Constructing Learning Trajectories: The Impact of Learning Progress on Task Similarity Preference

Brooke E. Jordan

Departmental Honors Thesis

Department of Psychology, The University of Texas at Austin

Faculty Mentor: Jacqueline Woolley, Ph.D.

Graduate Student Mentor: James Daly, M.A.

December 2023

Abstract

Prior research suggests that children are motivated to continue with learning activities when they experience improvements in performance, however it is less clear how these prior experiences entice children to pursue similar activities to further their learning. Research on children's learning preferences typically explores the impact of progression in learning on task selection as moderated by an additional motivation-based variable (e.g., intrinsic/extrinsic motivation, curiosity). However, there is little evidence on the extent to which task similarities may mediate the connection between learning progress and task preferences. As such, the present study investigated how learning progress and task feature similarities impact future learning preferences in 4- to 6-year-old children. To emulate the experience of learning progress, or a lack thereof, performance was experimentally manipulated such that half of the participants experienced improvements in their performance whereas the other half performed at a consistently high level. I found that there was not a significant difference between the distributions of similar versus novel future task selections between learning progress conditions. Additionally, while participants preferred tasks of character-based similarity over novel tasks across learning progress conditions, they preferred novel tasks more than tasks of procedure-based similarity. Both findings support the notion that task-feature similarity is not a strong motivator of future learning preferences.

Key Words: learning progress, intrinsic motivation, challenge preference, autonomous task selection

In the classroom, each child is unique in how they will learn, improve, and feel motivated to improve in their learning. Throughout the year, some students may receive consistently moderate grades and stagnate in their learning, whereas others may steadily grow and experience learning improvements. In designing curricula to engage students of all learning levels, it is important to consider how learning motivations differ between students who are not experiencing progress in their learning versus those who are. As we move toward providing children with more autonomy in determining their learning trajectories, it is crucial that we understand what specifically motivates children to pursue further learning, and how this may differ based on their prior learning experiences.

When discussing ways to maximize learning outcomes for children, it is important to understand the mediating factors that underlie the motivation to learn. In the US primary school learning environment, educators are beginning to adopt a more student-centered approach to teaching, in which students are granted more autonomy in the progression of their learning (Chong & Reinders, 2022; Lim & Park, 2022; Okada, 2023; Reeve et al., 2013; Reeve & Cheon, 2021; Waterschoot et al., 2019). In an autonomous learning environment, it is crucial to be able to present learning outcomes in a way that motivates students to *want* to further their learning about a concept. In this way, understanding the role of prior performance in motivating future learning about a concept is important for creating learning tools that motivate students at all levels of mastery. Studies of this nature tend to focus on the impact of progression in learning outcomes on choices of future task difficulty, but do not address task feature similarity as a potential mediator of this connection.

For the purposes of this study, learning progress will be defined as the experience of increased knowledge about a novel concept, wherein a person feels as though they have improved in performance. The term “learning progress” stems from literature on the Learning Progress Hypothesis: an AI-based theory of human behavior based in part on Vygotsky’s Zone of Proximal Development which posits that humans are intrinsically motivated toward intermediate difficulty learning tasks (i.e., tasks that are neither too easy nor too difficult) that further learning in a particular area (Oudeyer et al., 2007; Oudeyer et al., 2016). Studies of learning progress tend to focus on immediate within-trial decision-making to build a computational model of human behavior (e.g., Gottlieb & Oudeyer, 2018; Poli et al., 2022; Ten et al., 2021).

Research on progress-driven learning has focused mainly on the interaction between facets of intrinsic motivation (e.g., curiosity, uncertainty, personal interest, satisfaction) and learning outcomes (e.g., Gottlieb & Oudeyer, 2018; Oudeyer et al., 2016; Patall, 2012; Poli et al., 2022; Ten et al., 2021). Intrinsic motivation in the context of learning can generally be understood as a desire to learn for the inherent gratification of understanding a new concept independent of any material reward (Ryan & Deci, 2000a). In this way, intrinsically motivated learning requires that the learner feels personally enticed to pursue further knowledge about a concept. One method of promoting intrinsic motivation in learning is to allow students more autonomy in defining their own learning trajectories; a concept that has been implemented in primary education and continues to be a strong focus in the literature (Chong & Reinders, 2022; Lim & Park, 2022; Okada, 2023; Reeve et al., 2013; Reeve & Cheon, 2021; Waterschoot et al., 2019).

When discussing autonomy in this context, the term “autonomous task selection” is used in the present study to refer to one's freedom of choice in decision-making regarding future learning trajectories. Volition in defining one’s learning trajectory is believed to promote better learning outcomes with regard to internalization of learning content (Wangwongwiroj & Bumrabphan, 2021). Additionally, autonomy-supportive teaching – allowing students to determine their individual learning trajectories – is associated with positive learning outcomes (see Reeve & Cheon, 2021 for a review). Among these are increased classroom engagement, absorption of learning materials, academic achievement/grades, self-confidence, and vitality (Reeve & Cheon, 2021). Another factor to be considered in autonomous learning environments is a child's propensity to pursue challenge, or their challenge preference. Challenge preference has been measured using the learner's choice of task difficulty (e.g., Leonard et al., 2020; Leonard et al., 2022; Ten et al., 2021), or by using a standardized test of individual differences in the propensity to seek out challenges (e.g., Arend et al., 1979; Smiley & Dweck, 1994).

In the present study, I explored the extent to which future learning preferences are influenced by experiences of learning progress. Additionally, I examined whether the degree of task feature similarity (procedure- or character-based) and individual differences in challenge preference influence learning preferences. I hypothesized that, if a child experienced learning progress, they would be more likely to choose a similar future learning task. Conversely, if a child did not experience learning progress, they would be more likely to choose a novel future learning task. In the following section, I will review intrinsic motivation, learning progress, autonomous task selection, and challenge preference in depth, providing an understanding of how each relates to the present study.

