi.org/10.1017/ehs.2022.4

Sarooghi, H., Libaers, D., & Burkemper, A. (2015). Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors. *Journal of Business Venturing*, *30*(5), 714–731. https://doi.org/10.1016/j.jbusvent.2014.12.003

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

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 of London. Series B, Biological Sciences*, *371*(20150191). https://doi.org/10.1098/rstb.2015.0191

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. JHU Press.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Walter, V., Gampe, A., Carpenter, M., & Tomasello, M. (2014). Limitations to the cultural ratchet effect in young children. *Journal of Experimental Child Psychology*, *126*, 152–160. https://doi.org/10.1016/j.jecp.2014.04.006

Thibodeaux, J., Bock, A., Hutchison, L. A., & Winsler, A. (2019). Singing to the self: Children's private speech, private song, and executive functioning. *Cognitive Development*, *50*(January 2018), 130–141. https://doi.org/10.1016/j.cogdev.2019.04.005

Tomasello, M. (2020). *Mensch werden: eine Theorie der Ontogenese*. Suhrkamp Verlag.

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition*, *105*(2), 362–379. https://doi.org/10.1016/j.cognition.2006.10.003

Träuble, B., & Pauen, S. (2011). Cause or effect: What matters? How 12-month-old infants learn to categorize artifacts. *British Journal of Developmental Psychology*, *29*(3), 357–374. https://doi.org/10.1348/026151009X479547

van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Suci Utami, S., & Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science*, *299*(5603), 102–105. https://doi.org/10.1126/science.1078004 van Schaik, C. P., Ancrenaz, M., Djojoasoro, R., Knott, C. D., Morrogh-Bernard, H. C., Nuzuar, Odom, K., Atmoko, S. S. U., & van Noordwijk, M. A. (2009). Orangutan cultures revised. In S. A. Wich, S. S. U. Atmoko, T. M. Setia, & C. P. van Schaik (Eds.), *Orangutans: Geographic variation in behavioral ecology and conservation* (pp. 299–309). OUP Oxford.

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

von Hippel, W., & Suddendorf, T. (2018). Did humans evolve to innovate with a social rather than technical orientation? *New Ideas in Psychology*, *51*(June 2017), 34–39. https://doi.org/10.1016/j.newideapsych.2018.06.002

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar (eds.)). MIT Press. (Original work published 1934)

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

Whalley, C., Cutting, N., & Beck, S. R. (2017). The effect of prior experience on children's tool innovation. *Journal of Experimental Child Psychology*, *161*, 81–94. https://doi.org/10.1016/j.jecp.2017.03.009

Whiten, A., & Flynn, E. G. (2010). The Transmission and Evolution of Experimental Microcultures in Groups of Young Children. *Developmental Psychology*, *46*(6), 1694–1709. https://doi.org/DOI: 10.1037/a0020786

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., Manfra, L., & Diaz, R. M. (2007). "Should I let them talk?": Private speech and task performance among preschool children with and without behavior problems. *Early Childhood Research Quarterly*, *22*(2), 215–231. https://doi.org/10.1016/j.ecresq.2007.01.001

Zelazo, D. P., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., Sutherland, A., & Carlson, S. M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development*, *68*(3), 1–151.

List of Figures

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

**FAKULTÄT FÜR VERHALTENS-
UND EMPIRISCHE KULTURWISSENSCHAFTEN**

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg / Doctoral Committee of the Faculty of Behavioural and Cultural Studies of Heidelberg University

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe. / I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe. / I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname / First name Family name	Sabine Breyel
Datum / Date	25.02.2022
Unterschrift / Signature	Sabine Breyel

Scientific publications for the publication-based thesis

I. Manuscript

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development*, *58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

Cognitive Development 58 (2021) 101049

Contents lists available at ScienceDirect

Cognitive Development

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

The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task

Sabine Breyel *, Sabina Pauen

Department of Developmental and Biological Psychology, Heidelberg University, Hauptstraße 47-51, 69117, Heidelberg, Germany

ARTICLE INFO

Keywords:
Tool innovation
Problem solving
Preschool children
Cognitive development

ABSTRACT

This study focuses on tool innovation in preschoolers. It extends previous work by investigating task- and age-related changes in success rates, latency to success, and solution strategies. Three- to five-year-olds had 10 min to retrieve a toy from either a horizontal or a vertical tube by manufacturing a tool from multiple materials. In the Horizontal Tube Task, success rates were high irrespective of age. Solution strategies revealed mainly second-order innovation (i.e., manipulating materials while retrieving the toy). In the Vertical Tube Task, success rates increased with age and solution strategies reflected mainly first-order innovation (i.e., manufacturing the tool before trying to retrieve the toy). In the Vertical Tube Task, children needed more time to find a solution, with first-order innovation tending to be faster than second-order innovation. These findings demonstrate the potential of young preschoolers to innovate tools when given enough time, but also a high task-dependency of performance.

1. Introduction

Comparative research suggests that humans use tools far more often and in more sophisticated ways than any other species (Reynaud, Lesourd, Navarro, & Osiurak, 2016; Vaesen, 2012). As tool knowledge, brain development, and cognitive development seem to be linked (Gibson, 1993; Vaesen, 2012), psychologists soon became interested in studying age-related changes in tool understanding and tool use (e.g., Chen, Siegler, & Daehler, 2000; Connolly & Dalgleish, 1989; McCarty, Clifton, & Collard, 2001). Today, we know that even preverbal infants have a basic concept of tools in the sense of their form-function relations (e.g., Träuble & Pauen, 2007, 2011), and that toddlers can use tools spontaneously and rationally to solve a given task (Brown, 1990; Buttelmann, Carpenter, Call, & Tomasello, 2008; Reindl, Beck, Apperly, & Tennie, 2016).

Understanding the function of a tool or learning how to use it is substantially different from the process of making a new tool, however. Tool making does not only require the physical transformation of materials but also includes "the prior step of imagining the type of tool suitable" (Beck, Apperly, Chappell, Guthrie, & Cutting, 2011, p. 302) for the problem at hand. These two different processes have been called tool manufacture and tool innovation, respectively (Beck et al., 2011). Tool innovation involves (a) the ability to analyze a given problem to identify what kind of tool is needed, (b) the activation of knowledge about different materials, (c) thinking about potential transformations of the material provided for tool making, and (d) choosing the transformation option most suitable for the problem. Tool manufacture requires specific fine motor skills, but also the ability to adjust one's strategy in the case of a failed solution attempt. The latter may also become part of tool innovation (i.e., when it is necessary to rethink the tool design).

---

* Corresponding author.
E-mail addresses: sabine.breyel@psychologie.uni-heidelberg.de (S. Breyel), sabina.pauen@psychologie.uni-heidelberg.de (S. Pauen).

https://doi.org/10.1016/j.cogdev.2021.101049
Received 21 September 2020; Received in revised form 23 March 2021; Accepted 19 April 2021

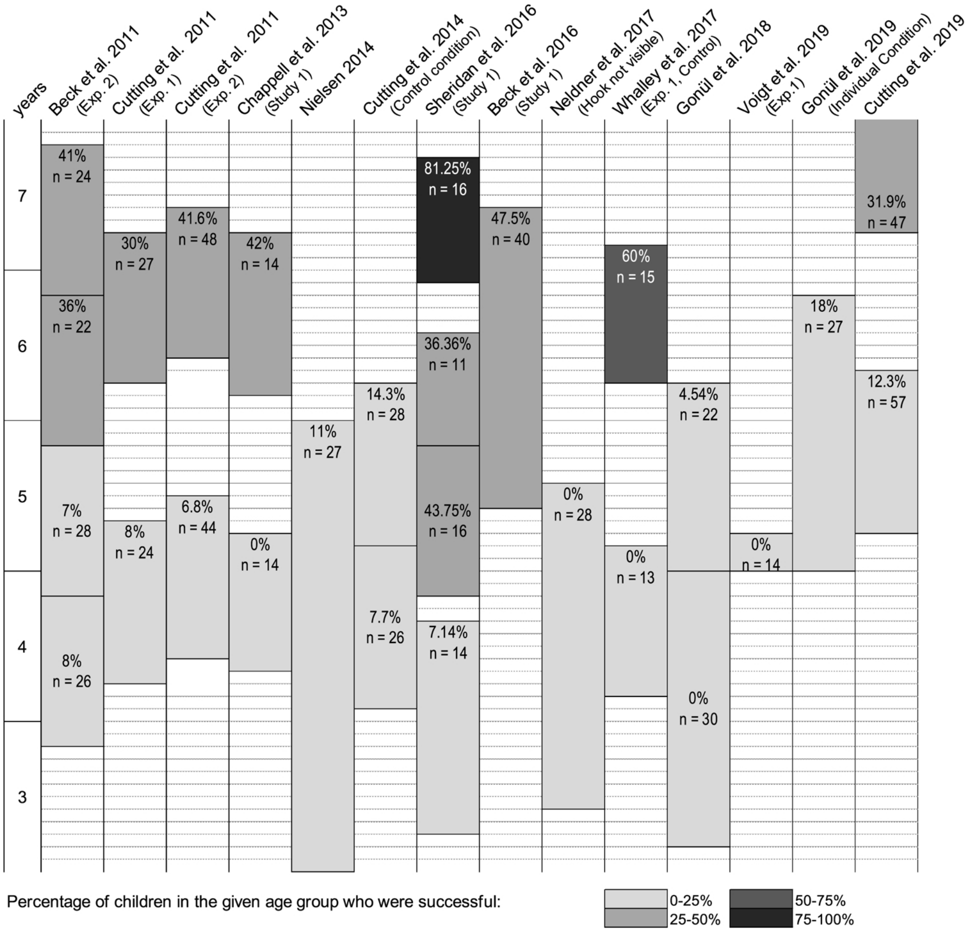


Fig. 1. Overview of developmental studies on the Hook Task: Age groups tested, sample sizes, and percentages of successful children in those experiments that were most similar to the procedure and original task version of Beck et al. (2011).

In their pioneering work, Beck et al. (2011) found only very low performance in an innovation task when testing children younger than eight years of age. The authors concluded that young children still seem to have general difficulties with tool innovation. However, over the past 10 years several studies have shown that some developmental progress can be observed even during preschool years and task demands play an important role in this context (e.g., Cutting, 2013; Cutting, Apperly, & Beck, 2011; Sheridan, Konopasky, Kirkwood, & Defeyter, 2016; Voigt, Pauen, & Bechtel-Kuehne, 2019). The present paper will summarize this work and identify critical variables that are likely to influence children's tool innovation and provide new data on preschool children's performance in two innovation tasks. In addition, we will suggest new ways to analyze children's solution attempts. Our main goal is to shed light on the very beginnings of tool innovation in human ontogeny.

1.1. *Tasks to study tool innovation in children*

The paradigms used to study tool innovation in children have been partly adapted from research with non-human animals (e.g., Mendes, Hanus, & Call, 2007; Tennie, Call, & Tomasello, 2009; Tennie, Call, & Tomasello, 2010; Weir, Chappell, & Kacelnik, 2002) and differ in terms of the way tools are used to solve the problem. In the *Floating Peanut Task,* a reward is placed on the bottom of a long narrow tube. The reward can be reached by filling the tube with water to make the object float. Only 0–17 % of four-year-old children were successful. Among six-year-olds, 8–50 % were able to solve the task (Ebel, Hanus, & Call, 2019; Hanus, Mendes, Tennie, & Call, 2011; Nielsen, 2013).

In the *Loop Task,* a reward is placed on a plastic platform inside a mesh-box and cannot be reached by hand. The participant had to form a loop with a piece of wooden wool, push the loop through the mesh to put it over a screw attached to the platform and pull the platform closer to the mesh to reach the reward. Four-year-old children were not able to find the solution independently within two minutes and only one-third of the children solved the task after watching a social demonstration of the solution (Tennie et al., 2009).

The *Shelf Task* and the *Bridge Task* require the participant to use a tool indirectly to change the structure of the apparatus to get access to the reward. A rubber ball with a reward attached to it had to be retrieved from a plastic box. The participant had to create a tool that prevents the ball from falling into the box (where it could no longer be reached) by building a shelf (Shelf Task) or a bridge (Bridge Task) to maneuver the ball to a little exit hole. 38 % of the four- to six-year-old children removed a stick from a piece of cardboard to use it as a shelf, and 65 % of the children straightened a bent strip to use it as a bridge (Cutting, 2013).

The most frequently used innovation tasks are the *Hook Task* or *Vertical Tube Task* (Beck et al., 2014; Beck, Williams, Cutting, Apperly, & Chappell, 2016, 2011; Chappell, Cutting, Apperly, & Beck, 2013; Cutting et al., 2011; Cutting, Apperly, Chappell, & Beck, 2014; Cutting, Apperly, Chappell, & Beck, 2019; Gönül, Takmaz, Hohenberger, & Corballis, 2018; Gönül, Hohenberger, Corballis, & Henderson, 2019; Neldner, Mushin, & Nielsen, 2017, 2019; Nielsen, Tomaselli, Mushin, & Whiten, 2014; Sheridan et al., 2016; Voigt et al., 2019; Whalley, Cutting, & Beck, 2017) and the *Unbending Task* or *Horizontal Tube Task* (Cutting et al., 2011; Cutting, 2013; Neldner et al., 2019; Voigt et al., 2019). In the Hook Task, children are asked to get a reward from the bottom of a narrow vertical tube. Children are presented with a pipe cleaner and some distractor items (e.g., a string) to manufacture a tool. The solution is to bend one end of the pipe cleaner into a hook and to pull out the target object by using this hook. In the Unbending Task (Cutting, 2013; Cutting et al., 2011), the reward is situated in the middle of a horizontally mounted tube and can be accessed from both sides of the tube. To solve this task, the child needs to straighten a u-shaped pipe cleaner to push the reward out of the tube. Since these tube tasks have been the major focus of previous research on tool innovation in children, we will describe these findings in more detail and summarize which task characteristics have been found to influence children's performance.

1.2. *Children's performance in the Hook Task*

Studies applying the Hook Task typically report how many children of a given age range find a solution within the given time frame of one to two minutes. Although this classical paradigm is in broad use, existing evidence does not allow to draw precise conclusions regarding the exact onset of the development of tool innovation yet (see Fig. 1). Overall, three- and four-year-old children reached very low success rates between 0% and 8%. At five to six years of age, success rates seemed to gradually increase but results were still quite mixed, ranging from 4% to 18 % (see an exception of 44 % in Sheridan et al., 2016). Even in the group of seven-year-olds, most studies reported success rates below 50 %, but one study also reports about 80 % (see Sheridan et al., 2016).

Four of the 13 studies listed in Fig. 1 conducted statistical age group comparisons in the range between three and six years of age (Beck et al., 2011; Cutting et al., 2011; Sheridan et al., 2016; Whalley et al., 2017). Beck et al. (2011) found no differences between four-year-olds ($N = 28$) and five-year-olds ($N = 26$), but the six-year-olds ($N = 22$) were more successful than the five-year-olds. Sheridan et al. (2016) found that three- to four-year-olds ($N = 14$) were less successful than five-year-olds ($N = 16$) but that performance did not increase between five-year-olds and six-year-olds ($N = 11$). Cutting et al. (2011) and Whalley et al. (2017) both looked at four- to five-year-olds and six- to seven-year-olds. Cutting et al. (2011) reported no difference in success rates in the first experiment (total $N = 51$) but a significant difference in the second experiment (total $N = 92$). This raises the question of whether the first experiment did not reveal age differences due to a lack of statistical power. However, other studies based on rather small sample sizes (see Beck et al., 2011; Whalley et al., 2017, N = 13–28 per age group) did find significant age differences, indicating that the underlying effects were quite strong.

In sum, existing evidence is still inconclusive when it comes to determining the onset and early development of tool innovation as well as the size of age effects. One reason for the diversity in existing findings may be that the definition and range of age groups differ strongly between studies. Although tool innovation skills seem to be low in young children and to increase only gradually throughout early childhood, research on children under the age of four years is still largely missing.

1.3. *Contrasting performance in the Hook Task and the Unbending Task*

While success rates for the Hook Task in four- to five-year-old children only rarely exceeded 20 % (see Sheridan et al. (2016) and Voigt et al. (2019) for exceptions), 23%–60% of same-aged children managed to build and successfully use a tool in the Unbending Task (Chappell et al., 2013; Cutting et al., 2011; Neldner et al., 2019). Even among three-year-old children, about one-third of the participants solved the task (29 %, Neldner et al., 2019).

So far, studies that directly compare children's performance between the Unbending Task and the Hook Task are rare and do not reveal a clear picture yet. Cutting et al. (2011) found that 33 % of four- to five-year-olds solved the Unbending Task compared to only 8% of the children solving the Hook Task. Similar results were reported by Chappell et al. (2013) for a same-aged sample, with 22–42 % of the children solving the Unbending Task and only 0–18 % of the children solving the Hook Task. In both studies, task differences have not been tested for statistical relevance. Voigt et al. (2019), Exp. 2 and 3, each $N = 29$) found that five-year-old children solved a version of the Unbending Task more often (87 %–100 %) and faster (116s–127 s) than a version of the Hook Task (64–85 %, 192–220 s) when multiple materials were available to build a tool, although these differences were non-significant. It is possible that a rather small task effect could not be revealed in this sample due to a lack of statistical power. In sum, existing findings suggest that the task with a horizontal tube is somewhat easier than the task with a vertical tube, but more studies are needed that account for the nature of the task (unbending, hook) while also taking children's age into account.

1.4. *Different tool-manufacturing options*

Previous research suggests that in addition to the nature of the task and the exact age of the children tested, other details of the setup also need to be taken into consideration, such as the tool-manufacturing options, the time given to work on the task, and the number (and kind) of materials provided. In their original version, the Hook Task and the Unbending Task both provided only a single

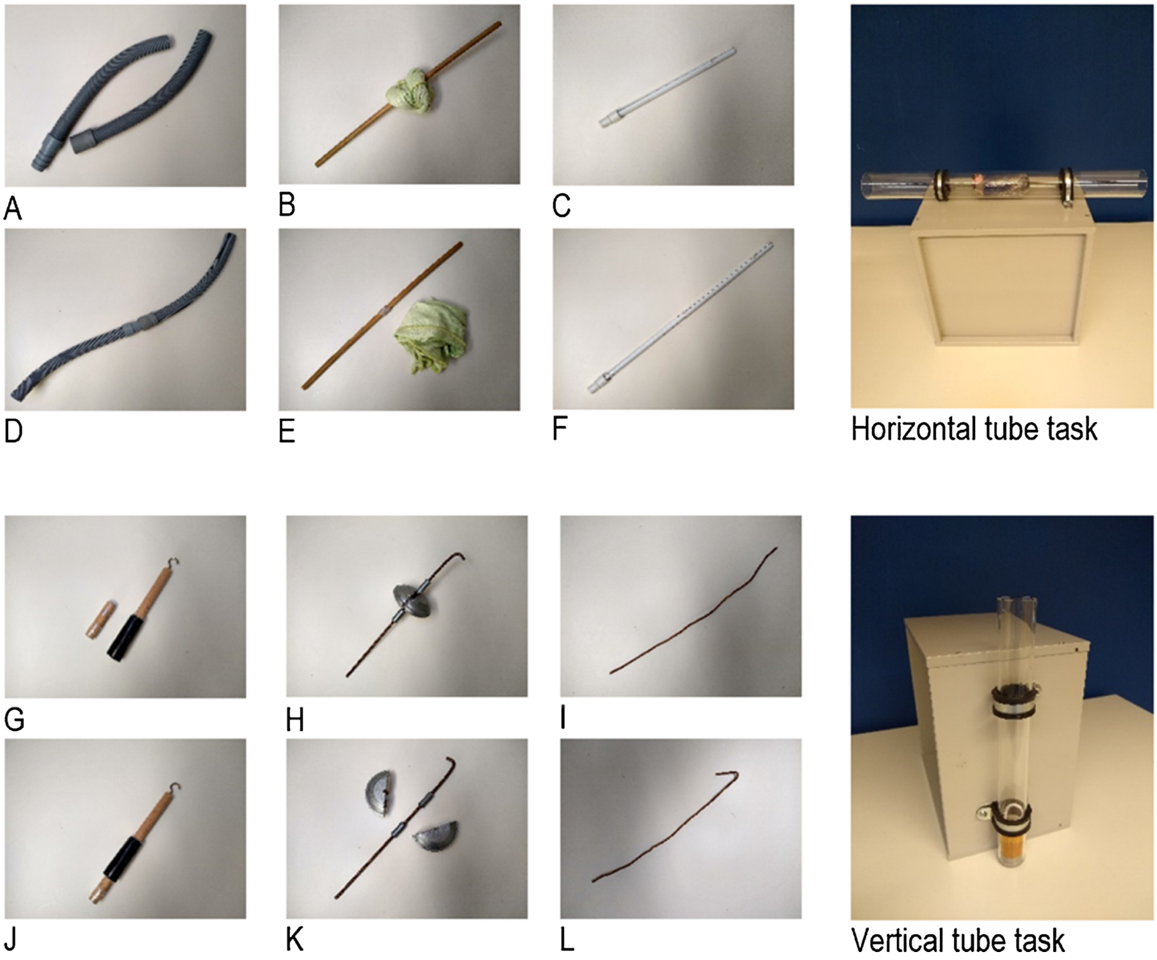


Fig. 2. Item sets for the Horizontal Tube Task (upper panel from left to right: add, subtract, reshape) in their original form (A–C) and their modified form (D–F). Item sets for the Vertical Tube Task (lower panel from left to right: add, subtract, reshape) in their original form (G–I) and their modified form (J–L).

manufacturing method to solve the problem (i.e., to reshape the pipe cleaner from one state to the other). This is important to note, as comparative research suggests that multiple tool making modes exist that may vary in difficulty depending on the kind of transformation required (Beck, 1980; Kacelnik, Chappell, Kenward, & Weir, 2006). When the mode of transformation affects tool making, it may also affect tool innovation. Four modes of tool making have been identified in wild animal research so far; they include *detach*, *subtract*, *add*, and *reshape* (Beck, 1980; Kacelnik et al., 2006). *Detach* is defined as the complete removal of a tool from another object or surface by severing the attachment between the two. *Subtract* refers to the removal of parts of an object to render it a functional tool. *Add* is defined as connecting two or more objects to form a tool and *reshape* means the change of an object's form or structure to create a tool.

Research on nonhuman primates suggests that detach is the most commonly found and easiest mode of tool making, followed by subtract, add, and lastly reshape (Bania, Harris, Kinsley, & Boysen, 2009; Beck, 1980; Shumaker, Walkup, & Beck, 2011; van Schaik et al., 2009). Consistent with these findings, developmental studies demonstrate that four- to five-year-old children used detach most frequently when different manufacturing methods were available (67 % detach vs. 7% subtract, 26 % alternative strategies and 0% add, or reshape, Voigt et al., 2019). Success rates were also highest when detaching was required to build a tool in a subtask (78 %–100 %, Cutting, 2013), compared to subtasks with the other modes of tool making (0–46 %, Cutting, 2013; Neldner et al., 2019).

However, it remains unclear to what extent the three strategies subtract, add and reshape differ in terms of their action complexity. Only one study directly compared success rates between three subtasks with subtract, add and reshape as manufacturing options in four- to seven-year-old children (Cutting, 2013, unpublished thesis). It did not reveal any significant difference in the frequency of using the three strategies neither for the Unbending Task (reshape: 43 %, add: 24 %, subtract: 26 %, total N = 133) nor for the Hook Task (subtract: 29 %, reshape: 11 %, add: 11 %, total N = 145). More research along these lines is thus needed before we can draw any final conclusions.

1.5. *Time for finding a solution*

The great majority of studies on tool innovation gave children only one to two minutes to work on the task (see Beck et al., 2011, 2016; Chappell et al., 2013; Cutting et al., 2011, 2014; Neldner et al., 2017; Nielsen et al., 2014). Recent findings suggest that young children may need far more time to explore different approaches and to generate a solution. According to Voigt et al. (2019), not a single five-year-old child found the solution to build a hook from a wire within the first minute, but within 10 min 21 % of all five-year-olds were able to solve the problem. The effect of time was even more pronounced when different manufacturing options (i.e., detach, add, subtract or reshape) were available: While only 28 % of the children found a solution within the first minute, 93 % were successful after 10 min. When only three manufacturing methods (subtract, add, reshape) were presented, 76 % of the children solved the innovation task within 10 min. These results imply that more time to work on the task helps young children to generate ideas for tool making, especially when different manufacturing options are offered.

1.6. *Number of materials for tool manufacture and quality of the solution attempt*

When provided with multiple materials, five-year-old children did not only use the intended modification modes like subtract (36 %) or add (27 %) but also found unexpected strategies (37 %) to innovate a tool (Voigt et al., 2019). Unexpected strategies included for example (a) opening the node of a piece of cloth attached to a stick (see also Fig. 2B upper panel) and wrapping the cloth loosely around the stick instead of completely subtracting the cloth from the stick, or (b) inserting multiple materials into the horizontal tube from one side to push the reward out at the opposite side. Comparable strategies for solving the Unbending Task have also been reported by Cutting (2013) and Neldner et al. (2019).

The evaluation of these different solution attempts is controversial. Classifications from the animal innovation literature (Burkart, Strasser, & Foglia, 2009; Ramsey, Bastian, & van Schaik, 2007; Whiten & van Schaik, 2007) suggest distinguishing between innovations resulting from a more incidental discovery or social learning (simple/weak innovation) and innovations involving more cognition and reflecting deliberate action of the individual (complex/invention innovation). Similar classifications based upon differences in cognitive complexity have not found their way into human innovation literature yet, since it seems difficult to draw any conclusion about the intentions underlying observed solution attempts. For example, while both Cutting (2013) and Neldner et al. (2019) did not classify pushing multiple materials into the tube as directed tool innovation because children “used the tools in their current state rather than innovating on them” (Neldner et al., 2019, p. 7), Voigt et al. (2019) argued that this behavior could still be regarded as innovation because children produced a useful behavior to successfully solve a novel problem (see also Carr, Kendal, & Flynn, 2016).

In this context it might be useful to take a closer look at the temporal or spatial succession of actions in the process of problem solving. We suggest differentiating between (a) tool manufacturing that is clearly separated (temporarily and/or spatially) from the event or place where the target problem is solved with this tool and (b) the combination of materials during an ongoing attempt to solve the target problem (i.e., performed at the same location). In the former case, children need to imagine a sequence of actions and their consequences before they start to actually apply the newly manufactured tool, whereas in the latter case they respond to immediate feedback of success and failure and adjust their handling of the material to solve the task without necessarily building a new tool. We suggest calling the former *first-order innovation* and the latter *second-order innovation*. Importantly, first-order innovation can follow previous unsuccessful attempts to solve the task as it does not imply that children have a correct idea of how to build a tool from the very beginning.

1.7. *Open issues in developmental research on tool innovation*

In sum we conclude that tool innovation increases throughout early childhood but that fine-grained age comparisons among preschool children are still missing. Task characteristics including the nature of the task, the number and selection of materials offered to manufacture a tool, and the amount of time to find a solution seem to affect children’s performance. However, their impact has not yet been studied with respect to its relevance for children of different ages. Moreover the quality of solution attempts and their related changes with age and specific task demands need to be explored in more detail. Only when corresponding data is available, it will be possible to better understand the onset of tool innovation in human ontogeny.

On the one hand, only very few studies have tested three-year-olds so far (see Fig. 1), suggesting that this age group is likely to mark the very beginnings of tool innovation. On the other hand, Voigt et al. (2019) recently found that by five years of age children are already very competent tool innovators in tube-related tasks when sufficient resources are available in terms of time and material. Thus, by studying three-, four-, and five-year-old children in the current study, we intend to cover the most relevant developmental steps in tool innovation during the preschool period. We will not only test the effect of age on innovation success but also the effect of the two task versions (Horizontal and Vertical Tube Task). For the first time, the type of innovation will be explored in more detail, categorizing children’s strategies as first- or second-order innovation. Finally, different manufacturing options (i.e., subtract, add, reshape) as well as extended time to work on the task (i.e., 10 min) will be available to increase the variability of possible solutions and to gain further insights into how young children approach tool innovation problems.