Intrinsic Motivation

Prior to discussing learning progress, it is important to understand the link between intrinsic motivation and learning. Intrinsic motivation for learning refers to an internal drive to learn, independent of any material gain (Ryan & Deci, 2000a). Conversely, extrinsic motivation is a drive to learn which is contingent upon the attainment of some reward (Ryan & Deci, 2000a). Ryan and Deci (2000a) posit that the two forms of motivation are not functionally dichotomous and instead exist along a continuum. In line with their perspective, I will use the term "intrinsic motivation" to mean any motivation without tangible reward, though not necessarily entirely devoid of extrinsic motivation (Ryan & Deci, 2000a; Ten et al., 2021).

The most common subtypes of intrinsic motivation studied in learning progress research are novelty and curiosity, predominantly epistemic curiosity (e.g., Goupil & Proust, 2023; Laversanne-Finot & Oudeyer, 2021; Oudeyer et al., 2007). Novelty refers to the general quality of being new, and as such, a novel concept is any concept of which a learner has no prior knowledge. Epistemic curiosity refers to an internal drive to seek out new knowledge about a novel concept with the goal of decreasing uncertainty and increasing understanding (Lowenstein, 1994). Both concepts are crucial to developing an understanding of the factors that influence learning progress.

Given an innate drive to attain knowledge, one might think that humans would be intrinsically motivated to seek complete novelty, but this is not the case. Empirically-based computational models of human behavior posit instead that autonomous agents tend to weigh the feasibility of achieving learning in a specific area in conjunction with intermediate novelty to decide what to learn next (Laversanne-Finot et al., 2021; Oudeyer et al., 2007; Oudeyer & Kaplan, 2007). In other words, adult learners tend to gravitate towards tasks that are partially similar to – and, by convention, partially novel in relation to – the initial task (Oudeyer & Kaplan, 2007). The development of this tendency appears to occur between the ages of 4-7, as supported by Bonawitz et al.'s (2012) finding that children aged 4-5 tended to consistently prefer novel tasks while children aged 6-7 based their exploratory behavior on prior beliefs about the tasks. In the present study, two degrees of novelty were included within the choices of future tasks to perform next. In each set of three options for future tasks, one embodied complete novelty (bearing no similarity to the prior task), while the other two incorporated partial novelty that was either procedure- or character-based. This will be discussed in more depth in the Autonomous Task Selection section of this review.

The motivation to seek novelty is inherent to epistemic curiosity (Laversanne-Finot & Oudeyer, 2021). Epistemic curiosity is inextricably linked to metacognition – one's understanding of one's own knowledge about a topic – which is a crucial factor in intrinsically motivated learning (Goupil & Proust, 2023). In other words, a person must be aware of the state of their knowledge about a concept to be able to experience epistemic curiosity. Curiosity as a whole, though especially epistemic curiosity, is an inseparable facet of learning progress that has been incorporated into many of the computational models of human behavior (Gottlieb & Oudeyer, 2018; Goupil & Proust, 2023; Poli et al., 2022; Ten et al., 2021).

The notion of an intrinsic bias toward attending to information that is neither too simple nor too complex has been explored in the context of learning trajectories (e.g., Berlyne, 1950). In adults, Baranes et al. (2014) found that intrinsically motivated exploration of learning environments is influenced by intermediate levels of novelty, the difficulty of the task, and the number of choices presented. More recently, Wade and Kidd (2019) established a bidirectional link between epistemic curiosity and learning in adult participants. While a connection between intrinsic motivation and learning trajectories has been established in adults, there is a lack of research into the extent to which this impact extends to children. In the following section, I will discuss intrinsic motivation in the context of learning progress and introduce key findings that drive the present study.

Learning Progress

The learning progress hypothesis conceptualizes the brain as a “predictive machine” that uses an intrinsically motivated decision-making algorithm to maximize learning (Oudeyer et al., 2007; Oudeyer et al., 2016). This “machine” functions by focusing on intermediate difficulty tasks while actively avoiding too easy, too difficult, or entirely unlearnable tasks. Research on learning progress has focused predominantly on the impact of various forms of motivation (e.g., curiosity, uncertainty, intrinsic/extrinsic) on learning progress (e.g., Gottlieb & Oudeyer, 2018; Oudeyer et al., 2016; Patall, 2012; Poli et al., 2022; Ten et al., 2021). As previously mentioned, Wade and Kidd (2019) found a bidirectional link between learning progress and epistemic curiosity, meaning that perception of knowledge about a concept drives curiosity for further learning about that concept. In other words, a person’s metacognition regarding their own prior knowledge about a task increases their curiosity for new information, which in turn drives better learning outcomes.

Studies of learning progress typically feature some novel task with a learnable “rule” which participants are instructed to decipher in conjunction with some manipulated variable, such as motivation or the ability to make learning progress. An example of this experimental design is Ten et al. (2021). Ten et al. (2021) examined the impact of intrinsic vs. extrinsic levels of motivation and continuous feedback on adult participants' exploratory behavior across four levels of difficulty: easy, medium, hard, and unlearnable. Participants were presented with four families of monsters, which they could choose to engage with of their own volition, and were tasked with deciphering each monster's preference for one of two food items, which was determined by a combination of superficial features (e.g., monsters with big eyes preferred cherries, while those with small eyes preferred carrots). Additionally, participants were given either intrinsically- or extrinsically-motivated instructions for learning the food preferences. Participants in the intrinsic motivation group were instructed to complete 250 trials of the task with no further requirements, while the extrinsic motivation group was instructed to maximize learning about the rule-governed food preferences of each family of monsters, and that their understanding of each rule would be assessed at the end of the study. Level of difficulty was varied as a function of the amount of features that determined food preference, such that only one feature governed food preference in the easy condition, two were varied but only one determined preference for the medium condition, and both of the varied features determined food preference for the hard condition. The unlearnable task had no discernable pattern, meaning there was no possibility for learning progress. This study found an impact of intrinsically- versus extrinsically-motivated learning instructions on exploratory behaviors between tasks in adults. More specifically, they found that the extrinsically-motivated group demonstrated greater in-game challenge preference but less learning achievement. This was due to unnecessary focus on the unlearnable task which prevented them from engaging with the learnable tasks. The present study used an adapted version of this methodology for the initial learning task.