We expect success rates to increase with age and the latency to success to decrease with age because previous findings indicate an improvement in tool innovation across early childhood. We also expect the Horizontal Tube Task to be solved more often and faster than the Vertical Tube Task, as previous studies suggested the Horizontal Tube Task to be easier than the Vertical Tube Task. The analyses of innovation quality and preferences for different manufacturing options are exploratory because this is the first study to

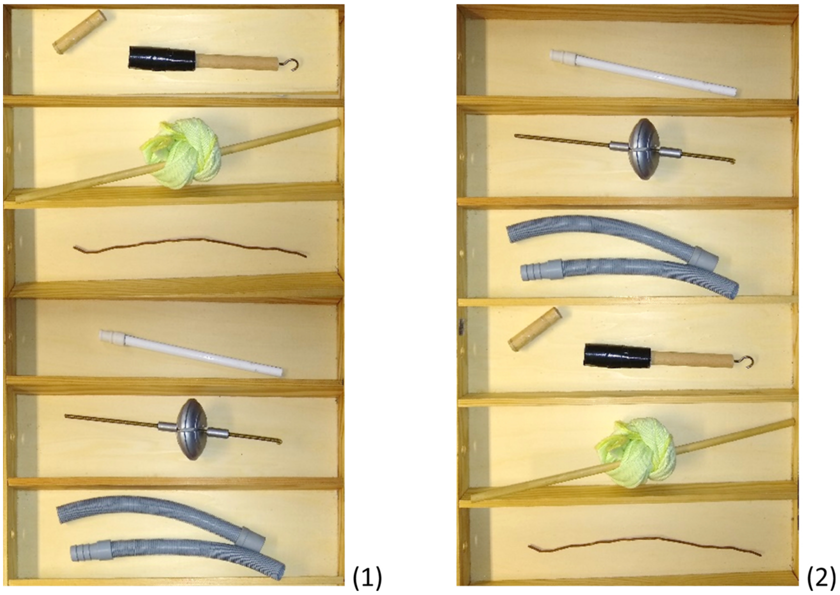


Fig. 3. Arrangements of item sets: Items were presented in a wooden box with six compartments that could be covered individually by pieces of black cloth. The two different item arrangements were counterbalanced across participants.

address this issue while also controlling for children's age and task type.

2. Material and methods

2.1. Participants

The study was conducted in a medium-sized town in Southwestern Germany with participants of European descent and an upper-middle-class socioeconomic background. More than half of the parents had a higher educational background (equivalent to a college degree or higher: 80.9 % of mothers, 77.5 % of fathers). Only a small percentage had either a middle (9% of mothers, 7.9 % of fathers) or a lower school education (1.1 % of mothers, 4.5 % of fathers), and data from 9 % of the mothers and 10.1 % of the fathers were missing. A total of N = 117 participants were recruited from a pool of healthy, full-term children. From this sample N = 28 children were excluded from further analyses for various reasons: Seven children (N = 2 three-year-olds, N = 2 four-year-olds, N = 3 five-year-olds) completed the task but used only one single unmodified material, thus showing no tool innovation defined as manufacturing a functional tool or combining tools in a new and useful way. The remaining N = 21 children (N = 7 three-year-olds, N = 12 four-year-olds, N = 2 five-year-olds) were excluded due to technical problems (N = 2), experimenter errors (N = 7), parental intervention (N = 1), broken material (N = 1), unwillingness to participate or to solve the task without help (N = 7), or interruptions of the testing procedure (bathroom visit; N = 3). The final sample included N = 89 children (42 girls, M = 4;5 months, range = 3;0–5;11). Approximately half of the children worked on the Horizontal Tube Task (N = 42) and the other half completed the Vertical Tube Task (N = 47). The three age groups (three-, four-, and five-year-olds) were equally distributed in each task group. Final sample sizes were kept similar to earlier studies on tool innovation.

The experiment was conducted according to the Declaration of Helsinki. All parents signed written informed consent forms before the start of the session. A female experimenter tested each child individually in a quiet laboratory room. One parent was sitting in a corner of the room and filled out different questionnaires (in N = 5 cases, parents were waiting in another room). When siblings were present they were engaged in a quiet task (e.g., reading a book, drawing). Parents and siblings who stayed in the testing room were instructed not to look at, help with, or comment on the participant's actions. If a participating child asked for help, the parent was instructed to say: "I am sure you can do this, you can try out anything you want."

2.2. Material

The materials used in this study were the same as in a precursor study by Voigt et al. (2019), requiring children to get a little cuddly toy out of a transparent tube. Children worked on one of two tube tasks[1] (see Fig. 2): In the Horizontal Tube Task, children needed a tool that was long enough to push out the toy from a transparent horizontal plexiglass© tube that was open on both sides. In the

---

[1] Some children worked on both tasks, but only data for the first task will be discussed in this report.

Vertical Tube Task, children needed a long tool with a hook-like ending to pull up the toy from the bottom of a plexiglass© tube with only one opening at the top. Each tube was mounted on a wooden box to keep it stable (see Fig. A1 for exact sizes). In both tasks, children received three potentially functional item sets and three distractor item sets. Each set consisted of one or two parts that were to be manipulated in a specific way to become functional (see Voigt et al., 2019, Exp. 3). The manufacturing methods required the child to either put two parts together (add), disconnect one part from the other (subtract), or modify the shape of the critical item (reshape).[2] The functional item sets for the Vertical Tube Task served as distractors for the Horizontal Tube Task and vice versa. Two arrangements of item sets were counterbalanced across participants and task versions (see Fig. 3).

2.3. Procedure

Following a short warm-up game (i.e., blowing balls of cotton wool into a goal), each child was presented with the six item sets, each located in a separate compartment of a large material box and covered by a black cloth. The experimenter lifted the cover of a given compartment, pointed to the specific item set and said: "You can take this. Have a closer look and try it out." If the child did not touch the material spontaneously, the experimenter repeated: "You are very welcome to take it out." Importantly, the experimenter did not provide any hints on how to handle the material. After 20 s or once the child lost interest in exploring the given material, the experimenter took back the items, re-established their unmodified state, put them back into the compartment, and covered it up again before the next item set was presented.

Once the child had been familiarized with each item set, the experimenter presented one of two empty familiarization tubes to highlight its affordance (i.e., two openings for the horizontal tube, one opening for the vertical tube). The horizontal tube was presented by saying: "This tube has two openings: You can reach in here[3] *(experimenter reaches into one opening with her hand),* and you can reach in here too *(experimenter reaches into the second opening with her other hand)*." The vertical tube was presented by saying: "This tube has one opening: You can reach in here (*experimenter reaches into the opening with her hand)*. It is closed on the other side: You cannot reach in here (*experimenter shows the closed end of the tube)*."

Next, the test tube was presented with the toy inside (i.e., placed in the middle of the horizontal tube or at the bottom of the vertical tube), and the uncovered box containing the six item sets was placed within reach of the child. The experimenter said: "Can you see the toy inside the tube? Let's get it out to have a closer look at it! I must do some work now, but you can stay here and try to get the toy out of the tube. I am sure you can do it! You can try out anything you like." While the child was working on the task, the experimenter sat down in a corner of the room pretending to be busy writing. In pre-defined situations the experimenter provided special prompts: (1) If the child did not begin to work on the task within two minutes or stopped working on the task for more than two minutes, the experimenter said: "I am still working on something. Just try to get the toy out of the tube on your own! I am sure you can do it." (2) If the child tried to retrieve the toy without using any of the material provided for at least three minutes, the experimenter directed attention to the item sets by saying: "You can also use any of these things to get the toy out of the tube. You can do it any way you like." (3) If the child unsuccessfully tried to solve the task with one specific item for at least three minutes, the experimenter guided the child's attention to the other item sets by saying: "Look, there are more things. You can use any of them." If the child failed to solve the task within the maximum time of 10 min, the observation stopped and the experimenter assisted the child in retrieving the toy to avoid frustration.

2.4. Coding

All sessions were videotaped and coded offline. The dependent measures were (a) success on the task (0 = failure, 1 = success), (b) innovation type (1 = first-order innovation, 2 = second-order innovation), (c) manufacturing method (add, subtract, reshape, unexpected strategy,[4] as defined by Voigt et al., 2019), and (d) latency to success (in seconds). Children were classified as successful if they retrieved the toy from the tube within 10 min without assistance by the experimenter, either by modifying a given item or by combining items. Based on the arguments raised in the introduction, we distinguished between different types of tool innovation: If an item was modified in a functional way before inserting it into the tube (e.g., connecting two pieces of an item set or removing a

---

[2] The manufacturing option detach, which has previously been used in Exp. 2 by Voigt et al. (2019), was not applied in this study because we focus on tool innovation and detach only requires the child to choose a ready-made tool rather than manufacturing a tool.

[3] We have adopted this expression from a previous study (Voigt et al., 2019) using the same material and procedure. To rule out the possibility that the phrase "reach in here" biased children (especially the younger ones) towards using their hands rather than a tool to reach for the toy, we coded how three-year-olds touched the tube for the first time. Of N = 28 three-year-olds, N = 19 children first touched a tool to use it on the tube (without reaching into the tube with their hands). Another N = 4 children first reached into the tube by hand, but quickly turned to the tools and continued by trying to find or manufacture a proper tool (without a prompt). Only N = 5 children used their hands for a longer time and thus received a prompt from the experimenter to focus their attention on the tool materials. Following that, all five children started to reach for the toy by using a tool or combination of tools. Thus 82% (23/28) of the youngest children revealed no sign of being biased towards using their hands. Nevertheless, future studies may use a more neutral instruction (e.g., "Here is an opening" vs. "Here are two openings") to avoid sending out potentially misleading signals to the children."

[4] Unexpected strategies included (a) pushing several items into the horizontal tube, (b) using a modified item set that was originally designed to be functional for the other tube task, (c) unexpected modifications on the item sets (e.g., wrapping up the cloth of the subtract-horizontal item set to make it fit into the tube, or (d) other functional combinations (e.g., pulling up a part of the add-vertical item set by using the hook of the subtract-vertical item set).

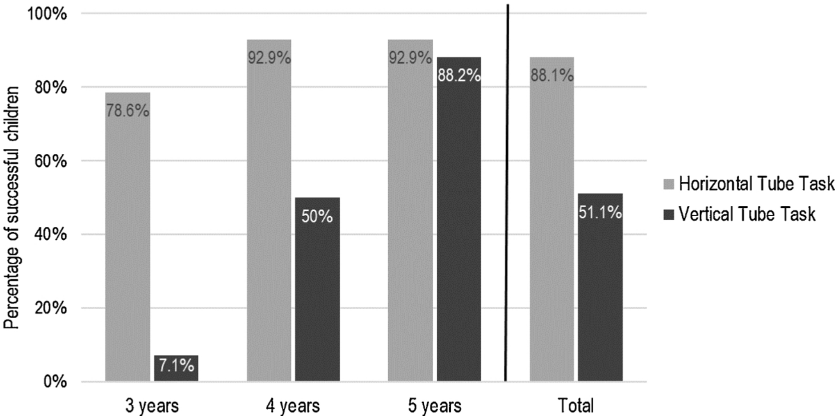


Fig. 4. Percentage of successful children in each age group and task version. For sample sizes see Table A1 in the Appendix A.

Table 1
Binary logistic regression predicting success in the Horizontal Tube Task (N = 42) and the Vertical Tube Task (N = 47).

	B	SE	Wald	df	*p*	Odds Ratio and 95 % CI		
						Lower	Odds Ratio	Upper
Horizontal Tube Task								
Age	.077	.057	1.785	1	.181	.965	1.080	1.208
Constant	−1.865	2.790	.447	1	.504		.155	
Vertical Tube Task								
Age	.201	.054	14.020	1	.000	1.101	1.223	1.358
Constant	−10.673	2.858	13.941	1	.000		.000	

non-functional part of an item before touching the tube with the material), it was coded as first-order innovation. If parts of an item set or different items were combined while they were already in use at the tube, it was coded as second-order innovation (e.g., pushing two or more items loosely into the horizontal tube, one after the other without connecting them, or pulling up a hooked item that was stuck in the vertical tube with another item, without connecting them beforehand). In addition, the latency to success was coded by measuring the time between providing the item sets (i.e., the moment the experimenter removed her hand from the box) and the removal of the toy from the tube.

A second coder rated 50 % of the videos to check for interrater agreement. Regarding children's success and the type of innovation they used, both coders agreed in all cases. The coders' agreement for the manufacturing method was also very good (Cohen's κ = .856). For latency to success the intraclass correlation coefficient was excellent (two-way mixed, absolute agreement, single measures: ICC = .999, 95 % CI = .9993–.9998).

2.5. *Analysis*

We used IBM SPSS Statistics (version 25) for statistical analyses. To determine the effect of task version on children's success rates and the use of first- vs. second-order innovation, we conducted Chi-Square tests and calculated Phi-coefficients (ϕ) as a measure of effects size, with ϕ defined as the square root of χ^2/N. To determine the effect of age on success and the type of innovation used in both task versions, binary logistic regression analyses were conducted, as this kind of analysis allowed us to consider children's age in a more fine-grained way than group comparisons do. For the outcome measure latency to success, we conducted linear regressions. Cox and Snell's R^2 and Nagelkerke's R^2 as well as Cohen's f^2 are reported as measures of effect size. For testing differences in latency to success between the two task versions and the type of innovation, non-parametric Mann-Whitney U tests were used since the data was not normally distributed. Pearson correlation coefficients (r) are reported as a measure of effect size. Cohen's classification (1988) is used to assess the magnitude of the effects. It distinguishes between weak effects (r / ϕ = .10 and f^2 = .02), medium effects (r / ϕ = .30 and f^2 = .15), and strong effects (r / ϕ = .50 and f^2 = .35).

3. Results

In a first step, we looked at the total sample to determine whether task version and age affected children's success. Then we focused only on those children who were able to solve the task (N = 61 in both tasks combined), exploring the effect of task version on the type of innovation used, as well as reporting which manufacturing methods children chose and how long it took them to find a solution. Finally, we provide data on the number and the type of prompts given in the two tasks to consider possible effects on children's performance.

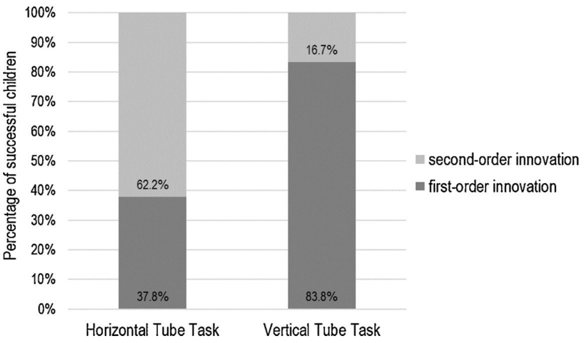


Fig. 5. Performance in the Horizontal Tube Task (N = 42) and Vertical Tube Task (N = 47): Percentage of children who solved the task individually within 10 min using first- or second-order innovation.

Table 2
Percentage of successful children in the three age groups showing first- or second-order innovation in the Horizontal and Vertical Tube Task.

		Age Group		
		3 years	4 years	5 years
Horizontal Tube Task				
	First-order innovation	9.1 % (1/11)	61.5 % (8/13)	38.5 % (5/13)
	Second-order innovation	90.9 % (10/11)	38.5 % (5/13)	61.5 % (8/13)
Vertical Tube Task				
	First-order innovation	0%	87.5 % (7/8)	86.7 % (13/15)
	Second-order innovation	100 % (1/1)	12.5 % (1/8)	13.3 % (2/15)

3.1. *Success rates*

Consistent with our hypothesis, significantly more children solved the Horizontal Tube Task (88.1 %) compared to the Vertical Tube Task (51.1 %), $\chi^2(1) = 12.44, p < .001$, resulting in a medium effect size of $\phi = .37$. To allow for a descriptive age comparison of success rates, Fig. 4 reports the percentage of successful children among three-, four-, and five-year-olds in each task.

To determine the effect of age in each task version, two binary logistic regression models were run using age (in months) as a predictor (see Table 1). The regression model for the Horizontal Tube Task was not significant and revealed only a poor fit to the data, $\chi^2(1) = 2.168, p = .141$, Cox and Snell's $R^2 = .05$, Nagelkerke's $R^2 = .097$. Age did not predict success. In contrast, the model for the Vertical Tube Task was significant and showed good fit to the data, $\chi^2(1) = 25.12, p < .001$, Cox and Snell's $R^2 = .414$, Nagelkerke's $R^2 = .552$, and a strong effect, $f^2 = 1.23$. The model correctly predicted 82.6 % of the unsuccessful children and 75 % of the successful children. If age increased by one month, the odds of being successful increased by 22.3 %.

3.2. *First-order and second-order innovation*

Since this study is the first to investigate the impact of task and age differences on the type of innovation in preschoolers, our analyses were primarily exploratory. The investigation of innovation types, manufacturing methods, and latency to success were based on successful children only, resulting in varying sample sizes per age group.

As depicted in Fig. 5, children used first-order innovation more often in the Vertical Tube Task than in the Horizontal Tube Task, whereas the reverse was true for second-order innovation, $\chi^2(1) = 10.44$, $p = .001$, N = 61, $\phi = .41$.

To determine whether the type of innovation was related to children's efficiency in working on the problem, we compared the latencies until success for first- and second-order innovation. Children using first- and second-order innovation did not differ in terms of latency in the Horizontal Tube Task, $\text{Median}_{\text{first-order}} = 94$ s, $\text{Median}_{\text{second-order}} = 144$ s, $U = 132.5$, $z = -.893$, N = 37, $p = .377$. However, children using first-order innovation were faster than children using second-order innovation in the Vertical Tube Task, $\text{Median}_{\text{first-order}} = 215.5$ s, $\text{Median}_{\text{second-order}} = 473.5$ s, $U = 8.0$, $z = -2.479$, N = 24, $p = .010$, revealing a rather strong effect, $r = .51$.

In the Horizontal Tube Task, the majority of three-year-old children showed more second-order than first-order innovation, whereas no clear preference could be observed for both older age groups (see Table 2). In the Vertical Tube Task, only one three-year-old child solved the task using second-order innovation. In the two older age groups, the descriptive results suggested a preference for first-order innovation. However, logistic regression analyses showed that age did not significantly predict the type of innovation used neither in the Horizontal Tube Task, $\chi^2(1) = .227, p = .634$, Cox and Snell's $R^2 = .006$, Nagelkerke's $R^2 = .008$, nor in the Vertical Tube Task, $\chi^2(1) = .505$, $p = .477$, Cox and Snell's $R^2 = .021$, Nagelkerke's $R^2 = .035$ (see Tables A2 and A3). It should be noted that age variance was strongly reduced in the Vertical Tube Task, as only one out of 14 three-year-olds and 8 out of 16 four-year-olds were successful and contributed to the analysis. Hence the lack of significance regarding first- and second-order innovation in this task could be due to a lack of statistical power.

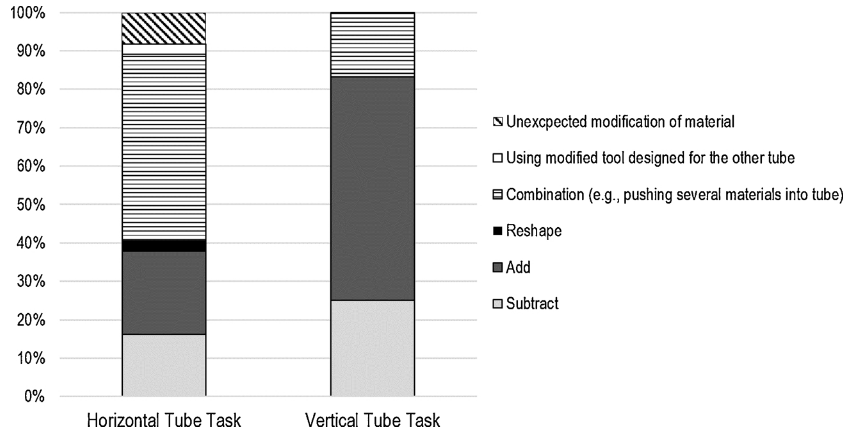


Fig. 6. Percentage of successful children using different manufacturing strategies to solve each task.

3.3. *Manufacturing methods*

3.3.1. *Approaches of children solving the task successfully*

An integrated analysis of both tasks showed that the vast majority of successful children either put two items firmly together to build a proper tool (add: 36.1 %) or combined different items in another way to solve the problem at hand (combination: 36.1 %). The strategy subtract was used by 19.7 % of the children. Interestingly, reshape was used by only one single child in the Horizontal Tube Task (1.6 %). Another four children used the material in an unexpected way, e.g., by wrapping the cloth from the horizontal subtract tool around the stick to make it fit into the horizontal tube (6.5 %) instead of subtracting the cloth. As illustrated in Fig. 6, manufacturing methods showed more variation in the Horizontal Tube Task than in the Vertical Tube Task.

Solving strategies also became more variable with children's age in both tasks (see Table A4 for percentages). Among the N = 37 successful children in the Horizontal Tube Task, 10 out of the 11 three-year-olds pushed different items into the tube; the other child used the strategy add. Successful four-year-olds (N = 13) either used subtract (N = 5), add (N = 4) or they put multiple materials into the tube (N = 3). One four-year-old child successfully used a tool designed for the other tube task. In the group of successful five-year-olds (N = 13), children pushed several items into the tube (N = 5), added two items (N = 3) or modified the subtraction tool in an unexpected way (N = 3; wrapping up the cloth around the stick to make it fit into the tube). One five-year-old child used the horizontal reshape tool and one child used the subtract method successfully. In sum, in the Horizontal Tube Task no clear preference for one manufacturing method was observed, but the repertoire of different strategies to solve the task – including unexpected ones – increased with age.

In the Vertical Tube Task, only one three-year-old was successful by unexpectedly combining different materials. Seven out of eight successful four-year-olds manufactured a tool by adding two items; the remaining child combined two materials to retrieve the toy. Within the group of five-year-olds (N = 15), the majority was split between add (N = 7) and subtract (N = 6). The remaining two children combined different materials. Reshape was not used in any of the age groups to solve Vertical Tube Task. In sum, the variability of strategies increased with age like in the Horizontal Tube Task. However, in contrast to the Horizontal Task most of the children used the intended manufacturing methods instead of unexpected combinations in the Vertical Tube Task.

3.3.2. *Approaches of children failing to solve the task*

A qualitative analysis of unsuccessful children's behavior was also highly informative regarding children's problem-solving abilities: Some children who worked on the Horizontal Tube Task (N = 5) took the correct approach by trying to push the toy, but they only used unmodified materials of which no part was long enough to push the toy out; they failed to combine different parts of the material in a functional way. Among unsuccessful children in the Vertical Tube Task (N = 23), a substantial number of children either built a correct tool or had a correct idea to build a proper tool but nonetheless failed to get the toy out. In three cases, children did not use the correctly built tool on the tube. In further two cases, children did not manage to remove the toy with the correct tool at the first try and stopped using this approach. In two other cases, children had an idea for the correct modification (bending the wire), but the hook was either too wide to fit into the tube or too small to pick up the toy. One child came up with a functional idea to modify an item set too late to apply the tool successfully within the time limit. Further two children tried to use the hooked part of a given item set but did not find a way to extend its length in order to reach the toy. Only two children did not modify any of the materials; nine children showed only irrelevant modifications and combinations. The remaining two children dropped one part of vertical add item set into the vertical tube and did not manage to get it out again. In sum, many children who failed to solve the Vertical Tube Task showed promising innovation attempts, but they had problems (a) thinking through all steps necessary for manufacturing a proper tool, (b) considering situational constraints, (c) handling the tool properly at the tube, or (d) to keep trying when first attempts failed.

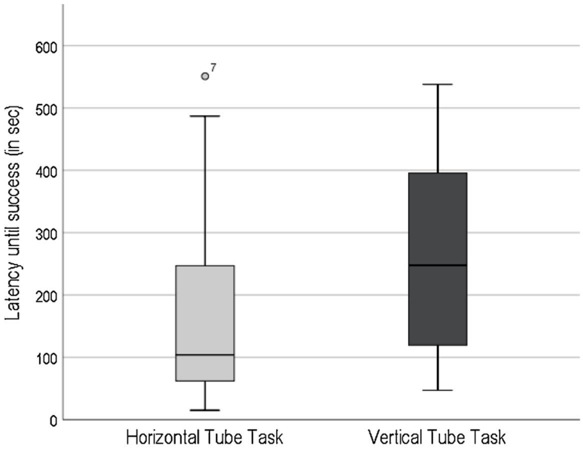


Fig. 7. The number of seconds it took successful children to get the toy out from the horizontal or vertical tube.

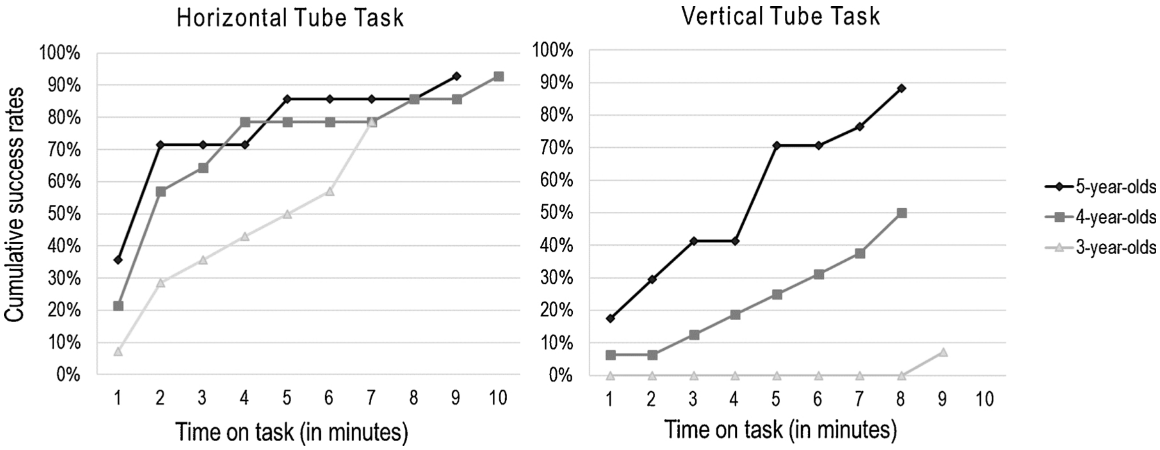


Fig. 8. Cumulative success rates (in %) for the Horizontal Tube Task (left panel) and the Vertical Tube task (right panel) per minute in the three age groups.

3.4. *Latency to success*

Consistent with the findings referring to higher success rates, the Horizontal Tube Task was also solved faster (median latency = 104 s, IQR = 191 s) than the Vertical Tube Task (median latency = 274.5 s, IQR = 288 s), $U = 596.0$, $z = 2.244$, N = 61, $p = .025$, $r = .29$. Overall, the range of latency to success was very high. Some children solved the task very fast whereas others needed more time to find a solution (see Fig. 7).

Fig. 8 presents the cumulative percentages of success rates over the course of 10 min for all three age groups (for exact percentages see Table A5). In the Horizontal Tube Task, 7.1 % of the three-year-olds, 21.4 % of the four-year-olds, and 35.7 % of the five-year-olds solved the task within the first minute (collapsed across age groups: 21.4 %). The success rates sharply increased within the second minute in all three age groups. After four to five minutes, the majority of older children had solved the task whereas we saw a gradual increase of successes in three-year-olds who reached the same level of performance as the older children after about seven minutes.

For the Vertical Tube Task, success rates started at lower levels in each age group. Collapsed across all ages, only 8.5 % of the children (N = 4) were successful within the first minute. In both older age groups, success rates increased over the next seven minutes. Only one three-year-old child solved the task after about nine minutes.

Linear regression analyses revealed that age had no significant effect on the time (in seconds) that children needed to find a solution, neither in the Horizontal Tube Task, $F(1,35) = 2.557$, $p = .119$, nor in the Vertical Tube Task, $F(1,22) = 1.822$, $p = .191$ (see Tables A6 and A7). As noted above, age variance was strongly limited in the Vertical Tube Task which might have affected this result.

3.5. *Number of prompts*

To determine whether the number of prompts varied between tasks or with children's age, we ran additional exploratory post-hoc analyses. We differentiated motivational prompts ("I am still working on something. Just try to get the toy out of the tube on your own! I am sure you can do it.") and item-related prompts referring to the material (“You can also use any of these things to get the toy out of the tube. You can do it any way you like.” and “Look, there are more things. You can use any of them.”). Although the caregivers accompanying the children were instructed not to interfere, they were allowed to respond with a motivational prompt (“I am sure you can do this, you can try out anything you want.”) if the child turned towards them and asked for their help. Prompts provided by the experimenter and the parent were summed up for further analyses.