More recently, Leonard et al. (2022) examined how children challenge themselves based on feedback given during the learning process. They did so using a simulated tree with a rigged pulley system fixed to one of the branches. Participants were tasked with using the pulley system to pick up an “egg” at the base of the tree and pull it over the limb. The pulley system was designed such that the researcher could choose when the participant failed at the task by cutting the magnetic connection between the egg and the pulley. Feedback was varied such that half of the participants had incrementally increasing feedback (i.e., their performance improved from 25% to 50% to 75% completion), while the other half had static feedback (i.e., their performance remained at 75% completion throughout). After completing the tree task, the participants were asked whether they would like to continue the game with the taller tree (the harder condition) or change to a shorter tree (the easier condition), with varied motivations driving the choice of future task. This study found that children were more likely to choose a harder follow-up task than an easier one when they were given incrementally improving feedback, as opposed to consistently neutral feedback.

Each of these studies primarily investigated the impact of some motivating factor on future learning preferences, however neither examined how task-feature similarities may have mediated those preferences. In Ten et al. (2021), they examined how intrinsic versus extrinsic motivation impacted exploratory behavior between tasks of varied difficulty, however they did not examine whether the specific task similarities could have influenced exploration between learning tasks. As for Leonard et al. (2022), while they more directly investigated the influence of task performance on future learning preferences, the only option for future task selection was one that was both procedurally and superficially similar to the initial task, disallowing an examination of the impact of varied task-feature similarity on future learning preferences.

While learning progress heuristics have been fairly well researched in adult participants, there is a lack of research into how this concept applies to children. Abdelghani et al. (2022) explored curiosity-driven learning in children but focused primarily on developing a curiosity training software, not analyzing learning progress. Similarly, though recent research by Leonard and colleagues (2020, 2022) evidenced an impact of perceived task-progression on future task selection in children, their studies did not investigate the differential effect of procedure- versus character-based task similarities on task selection. I will explore this knowledge gap further in the following section.

Autonomous Task Selection

Learning progress tends to be operationalized based on participants' ability to make self-informed choices between tasks – an ability referred to as “autonomous task selection”.

In a naturalistic setting, autonomous task selection consists of allowing the learner some degree of control over their independent learning trajectory. Autonomy in determining learning trajectory has long been theorized to be essential for intrinsically motivated learning and the ability to personally engage with extrinsic motivations (Ryan & Deci, 2000b). This idea stems from self-determination theory, which proposes essential psychological “needs” such as competence and autonomy (Ryan & Deci, 2000b). The need for competence refers to the inherent desire to learn (i.e., epistemic curiosity) whereas the need for autonomy describes the general desire to make decisions of one's own volition. Both concepts are proposed to be crucial for humans to engage with their learning, making autonomy in task selection an important factor in the present study (Adams et al., 2017; Ryan & Deci, 2000b).

In adults, autonomy in future learning direction, as defined by the ability to choose future tasks, promotes better internalization of learned material as well as greater levels of challenge preference in future learning tasks (Baranes et al., 2014; Ten et al., 2021; Wangwongwiroj & Bumrabphan, 2021). Adults also prefer to sample a variety of tasks when given the opportunity rather than adhere to one specific task (Baranes et al., 2014; Ten et al., 2021). This further supports the notion that humans tend to prefer both intermediate challenge and novelty when given autonomy in learning trajectory (Oudeyer et al., 2007; Oudeyer et al., 2016).

Recently, there has been a strong push towards autonomy-supportive teaching across all levels of education, in which teachers are encouraged to provide students with some degree of choice in learning (Chong & Reinders, 2022; Lim & Park, 2022; Okada, 2023; Reeve et al., 2013; Reeve & Cheon, 2021; Waterschoot et al., 2019). In adopting an autonomy-supportive teaching style, it is important to understand how to present learning choices such that they will entice learners to further their knowledge about the desired learning outcome. While research has examined adults' exploratory behavior with novel tasks, less research has investigated the effect of various forms of task similarities, namely procedure- and character-based task similarities, on future learning preferences.

Superficial ("Character-based") and Procedural ("Procedure-based") Task Similarities

Superficial task features refer to details of the stimuli that are irrelevant to performance of the task itself, but are still pertinent to the learner's engagement with the task (i.e., character, environment; Chen, 1996). For the purposes of this study, the superficial similarities between tasks were always character-based, thus the terms "superficial" and "character-based" will be used interchangeably for the remainder of this review. On the other hand, procedural task features refer to the actual operational activity of the task (i.e., sorting, counting; Chen, 1996). For instance, if the initial task features a monster deciding which food to eat (recall the methodology for Ten et al., 2021), a superficially similar, but procedurally different, future task would feature the same monster engaging in a new task (e.g., the same monster but the task is playing hide-and-seek). In contrast, a procedurally similar, but superficially different, future task would feature a different character (e.g., a robot) engaging in the same prior task of deciding which food to eat.