3.5.1. Task differences

To avoid that power-issues lead to potential null-effects, we collapsed the data across age groups. In the Horizontal Tube Task (N = 42) at total of 12 children (28.6 %) received at least one motivational prompt, and 9 children (21.4 %) received at least one item-related prompt. In the Vertical Tube Task (N = 47), 22 children (46.8 %) received at least one motivational prompt, and 11 children (23.4 %) received at least one item-related prompt. The number of motivational prompts tended to be somewhat higher in the Vertical Tube Task (mean = 1.85, SD = 3.21) than in the Horizontal Tube Task (mean = .98, SD = 1.92), but this difference failed to reach level of significance, $U = 1162.0$, $z = 1.648$, N = 89, $p = .099$, $r = .17$. The number of item-related prompts did not differ between tasks, $U = 1004.0$, $z = .192$, N = 89, $p = .848$, $r = .02$, $\text{mean}_{\text{horizontal}} = .31$, $\text{SD}_{\text{horizontal}} = .68$, $\text{mean}_{\text{vertical}} = .34$, $\text{SD}_{\text{vertical}} = .76$.

3.5.2. Age differences

We found medium to large negative correlations between children’s age and the number of prompts (Spearman’s correlation coefficients): With increasing age children needed less motivational prompts, $r_S = -.41$, $p < .001$, and less item-related prompts, $r_S = -.24$, $p < .05$.

4. Discussion

The current study aimed to systematically test the impact of age and task differences on tool innovation in three- to five-year-old children. Based on previous work, we expected the Horizontal Tube Task to be easier to solve than the Vertical Tube Task, and older children to be more successful than younger children. Next to age and task differences in success and latency to success, we examined differences in the manufacturing options used by successful children and their kind of solution attempt (first-order or second-order) in an exploratory approach. In an additional post-hoc analysis, we investigated whether the number of prompts varied between tasks and across age.

4.1. Task and age differences in innovation success

The pattern of results partly confirmed our hypotheses about task difficulty and age trends. As expected, age proved to predict success in the Vertical Tube Task, ranging from very poor performance in three-year-olds (7% success) to very good performance in five-year-olds (88 % success). In contrast, the performance was similarly high across age groups in the Horizontal Tube Task (78 - 92 %). These results contradict the prevalent idea that young children are poor innovators. We showed that five-year-olds are very competent in solving two different innovation tasks, thus confirming the findings reported by Voigt et al. (2019) who used the same study setup, including different manufacturing options and a 10-minute time frame for finding a solution. In this precursor study, 76 % of the five-year-olds solved one of the two innovation tasks (Exp. 3).

The high level of success we found in the Horizontal Tube Task contrasts previous research on the Unbending Task, reporting lower success rates ranging from 29 to 59% in three-year-old children (Neldner et al., 2019), and 23–73 % in four-to-five-year-old children (Chappell et al., 2013; Cutting et al., 2011; Neldner et al., 2019). One reason for the discrepancy may have to do with differences in the study design: In the original Unbending Task, children had to unbend a pipe cleaner to render it a functional tool within one minute (e.g., Cutting et al., 2011). In a modified version, children were presented with three separate subtasks with one item set to be either subtracted, added, or reshaped within one minute (Neldner et al., 2019). In the present study, children had six different item sets (three of which were functional) to build a tool using different strategies (i.e., subtract, add or reshape) within 10 min. If our experiment had been stopped after one minute as in other studies, only 21.4 % of the children would have been successful in the Horizontal Tube Task, thus being comparable with the lower success rates reported previously. Our latency analyses revealed that success rates increased substantially after the first minute. This points to the high relevance of offering children more time to solve a problem requiring tool innovation (see also Voigt et al., 2019).

The finding that even three-year-old children found it rather easy to solve the Horizontal Tube Task in the present case suggests that the ability to innovate simple tools to solve a new problem starts to develop very early. This is in line with related research on tool understanding and tool use in toddlers showing that understanding form-function relations and using this knowledge for guiding tool-related actions can already be observed well before the third year of life (Brown, 1990; Träuble & Pauen, 2007, 2011). To examine whether such knowledge is sufficient for toddlers to manufacture tools themselves, future studies should test the Horizontal Tube Task on even younger children.

In contrast to the Horizontal Tube Task, the Vertical Tube Task was only solved by about half of the sample. Moreover, the odds of being successful increased with age. One potential explanation for these age and task differences refers to the assumption that young children may initially experience a strong impulse to make physical contact with the target. In the case of the Horizontal Tube Task, using an item led them to push the object further away (i.e., towards the other opening), without being successful yet. Based on this first contact and visual feedback, some children tried to manufacture a proper tool by manipulating or combining material in the intended way, whereas the majority simply put more material into the tube to push the target out. A corresponding step-by-step approach is not suitable in the case of the Vertical Tube Task, however. Making contact with the target alone does not bring the child any closer to the solution because the toy remains stuck at the bottom of the tube even when it is reached with the help of a tool. This points to the problem of how to keep contact when reversing one's action (i.e., pulling the toy up). An inspection of the video material revealed that items that were not yet functional to solve the task (i.e., too short, lacking a hook) accidentally fell into the vertical tube quite often, thus forcing the child to first remove the dropped materials (e.g., by turning the tube upside down or pulling it up with another item) before being able to continue with the target task.

The Horizontal Tube Task has the advantage of providing immediate positive visual feedback regarding the chosen strategy, allowing for a step-by-step approach towards the solution, whereas the Vertical Tube Task requires the child to plan ahead, to deal with failed attempts more often, and to show a reversal in action (first approaching the target, then changing the direction). Consistent with this interpretation, we found that second-order innovation was observed significantly more often for the Horizontal than the Vertical Tube Task. The combination of systematic thinking, planning, and a more complex action sequence may render the Vertical Tube Task so difficult - especially for young children. Related research suggests that the ability to plan actions ahead shows major improvements within the preschool period. Four-year-olds were significantly better in planning their actions than three-year-olds when solving a grasping posture task (Jongbloed-Pereboom, Nijhuis-van der Sanden, Saraber-Schiphorst, Crajé, & Steenbergen, 2013). Moreover, five-year-olds already mastered two versions of the Tower of London Task (with and without intermediate moves), whereas planning accuracy was considerably lower in the version with intermediate moves in four-year-olds, presumably because this task version required participants to use a search-ahead strategy (Kaller, Rahm, Spreer, Mader, & Unterrainer, 2008). Changes in planning abilities may thus map onto children's increasing performance from three to five years of age in the Vertical Tube Task.

Next to planning abilities, the level of executive function skills might impact tool innovation (e.g., Beck et al., 2016; Cutting et al., 2011). There are good reasons to assume that inhibition, working memory, and cognitive flexibility as three main components of executive functions (see Miyake et al., 2000) help children to succeed in an innovation task. The ability to inhibit dominant responses allows the child to resist the impulse of immediately using items in their unmodified form and to consider alternatives before starting to act. Furthermore, inhibition enables the child to stop repeating an unsuccessful strategy. Cognitive flexibility comes into play to generate new, more successful ideas. Lastly, working memory is needed to keep the goal represented in mind and to update relevant knowledge about the tools and their properties. All three executive skills develop substantially between three and six years of life (Garon, Bryson, & Smith, 2008; Zelazo et al., 2003), further supporting the idea that increasing executive function skills could predict age differences in innovative problem-solving tasks that require a strategic approach. Although theoretical assumptions suggest a close relation between executive functions and innovative problem solving, existing data has failed to support this claim. Neither did an experimental manipulation of inhibitory control improve children's performance in the Hook Task (Chappell et al., 2013), nor did executive function measures predict children's success (Beck et al., 2016; Gönül et al., 2018). There are good reasons to continue exploring the role of executive functions in tool innovation, however. Existing studies used a task setup that neither left children much time nor different options to find a solution. To test the impact of executive functions on children's performance task setups are needed that allow children to choose between alternative options, as this requires them to remember and update the effects of different tools, to inhibit using inefficient strategies and to switch strategies following failed attempts.

4.2. First-order and second-order innovation

Our exploratory analysis showed that the different task characteristics affected the repertoire of solutions children spontaneously found. In the Horizontal Tube Task, more children displayed second-order innovation than first-order innovation, whereas the reverse was true for the Vertical Tube Task. First-order innovation was defined as tool manufacturing that took place before the target problem is solved with this tool, reflected by the original manufacturing methods add, subtract and reshape. Second-order innovation was defined as the combination of materials during an ongoing attempt to solve the target problem. In the Horizontal Tube Task, second-order innovation manifested as pushing two or more items into the tube one after the other. In the Vertical Tube Task, second-order innovation involved pulling up a hooked item that was stuck in the vertical tube with another item.

Although the current study is the first to discriminate these subtypes of innovation, the observation of alternate strategies has already been reported in previous work. For example, Voigt et al. (2019) stated that 37 % of five-year-old children displayed unexpected strategies similar to those observed in the present study, and Neldner et al. (2019) found that 23–35 % of three-year-olds, 30–46 % of four-year-olds, and 26–40 % of five-year-olds used alternate strategies in the add and reshape subtasks with the horizontal tube (no alternate strategy occurred in the subtract subtask). This clearly points to the need to take a closer look at such unexpected strategies.

As outlined before, basic cognitive functions like planning or executive functions develop substantially during the preschool period. These abilities are likely to be associated with first-order innovation because it requires the child to mentally represent and implement a solution strategy before beginning to work on the target task. Although the descriptive data suggested that the youngest age group tested here applied first-order innovation less often than older children, age was no significant predictor for innovation type in the present analyses. However, these analyses were only exploratory and based on a rather small subsample. Because only a few younger children were successful in the Vertical Tube Task, age variance was strongly reduced which is likely to explain the non-significant results. Due to the preliminary nature of our data, further studies are needed before any final conclusion about the development of first- and second-order innovation can be drawn.

In terms of latency to success, we found an advantage for first-order innovation in the Vertical Tube Task but not in the Horizontal Tube Task. It does not seem surprising that first- and second-order innovation both resulted in similar latencies in the Horizontal Tube

Task. Since the horizontal tube has two openings it does not really matter whether children firmly connect two items before inserting the newly created tool into the tube or whether they insert multiple items without previously connecting them. Both approaches lead to quick success. This is not the case in the Vertical Tube Task, however. In this task the tube has only one opening and the most efficient way to solve the task is to build a functional hook before trying to retrieve the toy (i.e., showing first-order innovation). When children put a hook-shaped item into the tube that is too short to reach the target, this item easily slips off their hand and falls down. In such a case the item needs to be lifted up by using a second item. Solving the task by using such a two-step approach (i.e., second-order innovation) is often time-consuming, which explains the longer latencies for second-order innovation compared to first-order innovation in the Vertical Tube Task.

4.3. Choice of manufacturing options

The observation of more second-order innovation in the Horizontal Tube Task and more first-order innovation in the Vertical Tube Task matches findings regarding manufacturing methods. In the Horizontal Tube Task, more than half of the children pushed several items into the tube or combined other items without manufacturing a tool first. Fewer manufacturing options would likely have led to lower rates of second-order innovation. In the Vertical Tube Task, about 80 % of the children displayed first-order innovation by subtracting or adding an item set to render it a functional tool.

We cannot make any statements about the relative difficulty of the different manufacturing methods as it has been suggested by corresponding animal literature (Bania et al., 2009; Beck, 1980; Shumaker et al., 2011; van Schaik et al., 2009), because we used a within-subject design and did not present each item set separately to determine children's specific skills in one action domain (i.e., add, subtract, and reshape). However, our data reveal interesting patterns of manufacturing options across the two task versions and across age groups which should be examined more closely by further research. On the one hand, we found more variation in solution strategies for the Horizontal than the Vertical Tube Task. This is in line with previous findings indicating that children used alternate strategies quite often in the Horizontal Tube Task (Cutting, 2013; Neldner et al., 2019; Voigt et al., 2019), but only rarely in the Vertical Tube Task (see Beck et al., 2011 for an exception in older children). On the other hand, we observed that the variability of solution strategies increased with age in both tasks. Whereas the three-year-old children predominantly solved the tasks by combining different items, four- and five-year-olds also used tool-manufacturing strategies like add or subtract. Which skills underlie this development is difficult to determine based on the current data. Older children may get better at inspecting the materials in a goal-oriented way and exploring their potential to be useful. An observation supporting this idea comes from a previous study in which older children showed more exploratory behaviors such as combining or bending items in a delay phase before the start of the innovation task, compared to younger children who tended to only look at or pick up the material without trying to manipulate it (Chappell et al., 2013). To examine whether this also holds for the time when children actually work on the task, more detailed analyses of the task progress are needed, including information on the number of items that are used, the time children take to explore each item before using it or the number of different tool actions performed before finding a solution.

Consistent with very low success rates in the original Hook Task (e.g., Beck et al., 2011), the current study showed that the manufacturing option reshape does not seem to be salient for young children, especially when other manufacturing options are available (see also Voigt et al., 2019). Only two children tried to modify the reshape tool for the Vertical Tube Task but failed in manufacturing or using the tool. Only one child successfully used the reshape option to solve the Horizontal Tube Task, and this was not the original reshape option (i.e., unbending) but rather making a stick longer by changing its form. It may well be that bending a wire to produce a specific shape (i.e., a hook of a specific size) is still difficult for preschoolers who then prefer other tool making options.

4.4. Latency to success

Consistent with our results regarding success rates, the Horizontal Tube Task was solved faster than the Vertical Tube Task. In a previous study applying the same experimental setup, the authors did not find significant differences in latency to success between both tasks (Voigt et al., 2019). Differences in sampling might explain these contrasting findings. Whereas Voigt et al. (2019) tested only five-year-old children of a very small age range (5;0–5;2), we included three- to five-year-old children ranging from 3;0 to 5;11.

In previous research, children were typically given only one to two minutes to solve the task. If this cutoff is set for the current study, we observe success rates of 21.4 % in the Horizontal Tube Task, and 8.5 % in the Vertical Tube Task. These success rates are overall similar to those found in previous research on the Unbending Task as well as the Hook Task (Chappell et al., 2013; Cutting et al., 2011; for exceptions with higher rates see Neldner et al., 2019; Sheridan et al., 2016). However, the analyses of latency to success in our study demonstrates that previous studies might have underestimated children's ability to solve innovative problems. In both tasks, we saw large increases in success rates with time. To measure problem solving in a more ecologically valid way, future studies should provide enough time for children to work out solutions. We suggest a time frame of 10 min. Providing more time could entail the disadvantage of straining both motivation and patience of unsuccessful children and increasing the drop-out rate.

While latencies are a common outcome variable in the animal literature on innovative problem solving (see Griffin & Guez, 2014 for an overview), they have not been reported in earlier studies on childhood innovation (see Hanus et al., 2011; Voigt et al., 2019 for exceptions). Related research on human problem solving showed that three-year-old children needed more time to solve a novel artificial fruit problem than five-year-old children (Flynn, Turner, & Giraldeau, 2016). In the current study, we did not find any significant relationship between the age of the child and latency to success. Given that only children who successfully solved the task were considered for this analysis, the subsample sizes were rather small, especially for the younger children in the Vertical Tube Task

(N = 1 for three-year-olds, N = 8 for four-year-olds). To draw solid conclusions about age effects regarding the time to solve the tube problem, larger samples of successful children are needed.

We would like to point out that the time needed until finding a solution may not only be related to cognitive processes, but also motor skills. In the present study, eight children (17 %) participating in the Vertical Tube Task did in fact manufacture a proper tool or had the correct idea to manufacture it, but failed to apply it, or stopped trying using it when it did not lead to immediate success. Several children also dropped material into the vertical tube while trying to connect to the toy and then needed time to remove this material from the tube before being able to start a new solution attempt, causing longer latencies or even failure: In two cases children were not able to remove the non-functional material from the tube within 10 min. Future studies – especially those applying the Vertical Tube Task – might consider differentiating between correct tool manufacturing and the successful application of a given tool (see Neldner et al., 2017) to examine the impact of fine motor skills on success in tool innovation tasks in more detail.

4.5. The impact of prompts on tool innovation

Although it is common in the innovation literature to give children prompts while they are working on innovation tasks, the possible effects of different prompts on children's behavior have rarely been discussed so far (but see Chappell et al., 2013 for an exception). In most existing studies children were directly prompted during the instruction that "these things may help" or that they "can use these things" to get the reward (Cutting et al., 2011; 2019; Gönül et al., 2018, 2019; Neldner et al., 2017, 2019; Nielsen et al., 2014; Sheridan et al., 2016). Furthermore, most studies provided more prompts while the child was working on the task. Examples are: "Can you think of how you might use these things to get the sticker?" (Cutting et al., 2011; Sheridan et al., 2016), "You can use whatever you want." (Nielsen et al., 2014), "Maybe you could use these things to help you." (Cutting et al., 2011, 2019) or "You can try anything." (Neldner et al., 2017). This handling of prompts is overall similar to the current study. We adapted a procedure from Voigt et al. (2019) that allows children to first figure out on their own that they need the materials to solve the task before a first prompt was provided (i.e., "You can use any of these things."). If necessary, children received more motivational prompts ("Just try to get the toy out of the tube on your own! I am sure you can do it.") and item-related prompts ("You can also use any of these things.", "There are more things. You can use any of them.").

In a post-hoc analysis, we found that children tended to get more motivational prompts in the Vertical Tube Task than in the Horizontal Tube Task, although this task difference was non-significant. This trend is likely to reflect the higher task difficulty in the Vertical Tube Task, leading younger children to get frustrated or to ask for support. However, as the low success rate for the Vertical Tube Task in young children reveals, we found no evidence suggesting that more motivational prompts increased children's performance. With respect to item-related prompts which support children in focusing on the given tool materials, our data shows that only very few prompts of this kind were given in each task. Thus, it seems unlikely that the prompting procedure explains the observed task differences. Interestingly, younger children received more prompts than older ones, but their overall performance was poorer than that of older children. This suggests that prompts do not necessarily lead to better performance. To explore how exactly different kinds of prompts affect the tool innovation process further studies are needed.

4.6. Promising directions for future studies

When planning new studies on tool innovation in children, more attention should be given to the processes leading to success or failure. One way to learn more about the intentions that guide children's tool innovation might be to analyze children's private speech. During coding in the current study, we observed that many children verbally described their actions and thoughts while working on the task. This self-directed form of speech has previously been documented across several cognitive activities, including problem solving, executive function tasks, or planning tasks, and is suggested to support young children's cognitive self-regulation (Benigno, Byrd, McNamara, Berg, & Farrar, 2011; Lidstone, Meins, & Fernyhough, 2011; Winsler, 2009). Private speech could thus reveal more details about the cognitive processes underlying young children's solution attempts. Next to cognitive processes, also motivational aspects of private speech seem to correlate with children's persistence in working on a task (Chiu & Alexander, 2000; Sawyer, 2017) and to interact with emotion regulation strategies (Day & Smith, 2013). Motivational and emotional processes have been largely ignored so far in tool innovation research even though they are certainly relevant for maintaining interest in the task and persisting in finding a solution. Especially those children who do not immediately find a solution have to regulate frustration and other negative emotions in order to continue working on the task.

As another way to put more focus on the process of tool innovation, we took a closer look at the different manufacturing strategies and introduced the differentiation between first- and second-order innovation. Such additional descriptions of children's behaviors can help to identify general developmental trends of tool manufacturing and innovation. However, we would like to point out again that drawing conclusions about children's innovative skills based on their choice of strategies alone seems to be problematic (see also Carr et al., 2016; Voigt et al., 2019). For example, inserting several short items into the horizontal tube may reflect incidental learning, but it may also be based on a complex plan, put into action step by step.

Instead of focusing on the source of innovation (e.g., intentionality vs. chance), Carr et al. (2016) proposed to focus on subsequent learning and the transmission of knowledge to other problems. If children actually learn the innovative behavior – no matter whether it was insightful or accidental – it is more likely that they repeat it and successfully pass on their knowledge to others. This might be called *transmitted tool innovation* compared to behavior that is limited to one situation or the repertoire of a single child. Diffusion chain experiments with children teaching other children may reveal how tool innovation turns into solid tool knowledge and how it is transmitted to others (see e.g., Tennie, Walter, Gampe, Carpenter, & Tomasello, 2014). Corresponding work could help us to better

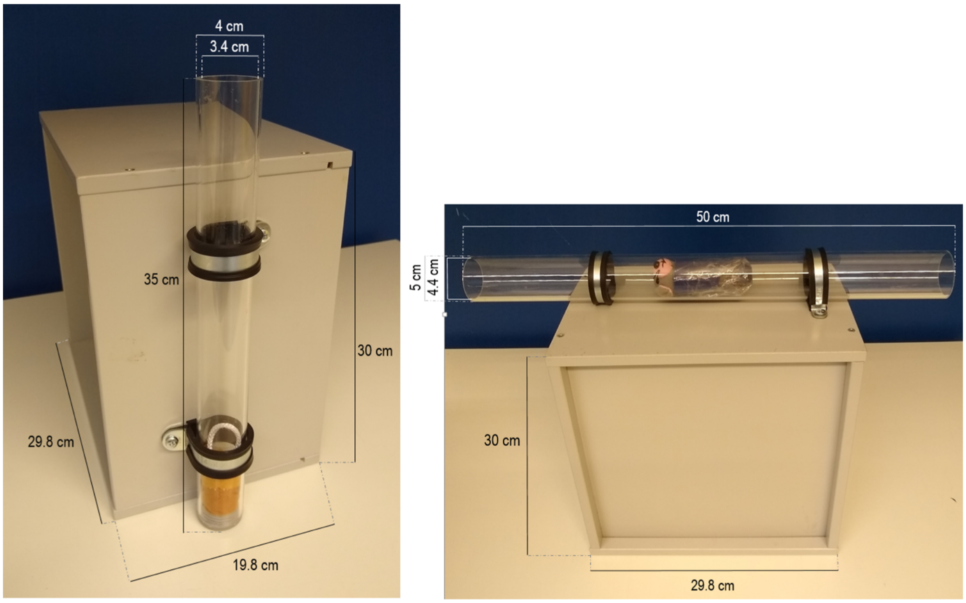


Fig. A1. Measures for the apparatuses of the Horizontal Tube Task (left) and Vertical Tube Task (right).

understand how early in life it is possible to participate in the cultural transmission of tool knowledge.

Furthermore, it may make sense to distinguish between tool innovation that is restricted to a specific problem and tool innovation that leads to a transfer of knowledge to new situations. The latter may be called *transferable tool innovation*. To test the development of transferable tool innovation, multi-trial experiments with different apparatuses or an adapted tool selection should be established. One study along these lines found that five- to six-year-old children were capable of transferring their tool-making knowledge from the original Hook Task to very similar tasks requiring the child to build a pipe cleaner hook (Beck et al., 2014). Younger children were less successful when encountering novel problems requiring similar solutions, probably reflecting a lack of analogical reasoning skills. More studies are needed to promote our understanding of tool innovation transfer in younger children and its possible limitations.

4.7. Conclusions

In this report, we systematically compared tool innovation in three-, four-, and five-year-olds in two different but related tasks (Horizontal and Vertical Tube Task). Whereas past research focused primarily on success rate as a dependent variable, the present report made a first step towards analyzing the processes leading to successful tool innovation by using latency to success as an additional dependent variable, by looking at age-related changes in the strategies that children use to create a functional tool, and by differentiating between first- and second-order tool innovation.

Our data suggest that the two tool innovation tasks have very different task demands, determining not only whether children find a solution but also which solution strategies they apply and how much help they need. The Horizontal Tube Task can be used to analyze the approaches of innovative problem solving across a broad age range, as even young children are able to solve this task. When applying the Vertical Tube Task to probe tool innovation in younger children, it seems important to focus more on children’s solution attempts – including unsuccessful ones – and exploring to what extent children ask for and receive support.

To promote the field of innovation research in children, less attention should be given to the question of whether children succeed or fail in a given task. Instead, future studies should focus more on the underlying cognitive and motivational mechanisms leading to success or failure. More research along these lines will enable us to better understand why and how humans become sophisticated masters of tool innovation and how this development can best be supported in children.

Funding

This study was supported by a stipend from the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes) awarded to the first author.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We thank the children and parents who participated in the study. We also thank Christiane Fauth and the Baby-Uni Heidelberg team for their help in recruiting participants, Anna Schleicher and Theresa Kloss for their help in data collection, and Sandra Schmitt for her assistance in reliability coding. We would also like to thank the anonymous reviewers for their valuable comments on this manuscript.

Appendix A

Table A1

Relative and absolute frequencies of success in both task versions and age groups.

	Age Group	N	Number of Successful Children	Success Rate
Horizontal Tube Task	3 years	14	11	78.6 %
	4 years	14	13	92.9 %
	5 years	14	13	92.9 %
	Total	42	37	88.1 %
Vertical Tube Task	3 years	14	1	7.1 %
	4 years	16	8	50 %
	5 years	17	15	88.2 %
	Total	47	24	51.1 %

Table A2

Binary logistic regression predicting the type of innovation (1 = first-order innovation, 0 = second-order innovation) in the Horizontal Tube Task (N = 37).

						Odds Ratio and 95 % CI		
Predictor	B	SE	Wald	df	*p*	Lower	Odds Ratio	Upper
Age	.016	.033	.226	1	.635	.952	1.016	1.084
Constant	−1.347	1.829	.542	1	.461		.260	

Model: $\chi^2(1) = .227$, $p = .634$; Cox and Snell's $R^2 = .006$, Nagelkerke's $R^2 = .008$.

Table A3

Binary logistic regression predicting the type of innovation (1 = first-order innovation, 0 = second-order innovation) in the Vertical Tube Task (N = 24).

						Odds Ratio and 95 % CI		
Predictor	B	SE	Wald	df	*p*	Lower	Odds Ratio	Upper
Age	.050	.070	.501	1	.497	.916	1.051	1.205
Constant	−1.331	4.121	.104	1	.747		.264	

Model: $\chi^2(1) = .505$, $p > .477$, Cox and Snell's $R^2 = .021$, Nagelkerke's $R^2 = .035$.

Table A4

Percentage of successful children using different manufacturing methods to solve each task.

			Age Group	
		3 years	4 years	5 years
Horizontal Tube Task		(N = 11)	(N = 13)	(N = 13)
	Subtract	–	38.5 %	7.7 %
	Add	9.1 %	30.8 %	23.1 %
	Reshape	–	–	7.7 %
	Pushing Several Materials Into Tube	90.9 %	23.1 %	38.5 %
	Unexpected Modification of Material	–	–	23.1 %
	Using Tool Designed for the Other Tube	–	7.7 %	–
Vertical Tube Task		(N = 1)	(N = 8)	(N = 15)
	Subtract	–	–	40 %
	Add	–	87.5 %	46.7 %
	Reshape	–	–	–
	Combination of Material	100 %	12.5 %	13.3 %

Table A5
Cumulative success rate (in %) per minute in the Horizontal and Vertical Tube Task.

		Minutes									
		1	2	3	4	5	6	7	8	9	10
Horizontal Tube Task	3 years	7.1	28.6	35.7	42.9	50	57.1	**78.6**			
	4 years	21.4	57.1	64.3	78.6	78.6	78.6	78.6	85.7	85.7	**92.9**
	5 years	35.7	71.4	71.4	71.4	85.7	85.7	85.7	85.7	**92.9**	
	Total	*21.4*	*52.4*	*57.1*	*64.3*	*71.4*	*73.8*	*81*	*83.3*	*85.7*	***88.1***
Vertical Tube Task	3 years	0	0	0	0	0	0	0	0	**7.1**	
	4 years	6.3	6.3	12.5	18.8	25	31.1	37.5	**50**		
	5 years	17.6	29.4	41.2	41.2	70.6	70.6	76.5	**88.2**		
	Total	*8.5*	*12.8*	*19.1*	*21.3*	*34*	*36.2*	*40.4*	*48.9*	***51.1***	

Note. Bold numbers indicate the final percentage of children who were successful within 10 min.

Table A6
Linear regression predicting latency to success (in seconds) in the Horizontal Tube Task.

	Unstandardized Coefficients and 95 % CI						
Predictor	Lower	B	Upper	SE	Beta	T	*p*
Age	−8.207	−3.616	.974	2.261	−.261	−1.599	.119
Constant	111.716	364.577	617.437	124.555		2.927	.006

ANOVA: $F(1,35) = 2.557$, $p = .119$, corrected $R^2 = .068$.

Table A7
Linear regression predicting latency to success (in seconds) in the Vertical Tube Task.

	Unstandardized Coefficients and 95 % CI						
Predictor	Lower	B	Upper	SE	Beta	T	*p*
Age	−13.797	−5.439	2.918	4.030	−.277	−1.350	.191
Constant	71.986	580.029	1088.072	244.973		2.368	.027

ANOVA: $F(1,22) = 1.822$, $p = .191$, corrected $R^2 = .076$.

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.cogdev.2021.101049.

References

Bania, A. E., Harris, S., Kinsley, H. R., & Boysen, S. T. (2009). Constructive and deconstructive tool modification by chimpanzees (Pan troglodytes). *Animal Cognition, 12*(1), 85–95. https://doi.org/10.1007/s10071-008-0173-0

Beck, B. B. (1980). *Animal tool behavior.* Garland STPM Publishing.