Research on task similarities mediating autonomous task selection has not compared learning preferences for tasks of superficial versus procedural similarity. Chen (1996) found superficial and procedural task similarities to be independently beneficial to learning outcomes in children aged 5-8, however they did not examine which task similarity was preferred more by the participants. As described above, Leonard et al. (2022) incorporated one future task that was both procedurally *and* superficially similar to the initial task, again disallowing an analysis of children's preferences in degree of task similarity. For this reason, in addition to the aforementioned novel aspects of the future tasks, the present study incorporated one superficially similar, but procedurally distinct (character-based) future learning task as well as one that is procedurally similar, but superficially distinct (procedure-based), in the autonomous task selection portion of the experiment.

Approaches to Measuring Challenge Preference

Challenge preference is often incorporated into computational models of learning progress (e.g., Leonard et al., 2020; Leonard et al., 2022; Ten et al., 2021). In this context, challenge preference refers to an individual's tendency to seek out challenges. Challenge preference can either be conceptualized as task-specific challenge preference (i.e., how much a participant challenges themself within the experiment) or individual differences in ubiquitous preference for challenge (i.e., how much a participant challenges themself in their daily life). The former is measured by experimentally manipulating the difficulty of future tasks and allowing participants to choose which they prefer (e.g., the self-challenge measure in Ten et al., 2021). The latter measures the participant's tendency to challenge themself on a generic task, and relates that general tendency to the experiment. Research on challenge preference tends to focus on task-specific challenge preference, without addressing pre-existing individual differences in propensity for self-challenge (e.g., Leonard et al., 2020; Leonard et al., 2022; Poli et al., 2022; Ten et al., 2021). As such, the present study used the measure of individual differences in children's challenge preference developed by Arend et al. (1979) and replicated by Smiley and

Dweck (1994). This measure will be described in more detail in the Materials, Measures, and Manipulations section.

Conclusions

In this review, I have outlined the importance of intrinsic motivation, learning progress, autonomous task selection, and challenge preference to the construction of learning preferences in humans. Intrinsic motivation has been strongly connected to learning progress and exploratory behaviors in novel environments (Baranes et al., 2014; Bonawitz et al., 2012; Wade & Kidd, 2019). Studies of learning progress have incorporated aspects of autonomous task selection and dynamic feedback to mimic the naturalistic experience of learning progress (e.g., Leonard et al., 2022; Poli et al., 2022; Ten et al., 2021). Broadly, autonomous task selection promotes better generalized learning outcomes when incorporated in US Montessori schools (Reeve et al., 2014; Reeve & Cheon, 2021).

While there is literature on how various forms of motivation impact adults' learning progress, less research has examined how this effect applies to children (Gottlieb & Oudeyer, 2018; Oudeyer et al., 2016; Poli et al., 2022; Ten et al., 2021). Additionally, while there has been a strong focus on the impact of perceived learning progress on immediate learning, there has been far less emphasis on its impact on future learning trajectories. As such, the present study investigated how learning preferences in 4- to 6-year-old children are influenced by the experience of learning progress. In other words, I investigated whether the extent to which a child observes improvements in task performance impacts how they decide what they want to learn next. Additionally, I investigated whether (a) the degree of similarity of the future task alternatives to the initial task, and/or (b) individual differences in challenge preference have an impact on future task selection.

The guiding hypothesis of the present study was that, if participants experienced progress in their learning about an activity, they would be motivated to explore future tasks that were similar to the initial task. Conversely, if participants did not experience learning progress, they would choose the entirely novel future task. More broadly, I hypothesized that given perceived learning progress, children would gravitate towards future learning that incorporated similar task features to the initial learning task. Examining learning progress in this context sheds light on how children develop feature-driven learning preferences based on perceptions of learning progress in an autonomy-centered learning environment. In the following section, I will describe how each of these concepts were measured in the present study.

Methods

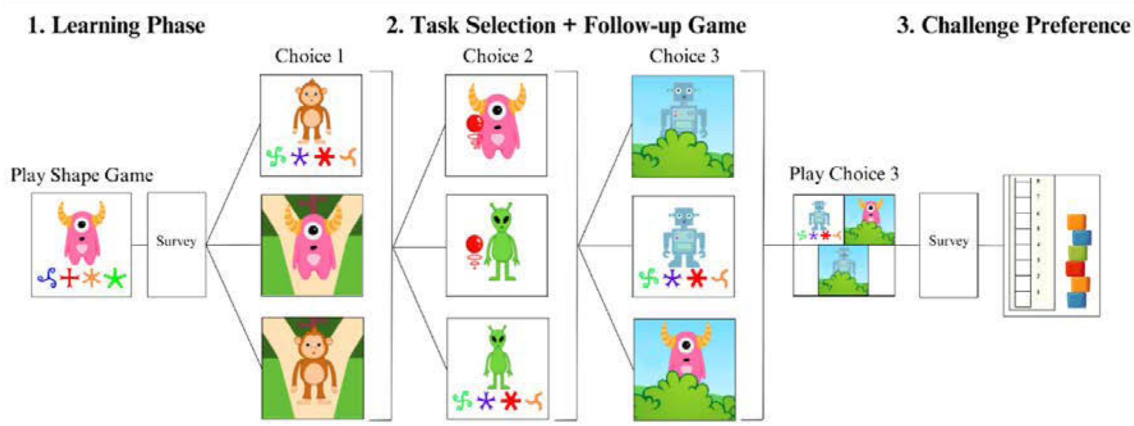
Design Overview

Given the lack of research on the mediational role of task-feature similarity on future learning preferences in children, the present study emulated an autonomous learning environment with future task choices containing varied feature similarities. The present study employed a between-participants design. For the primary research question, the independent variable was the experience of learning progress, and the dependent variable was the choice of future learning tasks. The secondary research question incorporated challenge preference as a secondary predictor variable. All methods and data analysis plans were pre-registered on AsPredicted, and an anonymous PDF of the published pre-registration can be accessed via this link: https://aspredicted.org/PK4_4P3. Participants were 4- to 6-year-old children.

To examine the impact of learning progress on preference for similar/dissimilar future learning tasks, this experiment consisted of three phases: a learning phase; a future task selection phase, which culminated in the participants engaging with the third future task of their choosing; and a general challenge preference measurement phase (see Figure 1). Prior to experimentation, the participants were randomly assigned to either the learning progress (LP) condition or the no learning progress (NLP) condition, as well as two additional counterbalancing conditions. In the LP condition, participants received experimentally manipulated feedback to demonstrate progress in their learning about a novel task, while in the NLP condition they received experimentally manipulated feedback to demonstrate no progress in their learning. The present study used two counterbalancing schemes to mitigate order effects. In one of the counterbalancing schemes, half of the participants played the initial game with the monster character, while the other half played the initial game with the robot character. In the other

Figure 1

Flowchart of the Experimental Design



Note. The flow of the experiment. 1. The learning phase, with either the learning progress (LP) or no learning progress (NLP) task and survey; 2. The three choices of future task, in which they play their third choice and complete a survey; 3. The measure of individual differences in challenge preference.

counterbalancing scheme, the participants received one of three possible orders of presentation of the future task options within each choice trial. In the learning phase, each participant was presented with a novel sorting task and asked to maximize correct responses while being presented with the experimentally manipulated feedback. The feedback ensured that half of the participants experienced learning progress, while the other half did not.

After a short survey to measure individual perceptions of learning progress, participants were presented with three opportunities to choose the task that they would like to complete next. Within each choice of future task, they had the option of either a task with procedure-based similarity, character-based similarity, or an entirely novel task. Participants then engaged with the third game they chose for a total of 24 trials to separate the initial game from the challenge preference measure, this time without summary feedback displaying their overall game performance. After the follow-up game, participants again responded to the same survey as the one following the initial learning phase, with slight wording adjustments to fit the task. Finally, participants were presented with a measure of individual differences in general challenge preference developed by Arend et al. (1979). This was used in a secondary analysis of the extent to which individual differences in challenge preference impact future task selection.

Participants

The final sample consisted of 110 4- to 6-year-old children (mean age = 65.4 months, s = 10.6, range = 48.5 – 83.8 months). Participant sex was equally split (n = 55 male, n = 55 female) and predominantly Non-Hispanic/Latinx (72.7% Non-Hispanic/Latinx, 21.8% Hispanic/Latinx, 5.45% did not report). The racial make-up was as follows: 55% White, 28% Multiracial, 11% Asian, and 6% did not wish to disclose. Participants were recruited from the University of Texas at Austin's Children's Research Center database, the Priscilla Pond Flawn Child and Family Laboratory at the University of Texas at Austin, and the University of Texas at Austin's Child Development Centers. The pre-registered sample size of n = 48 was based on a power analysis for independent samples of dichotomous data conducted using the "pwr" package in RStudio version 4.2.0 (R Core Team, 2022). These results suggested that a sample size of n = 48 would be required to detect an effect size of $w \geq 0.477$ at a significance level $\alpha = .05$ and power $1-\beta = .80$, for a chi-square test of independence. In the process of data collection, expectations for the number of participants that we could test were exceeded. As such, data collection was extended to the week of October 23-27, 2023, allowing leeway to ensure the sample included even distributions of participants across conditions. Following this modification, the final sample size for the present study was 110 4- to 6-year-old children. Children with learning disabilities (e.g., ADHD) and those who did not complete the future task selection portion of the experiment were excluded from analysis (n = 3 and n = 3, respectively). Conditions were randomly assigned prior to experimentation. Compensation for participation in the study consisted of a small toy or sticker.

Materials, Measures, and Manipulations

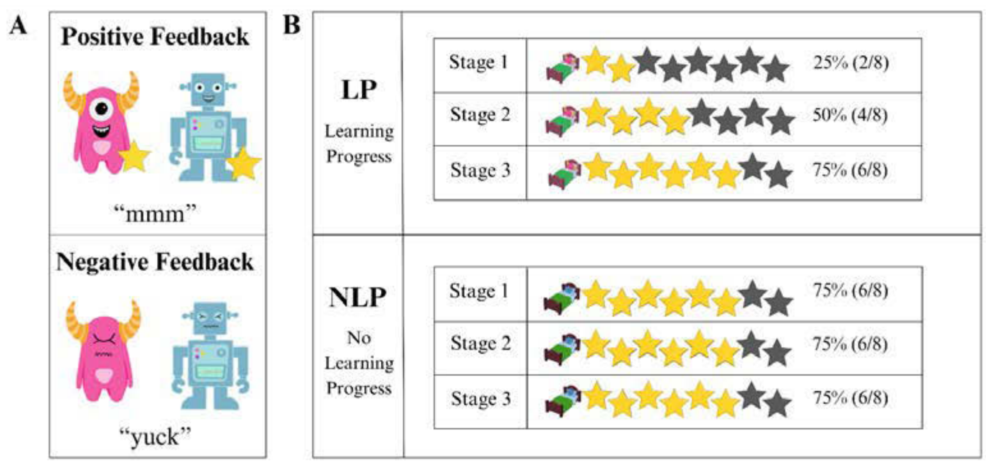
The computerized materials for this study were assembled using jsPsych, a library for creating web-based JavaScript experiments (de Leeuw, 2015). The full experiment code is available in the monster-shape repository of the Austin Thought Lab Github in the file shape-game, accessible via this link: https://github.com/AustinThoughtLab/monster-shape. The script used by the experimenters is located in Appendix A. The computerized portion of the experiment was administered via MacBook Airs (OS 13.5.1, Generation 8,2), and the self-challenge measure was administered using wooden blocks and an illustrated measuring board. More information regarding each of the measures is provided in the subsections below.