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., Cutting, N., Apperly, I. A., Demery, Z., Iliffe, L., Rishi, S., et al. (2014). Is tool-making knowledge robust over time and across problems? *Frontiers in Psychology, 5*(1395), 1–10. https://doi.org/10.3389/fpsyg.2014.01395

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 of London. Series B, Biological Sciences, 371*(1690), 20150190. https://doi.org/10.1098/rstb.2015.0190

Benigno, J. P., Byrd, D. L., McNamara, J. P., Berg, W. K., & Farrar, M. J. (2011). Talking through transitions: Microgenetic changes in preschoolers' private speech and executive functioning. *Child Language Teaching and Therapy, 27*(3), 269–285. https://doi.org/10.1177/0265659010394385

Brown, A. L. (1990). Domain-specific principles affecting learning and transfer in children. *Cognitive Science, 14*(1990), 107–133. https://doi.org/10.1207/s15516709cog1401_6

Burkart, J. M., Strasser, A., & Foglia, M. (2009). Trade-offs between social learning and individual innovativeness in common marmosets, Callithrix jacchus. *Animal Behaviour, 77*(5), 1291–1301. https://doi.org/10.1016/j.anbehav.2009.02.006

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2008). Rational tool use and tool choice in human infants and great apes. *Child Development, 79*(3), 609–626. https://doi.org/10.1111/j.1467-8624.2008.01146.x

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What is innovation, how does it develop, and who does it? *Child Development, 87*(5), 1505–1519. https://doi.org/10.1111/cdev.12549

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 of London Series B, Biological Sciences, 368*(1630), 20120409. https://doi.org/10.1098/rstb.2012.0409

Chen, Z., Siegler, R. S., & Daehler, M. W. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development, 65*(2), i–105. https://doi.org/10.4135/9781452286143.n496

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

Connolly, K., & Dalgleish, M. (1989). The emergence of a tool-using skill in infancy. *Developmental Psychology, 25*(6), 894–912. https://doi.org/10.1037/0012-1649.25.6.894

Cutting, N. (2013). *Children's tool making: From innovation to manufacture*. University of Birmingham. http://etheses.bham.ac.uk/3969/1/CUTTING12PhD.pdf.

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*(1), 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. https://doi.org/10.1016/j.cogdev.2019.100811

Day, K. L., & Smith, C. L. (2013). Understanding the role of private speech in children's emotion regulation. *Early Childhood Research Quarterly, 28*(2), 405–414. https://doi.org/10.1016/j.ecresq.2012.10.003

Ebel, S. J., Hanus, D., & Call, J. (2019). How prior experience and task presentation modulate innovation in 6-year-old-children. *Journal of Experimental Child Psychology, 180*, 87–103. https://doi.org/10.1016/j.jecp.2018.12.004

Flynn, E. G., Turner, C., & Giraldeau, L.-A. (2016). Selectivity in social and asocial learning: Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 371*, 20150189. https://doi.org/10.1098/rstb.2015.0189

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gibson, K. R. (1993). Tool use, language and social behavior in relationship to information processing capacities. In R. Gibson, Kathleen, & T. Ingold (Eds.), *Tools, language and cognition in human evolution* (pp. 251–269). Cambridge University Press. https://doi.org/10.1017/S0021937100024837.

Gönül, G., Hohenberger, A., Corballis, M., & Henderson, A. M. E. (2019). Joint and individual tool making in preschoolers: From social to cognitive processes. *Social Development, 28*(4), 1037–1053. https://doi.org/10.1111/sode.12373

Gönül, G., Takmaz, E. K., Hohenberger, A., & Corballis, M. (2018). The cognitive ontogeny of tool making in children: The role of inhibition and hierarchical structuring. *Journal of Experimental Child Psychology, 173*, 222–238. https://doi.org/10.1016/j.jecp.2018.03.017

Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms. *Behavioural Processes, 109*, 121–134. https://doi.org/10.1016/j.beproc.2014.08.027

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). https://doi.org/10.1371/journal.pone.0019555

Jongbloed-Pereboom, M., Nijhuis-van der Sanden, M. W. G., Saraber-Schiphorst, N., Crajé, C., & Steenbergen, B. (2013). Anticipatory action planning increases from 3 to 10 years of age in typically developing children. *Journal of Experimental Child Psychology, 114*(2), 295–305. https://doi.org/10.1016/j.jecp.2012.08.008

Kacelnik, A., Chappell, J., Kenward, B., & Weir, A. A. S. (2006). Cognitive adaptations for tool-related behaviour in New Caledonian Crows. In E. A. Wasserman, & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 515–528). Oxford University Press.

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

Lidstone, J., Meins, E., & Fernyhough, C. (2011). Individual differences in children's private speech: Consistency across tasks, timepoints, and contexts. *Cognitive Development, 26*(3), 203–213. https://doi.org/10.1016/j.cogdev.2011.02.002

McCarty, M. E., Clifton, R. K., & Collard, R. R. (2001). The beginnings of Tool use by infants and toddlers. *Infancy, 2*(2), 233–256. https://doi.org/10.1207/S15327078IN0202_8

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: Orangutans use water as a tool. *Biology Letters, 3*(5), 453–455. https://doi.org/10.1098/rsbl.2007.0198

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

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., et al. (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

Nielsen, M. (2013). Young children's imitative and innovative behaviour on the floating object task. *Infant and Child Development, 22*, 44–52. https://doi.org/10.1002/icd.1765

Nielsen, M., Tomaselli, K., Mushin, I., & Whiten, A. (2014). Exploring tool innovation: A comparison of Western and Bushman children. *Journal of Experimental Child Psychology*, 126. https://doi.org/10.1016/j.jecp.2014.05.008

Ramsey, G., Bastian, M. L., & van Schaik, C. P. (2007). Animal innovation defined and operationalized. *The Behavioral and Brain Sciences, 30*(4), 407–432. https://doi.org/10.1017/s0140525x07002373

Reindl, E., Beck, S. R., Apperly, I. A., & Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B: Biological Sciences, 283*(1825), 20152402. https://doi.org/10.1098/rspb.2015.2402

Reynaud, E., Lesourd, M., Navarro, J., & Osiurak, F. (2016). On the neurocognitive origins of human tool use: A critical review of neuroimaging data. *Neuroscience and Biobehavioral Reviews, 64*, 421–437. https://doi.org/10.1016/j.neubiorev.2016.03.009

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

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 of London Series B, Biological Sciences, 371*, 20150191. https://doi.org/10.1098/rstb.2015.0191

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: The use and manufacture of tools by animals*. JHU Press.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Call, J., & Tomasello, M. (2010). Evidence for emulation in chimpanzees in social settings using the floating peanut task. *PloS One, 5*(5). https://doi.org/10.1371/journal.pone.0010544

Tennie, C., Walter, V., Gampe, A., Carpenter, M., & Tomasello, M. (2014). Limitations to the cultural ratchet effect in young children. *Journal of Experimental Child Psychology, 126*, 152–160. https://doi.org/10.1016/j.jecp.2014.04.006

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition, 105*(2), 362–379. https://doi.org/10.1016/j.cognition.2006.10.003

Träuble, B., & Pauen, S. (2011). Cause or effect: What matters? How 12-month-old infants learn to categorize artifacts. *The British Journal of Developmental Psychology, 29*(3), 357–374. https://doi.org/10.1348/026151009X479547

Vaesen, K. (2012). The cognitive bases of human tool use. *The Behavioral and Brain Sciences, 35*, 203–262. https://doi.org/10.1017/S0140525X11001452

van Schaik, C. P., Ancrenaz, M., Djojoasoro, R., Knott, C. D., Morrogh-Bernard, H. C., Nuzuar Odom, K., Atmoko, S. S. U., & van Noordwijk, M. A. (2009). Orangutan cultures revised. In S. A. Wich, S. S. U. Atmoko, T. M. Setia, & C. P. van Schaik (Eds.), *Orangutans: Geographic variation in behavioral ecology and conservation* (pp. 299–309). OUP Oxford.

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

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

Whalley, C., Cutting, N., & Beck, S. R. (2017). The effect of prior experience on children's tool innovation. *Journal of Experimental Child Psychology, 161*, 81–94. https://doi.org/10.1016/j.jecp.2017.03.009

Whiten, A., & van Schaik, C. P. (2007). The evolution of animal 'cultures' and social intelligence. *Philosophical Transactions of the Royal Society B, 362*, 603–620. https://doi.org/10.1098/rstb.2006.1998

Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3–41). https://doi.org/10.1017/CBO9780511581533.003. Issue 1992.

Zelazo, D. P., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., et al. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3), 1–151.

II. Manuscript

Breyel, S., & Pauen, S. (2022a). *Private speech during problem-solving: Tool innovation challenges both preschoolers' cognitive and emotion regulation.* Manuscript submitted for publication in Journal of Cognition and Development.

Private Speech during Problem-Solving: Tool Innovation Challenges both Preschoolers' Cognitive and Emotion Regulation

Sabine Breyel* and Sabina Pauen

Department of Developmental and Biological Psychology, Heidelberg University

We have no conflicts of interest to disclose.

Correspondence details for the *corresponding author:

Sabine Breyel

Department of Developmental and Biological Psychology, Heidelberg University

Hauptstraße 47-51, 69117 Heidelberg, Germany

sabine.breyel@psychologie.uni-heidelberg.de

Private Speech during Problem-Solving: Tool Innovation Challenges both Preschoolers' Cognitive and Emotion Regulation

Abstract

The current study examined children's spontaneous private speech during the Vertical and the Horizontal Tube Task to shed light onto the cognitive, motivational, and emotional processes underlying tool innovation. Relations between private speech of N = 65 three- to five-year-old children and their task performance (i.e., success and latency to success) were analyzed using Bayesian statistics. Metacognitive and Cognitive Speech showed only weak associations with task success, but Cognitive Speech correlated positively with latency to success. The strongest evidence was found for Negative Speech: Children who expressed negative emotions more often and who evaluated the task as being very difficult needed more time to find a solution than children who used less Negative Speech of this kind. These findings indicate that cognitive skills and emotion regulation are closely related in preschoolers' tool innovation.

133 words

Keywords*:* private speech; tool innovation; cognitive development; emotion regulation; preschool age

Introduction

When children work on a challenging task, they often talk to themselves – a behavior referred to as "private speech" in the psychological literature (Winsler, 2009). Private speech occurs most frequently during preschool years and decreases again around age five to six years (Winsler & Naglieri, 2003). About 60 years ago, Vygotsky (1962) suggested that private speech provides an intermediate step between solving a given problem in a social dialogue with another person and solving the problem inside the head, thus promoting self-regulation in young children. In the meantime private speech has been documented across several cognitive activities, including memorizing information, shifting perspectives, inhibiting impulses, planning ahead, constructing, and categorizing (e.g., Benigno et al., 2011; Lidstone et al., 2011). The current study focuses on private speech in tool innovation as an advanced form of problem-solving.

In a tool innovation task children are asked to retrieve a target from an apparatus. For this purpose they need to create a functional tool from a selection of given materials. Current research shows that even preschool children can meet this challenge under certain conditions (Breyel & Pauen, 2021; Sheridan et al., 2016; Voigt et al., 2019). Previous studies primarily focused on the outcome (i.e., on whether children solved the task or not), thus leaving open the question of which mental processes contribute to success or failure. By analyzing children's private speech while working on a tool innovation task, we hope to shed more light on this issue. Before introducing the rationale of our study in more detail, we provide a brief summary of existing evidence on the development of tool innovation and private speech as well as the relation between private speech and performance in cognitive tasks.

Tool innovation in young children

Children begin to understand form-function relations in infancy (e.g., Träuble & Pauen, 2011) and they start to use tools to reach a given goal as toddlers (Brown, 1990; Reindl et al., 2016). *Making* a new tool to solve a problem requires more than that, however. For tool making children first need to imagine how an appropriate tool should look like (tool innovation) and then physically transform the given materials into this type of tool (tool manufacture, see Beck et al., 2011).

For a long time young children were thought of a being rather poor tool innovators (Beck et al., 2011; Cutting et al., 2019; Hanus et al., 2011; Nielsen, 2013). However, recent research has identified several factors that can either enhance or diminish young children's performance. For example, Sheridan and colleagues (2016) demonstrated that 44% of a sample of four- to five-year-olds solved a tool innovation task in a children's museum children, whereas other studies in the lab found much lower success rates (0–8%) for the same age-range (e.g., Beck et al., 2011; Chappell et al., 2013; Cutting et al., 2011). Evidence provided by Voigt et al. (2019) suggests that tool innovation depends on the time children are allowed to work on a given task and on the materials available to build a tool. While none of a group of five-year-olds solved a corresponding task within one minute, 21% of the children managed to build a proper tool within 10 minutes. When multiple materials were available to build a functional tool, even more (71–93%) children solved the task within 10 minutes. Further evidence by Breyel & Pauen (2021) suggests that tool innovation performance is task dependent. In their study three- to five-year-olds were asked to get a toy out of the center of a long transparent horizontal tube (Horizontal Task) or from the bottom of a long vertical tube (Vertical Task). To solve these tasks, children needed to build a tool from a set of different materials. Children of all age groups were highly competent in building a tool that was long enough to push the target out in the

Horizontal Task (average success rate 88%). However, when being asked to build a tool with a hook in order to lift up the target in the Vertical Task, only 7% of the three-year olds and 50% of the four-year-olds were successful within 10 minutes, but 88% of the five-year-olds succeeded.

In sum these findings suggest that young children's tool innovation performance varies with characteristics like the task environment, the type of apparatus, the availability of different materials for building a tool, the time offered to find a solution, as well as the age of the child. However, only little is known so far about the role of cognitive, motivational, and emotional processes that contribute to success or failure in tool innovation tasks.

Private speech in young children

According to Vygotsky's theory (1962, 1978) private speech is a precursor of inner speech which serves to guide one's thoughts and actions. First, children acquire language and become engaged in social talk (e.g., with their caregiver). In an intermediate step, they start talking overtly to themselves – thus revealing private speech. Finally, they begin to internalize this dialogue and becoming capable of silent inner speech.

In empirical investigations private speech utterances have been categorized according to their function (e.g., Diaz, 1992; Feigenbaum, 1992), their form (e.g., Winsler et al., 2003), their content (e.g., Chiu & Alexander, 2000; Krafft & Berk, 1998; Sawyer, 2017), and the level of internalization (Berk, 1986). The function of speech refers to the potential effects of the specific utterance for subsequent behavior (e.g., directing attention, regulating). The form of speech comprises structural aspects of language (e.g., question vs. statement, complete vs. fragmented). Content-related classifications focus on what the child is saying (e.g., reasoning vs. motivational statements) and are often intertwined with functional aspects. Lastly, the levels of speech internalization represent the assumed developmental stages from overt forms of speech to silent inner speech. Berk (1986) described this development with tree different levels of internalization: At level one, children show overt task-irrelevant speech. At level two, they produce task-relevant externalized speech. At level three, they reveal manifestations of inner speech (i.e., whispering, muttering, or lip movements without a clear expression of words) that could be referred to as partially internalized speech.

Sawyer (2017) introduced a coding scheme that incorporates the categorization based on both semantic content/function and internalization. At the content/function level, the author distinguishes between speech with cognitive, metacognitive, motivational, and playful content. At the internalization level, he focuses on partially internalized speech. In addition this scheme includes a valence dimension adapted from the self-talk literature (see also Winsler, 2009) to account for the specific effects of positive and negative statements on children's motivation and performance.

Existing work on private speech in young children has mainly focused on children's behavioral and cognitive self-regulation (for an overview see Winsler, 2009). Even though some researchers also investigated its role for emotion regulation (e.g., Day & Smith, 2013; Whedon et al., 2021) and task-oriented motivation (e.g., Atencio & Montero, 2009), studies that analyze the role of cognitive, emotional, and motivational self-regulation in parallel are still rare. Such an integrative analysis of private speech would be most beneficial for understanding the complex processes involved when children show tool innovation.

Private speech and problem-solving in young children

Previous studies used different tasks to elicit private speech in children and to investigate its relation to task performance and regulation skills. Mostly children were asked to perform planning tasks (Tower of London, e.g., Al-Namlah et al., 2012; Benigno et al., 2011), speech-action coordination tasks (Hammer Task, e.g., Aro et al., 2015; Winsler et al., 2007), hand-eye coordination tasks (Fishing Task, e.g., Chiu & Alexander, 2000; Sawyer, 2017), or tasks that require categorization and other executive functions like inhibition and working memory (e.g., Alarcón-Rubio et al., 2014; Thibodeaux et al., 2019). The majority of these studies examined the level of internalization of private speech. These findings as well as the few studies focusing on fully externalized private speech content (cognitive, metacognitive, motivational, and playful) and its relation to task performance will be summarized next.

Cognitive speech and performance

Manning et al. (1994) demonstrated that young children being rated as more academically advanced by their preschool teachers produced more cognitive private speech during classroom activities. Cognitive speech included focusing attention, describing actions, questioning and self-directing. However, in experimental tasks like the Jumping Task (Chiu & Alexander, 2000) or Fishing Task (Chiu & Alexander, 2000; Sawyer, 2017), children's cognitive speech was neither associated to their performance nor their persistence to solve the task. In the Jumping Task children were asked to jump as far as they could and it was measured how much time children spent on the task. In the Fishing Task children were asked to catch plastic or paper fish from an imaginary pond on the floor by using a magnetized rod. Some fish were difficult (Chiu & Alexander, 2000) or impossible to catch (Sawyer, 2017) due to their unique weight distribution. This difficult or insoluble part of the task was introduced to measure children's persistence. The remaining number of caught fish served as a measure for performance. Since these tasks mainly require motor skills and coordination, cognitive speech might play a minor role for being successful. How cognitive speech relates to performance in more cognitively challenging tasks (e.g., tool innovation) remains to be tested, however.

Metacognitive speech and performance

In the Fishing Task used by Sawyer (2017), a higher absolute and relative amount of metacognitive speech was positively correlated with three- to five-year-old's performance, but not with their persistence. In contrast Chiu and Alexander (2000) found that children who revealed a higher proportion of metacognitive speech were more persistent in solving their Fishing Task and Jumping Task. A higher proportion of metacognitive speech was also correlated to increased willingness to complete a challenging puzzle without help from an adult in a study by Manning et al. (1994). The authors measured children's private speech during normal classroom activities and demonstrated that children who learned more autonomously (i.e., focus on tasks, needing little or no help) produced more metacognitive speech, including corrections, coping and reinforcing utterances. Together these studies suggest that metacognitive speech is positively linked to task outcomes while the specific associations with performance and persistence vary with task demands.

Motivational speech and performance

Existing evidence regarding motivational speech still seems inconclusive. On the one hand, de Dios and Montero (2003) showed that five-year-olds who persisted longer in a medium difficult Puzzle Task used a higher proportion of motivational private speech and Sawyer (2017) found that three- to five-year-olds who persisted longer and performed better in the Fishing Task used more positive motivational utterances. On the other hand, the same study revealed that a higher proportion of motivational speech in relation to other forms of private speech was related to lower overall performance, thus indicating that talking to oneself can also serve to compensate for low task-oriented motivation (Sawyer, 2017). Hence, more evidence is needed before we can draw any conclusion regarding the role of motivational speech for young children’s performance in challenging tasks.

Playful speech and performance

To our knowledge only one study has directly examined playful private speech so far. Sawyer (2017) found that children using more playful speech during the Fishing Task were more persistent in catching the impossible fish, but no corresponding relation was found regarding the number of possible fish caught. Thus, playful speech seemed to affect children’s persistence but not performance.

Partially internalized speech and performance

In three studies partially internalized speech was found to be unrelated to three- to eight-year-old children’s performance in cognitive or motor tasks like the Fishing Task (Sawyer, 2017), executive function tasks like the Dimensional Change Card Sort (Thibodeaux et al., 2019), or performance in IQ tests and classroom assessments (Berk, 1986). In two other studies with five- to seven-year-olds more partially internalized speech was associated with better planning skills (Tower of London Task, Fernyhough & Fradley, 2005) and better executive shifting abilities (Dimensional Change Card Sort, Alarcón-Rubio et al., 2014). Furthermore, Winsler and Naglieri (2003) found that children’s cognitive achievement (Woodcock-Johnson-Revised measures math, spelling, and reading) moderated the effect of partially internalized speech on performance in five- to seven-year-olds. Low-achievers performed better in a Trail-Making Task when using partially internalized speech than when remaining silent. For high-achieving children, no such effect was found. Results on the association between performance on cognitive tasks and the use of partially internalized speech are thus inconclusive so far and evidence referring to younger preschool children is still sparse.

In sum, only few studies focused on the association between private speech content and task performance in young children. To our knowledge no study has examined children's private speech during open-ended problem-solving as in tool innovation tasks, i.e., when the solution or the means to achieve it are still unknown. When working on a corresponding task, children first need to represent the target state. They also need to realize that this state can only be reached with the help of a tool. Finding the right means to produce a tool requires them to activate knowledge about the materials available and to generate ideas about how to combine or manipulate them to manufacture a suitable tool. In case the chosen approach proves to be inappropriate to achieve a given goal, children need to evaluate their strategy and to decide whether to keep on trying or to find an alternative approach. During this process they may experience feelings of frustration and disappointment. Hence, tool innovation requires self-regulation at multiple levels, including the regulation of cognitive, motivational, and emotional states. These regulation processes are likely to be reflected in children's private speech.

The current study

Our main intention is to provide first data on how three- to five-year-old children try to solve tool innovation tasks by analyzing their private speech. More specifically, we investigate how different private speech categories are related to success rate and latency to success in two versions of the Tube Task (Horizontal vs. Vertical) in three- to five-year-old children, using a recently introduced task version (Voigt et al., 2019).

Based on the existing evidence (Chiu & Alexander, 2000; Manning et al., 1994; Sawyer, 2017) as well as theoretical considerations, we expect a positive relation between cognitive speech, metacognitive speech and task outcomes (i.e., success rate, latency to success) in the cognitively challenging tool innovation task, as these types of speech might help young children to analyze the problem, to plan their goal-directed actions, to monitor errors, and to reflect on progress in finding a solution. With respect to other categories of private speech (i.e., motivational, playful, partially internalized speech), our analyses are explorative since no directed hypotheses regarding their effect on problem-solving can be derived from the existing literature.

Method

Participants

The current study analyzes the private speech of children who took part in a tool innovation study conducted by Breyel and Pauen (2021). This study was conducted in a medium-sized town in southwestern Germany. Mothers were initially contacted in birth clinics. Once the child reached the appropriate age, families were contacted via an invitation letter to explain the study rationale. Written consent for participation was provided before the start of the testing.

For the purpose of the present study, the existing video recordings were transcribed in order to analyze private speech. From the original sample of N = 89 children, N = 16 children remained silent throughout the testing session (N = 3 three-year-olds, N = 6 four-year-olds, N = 7 five-year-olds) and N = 8 children produced social speech only (N = 4 three-year-olds, N = 2 four-year-olds, N = 2 five-year-olds). Since we were specifically interested in private speech, these N = 24 children were excluded from the following analyses, resulting in a final sample of N = 21 three-year-olds (9 girls, M = 3;6, range = 3;0 – 3;11), N = 22 four-year-olds (11 girls, M = 4;4, range = 4;0 – 4;11), and N = 22 five-year-olds (11 girls, M = 5;5, range = 5;0 – 5;11). Half of the sample participated in the Horizontal Task (N = 33) and the other half worked on the Vertical Task (N = 32). All children were full-term born and typically developed. They were of European descent and had an upper-middle-class socioeconomic family background, with most parents reporting to have a higher education (equivalent to a college degree or higher: 81.5% of mothers, 78.5% of fathers). Only a small percentage of parents had either a middle (7.7% of mothers, 6.2% of fathers) or a lower school education (1.5% of mothers, 4.6% of fathers), and corresponding data from N = 6 (9.2%) of the mothers and N = 7 (10.1%) of the fathers were missing.

Material and Procedure

The materials and the procedure of the two versions of the Tube Task are described in detail in Breyel and Pauen (2021). While each child was tested individually in a quiet room in the lab, the parent filled out questionnaires and was instructed not to look at what the child was doing. If the child asked for help, parents were instructed to say: "I am sure you can do this, you can try out anything you want."

Following a short warm-up game (i.e., blowing balls of cotton wool into a goal), children were allowed to explore six item sets (see Figure 1). These were later presented to solve the innovation task. For each task three item sets could be rendered functional tools by (a) combining two parts, (b) by subtracting one part from another, or (c) by changing the shape of a given item. The other three sets served as distractors, respectively. Following the exploration phase the experimenter presented an empty familiarization tube to highlight its affordance (i.e., two openings for the horizontal tube, one opening on top for the vertical tube). Next, the test tube with a cuddle toy stuck inside (horizontal tube: in the middle; vertical tube: at the bottom) was placed within reach of the child. The six previously familiarized item sets were presented next to it. In the Horizontal Task, children needed a tool that was long enough to push the toy out of one of the two open sides. In the Vertical Task, children needed a tool that was long enough and had a hook to pull up the toy (see Figure 1).

Each child worked on only one of these two task versions. The experimenter said: “Can you see the toy inside the tube? Let’s get it out to have a closer look at it! I must do some work now, but you can stay here and try to get the toy out of the tube. I am sure you can do it! You can try out anything you like.” While the child was working on the task, the experimenter sat down in a corner of the room, pretending to be busy writing. If the child did not start or did not continue working on the task for at least two minutes, the experimenter repeated the initial instruction. If the child did not use any of the provided materials or remained unsuccessfully while trying to use one specific item of the material set for at least three minutes, the experimenter guided the children’s attention to the other items by saying: “You can also use any of these things to get the toy out of the tube. You can do it any way you like.” If the child failed to solve the task within the maximum time of 10 minutes, the observation stopped and the experimenter assisted the child in retrieving the toy in order to avoid frustration.

Coding

Private Speech

Children’s speech was transcribed from video recordings. Following suggestions by Winsler et al. (2005), a string of speech was coded as a discrete utterance if it was separated by at least two seconds from the next verbalization. An utterance was classified as private speech if it was not addressed to any person in the room, as indicated by physical or visual contact, using personal pronouns, or by calling the person’s name (Winsler et al., 2005). If an utterance was addressed to someone else, it was coded as social speech and not considered for further analyses.

Private speech utterances were assigned to a subcategory of five main domains as introduced by Sawyer (2017): Cognitive Speech, Metacognitive Speech, Motivational Speech, Playful Speech, and Partially Internalized Speech (see Table 1 for examples). We adapted the corresponding coding scheme by adding the subcategory “Expression of New Ideas” to the Cognitive Speech category, as we were also interested in the role of private speech in coming up with new ideas for tool innovation. This adapted coding scheme allowed us to analyze semantic content as well as the valence of speech. Frequencies (utterances per minute) and proportions (percentage of total utterances) were calculated for each (sub)category.

One primary rater coded all transcripts and a second rater coded 25% of the transcripts independently. The raters agreed substantially according to Landis and Koch (1977) regarding the five main categories ($\kappa = .768$) as well as the subcategories ($\kappa = .687$).

Performance in the Tool Innovation Task

Children’s performance was videotaped and coded offline. The binary outcome measure was whether children successfully managed to retrieve the toy within 10 minutes by using the material provided or not. A second rater coded 50% of the videos and both raters agreed in all cases regarding children’s success. For successful children latency to success was defined as the time in minutes elapsing between the presentation of the task and the removal of the toy from the tube. The intraclass correlation coefficient for latency to success was excellent (two-way mixed, absolute agreement, single measures: ICC = .999, 95 % CI = .9993–.9998).

Analysis

We used JASP 0.14.1 for statistical analyses and applied Bayesian statistics to analyze the data. Since none of the speech variables was normally distributed, we used non-parametric tests. Bayes factors (BF_{10}) were interpreted following the guidelines proposed by Goss-Sampson (2020; see Table S.1). For all tests the null hypothesis postulates that there is no difference in speech variables (frequency and proportion) between the groups ($H_0 : \delta = 0$) and the two-sided alternative hypothesis states that the groups differ in terms of speech production ($H_1 : \delta \neq 0$). In case of directed hypotheses regarding the task performance, the one-sided alternative hypothesis states that unsuccessful children have lower values than successful children.

First we report preliminary analyses on the associations between gender, age, task version and children's private speech. For gender and task differences, we conducted Mann-Whitney tests and to examine age-related changes we used correlational analyses (Kendall's τ). For the main analyses we conducted Mann-Whitney tests to examine our hypotheses about differences between successful and unsuccessful children regarding utterances per min and proportions of speech. To minimize the variation of Bayes factors for the Mann-Whitney test, we increased the number of iterations to 10,000 (see also Goss-Sampson, 2020). To examine relations between private speech and latency until success, we used correlational analyses (Kendall's τ). When there was at least moderate evidence in favor of the alternative hypothesis regarding the main speech category ($BF_{10} \geq 3$), we conducted follow-up tests on the respective subcategories and report the posterior distribution of the standardized effect size δ (i.e., the population version of Cohen's d). Detailed test results and the dataset are available in the supplementary material.

Results

Preliminary Analyses: Gender, Task Version and Age

Regarding gender differences, Bayes factors indicated anecdotal to moderate evidence in favor of H_0 for the number of speech utterances per min (BF_{10} = 0.27-0.72) and the proportion of utterances (BF_{10} = 0.27-0.80). Thus, the data suggest that boys and girls did not differ in terms of their speech behavior. Regarding the two task versions Bayes factors indicated anecdotal to moderate evidence in favor of H_0 (BF_{10} = 0.29–0.53), except for the frequency (BF_{10} = 1.92) and the proportion of Motivational Speech (BF_{10} = 8.05, median δ = -0.65, 95% CI: [-1.16,-0.15]) as well as the frequency (BF_{10} = 8.71, median δ = -0.66, 95% CI: [-1.18,-0.16]) and the proportion of Negative Speech (BF_{10} = 16.60, median δ = -0.73, 95% CI: [-1.25,-0.22]). These results suggest that children expressed more negative utterances in the Vertical Task than the Horizontal Task. On the subcategory level this task difference was found for Negative Emotional Reactions (speech per min: BF_{10} = 3.14, median δ = -0.56, 95% CI: [-1.09,-0.06]; proportion of speech: BF_{10} = 4.30, median δ = -0.61, 95% CI: [-1.14,-0.11]), but neither for Negative Self Evaluation (BF_{10} = 0.45–0.46) nor for Negative Task Evaluation (BF_{10} = 0.36–0.38).

The analyses revealed anecdotal evidence for a positive correlation between age in months and Cognitive Speech per minute (τ = .17, BF_{10} = 1.25) and Metacognitive Speech per minute (τ = .20, BF_{10} = 2.62). For the remaining categories Bayes factors indicated moderate evidence in favor of the hypothesis that speech frequency and age were not related (τ = -.12 – .07, BF_{10} = 0.17–0.25), except for an inconclusive result for Partially Internalized Speech (τ = -.16, BF_{10} = 1.00). Regarding the proportion of speech, there was strong evidence in favor of a negative correlation between Partially Internalized Speech and age (τ = -.26, BF_{10} = 19.25). Among the subcategories of Partially Internalized Speech the negative correlation with age revealed very strong evidence for Extremely Low Volume (τ = -.28, BF_{10} = 39.57) but only anecdotal evidence for Grammatically Fragmented Speech (τ = -.18, BF_{10} = 1.36). For Semantically Condensed Speech the evidence spoke anecdotally in favor of the H_0 (τ = -.12, BF_{10} = 0.41). For the remaining categories data were more likely to match the hypothesis that speech was unrelated to age (τ = -.08–.16, BF_{10} = 0.16–0.83).

Main analyses

We first report success rates for each of the three age groups and both task versions separately, followed by the speech analyses results regarding task performance. As indicated by Table 2, the Horizontal Task was solved by the majority of children from all three age groups whereas success rates increased substantially with age in the Vertical Task. These results based on our subsample of privately speaking children are largely in line with the findings of the complete sample reported by Breyel & Pauen (2021).

Private Speech and Success

Utterances per Minute. We found that successful children used more Cognitive Speech (*Mdn* = 1.14, *IQR* = 2.29, W = 243.50, BF_{10} = 2.03, median δ = -0.44, 95% CI: [-0.97,-0.05]) and more Metacognitive Speech per min (*Mdn* = 0.58, *IQR* = 1.34, W = 277.00, BF_{10} = 1.23, median δ = -0.37, 95% CI: [-0.90,-0.03]) than unsuccessful children ($Mdn_{\text{cognitive/min}}$ = 0.50, *IQR* = 0.95, $Mdn_{\text{metacognitive/min}}$ = 0.10, IQR = 0.55). However, Bayes factors indicated only anecdotal evidence in support for this assumption. For the remaining private speech categories (i.e., Motivational, Playful, Partially Internalized Speech) we found anecdotal to moderate evidence for the null hypothesis (BF_{10} = 0.29–0.57, see Figure 2).

Proportion of Utterances. Analyses revealed anecdotal evidence for the hypothesis that successful children used a higher proportion of Cognitive Speech (*Mdn* = 33.33, *IQR* = 29.46) than unsuccessful children (*Mdn* = 20.00, *IQR* = 33.33, W = 266.50, BF_{10} = 1.75, median δ = -0.42, 95% CI: [-0.93,-0.04]). For the other categories, including Metacognitive Speech (BF_{10} = 0.67), there was anecdotal evidence for H_0 (no differences in success rates, BF_{10} = 0.39–0.73).

Private Speech and Latency to Success

Only successful children who also produced utterances while solving the tool innovation task were included in this part of the analysis (N = 48). Overall, latency to success ranged from 0.25 min to 9.18 min (*M* = 3.57, *SD* = 2.57), thus revealing high interindividual variability.

Utterances per Minute. The number of Cognitive Speech utterances per min was negatively related to the time children needed to solve the task, τ = -.30, BF_{10} = 27.37. On the subcategory level of Cognitive Speech no relations with latency were found (τ = -.09-.06, BF_{10} = 0.13-0.47). In contrast to our initial expectation, the data did not speak in favour of a correlation between Metacognitive Speech and latency to success, τ = -.14, BF_{10} = 0.84. For the remaining speech categories moderate evidence for the null hypothesis was found (no association between latency to success and speech, τ = -.09–.09, BF_{10} = 0.22-0.27), with one exception: Interestingly, the more Negative Speech children produced, the more time they needed to solve the task, τ = .32, BF_{10} = 26.49. The subcategories Negative Emotional Reactions (τ = .27, BF_{10} = 5.88) and Negative Task Evaluations (τ = .34, BF_{10} = 51.37) were driving this effect, whereas Negative Self Evaluations were not associated with latency to success (τ = .15, BF_{10} = 0.61).

Proportion of Utterances. Unlike in the case of utterances per minute, the Bayes factor for the relation between the proportion of Cognitive Speech and latency to success was inconclusive, τ = -.15, BF_{10} = 1.03. Similar to the results regarding utterances per minute, no association was evident between Metacognitive Speech and latency to success, τ = -.01, BF_{10} = 0.21. We found moderate evidence for a correlation between the proportion of Motivational Speech and the time children needed to solve the task (τ = .26, BF_{10} = 4.78). The driving force in Motivational Speech was again Negative Speech which was positively correlated to latency to success (τ = .44, BF_{10} = 3044.91, see Figure 3). On the subcategory level we found very strong evidence for a correlation between latency to success and Negative Emotional Reactions (τ = .34, BF_{10} = 61.52) and decisive evidence for the correlation between latency to success and Negative Task Evaluations (τ = .38, BF_{10} = 246.09). Furthermore, anecdotal evidence pointed to a relation between the proportion of utterances regarding Agentic Desire (i.e., the wish to reach the goal) and latency to success, τ = .23, BF_{10} = 2.26.

For the remaining subcategories of Motivational Speech as well as for Playful Speech and Partially Internalized Speech there was anecdotal or moderate evidence for the null hypothesis (no association between latency to success and speech, τ = -.08–.16, BF_{10} = 0.20-0.65).

Summary of Results

In line with our expectations we found that Cognitive and Metacognitive Speech frequency were associated with better problem-solving performance in terms of success rates among preschool children. It should be noted, though, that the evidence was only anecdotal and that effect sizes were rather small. For all other private speech categories anecdotal or moderate evidence spoke against the assumption of any systematic relation with success rate, indicating that only Cognitive and Metacognitive Speech seem to be relevant when it comes to predicting success rates in our tool innovation tasks. Even though successful children who used Cognitive Speech more often were also faster in solving the given task, they did not reveal a greater proportion of cognitive speech utterances in total private speech. Contrary to our hypotheses no evidence for a systematic relation between Metacognitive Speech and latency to success was found. At the same time analyses revealed that successful children who showed larger proportions of private speech in terms of complaining about high task difficulty, expressing negative emotions, and the desire to get the target needed more time to find the solution than children revealing lower proportions of these speech categories (and vice versa).

Discussion

This study is the first to explore young children’s private speech in a tool innovation task. We tested three- to five-year-olds with one of two versions of a tube task (Vertical Task, Horizontal Task) and categorized their utterances according to an adapted scheme originally introduced by

Sawyer (2017). In the following paragraphs we first discuss our preliminary results regarding age-related changes and task-related differences in private speech production before focusing on the main results on the associations between private speech and task performance.

Private Speech and Age

Differing from previous findings (Berk, 1986; Winsler, 2009), we found no evidence for an increase in Partially Internalized Speech with age but rather a slight decrease of whispering and muttering in older children. In the literature Partially Internalized Speech is mostly operationalized according to Berk (1986) as manifestations of inner speech, such as muttering or lip movements without a clear expression of words. Since we did not code silent lip movements in our study, but rather focused on vocal expressions only, we might have underestimated the use of Partially Internalized Speech in older children.

Furthermore we found anecdotal evidence suggesting that Cognitive and Metacognitive Speech increase with age. Previous private speech research did not find any significant positive correlation between age and Cognitive and Metacognitive Speech (Chiu & Alexander, 2000; Sawyer, 2017). However, in these previous studies private speech was measured while children worked on tasks mainly requiring motor skills and coordination (i.e., Jumping Task and Fishing Task). In contrast, our innovation task puts more demands on (meta)cognitive skills as it requires children to think about an open-ended problem. With age children become better in cognitive and metacognitive skills like the inhibition of irrelevant information or actions, the flexible shifting between thoughts or actions, working memory (Garon et al., 2008; Zelazo et al., 2003), and action planning (Jongbloed-Pereboom et al., 2013; Kaller et al., 2008). These age-related improvements might become more visible in children’s overt speech when the task at hand requires the respective skills to a higher extent. Regarding other types of fully externalized private speech, we found no systematic relations with age between three and five years of age, suggesting that the increase observed for Cognitive and Metacognitive Speech is unlikely to reflect a general tendency of older children to talk more while working on the given task.

Private Speech and Task Difficulty

Several studies suggest that more private speech occurs during moderately difficult tasks and less private speech occurs during easy or very difficult tasks (e.g., Behrend et al., 1989; Fernyhough & Fradley, 2005). Breyel and Pauen (2021) showed that the Horizontal Task was very easy to solve for most three- to five-year-olds, but that the Vertical Task was more difficult for younger than older children. This pattern proved to be the same in the subsample of privately talking children analyzed in the current study. Breyel and Pauen (2021) suggested that preschoolers participating in the Vertical Task need to plan ahead more than in the Horizontal Task because the Vertical Task requires them to first approach the target (i.e., lowering the tool with a hook towards the target) and then reverse their action to be successful (i.e., pull up the target). This two-step action plan is more complex than simply pushing a target in one direction until it falls out of the tube. However, these assumed task-dependent planning demands were not reflected by children's overt speech in the current study. Rather than showing more Cognitive and Metacognitive Speech in the Vertical than the Horizontal Task, children produced more Negative Emotional Reactions both in absolute and relative terms in the Vertical Task. Interestingly, the frequency and proportion of other subcategories of Negative Speech like Negative Self Evaluations and Negative Task Evaluations did not differ between the two task versions. If a task is difficult and children do not find a solution, one would expect them to express their incompetence ("I can't do it.") or to mention task difficulty ("It's too difficult.") as a potential cause for failure. However, most preschoolers in our sample expressed their negative feeling in a rather unspecific way (e.g., “Oh no!”, “What a bummer”) without making any causal attribution. Research on the development of attributions indicates that children do not differentiate consistently between causes of success and failure (i.e., task difficulty, effort, ability, luck) until primary school age (Frieze, 1981). This may suggest that most of the preschoolers tested in the current study were not yet capable of localizing the origin of their failed attempts, increasing the difficulty to evaluate their failure cognitively or to regulate associated negative emotions effectively. In any case, our findings suggest that higher task difficulty increases the risk that young children experience negative feelings because they fail to find the solution quickly, thus calling for regulation on the emotional level.

Private Speech and Task Performance

In line with our predictions successful children used more Cognitive Speech (in terms of frequency and proportion) than unsuccessful children, but group differences were rather small. A higher frequency of Cognitive Speech was also associated with shorter time needed to find a solution among successful children. Previous research did not find similar associations with task success (Chiu & Alexander, 2000; Sawyer, 2017). However, the corresponding tasks mainly required motor skills and hand-eye-coordination whereas the present tool innovation tasks also required open-ended problem-solving, planning and creativity. Thus, cognitive speech – involving thinking about the problem, planning, verbalizing a strategy, and producing new ideas – might play a more important role when solving tool innovation tasks compared to other tasks used to elicit private speech.

The frequency (not the proportion) of metacognitive utterances was slightly higher among successful than unsuccessful children in our sample. These results are in line with findings by Sawyer (2017) who demonstrated that Metacognitive Speech was positively correlated with three- to five-year-old's performance in the Fishing Task. Surprisingly, Metacognitive Speech was unrelated to latency to success in the current study. In a previous study by Chiu and Alexander (2000) children using a higher proportion of Metacognitive Speech were more persistent in gaining better results. It should be noted, though, that this persistence variable was independent of the task outcome whereas latency to success in the current study excluded unsuccessful children, thus disregarding the time they persisted in working on the task. For future studies it might be interesting to include a persistence variable in the innovation task to examine children's intrinsic motivation and its relation to (metacognitive) private speech.

Children who are engaged in problem-solving not only need to regulate their thoughts, but also their emotional and motivational states. Recognizing that an item is not suitable for solving the problem at hand may elicit feelings of irritation and frustration, eventually leading to a loss of motivation. To be successful in problem-solving (especially in a tool innovation task), children need to deal with failed attempts and to remain focused on the given task. In contrast to a previous study suggesting that children persisted longer and performed better in a Fishing task when they produced more positive speech (Sawyer, 2017), we did not find any relation between the expression of positive emotions and performance in our tool innovation task. Rather, the expression of negative emotions and negative evaluations of the task were associated with longer latency to success. Due to the correlational nature of our analysis, we cannot draw any conclusions about the causal direction of this relation. Either children using more Negative Speech needed more time to figure out a solution, or children who needed longer to find a solution expressed negative feelings more often. Since we found no difference in Negative Speech between successful and unsuccessful children, we can at least conclude that using more Negative Speech did not reduce the likelihood of finding a solution.

Nonetheless we should keep in mind that negative emotions can impede children's general performance. Day and Smith (2013) found that the amount of Negative Speech moderated the relation between emotion regulation strategies and the expression of anger in the Locked-Box Frustration Task. For children using low or moderate levels of Negative Speech more distraction was associated with less anger. For children who used high levels of Negative Speech distraction and anger were unrelated, however. Thus, high levels of negative feelings and their verbal expression seem to be associated with difficulties in effectively dealing with frustration. This may also be relevant in the given context, because the vast majority of existing studies on tool innovation granted children much shorter times to solve a given problem than we did here. On the one hand, offering children more time to work on a given task allows them to try out different approaches or to practice a strategy – probably one reason why success rates are increased (Voigt et al., 2019; Breyel & Pauen, 2021). On the other hand, providing more time also increases the risk that children who do not (quickly) find a solution become frustrated. In that case emotion regulation may become crucial for predicting success or failure.

Limitations and Future Directions

Only those children from the original sample who produced private speech were considered in the current analyses, resulting in comparably small subgroups for both two task versions and for each of the three age groups. This may explain why many Bayesian tests revealed only anecdotal to moderate evidence regarding support for or against speech-performance relations. One way to overcome this problem would be to apply Sequential Bayes Factors (SBF) that examine the evidence for hypotheses during ongoing sampling. While increasing the sample size after having already analyzed the data is heavily criticized in frequentist null-hypothesis testing, it is considered as unproblematic for the interpretation of results in Bayesian statistics (Schönbrodt et al., 2017). In an SBF procedure researchers define a threshold for Bayes factor values (e.g., $BF_{10} = 10$ for H_1 and $BF_{10} = 0.1$ for H_0, indicating strong evidence) and sampling continues until this threshold is reached. A recent study showed that this procedure can be highly efficient and informative in the domain of developmental psychology with its special demands regarding the recruitment and testing of young children (Mani et al., 2021). In the present case it was not possible to apply this strategy for practical reasons (i.e., limited resources and restrictions related to COVID-19). Thus, further studies with larger samples are needed before any conclusions can be drawn.

We operationalized innovation performance as success vs. failure and additionally assessed latency to success. A more fine-grained coding of children's approaches to the task in combination with children's speech would increase the depth of analyses. For example, future studies could investigate to what extent spontaneous speech might support ongoing or upcoming attempts to solve the task. This rationale is similar to the so-called item-based metric used in some private speech studies with tasks that consist of multiple items or parts (e.g., Benigno et al., 2011; see also Winsler et al., 2005) to assess the effects of speech on performance on an individual level.

When examining the association between private speech production and performance in cognitive tasks, it seems important to control for children's general cognitive competences and language skills. Winsler and Naglieri (2003) as well as Aro and colleagues (2015) have shown that five-year-olds who received lower scores in cognitive tests benefited more from private speech production while doing another cognitive task than children who received high cognitive scores. Thus, talking privately supported children with lower competence, whereas private speech had no effect or even impeded performance in highly competent children. Furthermore, Bono and Bizri (2014) reported that five-year-olds with higher language skills produced less task-relevant private speech during a challenging LEGO building activity than children with lower language skills. Private speech production mediated the relation between language skills and self-control, suggesting that children who have more language expertise may be already more advanced in their development of inner speech. They no longer need to talk out loud while working on a challenging task which might help them to focus even more on regulating their behavior.

The role of general language skills might be of special interest when investigating the effect of private speech on innovative problem-solving. Beck et al. (2016) demonstrated that children's receptive vocabulary skills predicted success in the Hook Task. This task is very similar to the Vertical Task used in the current study but provides only one potential solution (i.e., bending a pipe cleaner to form a hook). The authors suggested that the language measure might also represent a broader association between innovation and general intelligence. Future studies should thus include measures of children's cognitive achievements (including language skills) to investigate the complex interplay of cognitive development, language development, and the role of private speech in young children's problem-solving.

Conclusions

We investigated children's private speech in a tool innovation task to shed light onto the cognitive, emotional and motivational processes that might lead to success or failure in creative problem-solving. Our data suggest that young children are not only challenged in cognitive terms when trying to solve a tool innovation problem, but also need to regulate negative emotions when facing failed solution attempts. Dealing with negative emotions requires attentional and regulatory capacities that are actually needed to figure out solutions. For practitioners in childcare settings and parents, our findings suggests that it can be very useful to listen to children’s talk while they solve difficult tasks and to provide emotional (rather than only task-oriented) support if they experience frustration. In addition, our findings point to the need of more studies that analyze children’s private speech during complex problem-solving tasks in order to better understand the associations between cognitive competence, language skills and self-regulation skills in preschoolers.

Funding

This study was supported by a stipend from the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes) awarded to the first author.

Acknowledgements

We thank all families for participating in the study, the Baby-Uni Heidelberg team for their help in recruiting participants, and Sandra Schmitt for assistance in coding.

Declaration of interest statement

The authors have no conflicts of interest to disclose.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

References

Al-Namlah, A. S., Meins, E., & Fernyhough, C. (2012). Self-regulatory private speech relates to children’s recall and organization of autobiographical memories. *Early Childhood Research Quarterly*, *27*(3), 441–446. https://doi.org/10.1016/j.ecresq.2012.02.005

Alarcón-Rubio, D., Sánchez-Medina, J. A., & Prieto-García, J. R. (2014). Executive function and verbal self-regulation in childhood: Developmental linkages between partially internalized private speech and cognitive flexibility. *Early Childhood Research Quarterly*, *29*(2), 95–105. https://doi.org/10.1016/j.ecresq.2013.11.002

Aro, T., Poikkeus, A. M., Laakso, M. L., Tolvanen, A., & Ahonen, T. (2015). Associations between private speech, behavioral self-regulation, and cognitive abilities. *International Journal of Behavioral Development*, *39*(6), 508–518. https://doi.org/10.1177/0165025414556094

Atencio, D. J., & Montero, I. (2009). Private speech and motivation: The role of language in a sociocultural account of motivational processes. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 201–223). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511581533.017

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

Behrend, D. A., Rosengren, K., & Perlmutter, M. (1989). A New Look at Children’s Private Speech: The Effects of Age, Task Difficulty, and Parent Presence. *International Journal of Behavioral Development*, *12*(3), 305–320. https://doi.org/10.1177/016502548901200302

Benigno, J. P., Byrd, D. L., McNamara, J. P., Berg, W. K., & Farrar, M. J. (2011). Talking through transitions: Microgenetic changes in preschoolers' private speech and executive functioning. *Child Language Teaching and Therapy*, *27*(3), 269–285. https://doi.org/10.1177/0265659010394385

Berk, L. E. (1986). Relationship of Elementary School Children's Private Speech to Behavioral Accompaniment to Task, Attention, and Task Performance. *Developmental Psychology*, *22*(5), 671–680. https://doi.org/10.1037/0012-1649.22.5.671

Bono, K. E., & Bizri, R. (2014). The role of language and private speech in preschoolers' self-regulation. *Early Child Development and Care*, *184*(5), 658–670. https://doi.org/10.1080/03004430.2013.813846

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development*, *58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

Brown, A. L. (1990). Domain-specific principles affecting learning and transfer in children. *Cognitive Science*, *14*(1990), 107–133. https://doi.org/10.1207/s15516709cog1401_6

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 of London. Series B, Biological Sciences*, *368*(1630), 20120409. https://doi.org/10.1098/rstb.2012.0409

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

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. (2019). Is tool modification more difficult than innovation? *Cognitive Development*, *52*. https://doi.org/10.1016/j.cogdev.2019.100811

Day, K. L., & Smith, C. L. (2013). Understanding the role of private speech in children's emotion regulation. *Early Childhood Research Quarterly*, *28*(2), 405–414. https://doi.org/10.1016/j.ecresq.2012.10.003

de Dios, M. J., & Montero, I. (2003). The Motivational Function of Private Speech: An Experimental Approach. *Biennial Meeting of the Society for Research in Child Development*. http://www.gpo.gov/fdsys/pkg/ERIC-ED483019/pdf/ERIC-ED483019.pdf

Diaz, R. M. (1992). Methodological concerns in the study of private speech. In R. M. Diaz & L. E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 55–81). Erlbaum.

Feigenbaum, P. (1992). Development of the syntactic and discourse structures of private speech. In R. M. Diaz & L. E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 181–198). Erlbaum.

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

Frieze, I. H. (1981). Children's attributions for success and failure. In S. S. Brehm, S. M. Kassin, & F. X. Gibbons (Eds.), *Developmental social psychology* (pp. 51–71). Oxford University Press.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Goss-Sampson, M. (2020). *Bayesian Inference in JASP: A guide for students*. https://doi.org/10.17605/OSF.IO/CKNXM

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). https://doi.org/10.1371/journal.pone.0019555

Jongbloed-Pereboom, M., Nijhuis-van der Sanden, M. W. G., Saraber-Schiphorst, N., Crajé, C., & Steenbergen, B. (2013). Anticipatory action planning increases from 3 to 10 years of age in typically developing children. *Journal of Experimental Child Psychology*, *114*(2), 295–305. https://doi.org/10.1016/j.jecp.2012.08.008

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

Krafft, K. C., & Berk, L. E. (1998). Private speech in two preschools: Significance of open-ended activities and make-believe play for verbal self-regulation. *Early Childhood Research Quarterly*, *13*(4), 637–658. https://doi.org/10.1016/S0885-2006(99)80065-9

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Lidstone, J., Meins, E., & Fernyhough, C. (2011). Individual differences in children’s private speech: Consistency across tasks, timepoints, and contexts. *Cognitive Development*, *26*(3), 203–213. https://doi.org/10.1016/j.cogdev.2011.02.002

Mani, N., Schreiner, M. S., Brase, J., Köhler, K., Strassen, K., Postin, D., & Schultze, T. (2021). Sequential Bayes Factor designs in developmental research: Studies on early word learning. *Developmental Science*, *January*, 1–12. https://doi.org/10.1111/desc.13097

Manning, B. H., White, C. S., & Daugherty, M. (1994). Young children's private speech as a precursor to metacognitive strategy use during task engagement. *Discourse Processes*, *17*(2), 191–211. https://doi.org/10.1080/01638539409544866

Nielsen, M. (2013). Young Children's Imitative and Innovative Behaviour on the Floating Object Task. *Infant and Child Development*, *22*, 44–52. https://doi.org/10.1002/icd.1765

Reindl, E., Beck, S. R., Apperly, I. A., & Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1825), 20152402. https://doi.org/10.1098/rspb.2015.2402

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

Schönbrodt, F. D., Wagenmakers, E., Zehetleitner, M., & Perugini, M. (2017). Sequential Hypothesis Testing With Bayes Factors : Efficiently Testing Mean Differences. *Psychological Methods*, *22*(2), 322.

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 of London. Series B, Biological Sciences*, *371*(20150191). https://doi.org/10.1098/rstb.2015.0191

Thibodeaux, J., Bock, A., Hutchison, L. A., & Winsler, A. (2019). Singing to the self: Children's private speech, private song, and executive functioning. *Cognitive Development*, *50*(January 2018), 130–141. https://doi.org/10.1016/j.cogdev.2019.04.005

Träuble, B., & Pauen, S. (2011). Cause or effect: What matters? How 12-month-old infants learn to categorize artifacts. *British Journal of Developmental Psychology*, *29*(3), 357–374.

https://doi.org/10.1348/026151009X479547

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

Vygotsky, L. S. (1962). *Thought and language* (E. Hanfmann & G. Vakar (eds.)). MIT Press. (Original work published 1934)

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (eds.)). Harvard University Press. (Original work published 1930-1935)

Whedon, M., Perry, N. B., Curtis, E. B., & Bell, M. A. (2021). Private speech and the development of self-regulation: The importance of temperamental anger. *Early Childhood Research Quarterly*, *56*(June), 213–224. https://doi.org/10.1016/j.ecresq.2021.03.013

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., Abar, B., Feder, M. A., Schunn, C. D., & Rubio, D. A. (2007). Private speech and executive functioning among high-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1617–1635. https://doi.org/10.1007/s10803-006-0294-8

Winsler, A., De León, J. R., Wallace, B. A., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behaviour. *Journal of Child Language*, *30*(3), 583–608. https://doi.org/10.1017/S0305000903005671

Winsler, A., Fernyhough, C., McClaren, E., & Way, E. (2005). *Private speech coding manual. Unpublished manuscript*. George Mason University, Fairfax, VA, USA. http://winslerlab.gmu.edu/PSCodingManual.pdf

Winsler, A., & Naglieri, J. (2003). Overt and Covert Verbal Problem-Solving Strategies: Developmental Trends in Use, Awareness, and Relations with Task Performance in Children Aged 5 to 17. *Child Development*, *74*(3), 659–678. https://doi.org/10.1111/1467-8624.00561

Zelazo, D. P., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., Sutherland, A., & Carlson, S. M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development*, *68*(3), 1–151.

Table 1. Speech (sub)categories adapted from Sawyer (2017) with examples of the current study.

Speech (Sub)Category	Example Utterance
Cognitive Speech	
Thinking	*Looking at two items*: „Which one is longer?"
Planning	*While choosing items:* „First, I take this one and then the other thing."
Self-Guiding	„Ok, try this one."
Strategic	„The hook can get there and catch that (*object*) and fish it out."
Expression of New Idea[a]	*After first unsuccessful approach:* „I have an idea…"
Metacognitive Speech	
Solving Speech (+)	*Hooking up the toy:* „Now I've got it!"
Monitoring Errors	*After item fell into tube*: „Now that fell down, too."
Reflecting on Task Progress	*During an ongoing approach with a long item:* „I've already pushed it a little further."
Motivational Speech	
Agentic	„Come on out!"
Positive Emotional Reaction (+)	„Hihi! Hahaha!" (*smiling, laughing*)
Positive Self Evaluation (+)	„I'm sure I can do it."
Positive Task-Evaluation (+)	„This is so easy!"
Self-Encouragement (+)	„Come on, go again."
Negative Emotional Reaction (-)	„Oh, man, what a bummer!"
Negative Self Evaluation (-)	„I will never get it out."
Negative Task-Evaluation (-)	„It's pretty difficult."
Playful Speech	
Entering Pretend Scenario[b] (+)	„The mouse is too strong."
Addressing objects as living (+)	*Addressing the toy*: „I will get you, you will see!"
Word Play/Singing (+)	„Ohnemanna mull mull" (*fantasy words*)
Partially Internalized Speech	
Extremely Low Volume	(*muttering, whispering*)
Grammatically Fragmented	„This .. Ey!"
Semantically Condensed	„Maybe."