Learning Progress Manipulation

To measure learning progress, I used an adapted version of the monster task from Ten et al. (2021). For each trial of the initial computer game, participants were presented with either a monster or a robot – counterbalanced between participants – and four possible shapes, which they were tasked with matching to the character's shape preferences (Figure 1). Each time they selected a shape, the character would either say "Mmm," accompanied by an animation in which a star rolls up to the monster, or "Yuck," with a shaking animation (see Figure 2A). The goal of the task was to feed the monster as many of the shapes that he likes to eat (i.e., that elicits the "Mmm" and star reaction) before he goes to bed each night, for a total of three nights, or stages. At the end of each night, the participants were presented with experimentally manipulated feedback demonstrating how many stars they received that night (see Figure 2B). This feedback was manipulated such that participants in the LP condition got 25% of the choices correct (or 2/8 stars) on the first night, 50% (or 4/8 stars) on the second night, and 75% (or 6/8 stars) on the third night, mimicking the experience of learning progress. These percentages were modeled on Leonard et al.'s (2022) study, in which identical percentages were used to emulate learning progress. Participants in the NLP condition consistently got 75% correct (or 6/8 stars), mimicking a lack of progression in learning. At the end of the third night, the fabricated feedback from each of the nights was presented in tandem and explained to the participant to further emphasize the progression of their learning.

Figure 2

Experimentally Manipulated Feedback Presented to Participants



Note. **A**. The randomized feedback after each shape selection throughout the learning phase. **B**. The experimentally manipulated percent correct summary feedback presented at the end of each stage.

Survey

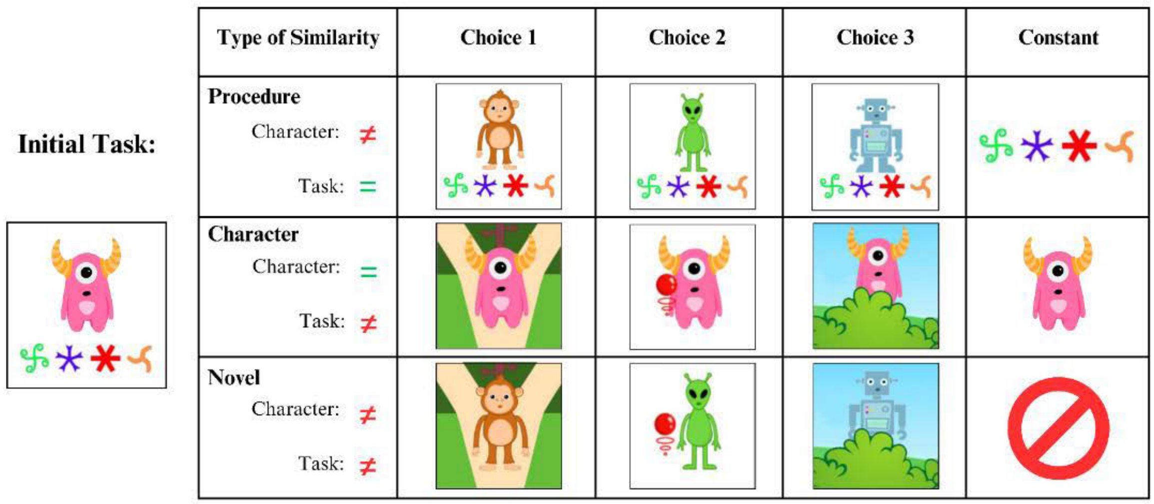
To investigate participants' subjective experience of the task, I conducted a survey directly following the initial and follow-up tasks. The survey consisted of four questions read out by the experimenter, all of which are listed in the script (Appendix A). The first question asked the participants how much they liked the game, with the intention of recording their subjective experience of the game as a whole. To directly assess the success of the manipulation, the second question asked participants to consider how much they felt as though they learned, with the options of either "A lot," "A little," or "Not at all." They were also asked why they felt this way, and their responses to this question were manually coded *{note to Dr. Jones: we haven't manually coded yet so I'll elaborate on the process once we do}*. The last question, serving as a manipulation check, asked how many stars they got throughout the game with the options of "More each day", "Same each day", or "Less each day". If they reported their performance accurately, all participants in the LP condition would say "More each day", while all participants in the NLP condition would choose "Same each day".

Future Task Preference Measure

Following either the experience of learning progress or no learning progress and the follow-up survey, the participants' choices of future tasks were measured. As depicted in Figure 3, at each choice opportunity, participants were presented with three task options: one of procedure-based similarity (same task, different character), one of character-based similarity (same character, different task), or one that was entirely novel (different character, different task). Participants were given three opportunities for future task selection. The decision to have three independent choice trials was made as this amount would provide the most amount of data while still allowing participants to anchor their selections on their experience of LP/ NLP in the

Figure 3

The Type of Similarity of Each Future Task Selection to the Initial Task



Note. The initial character was counterbalanced such that roughly half of the participants started with a monster whereas the other half started with a robot. The future task choices were adjusted accordingly.

initial task, without straying too far away from the initial task manipulation. Each choice was coded and analyzed against the experimental condition, degree of similarity to the initial task, and individual differences in challenge preference.

Second Computer Task

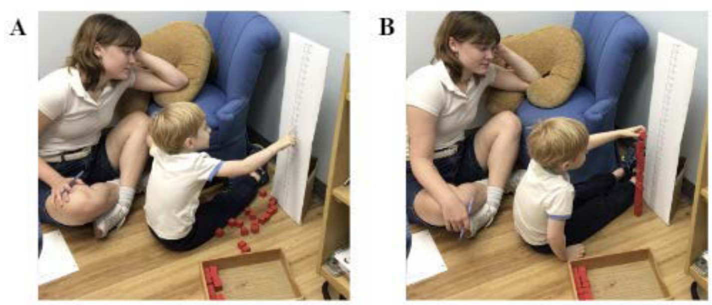
In the second computer game, participants engaged with their third choice of future task, which was either the same shape task with a different character, or a novel hiding game with the same character as the initial game for 24 total trials. Regardless of which initial condition participants were assigned to, all participants received continuous experimentally manipulated learning progress feedback in the same increments as the initial task (i.e., 2/8 stars for the first 8 trials, 4/8 stars for the next 8 trials, and 6/8 stars for the final 8 trials), though in this game the trials were not split into stages. No summary feedback was presented to the participants nor was it explained how many trials they got correct/incorrect. This decision was made to allow for an exploratory analysis of the impact of the summary feedback on the survey responses. Upon completion of this game, the participants were directed to a survey that was equivalent to the initial survey, with minor wording alterations to fit the differences in task and lack of summary feedback.