Note. (+) Subcategories that were scored as Positive Speech, (-) Subcategories that were scored as Negative Speech. [a] We added this subcategory to the scheme by Sawyer (2017). [b] We collapsed Sawyer's (2017) subcategories of Entering Pretend Scenario and Creating Pretend Scenario since this kind of talk was rare and very similar in meaning.

Table 2. Number of successful children (%) split by age and task version.

Age Group	Horizontal Task (N = 33)		Vertical Task (N = 32)	
	n	Successful (%)	n	Successful (%)
3-year-olds (N = 21)	10	8 (80%)	11	1 (9.1%)
4-year-olds (N = 22)	12	12 (100%)	10	7 (70%)
5-year-olds (N = 22)	11	10 (90.9%)	11	10 (90.9%)

Figure 1. Functional item sets for the Horizontal Task (A) in their original form (upper panel) and in their modified form (lower panel) and functional item sets for the Vertical Task (B) in their original form (upper panel) and their modified form (lower panel).

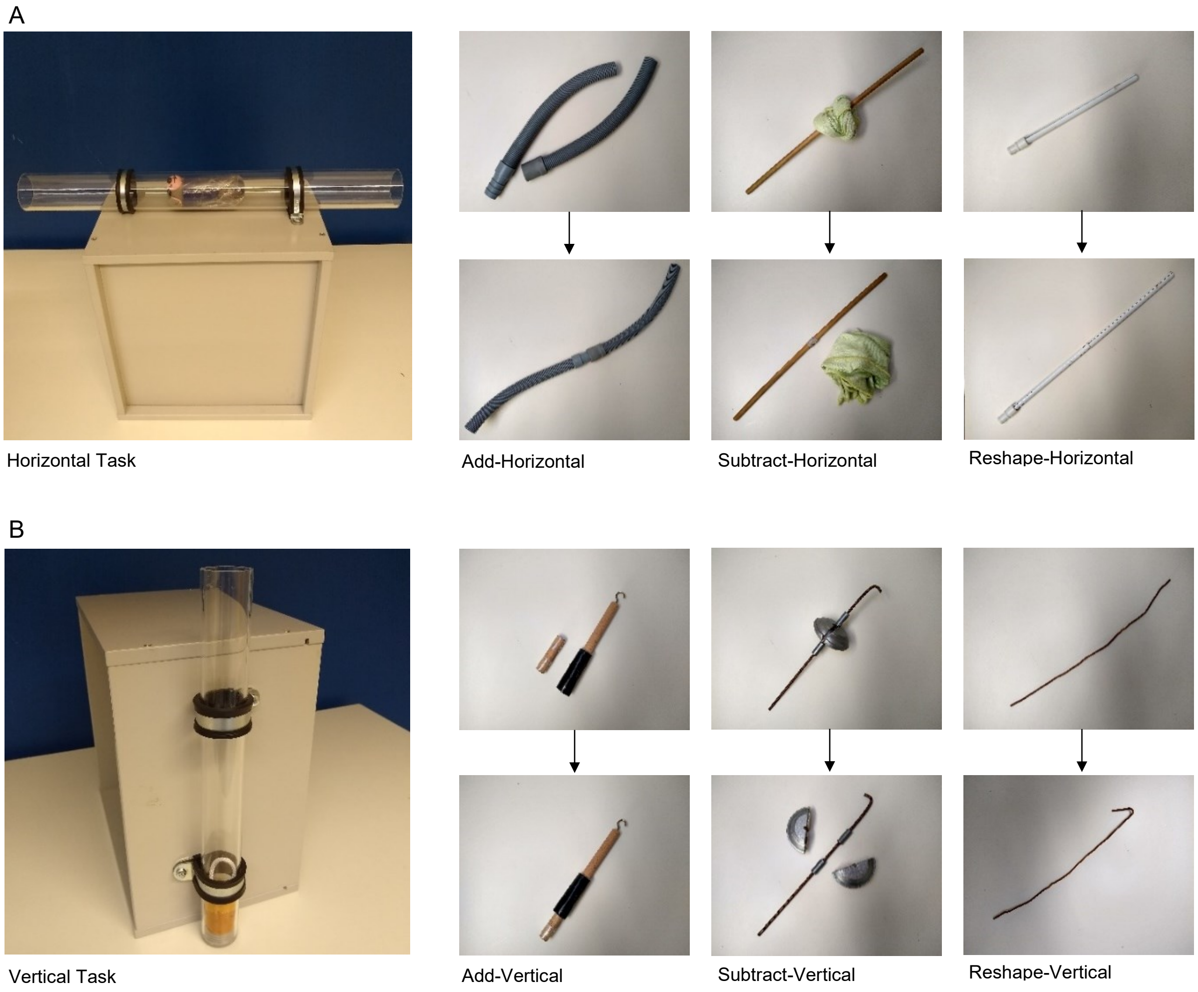
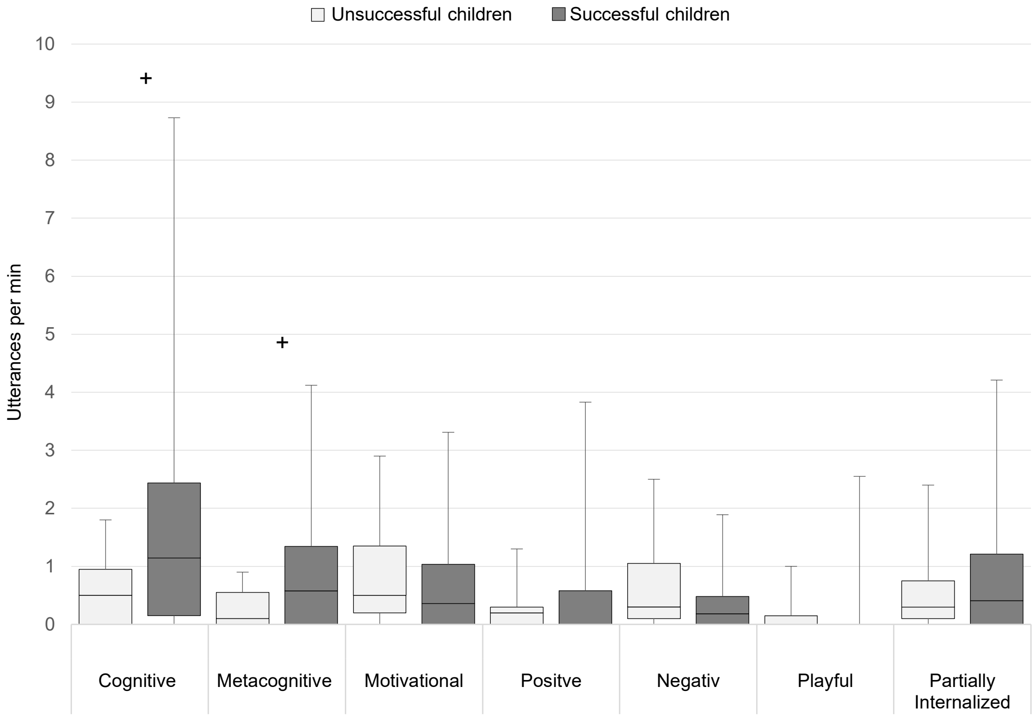
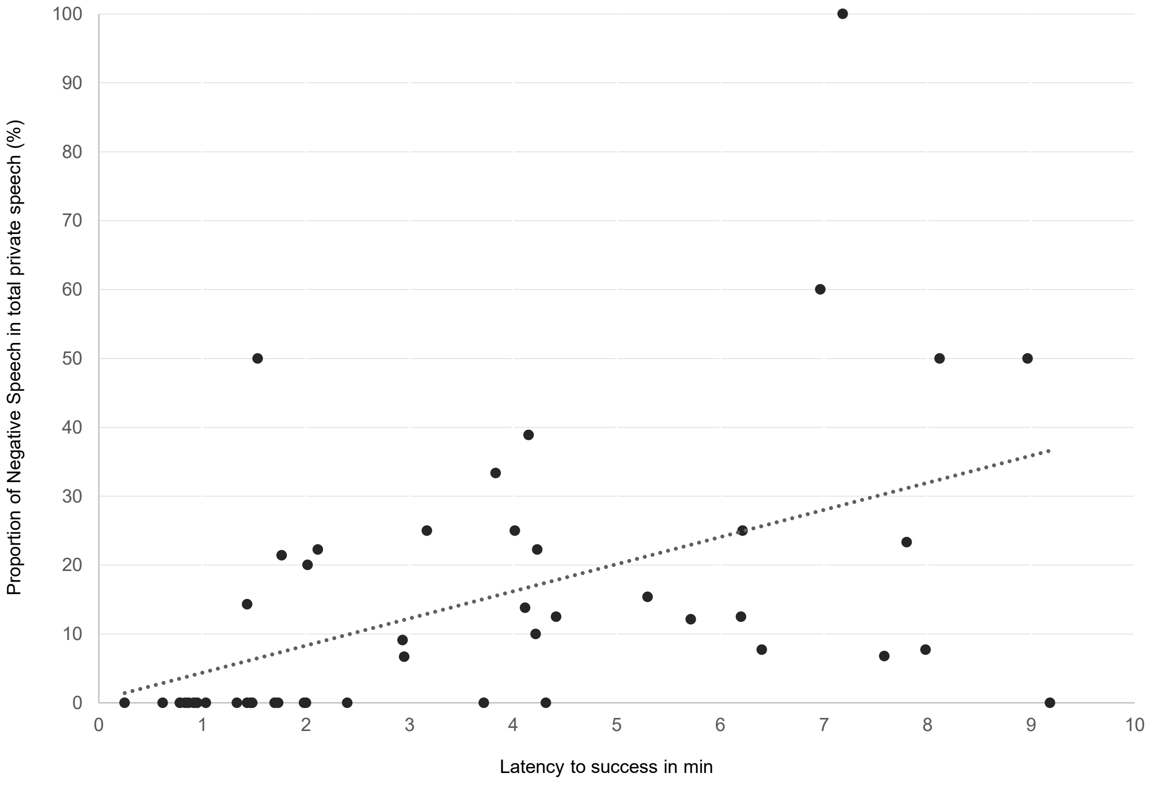


Figure 2. Boxplots showing utterances per min in the private speech categories stated by unsuccessful and successful children.



Note. $^{+}$ BF$_{10}$ = 1-3 (anecdotal evidence for H1).

Figure 3. Correlation between proportion of Negative Speech and latency to success (N = 48 successful children).



Supplementary Material for

Private Speech during Problem-Solving: Tool Innovation Challenges both Preschoolers' Cognitive and Emotion Regulation

Content:

- Additional information on interpretation of Bayes factors (Table S1)
- Descriptive statistics and test parameters for private speech utterances per min and proportion of private speech split by gender (Tables S2 and S3) and task version (Tables S4 and S5)
- Correlations between private speech and latency to success and age (Tables S6 and S7)
- Descriptive statistics and test parameters for private speech utterances per min and proportion of private speech split by task success (Tables S8 and S9)

Table S.1

Interpretation of Bayes Factors (Goss-Sampson, 2020)

Bayes Factor (BF_{10})	Evidence	In favor of
> 100	Decisive	H1
30–100	Very strong	H1
10–30	Strong	H1
3–10	Moderate	H1
1–3	Anecdotal	H1
1	No evidence	Neither
1–0.33	Anecdotal	H0
0.33–0.1	Moderate	H0
0.1–0.03	Strong	H0
0.03–0.01	Very strong	H0
< 0.01	Decisive	H0

Table S.2

Number of utterances per min (median and interquartile range) split by gender

Private Speech Category	Boys (N = 34)		Girls (N = 31)		W	BF_{10}	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	*IQR*	*Mdn*	*IQR*				
Cognitive/min	0.96	2.08	0.90	1.74	525.50	0.27	0.03	-0.43, 0.49
Metacognitive/min	0.56	1.18	0.27	0.99	600.50	0.35	0.18	-0.29, 0.65
Motivational/min	0.32	1.19	0.50	0.90	493.00	0.27	-0.04	-0.51, 0.41
Positive/min	0.00	0.48	0.20	0.63	451.50	0.33	-0.15	-0.64, 0.32
Negative/min	0.12	0.48	0.30	0.51	490.00	0.27	-0.05	-0.52, 0.42
Playful/min	0.00	0.00	0.00	0.19	390.00	0.72	-0.36	-0.94, 0.18
Partially Internalized/min	0.37	1.08	0.30	0.93	491.00	0.29	-0.11	-0.58, 0.35

Note. Bayes factors based on data augmentation algorithm with 5 chains of 10,000 iterations. For all tests, the alternative hypothesis specifies that boys ≠ girls.

Table S.3

Proportion of utterances in total private speech (median and interquartile range) split by gender

Private Speech Category	Boys (N = 34)		Girls (N = 31)		W	BF_{10}	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	IQR	Mdn	IQR				
Cognitive %	26.39	38.47	32.20	25.23	464.50	0.34	-0.17	-0.64, 0.29
Metacognitive %	19.09	26.85	16.67	23.83	618.50	0.51	0.27	-0.19, 0.77
Motivational %	19.09	33.33	24.14	34.76	475.50	0.30	-0.12	-0.58, 0.34
Positive %	0.00	7.55	6.06	18.89	433.00	0.42	-0.23	-0.73, 0.25
Negative %	12.50	24.31	7.89	24.17	524.50	0.27	0.00	-0.47, 0.47
Playful %	0.00	0.00	0.00	2.98	390.00	0.80	-0.38	-0.97, 0.16
Partially Internalized %	11.44	37.22	18.42	23.83	492.00	0.28	-0.07	-0.54, 0.38

Note. Bayes factors based on data augmentation algorithm with 5 chains of 10,000 iterations. For all tests, the alternative hypothesis specifies that boys ≠ girls.

Table S.4

Number of utterances per min (median and interquartile range) in the Horizontal Task and Vertical Task

Private Speech (*Sub*)Category	Horizontal Task (N = 33)		Vertical Task (N = 32)		W	BF_{10}	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	*IQR*	*Mdn*	*IQR*				
Cognitive/min	1.25	2.73	0.80	0.93	626.00	0.53	0.28	-0.19, 0.77
Metacognitive/min	0.52	1.40	0.47	0.74	610.00	0.42	0.23	-0.24, 0.72
Motivational/min	0.26	0.75	0.50	1.08	370.00	1.92^{+}	-0.48	-0.97, 0.002
Positive/min	0.00	0.58	0.17	0.55	481.50	0.29	-0.10	-0.58, 0.38
Negative/min	0.00	0.38	0.38	0.67	299.00	8.71*	-0.66	-1.18, -0.16
Neg. Emotional Reaction/min	0.00	0.16	0.21	0.37	310.00	3.14*	-0.56	-1.09, -0.06
Neg. Self Evaluation/min	0.00	0.00	0.00	0.00	454.00	0.45	-0.24	-0.81, 0.30
Neg. Task Evaluation/min	0.00	0.00	0.00	0.00	464.00	0.36	-0.16	-0.70, 0.35
Playful/min	0.00	0.00	0.00	0.00	497.00	0.32	-0.07	-0.61, 0.46
Partially Internalized/min	0.33	1.18	0.37	0.71	478.50	0.31	-0.13	-0.60, 0.32

Note. Bayes factors based on data augmentation algorithm with 5 chains of 10,000 iterations.

$^{+}$ Anecdotal evidence for H1, * moderate evidence for H1.

Table S.5

Proportion of utterances in total private speech (median and interquartile range) in the Horizontal Task and Vertical Task

Private Speech (*Sub*)Category	Horizontal Task (N = 33)		Vertical Task (N = 32)		W	BF$_{10}$	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	*IQR*	*Mdn*	*IQR*				
Cognitive %	36.36	50.00	24.72	22.11	651.00	0.61	0.31	-0.15, 0.80
Metacognitive %	20.00	30.00	16.67	23.23	603.00	0.40	0.22	-0.25, 0.70
Motivational %	14.29	25.00	30.56	27.16	295.50	8.05*	-0.65	-1.16, -0.15
Agentic Desire %	0.00	0.00	0.00	3.45	374.50	1.08[+]	-0.45	-1.03, 0.10
Positive %	0.00	13.56	5.41	18.33	477.00	0.31	-0.12	-0.61, 0.36
Negative %	0.00	13.79	20.60	21.94	271.00	16.60**	-0.73	-1.25, -0.22
Neg. Emotional Reaction %	0.00	3.85	7.18	20.25	298.00	4.81*	-0.61	-1.14, -0.11
Neg. Self Evaluation %	0.00	0.00	0.00	0.00	456.00	0.46	-0.23	-0.80, 0.31
Neg. Task Evaluation %	0.00	0.00	0.00	4.29	456.00	0.38	-0.19	-0.72, 0.33
Playful %	0.00	0.00	0.00	0.00	496.00	0.32	-0.09	-0.63, 0.44
Partially internalized %	13.56	44.44	20.00	23.60	464.50	0.33	-0.16	-0.63, 0.30

Note. Bayes factors based on data augmentation algorithm with 5 chains of 10,000 iterations.

[+] Anecdotal evidence for H1, * moderate evidence for H1, ** strong evidence for H1.

Table S.6

Correlations between utterances per min, latency to success and age

Private Speech (*Sub*)Category	Latency to Success in min (N = 48)		Age in months (N = 65)	
	τ	BF_{10}	τ	BF_{10}
Cognitive/min	-.30 [a]	27.37**	.17	1.25+
Thinking/min	-.09	0.47		
Planning/min	$5.34e^{-3}$	0.18		
Self Guiding/min	.03	0.15		
Strategic/min	.06	0.13		
Expressing New Idea/min	.05	0.13		
Metacognitive/min	-.14 [a]	0.84	.20	2.62+
Motivational/min	.08	0.25	.07	0.23
Positive/min	-.09	0.27	.06	0.21
Negative/min	.32	26.49**	-.02	0.17
Neg. Emotional Reaction/min	.27	5.88*		
Neg. Self Evaluation/min	.15	0.61		
Neg. Task Evaluation/min	.34	51.37**		
Playful/min	.09	0.27	-.08	0.25
Partially Internalized/min	-.05	0.97	-.16	1.00

[a] Alternative hypothesis specifies that the correlation is negative.

+ Anecdotal evidence for H1, *moderate evidence for H1, ** strong evidence for H1.

Table S.7

Correlations between the proportion of the private speech category, latency to success and age

Private Speech (*Sub*)Category	Latency to Success in min (N = 48)		Age in months (N = 65)	
	τ	BF_{10}	τ	BF_{10}
Cognitive %	-.15[a]	1.03[+]	.15	0.68
Metacognitive %	-.01[a]	0.21	.16	0.83
Motivational %	.26	4.78*	.10	0.30
Agentic Desire %	.23	2.26[+]		
Positive %	-.05	0.21	.02	0.17
Negative %	.45	3044.91****	-.002	0.16
Neg. Emotional Reaction %	.34	61.52***		
Neg. Self-Evaluation %	.16	0.65		
Neg. Task-Evaluation %	.38	246.09****		
Playful %	.09	0.29	-.08	0.26
Partially Internalized %	.04	0.20	-.26	19.25**
Extremely Low Volume %			-.28	39.57**
Gramatically Fragmented %			-.18	1.36[+]
Semantically Condensed %			-.12	0.41

[a] Alternative hypothesis specifies that the correlation is negative.

[+] Anecdotal evidence for H1, * moderate evidence for H1, ** strong evidence for H1, *** very strong evidence for H1, **** decisive evidence for H1.

Table S.8

Number of utterances per min (median and interquartile range) for unsuccessful and successful children

Private Speech Category	No Success (N = 17)		Success (N = 48)		W	BF_{10}	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	IQR	Mdn	IQR				
Cognitive/min[a]	0.50	0.90	1.14	2.21	243.50	2.03[+]	-0.44	-0.97, -0.05
Metacognitive/min [a]	0.10	0.50	0.58	1.30	277.00	1.23[+]	-0.37	-0.90, -0.03
Motivational/min	0.50	1.10	0.36	1.01	485.0	0.32	0.13	-0.36, 0.63
Positive/min	0.20	0.30	0.00	0.58	429.0	0.29	0.04	-0.48, 0.55
Negative/min	0.30	0.90	0.18	0.48	512.0	0.57	0.30	-0.20, 0.83
Playful/min	0.00	0.10	0.00	0.00	466.0	0.37	0.16	-0.39, 0.73
Partially internalized/min	0.30	0.60	0.41	1.20	414.0	0.28	0.04	-0.45, 0.54

Note. Bayes factors based on data augmentation algorithm with 5 chains of 10,000 iterations.

[a] Directional alternative hypothesis specifies that values of unsuccessful group are lower than values of successful children.

[+] Anecdotal evidence for H1.

Table S.9

Proportion of utterances in total private speech (median and interquartile range) for unsuccessful and successful children

Private Speech Category	No Success (N = 17)		Success (N = 48)		W	BF_{10}	Median Effect Size δ	95% CI for δ (lower, upper)
	Mdn	IQR	Mdn	IQR				
Cognitive %[a]	20.00	33.33	33.33	29.46	266.5	1.75[+]	-0.42	-0.93, -0.04
Metacognitive %[a]	5.88	20.00	20.00	28.93	307.5	0.67	-0.28	-0.77, -0.02
Motivational %	33.33	23.16	16.78	34.72	529.5	0.53	0.29	-0.21, 0.82
Positive %	5.26	20.00	0.00	14.34	471.5	0.39	0.20	-0.31, 0.72
Negative %	20.37	31.94	7.69	22.22	542.5	0.73	0.35	-0.15, 0.88
Playful %	0.00	2.63	0.00	0.00	469.5	0.39	0.19	-0.36, 0.77
Partially Internalized %	20.00	35.35	14.28	34.09	505.0	0.52	0.28	-0.22, 0.82

Note. Mann-Whitney Test. Result based on data augmentation algorithm with 5 chains of 10,000 iterations.

[a] Directional alternative hypothesis specifies that values of unsuccessful group are lower than values of successful children.

[+] Anecdotal evidence for H1.

III. Manuscript

Breyel, S., & Pauen, S. (2022b). *How prior experience in analogous tasks affects three-year-olds' tool making.* Manuscript submitted for publication in Cognitive Development.

How Prior Experience in Analogous Tasks Affects Three-year-olds' Tool Making

Sabine Breyel[a] & Sabina Pauen[a]

[a]Department of Developmental and Biological Psychology, Heidelberg University

Hauptstraße 47-51, 69117 Heidelberg, Germany

Declarations of interest: none

Correspondence concerning this article should be addressed to Sabine Breyel

Email: sabine.breyel@psychologie.uni-heidelberg.de

Department of Developmental and Biological Psychology, Heidelberg University

Hauptstraße 47-51, 69117 Heidelberg, Germany

Abstract

Research on analogical reasoning shows that young preschoolers have the basic skills to generalize knowledge about tool use across problems. However, little is known so far about the role of analogical transfer in the domain of tool making. The current study examined the effects of prior experience with two analogous training tasks on three-year-olds' performance in a tool innovation task (Vertical Tube Task, VTT). Children in the Training Condition (N = 25) showed significant improvement in terms of their ability to transfer the correct solution approach across different training tasks. At the same time, most children failed to implement this solution, resulting in success rates comparable to a Control Condition without prior training (N = 25). We discuss the implications of the observed discrepancy between generating a proper solution and implementing it for studies on the early development of tool innovation skills.

142 words

Keywords: tool making; innovation; analogical thinking; transfer; cognitive development

Introduction

Children are surrounded by many artifacts designed to manage everyday life. Already during the first year of life infants begin to understand the relation between the form and function of objects (Träuble & Pauen, 2007, 2011). Starting in their second year of life, they learn how to use artifacts properly themselves, for example, to eat with spoons or use hairbrushes (Connolly & Dalgleish, 1989; McCarty et al., 2001). Beyond using objects in everyday life, 18-months-old toddlers also understand that novel rake-like tools will pull objects into their reach, and they can transfer this tool knowledge to a new set of tools (Brown, 1990; Rat-Fischer et al., 2012). In contrast to the early emerging skills of functional understanding and tool use, children's ability to *make* tools seems to develop rather late. Until school age, most children have been found to struggle with constructing a new tool for solving a problem, especially when they need to figure out the appropriate tool design on their own (Beck et al., 2011; Frick et al., 2017). This kind of independent tool making has been termed 'tool innovation' (Chappell et al., 2013).

Most studies exploring the development of tool innovation use the so-called Hook Task (Beck et al., 2011). In this task, children are asked to get a reward out of a transparent vertical tube within one minute. Since the reward is placed at the bottom of the tube and cannot be reached by hand, a long tool with a hook is needed. In most studies, the only possible solution to build such a tool is to bend one end of a straight pipe cleaner into a hook. Success rates among three- to four-year-old children in the Hook Task were usually very low (i.e., 0 to 8%) and even among five- to six-year-olds only 4 to 18% of the children solved the problem (Beck et al., 2011; Chappell et al., 2013; Gönül et al., 2018; see Sheridan et al., 2016 for an exception of higher success rates; Whalley et al., 2017). However, Voigt et al. (2019) demonstrated that five-year-olds performed much better when provided with more time and different options to produce a suitable tool. The authors reported that 76–93% of the children solved the adapted version of the Hook Task, the Vertical Tube Task (VTT), within ten minutes. Breyel & Pauen (2021) replicated this finding for five-year-olds (93% success rate), but they also reported that only 50%

of the four-year-olds and 7% of the three-year-olds managed to build a suitable tool. The availability of time and different materials did not seem to help younger children in finding a solution and successfully implementing it, thus raising the interesting question of what renders the VTT so difficult for younger children.

In contrast to tool innovation, ‘tool manufacture’ does not require individuals to innovate the necessary tool design because it involves prior instruction on how to make the tool (Chappell et al., 2013). After having observed an adult who demonstrated appropriate tool making, the majority of preschoolers could solve the subsequent task themselves (e.g., Beck et al., 2011; Chappell et al., 2013; Cutting et al., 2011). Thus, young children are able to socially learn tool manufacture. But would they also be able to transfer their newly acquired tool-manufacturing skills either to new material or to a structurally similar task, thus going beyond imitation? This would require a deeper functional understanding of the problem structure as well as analogical reasoning skills. Although analogical reasoning has been suggested to play an important role in successful tool making (Beck et al., 2016), there is hardly any empirical investigation on the links between the two domains (for exceptions see Beck et al., 2014; Gerson et al., 2018). The current study aims at understanding whether analogical reasoning might help young children to transfer and adapt tool-making skills to different tasks. Before explaining the rationale of our study, we provide an overview of the literature on young children’s analogical reasoning skills and their transfer skills in the domain of tool making.

The Development of Analogical Reasoning

Drawing analogies involves mapping knowledge from a base to a target domain (Gentner, 1983). This mapping process involves recognizing similarities between sets of relations despite apparent differences between the objects that build these relations. For example, the metaphor ‘a cloud is like a sponge’ wants to illustrate the similar function of a cloud and a sponge to hold water and give it back later (Gentner, 1988). However, young children initially have a strong preference for object similarity and will rather interpret the metaphor based on the appearance of clouds and sponges, e.g., they have both round shapes (Gentner, 1988). As children gain more domain-specific knowledge about objects and their relations, this preference shifts, and relational similarities become more prominent (Rattermann & Gentner, 1998).

The beginnings of relational thinking in problem-solving situations can be traced back to early toddlerhood. In a very simple “out of reach” paradigm, Brown (1990) showed that children in their second year of life quickly learn which tool to select from a given set to get an interesting toy placed out of reach. Toddlers transferred this knowledge to a different-looking set of tools with the same functional properties. Thus, they did not pick a tool that resembled the original in superficial characteristics like color and pattern but rather focused on critical function attributes like length, rigidity, and shape.

In older children with more advanced language skills, a common approach to studying analogical reasoning is to measure children’s problem-solving skills after listening to a similar problem in a story (Brown et al., 1986; e.g., Holyoak et al., 1984). In such a story, the protagonist faces a problem and finds a solution by constructing a tool. Often, the child is allowed to re-enact the solution scene with props together with an experimenter. Finally, the target problem and various materials for tool making are presented and the child is asked for a solution. Findings from these studies suggest that three- to five-year-olds are able to transfer the solution from the story to the target problem, but only if they represent the relevant goal structure of the two situations (Brown et al., 1986). Those children who recalled the relevant details of the analogy story (either spontaneously or prompted) were more likely to solve the target problem (65−80% success) than children not recalling these details (about 20% success). Similarly, Crisafi and Brown (1986) showed that the majority of three-year-olds were able to transfer their knowledge from one task to another after receiving the hint that both tasks were somehow similar. First, all children learned how to operate an apparatus in order to extract sweets. Then they were asked to work on a similar apparatus. Among those children who received the similarity hint, 80–100% were able to transfer their previously acquired knowledge and succeeded in the second task as well, whereas only 10–40% of the group without hint were successful. Furthermore, Brown and Kane (1988) demonstrated that children of this young age can also be taught to develop a mind-set to look for analogies in tool use problems involving actions like stacking objects to reach something high or using a long object to reach out for something far away. However, contrary to older children, they needed at least two instances to exercise analogical transfer or some kind of instruction to reflect upon the similarities between the problems (e.g., telling a puppet how to solve the problem).

In sum, these studies on analogical reasoning demonstrate the high potential of young children to use relational mapping successfully. Only very few children show this ability spontaneously below the age of five to six years, but even three-year-olds seem able to solve analogical transfer tasks when supported by similarity prompts or the possibility to exercise transfer.

Analogical Reasoning and Transfer in Tool Making

Analogical reasoning could facilitate tool making because it helps to transfer knowledge about (causal) relations between familiar objects to a current problem and novel materials. Moreover, when structurally similar problems arise in a future situation, it might be helpful to map the tool-making knowledge from the previous to the new problem in order to solve it more efficiently or find even other innovative solutions. There is some evidence suggesting that young children can indeed benefit from prior experiences with the tool material when solving a Hook Task. For example, after observing an adult bending a pipe cleaner into a hook, 30–60% of three- to four-year-olds successfully built the same tool in the Hook Task, compared to only 0–7% of same-aged children solving the target problem without seeing a tool manufacturing demonstration (Gönül et al., 2018; Sheridan et al., 2016). Furthermore, children have been shown to profit from using – instead of just seeing – a premade pipe cleaner hook once before working on the target task (Whalley et al., 2017). In these cases, object similarity is at its maximum in terms of both the tool material and the apparatus. The corresponding training studies thus provided direct training rather than analogical training.

Beck et al. (2014) tested whether children can also generalize their prior experiences about tool making across problems that differ in non-functional characteristics (e.g., shape and color of apparatus). They found that preschoolers showed large improvements from near floor to near ceiling performance, thus indicating that five- to six-year-olds are well able to transfer knowledge across similar but not identical problems. Three- to four-year-olds' performance also improved across trials but at much lower rates. This indicates considerable improvement in transfer skills during early childhood.

In a second experiment, Beck and colleagues (2014) examined whether children can transfer knowledge about tool making from one manufacturing option (i.e., bending a pipe cleaner into a hook) to another (i.e., connecting two pieces of dowel to form a hook). The majority of the five- to seven-year-olds did not succeed under these circumstances, suggesting that the generalization of tool making across different tool manufacturing modes is still difficult at that age. Gerson et al. (2018) explored this generalization of tool making in more detail. Across three learning trials, four- to seven-year-olds either practiced bending different materials into a hook (i.e., pipe cleaner vs. bendable wands; High Alignability Condition) or they were presented with the same material repeatedly (i.e., pipe cleaners only; Low Alignability Condition). In a fourth generalization trial, children were then tested with material that required a different manufacturing strategy to form a hook, i.e., combining two pieces of dowel to form an l-shape. In both conditions, success rates increased from the first trial (10–20%) to the third trial (90–100%), suggesting that children learned equally well to reshape the material into a hook. In the generalization trial, six-year-olds performed slightly better after training with variable materials (High Alignability Condition) than with more homogenous materials (Low Alignability Condition). However, the possibility to make analogical comparisons during training did not improve the generalization performance of younger children.

The Current Study

The reported studies on analogical transfer in tool innovation tasks suggest that children have great difficulty generalizing their tool-making knowledge across different manufacturing methods, but that they can transfer their knowledge across similar problems requiring the same manufacturing method from about five years of age on (Beck et al., 2014; Gerson et al., 2018). However, empirical evidence on how well children younger than five years can engage in the latter 'close' analogical transfer is still inconclusive (Beck et al., 2014). This contrasts with findings from the analogical reasoning literature indicating that children as young as three years have the basic capacities to understand analogies in tool use problems when they receive some support to focus on similarities between the tasks or have the opportunity to exercise (Brown et al., 1986; Brown & Kane, 1988; Crisafi & Brown, 1986). This evidence rather suggests that even children below five years should be able to engage in analogical transfer in the domain of tool making. The current study aims at exploring this issue by testing whether three-year-olds' performance in a tool-making task is facilitated by previous training with analogous tasks.

As the target tool-making task, we used the Vertical Tube Task (VTT) that was originally introduced by Voigt et al. (2019). Similar to the Hook Task, children are asked to get a reward (here: a little cuddly toy) from the bottom of a narrow transparent vertical tube. Since they can not reach the toy with their hand, they need to manufacture a tool that is narrow enough to enter the tube, long enough to reach the toy, and which has a hook at its end to lift up the toy. The item selection of the VTT offers three different methods to manufacture such a tool, including 'Add' (i.e., connect two sticks, one equipped with a hook), 'Subtract' (i.e., remove a non-functional part from a long hook), and 'Reshape' (i.e., bend a wire at one end to form a hook).

Using a between-subject design, we compared VTT performance between children who previously participated in a two-phase training to manufacture a hook in different task settings and children without corresponding training experience. For the training phase, we focused on the manufacturing method Reshape because children have been shown to rarely use this method spontaneously (see Breyel & Pauen, 2021; Voigt et al., 2019). That way, we could check whether the analogical training has an effect not only on the success rate but also on the choice of the manufacturing method. Furthermore, we can better compare our findings to existing work which also applied Reshape as a manufacturing method (see Beck et al., 2014; Gerson et al., 2018; Whalley et al., 2017). Three-year-olds were tested because this age group showed only very low success rates in the VTT (Breyel & Pauen, 2021), thus raising the question of whether they can profit from analogical training of tool making. In general, we expect children who participate in the Training Condition to show better performance and more frequent use of the Reshape strategy than children in the Control Condition without training. In order to shed light on potential problems that toddlers may face in this task context, performance will be assessed by a range of different parameters, including success, latency to success, manufacturing method, type of innovation, first item selected, and the duration of different attempt categories (see coding section for details).

Method

Power Analysis

A previous study examined the effect of prior training experience on young children's subsequent tool-making (Whalley et al., 2017) and serves as a reference for the present power calculation. The authors found that significantly more children were successful in forming a pipe cleaner hook (60%) after practicing with a premade hook compared to those without corresponding experience (0%). We calculated the size of this effect based on the success rates in both conditions (Phi = .641). Since we included a target tool demonstration plus two training tasks before introducing the VTT, we expected to find a similarly high difference in success rates between the training and a control group. A power analysis (G*Power: Chi-Square-Test, alpha = .05, power = .80, df = 1) suggested that a total sample size of N = 50 should be sufficient to detect a medium to large effect (w = .04).

Participants

The study was conducted in a medium-sized town in southwestern Germany. The final sample consisted of N = 25 children in the Control Condition (13 girls, mean = 3 years 6 month, range = 3;0–3;11) and N = 25 children in the Training Condition (13 girls, mean = 3 years 6 month, range = 3;0–3;11). Additional N = 9 children have been tested but had to be excluded because they refused to participate (n = 3), parents helped them to solve the task (n = 1), the procedure was interrupted by bathroom visits (n = 2), or testing had to be stopped because the child no longer wanted to work on the task (n = 3). All participants were full-term born and typically developing children of European descent coming from an upper-middle-class socioeconomic background. The majority of the mothers (84 %) and fathers (80 %) had a high educational degree (equivalent to a college degree or higher). A smaller percentage reported a middle (16 % of mothers, 14 % of fathers) or lower school education (0 % of mothers, 2 % of fathers). Data from two fathers were missing (4 %).

Before testing started, parents gave their consent for the child to participate in the experiment, which was conducted according to the Declaration of Helsinki. While each child was tested individually in a quiet room in the lab, the parent filled out various questionnaires and was instructed not to look at what their child is doing. If the child asked for help, parents were instructed to say: “I am sure you can do this, you can try out anything you want.”

Material and Procedure

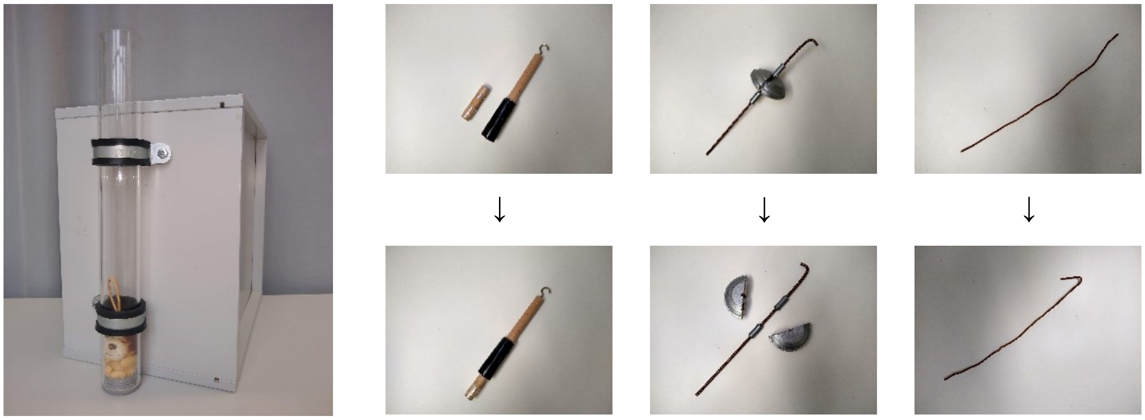
Following a short warm-up phase of free play between experimenter and child, the session began. Procedures differed between conditions.

Control Condition

In the Control Condition, children immediately started with the VTT. Following the procedure described by Breyel and Pauen (2021), they first explored the materials to be presented later for manufacturing a tool. Six items were presented in a wooden box with separate compartments. The cover of each compartment was opened separately and the child was allowed to explore the material for a maximum of 20 seconds before it was closed again. Three of the six items were distractors since they could not be used to solve the VTT. The other three items were potentially functional. However, they had to be modified in different ways to render them functional tools. The Reshape-item had to be bent to form a hook at one end. The Add-item consisted of two short sticks (one with a pre-formed hook at its end) which had to be connected to manufacture a tool long enough to reach the toy in the tube. The Subtract-item had a non-functional part that needed to be removed so that the tool could fit into the tube. After the child explored each item separately, the experimenter took the wooden box out of the child's sight and changed the arrangement of the items by switching their position in the compartments (see supplementary material Figure S1). Item arrangements for the exploration and the test phase were counterbalanced across participants.

Figure 1

Relevant Items and Their Necessary Transformations for the Vertical Tube Task



Apparatus Add-item Subtract-item Reshape-item

Note. The upper panel illustrates the items in their original form, and the lower panel shows the correctly modified tools.

Next, the apparatus was introduced (see Figure 1). It consisted of a small and narrow transparent tube containing a cuddly toy at the bottom. The tube was open only on its upper end and mounted on a wooden box to keep it upright. The experimenter said: “Look, there's a little lion stuck in there. We have to get the lion out.” Then, the experimenter placed the uncovered box containing the items within the child’s reach and said: “I have to do some work, but you can try on your own to get the lion out. You can try anything you want. I'm sure you can do it.” If the child did not use any of the items for three minutes, the experimenter again pointed out that all materials could be used and that the child could try out anything to get the toy out. If the child only used distractor items for three minutes, the experimenter suggested trying out a different item. If a child stopped working on the task for more than two minutes, the experimenter encouraged the child to continue: “I am still working on something. You can try to get the toy out of the tube on your own! I am sure you can do it.” If an item fell into the tube (i.e., blocking access to the target) and the child was not able to remove it within three minutes, the experimenter removed the item out of the child’s sight, and then the child was allowed to continue working on the task. Once the child retrieved the toy from the tube or after 10 minutes had elapsed, the observation stopped. When children did not succeed within 10 minutes, the experimenter helped them to find a solution.

Training Condition

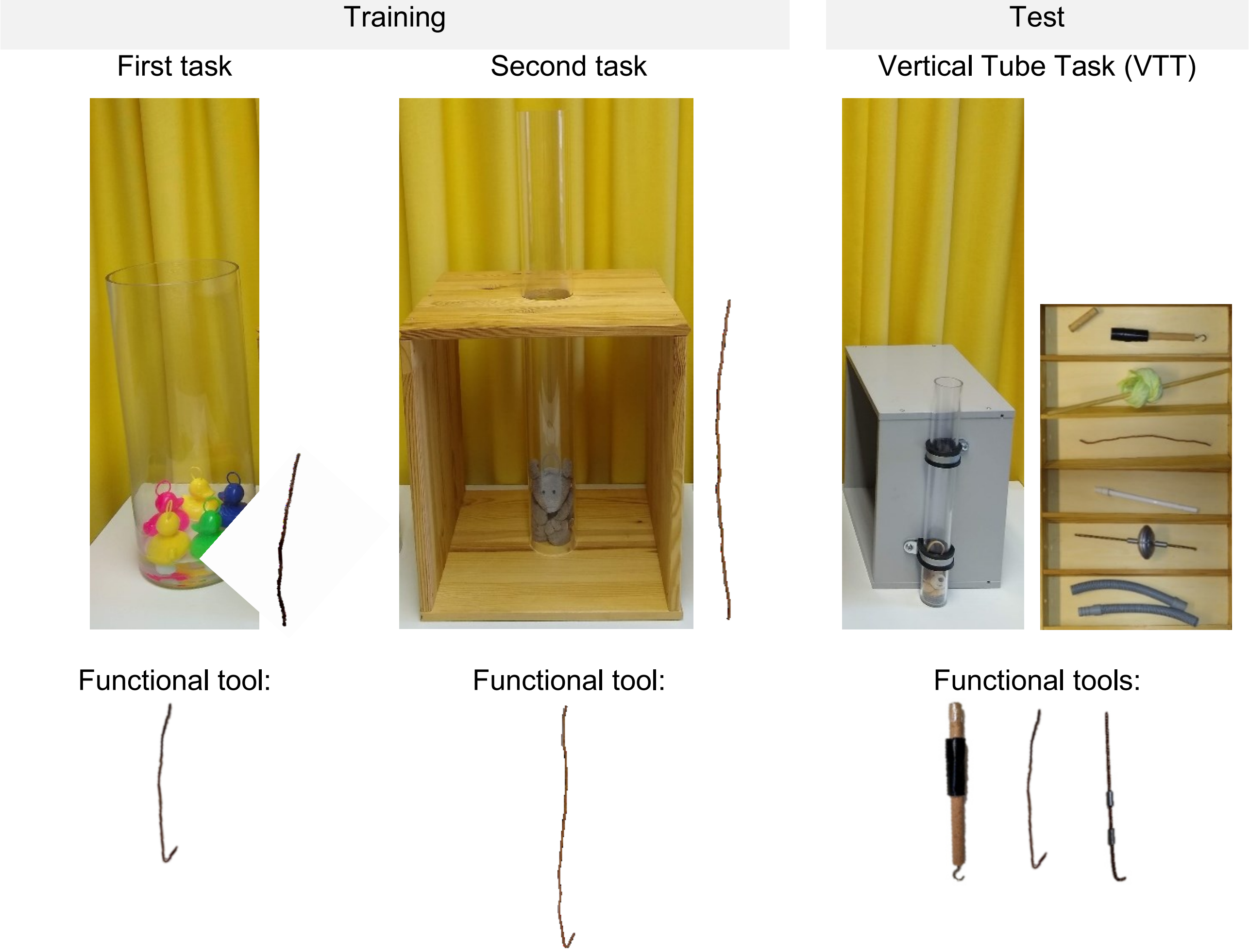
The Training Condition started by showing the child a small metal hook and asking whether the child knew what this was. If the child could not label the object, the experimenter said that this was a hook (hook introduction). In the next step (bending demonstration), a straight piece of soft wire was presented and the experimenter demonstrated how to bend a hook with this material, saying: “Look, what you can do with this: I hold it tight and then bend it around my finger. Now, I’ve got a small hook.” After reestablishing the original shape of the wire, it was the child’s turn (bending practice). If a child did not manage to bend a small hook, the experimenter encouraged him or her to try it again. This procedure was introduced to check whether children were able to manufacture a hook following an adult demonstration.

Next, the first analogical training task was administered. Five toy ducks had to be retrieved from a glass cylinder filled with water (see Figure 2). A straight piece of soft wire was placed next to the apparatus and the child was asked to get the ducks out without touching them with their hands. If the child did not use the wire spontaneously, the experimenter encouraged the child to try it out. If the child did not spontaneously bend the wire into a hook, the experimenter provided a hint: “Maybe you need a hook to get the ducks out.” If the child was not able to build an appropriate hook, the experimenter supported tool manufacturing. This task differed from the test task (VTT) in structural terms because the cylinder was much wider than the tube and filled with water, and multiple toys could be retrieved.

In the second analogical training task (see Figure 2), a large cuddly toy was placed at the bottom of a long transparent tube that was surrounded by a wooden rack. A long straight piece of soft wire was placed next to the apparatus and the child was again asked to get the toy out. The steps to support the child in solving the task were the same as in the first task (first step: verbal prompt, second step: help manufacturing a tool). The second training task closely resembled the test task in structural ways but was still different regarding the appearance and properties of the apparatus (i.e., the tube was much larger and mounted differently on a different-looking rack and the target object was a different kind of toy animal).

Figure 2

Apparatuses and Items for Training and Test Phase



Note. In the Control Condition, only the test task (VTT) was presented.

Coding

As the main goal of the present study was to learn more about the process of analogical transfer and potential causes of failure of younger children to show tool-manufacturing transfer, we used a differentiated coding system to analyze different aspects of performance.

Training Tasks

In the Training Condition, we coded whether children transferred the idea to build a hook (i.e., the child builds a hook without prompts from the experimenter) in each of the tasks and whether they succeeded independently or needed support from the experimenter.

Test Task: Vertical Tube Task (VTT)

Task Success. Children were coded as being successful if they removed the target from the tube within 10 minutes. When children did not need any of the prompts described in the procedure section, we coded this as an “independent success”. If children received any prompts, we regarded this as “supported success”.

First Item. To capture whether children instantly recognized that they need a hook, we coded which item was first inserted into the tube, differentiating between items with a hook (i.e., Reshape modified, Subtract, Add) and items without a hook (i.e., Reshape unmodified, distractor items).

Latency to Success. Latency to success was measured as the time between the start of the task (i.e., after the items were placed in reach of the child next to the test apparatus) and the complete removal of the target from the tube.

Manufacturing Method. The manufacturing method was categorized as Reshape (bending the wire into a hook), Add (prolonging the short wooden hook with a stick), Subtract (removing the non-functional part from the metal hook to make it fit into the tube), or a combination of items (i.e., pulling up the Add-item with the unmodified Subtract-item; see also Breyel & Pauen, 2021).

Type of Innovation. We categorized the solution by type of innovation in the following way (see Breyel & Pauen, 2021, p. 5): First-order Innovation was defined as “tool manufacturing that is clearly separated (temporarily and/or spatially) from the event or place where the target problem is solved with this tool”. Second-order Innovation was defined as “the combination of materials during an ongoing attempt to solve the target problem (i.e., performed at the same location)”. We further differentiated First-order Innovation into two subcategories regarding the underlying manufacturing process: If children used the original unmodified version of an item first and then adjusted their attempt by modifying the material, we called this S*tep-by-Step Tool Innovation*. If children modified the item set right away (without trying the original form of the item first) to use the functional tool, we called this *Direct Tool Innovation.* That way, we coded children’s tool behavior as an indicator for planning ahead which was assumed to be increased in Direct Tool Innovation.

Duration of Solution Attempts. An attempt was defined as a goal-directed activity to reach for or remove the target from the tube. An attempt started when the item or the child’s hand passed the opening of the tube and it ended when the item or hand was removed from the tube again or when the item fell into the tube. Based on an inspection of the videos, corresponding activities were classified as attempts (1) with modified items, (2) with unmodified (potentially functional) items, (3) with distractor items, (4) attempts to reach an item that previously fell into the tube, (5) and attempts to reach for the toy by hand. Since working times differed between children, we calculated relative durations by summing up all single attempt durations within a category and dividing this sum by total working time of the child.

Reliability

For the variables Success, First Item, Manufacturing method, and Innovation type, both coders agreed in all cases (Cohen’s Kappa = 1). For Latency to success and the relative duration of attempts, the second coder rated 13 children (25%), and the interrater agreement was excellent, ICC = .999 (two-way mixed, absolute agreement, single measures, 95 % CI = .999 – 1.000).

Hypotheses

In general, we expect children in the Training Condition to show better performance in the VTT than children participating in the Control Condition. More precisely, we expect that in the trained group, more children will solve the test task and they will be faster in finding a solution. We also expect children in the Training Condition to use the Reshape-item longer and in more functional ways (i.e., bending a hook) and to solve the task more often with the Reshape tool than children in the Control Condition. Regarding performance within the Training condition across the two tasks, we predict independent success rates to increase as a consequence of learning and practicing experiences.

Analysis

IBM SPSS Statistics (version 27) was used for statistical analyses. To test for differences between the two conditions we conducted Chi-squared tests regarding parent education and child gender, and we applied a Mann-Whitney-U test concerning child age since this variable was not normally distributed. For the within-group comparisons in the Training Condition, we conducted Cochran-Q tests and followed up on significant differences with pairwise McNemar’s tests, correcting the significance level for multiple tests according to Bonferroni. For the between-group comparisons regarding success rate and solution type, we conducted Chi-squared tests. If the expected cell frequencies were less than five, Fisher’s exact tests were applied. For 2 x 2 contingency tables (df = 1), we report the Chi-square- and p-values with continuity corrections due to Yates. Phi-coefficients (ϕ) were calculated as the square root of $\frac{\chi^2}{N}$ and are reported as a measure of effects size. Regarding latency to success and the relative attempt duration, we conducted Mann-Whitney-U tests since the data was not normally distributed. Pearson correlation coefficients (*r*) were calculated as the absolute value of $\frac{z}{\sqrt{n}}$ and are reported as a measure of effect size. Cohen's classification (1988) was used to assess the magnitude of the effects. It distinguishes between weak effects ($r / \phi = .10$), medium effects ($r / \phi = .30$), and strong effects ($r / \phi = .50$).

Results

Preliminary Analyses

The Training and Control Condition did not differ in terms of the age distribution of participants ($U = 301.00$, $z = -0.223$, $n = 50$, $p = .823$, $r = .03$), gender distribution ($\chi^2(1) = 0.00$, $p = 1.00$), and parent school education (mother: $\chi^2(1) = 1.34$, $p = .247$; father: $\chi^2(2) = 2.25$, $p = .416$). Hence, both groups were comparable in terms of important background variables.

Differences Between Conditions in the VTT

Success Rates

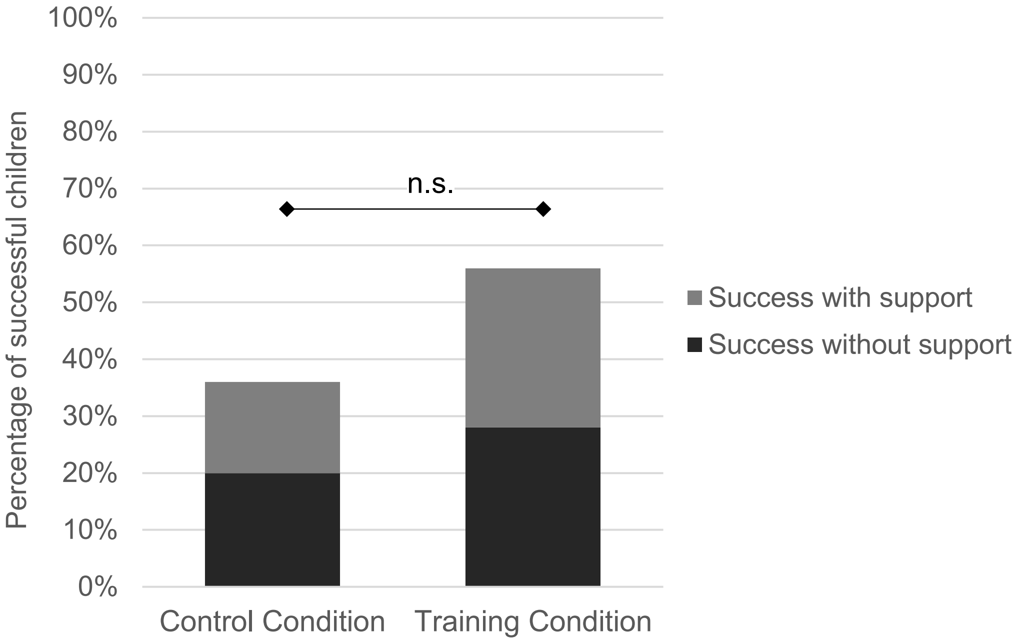
On a descriptive level, more children in the Training Condition solved the VTT than in the Control Condition (see Figure 3). However, this difference failed to reach the level of significance when looking only at independent success rates ($\chi^2(1) = .110$, $p = .741$, $\Phi = 0.05$), and when looking at supported success rates ($\chi^2(1) = 1.29$, $p = .256$, $\Phi = 0.16$).

First Item

Significantly more children in the Training Condition (84%) used an item with a hook (in the modified or original form) on their first solution attempt than in the Control Condition (20%), $\chi^2(1) = 18.03$, $p < .001$, $n = 50$, $\Phi = 0.60$. Furthermore, 56% (14 of 25) of the children in the Training Condition used the Reshape-item in a modified form (i.e., they formed a hook) on their first solution attempt, whereas none of the children in the Control Condition did, $\chi^2(1) = 16.77$, $p < .001$, $n = 50$, $\Phi = 0.58$.

Figure 3

Success rates in the VTT in the Control and Training Condition



Duration of Attempts

Children in both conditions spent an equal proportion of time with attempts to solve the task (see Table 1, *p* = .884). Trained children spent more time attempting to solve the task with modified material (*p* < .001, *r* = 0.69), but less time with unmodified material (*p* < .001, *r* = 0.52), distractor items (*p* < .01, *r* = 0.43), or by using their hands (*p* < .05, *r* = 0.35). The difference regarding the time spent with modified material was driven mainly by the use of the modified Reshape-item in the Training Condition, *p* < .001, *r* = 0.82. Children in both conditions spent a similar proportion of time to reach for items that had fallen into the tube, *p* = .818.

Table 1

Differences Between Conditions Regarding Attempt Durations (Relative to Total Working Time)

Attempt type	Control		Training		*p*	*U*	*z*	*r*
	Mean	SD	Mean	SD				
Total	0.42	0.15	0.43	0.18	.884	320.0	0.146	0.02
Unmodified	0.14	0.16	0.03	0.04	< .001	122.5	-3.694	0.52
Modified	0.06	0.11	0.30	0.20	< .001	560.0	4.856	0.69
Reshape	$2.72e^{-3}$	$1.36e^{-3}$	0.24	0.22	< .001	586.0	5.781	0.82
Add	0.03	0.10	0.05	0.13	.523	335.5	0.639	0.09
Subtract	0.02	0.05	$8.22e^{-3}$	0.02	.329	276.0	-0.977	0.14
Get fallen item	0.05	0.07	0.04	0.06	.818	301.0	-0.230	0.03
Distractor item	0.11	0.08	0.05	0.05	< .01	159.0	-3.011	0.43
Hand	0.05	0.13	$3.32e^{-3}$	$9.35e^{-3}$	< .05	200.5	-2.454	0.35

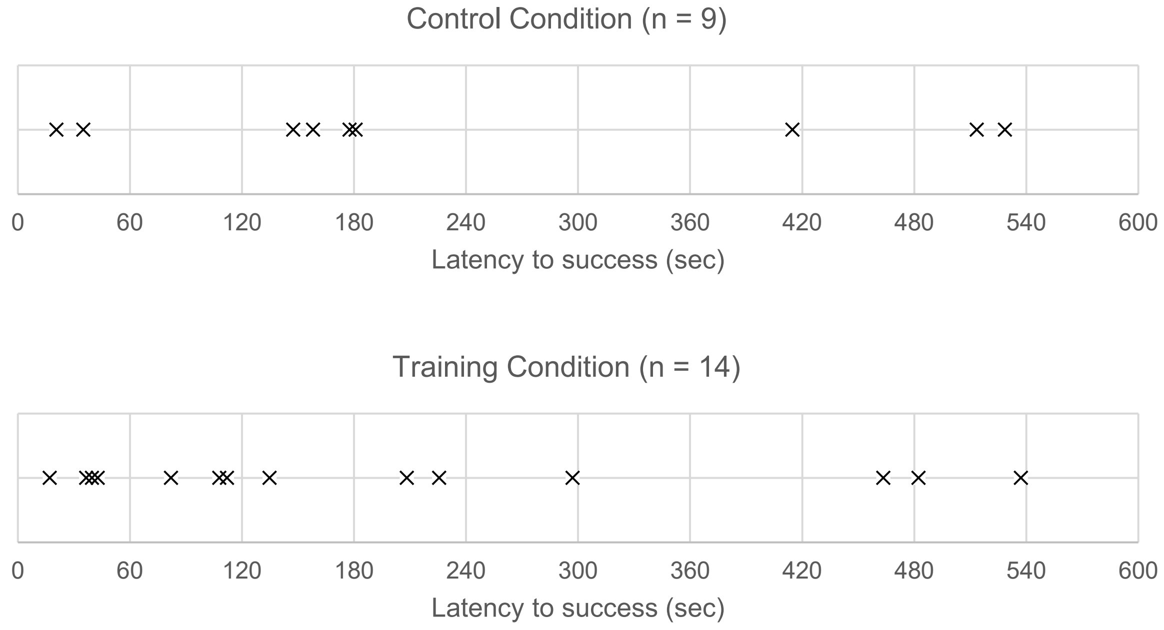
Note. Mann-Whitney U tests.