Challenge Preference Measure

Individual difference in challenge preference was measured using a tower-building task developed by Arend et al. (1979). In this task, participants were presented with 30 2.54 cm wooden blocks, and were asked to estimate how tall they believed that they could build each succeeding tower (see Figure 4). For a total of six trials, three measures were collected: their height estimate for the tower, measured in hypothetical number of blocks tall; their actual performance, measured in the number of blocks tall prior to the end of the trial; and their reason for stopping (e.g., the tower fell, they stopped building). Challenge preference was determined by taking the mean difference between their performance estimate on each trial and their actual performance on the prior trial, as captured by the following equation:

Figure 4

A Participant Performing the Self-Challenge Measure



Note. **A**. Participants provided a height estimate for their first tower. **B**. Participants built their first tower.

$$SC = \frac{\sum_{t=2}^{6} (estimate_t - actual_{t-1})}{n_{towers\ total} - 1}$$

Exclusion Criteria

Participants were excluded from analysis if they did not complete the follow-up task selection portion of the experiment (*n* = 4), if they explicitly stated that they believed their performance feedback was predetermined (e.g., "I think this game is rigged."; *n* = 0), if they fell outside of the target age range of 4 to 6 years old (*n* = 0), or if they had been diagnosed with a learning disability (e.g., ADHD, ASD; *n* = 3). All exclusion criteria were pre-registered and there were no additional exclusions.

Procedures

Prior to the participant's arrival at the Austin Thought Lab, the experimenter entered the participant's predetermined condition into the computer game to ensure the participant was tested on the correct condition. The experimenter also prepared a printed version of the script (Appendix A), as well as a sheet designed for documenting the data from the challenge preference measure, including a sheet explaining how to record specific occurrences with the task (Appendix B).

Upon arrival, the legal guardian(s) of the participant completed an IRB-approved consent form. Once informed consent was collected from the guardians, the participant was taken to a testing room and asked if they were ready to begin the experiment. With verbal assent from the participant, the researcher began the experiment on the laptop. To begin, a recording explained the game to the participant (Appendix A). During the instructions, the experimenter modeled one correct and one incorrect selection, at the prompting of the recording. The researcher then asked the participant if they understood the task, and if they were ready to play.

During the learning phase, participants were instructed to point to the screen and have the experimenter click the shapes for them. If the participant was confident with the touchpad, they were allowed to progress on their own up to the survey. The participants played through the entire game, with pre-recorded summary feedback played after every 8 trials. As all feedback describing their learning progress was pre-recorded, the experimenter did not intervene until the survey portion of the initial task.

After completing all three stages of the initial learning task, the condition-specific summary feedback was presented and explained to the participant in a pre-recorded message. The researcher then administered the follow-up survey, reading off each question and allowing time for the participant to respond. After the survey, the researcher advanced to the task selection portion of the study, which was also entirely pre-recorded. Between each choice of future task, the experimenter asked the participant why they chose that specific task. Participants were then allowed to play their third choice of follow-up task for a total of 24 trials. After these 24 trials, the experimenter administered the same survey as the initial learning phase survey, with minor wording alterations to fit the task (Appendix A).

After the second survey, the laptop was set aside and the participant was asked to move to the floor for the remainder of the experiment. The experimenter then explained the self-challenge task to the participant following the script. Once comfortable with the demands of the task, the participant built six towers, each time providing an estimate of how tall they believed that they could build the next tower. Upon completion, the participant was allowed to choose a toy/sticker from the laboratory's treasure chest, and the guardian(s) were debriefed on the experimental manipulation.

Statistical Analysis

All analyses were conducted in RStudio 4.2.0. Given that the data were predominantly categorical, the main statistical tests used were chi-square tests of independence and logistic regressions. For the primary research question, I conducted a chi-square test of independence to examine whether condition predicted choices of future tasks. To test if self-challenge – either independently or in conjunction with condition – predicted future task choices, a logistic regression was used.

As a manipulation check, an ordinal logistic regression was conducted to see if the participant's condition predicted how many stars they thought that they received throughout the initial game (more, the same, or less each day). If the manipulation worked, condition should predict their response to the stars question, as participants in the LP condition all received more stars each day, while participants in the NLP condition all received the same amount of stars each day. To ensure the efficacy of counterbalancing the order of presentation of the choices to each participant during experimentation, I used a chi-square test of independence to determine whether order predicted choices of future tasks. Similarly, to test if the initial character (Monster vs. Robot) had an impact on their choices, a chi-square test of independence was conducted. This test was chosen for both confounding variable checks as the predictor and outcome variables in both cases were nominal.

Results

Confounding Variables Check

Though both initial character as well as the order of presentation of future task options were counterbalanced, multiple analyses were run to rule out potential confounding variables. First, a chi-square test of independence found that children who played with the monster first did not make significantly different choices than those who played with the robot first, $X^2(3, N = 110) = 0.83$, $p = .841$. Similarly, a second chi-square test revealed that there was not a significant effect of the order of presentation of future tasks within choice trials on future task selections, $X^2(6, N = 110) = 2.59$, $p = .858$.