Latency to Success Among Successful Children **(N = 23)**

Descriptively, successful children in the Training Condition were faster in finding a solution (Mdn = 123.36s, IQR = 226.68, n = 14) than successful children in the Control Condition (Mdn = 177.60, IQR = 267.24, n = 9), but this difference did not reach level of significance, U = 55.00, z = -.504, *p* = .643. As Figure 4 illustrates, n = 17 children (collapsed across both conditions) solved the task within the first five minutes.

Figure 4

Latency to Success Split by Conditions



Note. Each cross represents a successful child.

Manufacturing Method and Innovation Type Among Successful Children

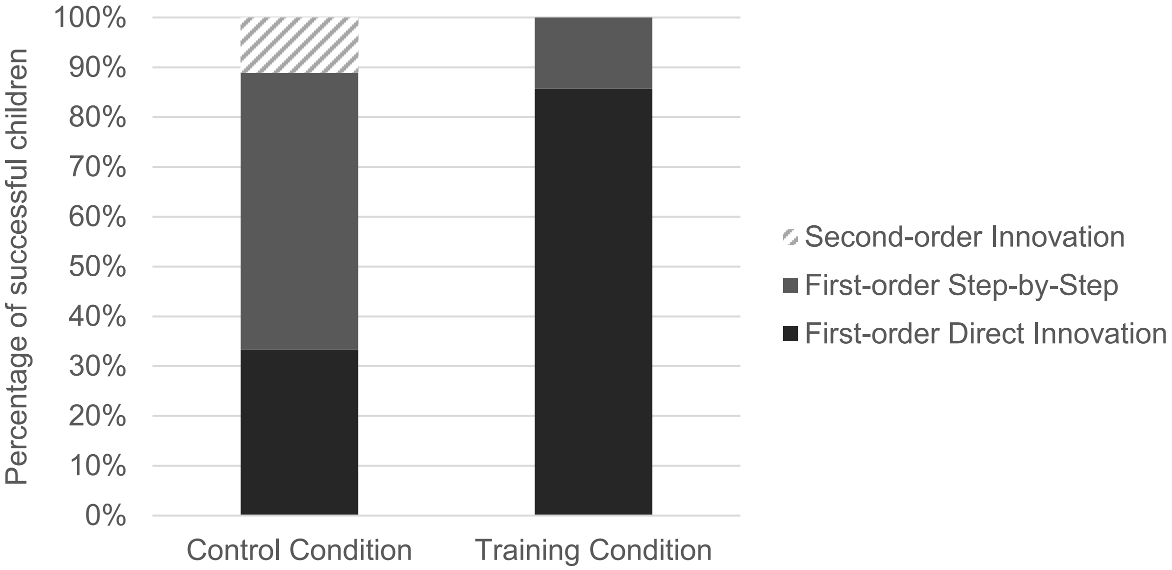
In the Control Condition, Subtract was the preferred strategy to build a tool (n = 5), closely followed by Add (n = 3). One child used a combination of the Add- and Subtract-item. Not a single child in this condition solved the task using the Reshape option. In the Training Condition, all three manufacturing options were used about equally often (Reshape and Add each n = 5, Subtract n = 4). The differences between conditions regarding manufacturing methods were not significant, $\chi^2(3) = 5.54$, $p = .119$, $\Phi = 0.49$.

Regarding the type of innovation, only one child of the Control Condition showed Second-order Innovation by pulling up the unmodified longer part of the Add-item with the Subtract-item. Among those who revealed First-order Innovation, 12 out of 14 children in the Training Condition (85.7%) and 3 out of 8 children in the Control Condition (37.5%) showed

Direct Tool Innovation, respectively (see Figure 5). A post-hoc analysis suggested that this group difference was significant, $\chi^2(2) = 6.57$, $p < .05$, $\Phi = 0.53$.

Figure 5

Type of Innovation in the Control (n = 9) and Training Condition (n = 14)



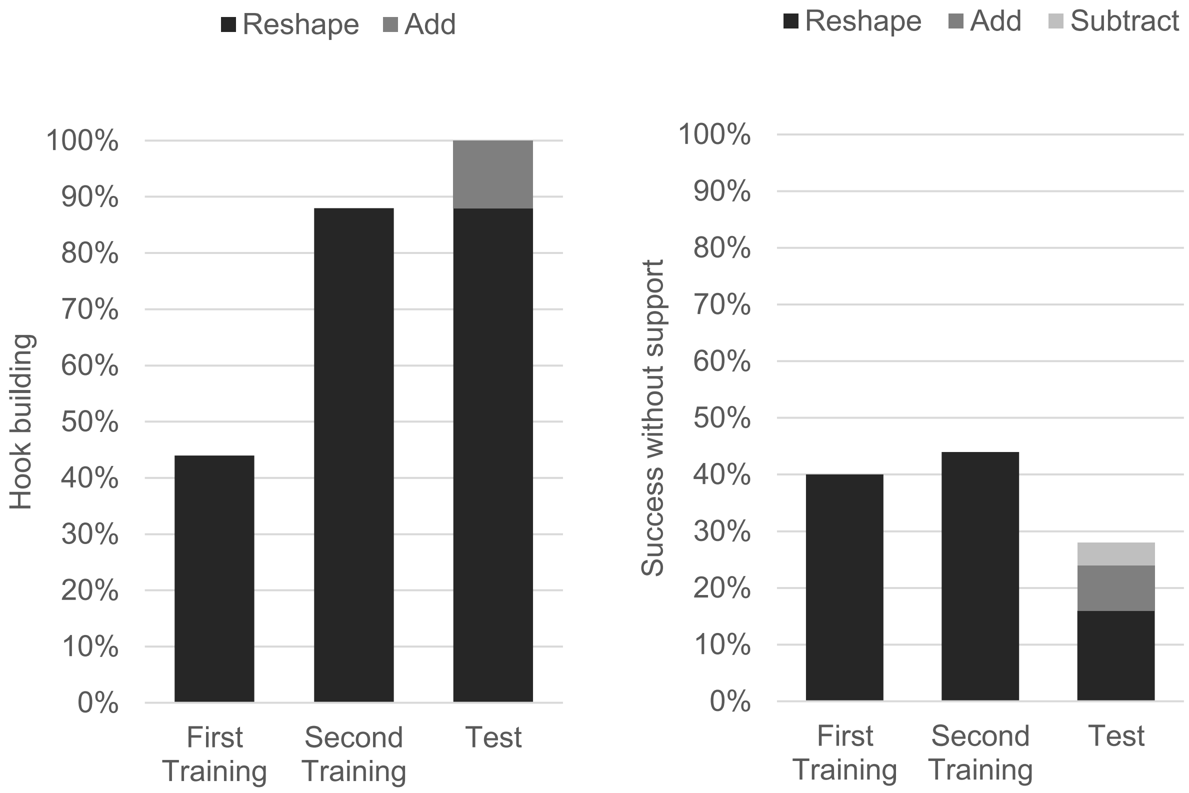
Performance Within the Training Condition

All 25 children bent the wire into a hook following the experimenter demonstration at the start of the training phase, with n = 12 succeeding on their first attempt and n = 13 needing two or more attempts until the hook had a proper shape. Hence, three-year-olds were able to form a hook with the material provided. A Cochran-Q test indicated that the proportion of children building a hook spontaneously differed between the three problem-solving tasks (first training task, second training task, VTT), $\chi^2(2) = 23.29$, $p < .001$. Follow-up tests showed that the increase in hook building from the first task to the second task ($p = .001$) and from the first to the third task ($p < .001$) were statistically significant, whereas there was no difference between second training and test ($p = .978$).

In sharp contrast, independent success rates were rather low and did not change significantly across the three tasks, $\chi^2(2) = 2.167$, $p = .338$ (see Figure 6). The majority of children (i.e., 18 of 25, 72%) needed support from the experimenter in either one or even both analogical training tasks in order to succeed (see Figure S2 in the supplementary material for more details on the kind of support needed). Together, these findings reveal that the majority of three-year-old children were able to learn how to bend a hook during the training phase and also transferred this approach across tasks. At the same time, only very few children were able to apply this knowledge successfully without help.

Figure 6

Percentage of Children in the Training Condition (n = 25) Building a Hook (left panel) and Solving the Respective Task Without Support (right panel)



Note. The manufacturing methods Add and Subtract were only available in the test task (VTT).

Descriptive Analysis: Reasons for Failure

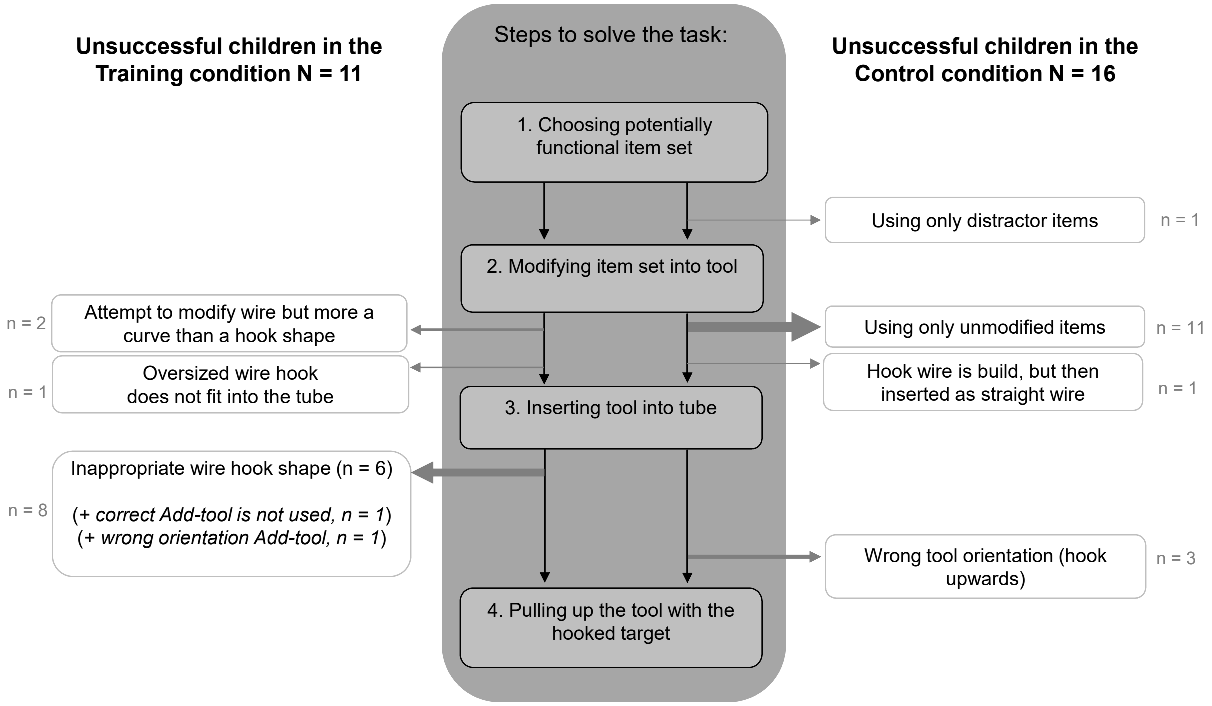
Given the unexpected result that success rates were not significantly higher in the Training Condition compared to the Control Condition even though children seem to have transferred the idea to build a hook across tasks, we describe in more detail why children in each condition failed at the VTT. In this context, it seems useful to differentiate action steps associated with success in the VTT: First, children need to choose a potentially functional item among multiple sets. Next, the material needs to be modified into a functional tool. Following that, the new tool needs to be inserted into the tube. Finally, the tool needs to be connected to the small loop attached to the toy in order to lift it. Figure 7 summarizes this process and reports how many children who failed in the VTT struggled during specific action steps.

As illustrated, children in the Training Condition had no problems with choosing a correct item but they struggled with producing an appropriate shape of a hook that needed to fit well into the tube and be narrow enough to catch the loop of the cuddly toy. Even though 10 of the 11 unsuccessful children tried to adjust the shape of the hook at least once after the first failed attempt, they were still not able to create the appropriate size and shape.

In the Control Condition, problems occurred much earlier in the task process, as the majority of the children used the items only in their original form (i.e., without manipulating them or manufacturing a tool). The minority of children who built a functional tool did not seem to realize its functionality as it was either deconstructed again before inserting it ($n = 1$) or inserted upside down ($n = 3$).

Figure 7

Action Sequence to Success and Reasons for Failure in the VTT



Discussion

The current study examined young children's capacity to transfer tool-making skills to analogous tasks. We found that three-year-olds were able to learn how to make a hook following the demonstration of an adult and spontaneously transferred this idea to similar tasks. However, despite having identified a correct solution approach, trained children were not more successful or faster in solving the test task compared to a control group with no prior tool-making experience.

The lack of significant differences in success rates between the two conditions may partly be due to the fact that the success rate in the Control Condition was surprisingly high (i.e., 36%). Based on the results of a previous study we had expected to find a comparably low success rate (i.e., 7%) among three-year-olds presented with the Vertical Tube Task (Breyel &

Pauen, 2021). This raises the question of what may account for this discrepancy in findings. Differing from Breyel and Pauen (2021), the experimenter intervened if an item fell into the tube and the child did not manage to remove it for three minutes in the present study. However, there was only one child in the Control Condition receiving this kind of support among the successful children. Thus, it seems unlikely that this difference can account for the higher success rate. Given the fact that all other aspects of the procedure were kept identical, differences in the sample need to be considered. The distribution of age, gender, and parent education in the two studies were comparable, however. As it has been shown that receptive vocabulary predicted children’s success in the Hook Task (Beck et al., 2016), this measure should be controlled in upcoming studies. In any case, the current results of the Control Condition call for replication before drawing well-reasoned conclusions about three-year-olds’ (lack of) tool innovation skills.

Looking at individual success rates within the training group across all three tasks, we found that the low performance of three-year-olds was comparable to the low success rates among three- to four-year-olds reported by Beck and colleagues (2014). However, as evidenced by a more detailed analysis of performance indicators, this does not necessarily reflect poor transfer skills. Rather, our findings reveal that children participating in the Training Condition transferred the idea to create a hook to similar tasks. Interestingly, their rate of spontaneous hook building was comparable in magnitude to that reported for older children’s task success rates (Beck et al., 2014; Gerson et al., 2018). Our finding is consistent with research on analogical problem solving, indicating that children as young as three years of age can transfer knowledge about using tools to similar tasks given some practice (e.g., Brown & Kane, 1988). Regarding knowledge transfer about tool manufacture, however, children's personal experience with manipulating the tool in a goal-directed way seems to play an important role. In line with this assumption, we found that only 40% of the children spontaneously bent a hook when presented with a goal (i.e., to remove the toys) in the first task, although they copied the bending action from the experimenter directly beforehand. Other studies confirm that coordinating knowledge about relevant material properties and necessary transformations is very difficult for children below five years of age, at least without getting the chance to use the ready-made tool themselves in advance (Cutting et al., 2014, Whalley et al., 2017).

The discrepancy between transferring the correct solution approach and the inability to implement it also becomes apparent when looking at the differences between the Training and Control Condition. Although the trained children chose the Reshape-item for the first attempt significantly more often, used it longer, and tried to modify it more often throughout the task, most children were not able to form a suitable hook shape with the Reshape-item to solve the task. We conclude from these data that children's failure was not due to a lack of idea generation or construction of the basic tool shape. Rather, appropriately adjusting the tool to the task demands seems to be a major challenge. This interpretation is consistent with another recent finding that children rarely adjusted a ready-made but oversized hook to solve the Hook task (Cutting et al., 2019). A potential explanation for why three-year-olds still struggle to modify the tool according to task demands might be that they mainly focus on the function of the hook and fail to consider the match between the tool's and the tube's size and form (see also Casler et al., 2011; Casler & Kelemen, 2007).

Concerning successful children, we found that 36% of the trained children (9 out of 25) solved the task using a manufacturing method they had not previously practiced (i.e., Add or Subtract), demonstrating the ability to abandon an unsuccessful strategy and successfully look for different approaches. This seems noteworthy, as previous work reported that transferring knowledge from one tool manufacturing method to another is difficult even for five- to seven-year-olds (Beck et al., 2014; Gerson et al., 2018). A facilitating factor for this kind of 'far transfer' in the current study may be that the items for the tool manufacturing strategies Add and Subtract had a visible hook (see Figure 1). As has recently been demonstrated, it seems to be easier for children to come up with the necessary modification (i.e., straightening out a curled part of the hook tool) when the hook itself is already visible (Neldner et al., 2017). Future studies should therefore examine to what extent material affordances influence the ability to transfer tool-manufacturing skills.

The majority of successful children in the Training Condition applied Direct Tool Innovation, i.e., they innovated a tool without trying out the original, unmodified version of the item first. In contrast, the majority of successful children in the Control Condition showed Step-by-Step Tool Innovation: First, they inserted the original version of an item (e.g., the Add-item), then they recognized it was not suitable (i.e., it was too short), and then they modified the item to render it a functional tool (i.e., connecting the stick to the short hook). Neldner (2020) suggested that the tool innovation process consists of an identification phase, an ill-structured problem phase, and an execution phase. In the identification phase, the problem and the affordances of the material are explored, followed by developing an innovation plan based on the imagination of an appropriate tool and the modifications necessary to achieve it. Our observations suggest that prior experiences with tool making in analogous tasks might enable children to skip the step of behaviourally exploring the affordances of the material because children have already developed a mental representation of what kind of tool they need and how it should look like. Further studies should examine the proposed phases of the tool innovation process more closely and identify which experiences might facilitate children’s performance in each phase.

Limitations and Future Directions

Designing appropriate tasks to study tool-making transfer is challenging. Based on the current work, we summarize which aspects should be considered by future studies. In our experiment, the training tasks and the test task differed in terms of the support provided by the experimenter. The procedure of the test phase was kept similar to previous studies using the VTT (see Breyel & Pauen, 2021; Voigt et al., 2019) to allow for comparisons across studies. In the training tasks, the experimenter had a more interactive role and thus offered more help to ensure that children were motivated to learn how to manufacture a hook. Given differences in prompting between the training- and the test phase, it is difficult to directly compare the amount of support children needed in each phase. Especially when working with younger children, future studies should standardize prompting rules for all phases of the experiment.

The design of new apparatuses and the choice of tool material need to be considered carefully. We intended to develop two training tasks that allow children to move and control the hook tool more easily inside the apparatus. This was achieved by using larger and wider tubes compared to the small, narrow tube of the VTT. However, these modifications seemed to entail other difficulties that we did not anticipate (see also Beck et al., 2014 for similar observations). For example in the second training task, the long part of the wire hook tool was often deformed when children tried to insert it into the tube, making it difficult or even impossible to control its movement inside the tube. Thus, before even getting the chance to catch the loop attached to the target toy, fine motor and coordination skills became crucial to successfully apply the tool. Even though the soft wire was easily deformed unintendedly, many children struggled with exerting the necessary force to bend the material far enough to form a small hook. In sum, our observations illustrate how specific features of the apparatus and the available items for tool manufacture can alter performance, thus calling for caution when comparing performance across different studies.

Finally, our findings raise questions about the optimal duration of the test phase. Most previous studies used a time frame of one (e.g., Beck et al., 2011; Cutting et al., 2011; Neldner et al., 2017) to two minutes, only (Nielsen et al., 2014; Whalley et al., 2017). Voigt et al. (2019) argued that this may not be enough for young children “to explore and play with ideas and to get into the problem” (p. 67). The authors found that providing five-year-olds with more time to work on innovation tasks led to higher success rates – however, only in combination with providing other manufacturing options than only Reshape. In the current study, we also presented different manufacturing options. A total of N = 17 (out of 23) successful participants solved the task within the first five minutes. Based on this finding, one could argue that less than 10 minutes are sufficient to discover the majority of successful innovators. However, the analysis of unsuccessful children's approaches and the strategies of "late solvers" is important to learn more about the innovation process and its underlying cognitive skills, as the present report illustrates. Corresponding research can deepen our understanding about which challenges young children are facing during tool innovation and which strategies they apply to deal with failed attempts.

Conclusion

Motivated by findings about children's analogical transfer skills in tool use, the present study investigated whether young children's tool *making* could be facilitated by providing them with practical experience in analogous tasks. The results revealed that while three-year-olds were able to transfer the correct solution approach to similar tool-making tasks, most of them failed to implement this approach successfully. Future research should aim for more differentiated assessments of young children's performance to promote our understanding of how they learn to coordinate cognitive representations of proper tools with perceptual feedback about their actions when trying to adapt chosen strategies to the specific demands of the problem.

Funding

This study was supported by a stipend from the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes) awarded to the first author.

Acknowledgments

We thank all families for participating in the study, the Baby-Uni Heidelberg team for their help in recruiting participants, and Dorothea Kieser for assistance in coding.

References

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., Cutting, N., Apperly, I. A., Demery, Z., Iliffe, L., Rishi, S., & Chappell, J. (2014). Is tool-making knowledge robust over time and across problems? *Frontiers in Psychology*, *5*(1395), 1–10. https://doi.org/10.3389/fpsyg.2014.01395

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 of London. Series B, Biological Sciences*, *371*(1690), 20150190. https://doi.org/10.1098/rstb.2015.0190

Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three- to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development*, *58*(101049), 1–20. https://doi.org/10.1016/j.cogdev.2021.101049

Brown, A. L. (1990). Domain-specific principles affecting learning and transfer in children. *Cognitive Science*, *14*(1990), 107–133. https://doi.org/10.1207/s15516709cog1401_6

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523. https://doi.org/10.1016/0010-0285(88)90014-X

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development*, *1*(2), 103–121. https://doi.org/10.1016/S0885-2014(86)80014-4

Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental Psychology*, *47*(3), 857–866. https://doi.org/10.1037/a0021174

Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition*, *103*(1), 120–130. https://doi.org/10.1016/j.cognition.2006.02.006

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 of London. Series B, Biological Sciences*, *368*(1630), 20120409. https://doi.org/10.1098/rstb.2012.0409

Connolly, K., & Dalgleish, M. (1989). The emergence of a tool-using skill in infancy. *Developmental Psychology*, *25*(6), 894–912. https://doi.org/10.1037/0012-1649.25.6.894

Crisafi, M. A., & Brown, A. L. (1986). Analogical Transfer in Very Young Children : Combining Two Separately Learned Solutions to Reach a Goal. *Child Development*, *57*(4), 953–968.

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. (2019). Is tool modification more difficult than innovation? *Cognitive Development*, *52*. https://doi.org/10.1016/j.cogdev.2019.100811

Frick, A., Clément, F., & Gruber, T. (2017). Evidence for a sex effect during overimitation: boys copy irrelevant modelled actions more than girls across cultures. *Royal Society Open Science*, *4*(12), 170367. https://doi.org/10.1098/rsos.170367

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, *7*(2), 155–170. https://doi.org/10.1016/B978-1-4832-1446-7.50026-1

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*(1), 47–59. https://doi.org/10.2307/1130388

Gerson, S. A., Burdett, E. R. R., & Beck, S. R. (2018). Preschoolers use analogy to facilitate innovative problem-solving. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1711–1716). Cognitive Science Society. https://cogsci.mindmodeling.org/2018/papers/0329/0329.pdf

Gönül, G., Takmaz, E. K., Hohenberger, A., & Corballis, M. (2018). The cognitive ontogeny of tool making in children: The role of inhibition and hierarchical structuring. *Journal of Experimental Child Psychology*, *173*, 222–238. https://doi.org/10.1016/j.jecp.2018.03.017

Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of Analogical Problem-Solving Skill. *Child Development*, *55*(6), 2042–2055.

McCarty, M. E., Clifton, R. K., & Collard, R. R. (2001). The beginnings of tool use by infants and toddlers. *Infancy*, *2*(2), 233–256. https://doi.org/10.1207/S15327078IN0202_8

Neldner, K. (2020). *The development and evolution of tool innovation in human children (Homo sapiens) and chimpanzees (Pan troglodytes): a cross-cultural and comparative investigation* [PhD Thesis, The University of Queensland]. https://doi.org/10.14264/uql.2020.949

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

Nielsen, M., Tomaselli, K., Mushin, I., & Whiten, A. (2014). Exploring tool innovation: A comparison of Western and Bushman children. *Journal of Experimental Child Psychology*, *126*. https://doi.org/10.1016/j.jecp.2014.05.008

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2012). The emergence of tool use during the second year of life. *Journal of Experimental Child Psychology*, *113*(3), 440–446. https://doi.org/10.1016/j.jecp.2012.06.001

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13*(4), 453–478. https://doi.org/10.1016/S0885-2014(98)90003-X

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 of London. Series B, Biological Sciences*, *371*(20150191). https://doi.org/10.1098/rstb.2015.0191

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition*, *105*(2), 362–379. https://doi.org/10.1016/j.cognition.2006.10.003

Träuble, B., & Pauen, S. (2011). Cause or effect: What matters? How 12-month-old infants learn to categorize artifacts. *British Journal of Developmental Psychology*, *29*(3), 357–374. https://doi.org/10.1348/026151009X479547

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

Whalley, C., Cutting, N., & Beck, S. R. (2017). The effect of prior experience on children's tool innovation. *Journal of Experimental Child Psychology*, *161*, 81–94. https://doi.org/10.1016/j.jecp.2017.03.009

Supplementary material

How Prior Experience in Analogous Tasks Affects Three-year-olds' Tool Making

Breyel, S., Pauen, S.

Department of Developmental and Biological Psychology, Heidelberg University

Contents:

Supplementary Method

- Arrangements of the items in the VTT: Figure S1

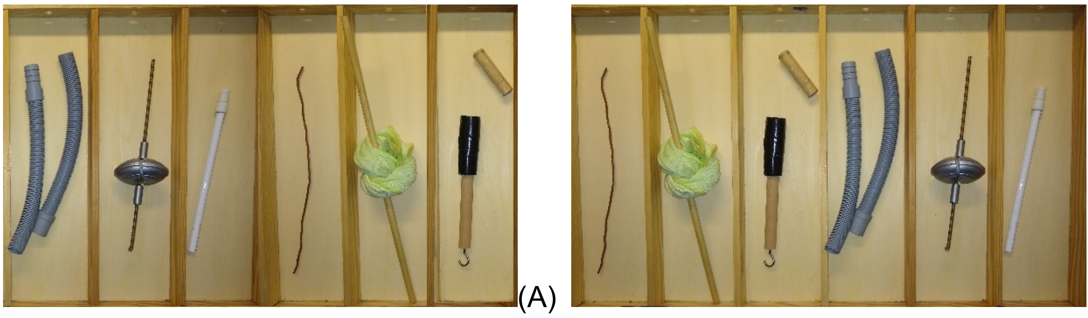
Supplementary Results

- Performance Within the Training Condition: Amount of Support: Figure S2

Supplementary Method

Figure S1

Two Item Arrangements for the VTT



(A) (B)

Note. A) From left to right: Distractor item 1, Subtract item, Distractor item 2, Reshape item, Distractor item 3, Add item. B) From left to right: Reshape item, Distractor item 3, Add item, Distractor item 1, Subtract item, Distractor item 2.

Supplementary Results

Performance Within the Training Condition: Amount of Support

Of the 25 children, n = 7 were successful without any support in both training tasks. Additional four children needed help in the first task but succeeded independently in the second task. Three children solved the first task without support but needed assistance in the second task. The remaining n = 11 children received help in both tasks, with the majority needing more help in the second task than in the first task (n = 10). This pattern of received support indicates that especially the second task was difficult to solve independently. Figure S2 shows how many children received which level of support in each training task.

Figure S2

Level of Support in each Training Task