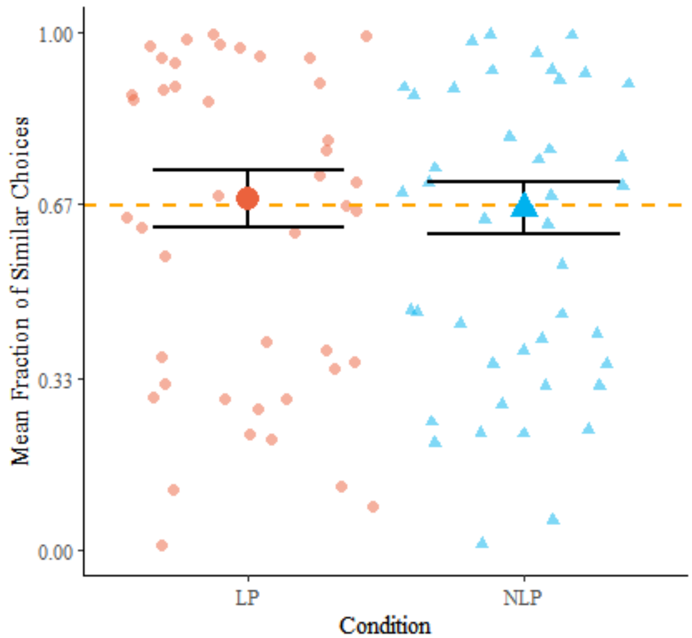
Learning Progress and Future Task Selections

The primary between-participants comparison was whether the participant's learning progress condition had an impact on their choices of similar versus novel future tasks. The choice data were analyzed in two different ways, both yielding similar results. The first method of analyzing the choice data was to combine all choices into a fraction of how many similar versus novel future task selections the participant made across all three choices. Each participant received a score of 0/3, 1/3, 2/3, or 3/3 depending on how many choices out of the three were similar to their initial game. The fractions were averaged across participants in each condition, as seen in Figure 5. A Fisher's Exact test ($p = .221$) did not indicate a significant difference between groups in choices of similar versus novel future tasks.

The second method of analysis used only their first choice of future task, as this choice was most closely connected to their performance feedback. The count data alone demonstrate a remarkably similar distribution of first choice selections between participants, as portrayed in Table 1. A Welch's 2-Sample t-test confirmed that there was not a significant difference between conditions in first choice similarity, $t(108) = .21$, $p = .837$. As such, regardless of how the data are categorized, there was no impact of condition on similar versus novel task selections.

Figure 5

Mean Fraction of Similar Future Task Selections by Condition



Note. Means were calculated using the fraction of similar tasks participants chose across all three choices, separated by condition. The orange line indicates the adjusted chance value, as ⅔ of future task choices were similar, whereas only ⅓ were novel. Error bars were constructed using a 95% confidence interval.

Table 1

Count of First Choice by Similarity and Condition

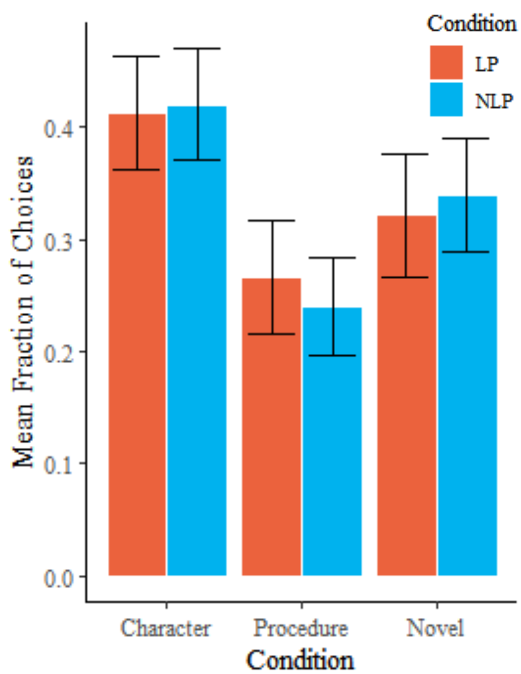
	Novel	Similar
LP	16	39
NLP	17	38

Note. The "Similar" category collapses across two choices, so there was a ⅔ chance of choosing a similar task.

A secondary analysis of the choice data was run to see if children were more inclined to choose specific aspects of task-feature similarity (Figure 6). Splitting similarity into character-based and procedure-based similarity, a chi-square test of independence between all three types of similarity indicated a significant difference between them, $X^2(2, N = 330) = 12.8$, $p = .002$. To determine where the difference occurred, three pairwise chi-square tests were run between the total number of character-based, procedure-based, and novel future task selections. The alpha level for these analyses was adjusted to $p = .017$ in alignment with a Bonferroni correction for pairwise tests. The results suggested that, across conditions, participants chose tasks of character-based similarity significantly more than those of procedure-based similarity, $X^2(1, N = 330) = 12.8$, $p < .001$, though the difference between character based similarity and entirely novel tasks failed to achieve significance with the Bonferroni correction, $X^2(1, N = 330) = 4.04$, $p = .044$. The difference between choices of procedurally similar versus novel tasks was not significant, $X^2(1, N = 330) = 2.53$, $p = .112$. Analyzing just the first choice of future task with these categories of

Figure 6

Mean Choice Similarity of Future Task Selections to Initial Task by Condition



Note. Means were calculated using the fraction of character-based, procedure-based, or novel future tasks participants chose across all three choices, separated by condition. Error bars were constructed using a 95% confidence interval.

similarity yielded equivalent results, $X^2(1, N = 110) = 9.47$, $p = .002$ (character: procedure); $X^2(1, N = 110) = 4.25$, $p = .039$ (character: novel); and $X^2(1, N = 110) = 1.1$, $p = .294$ (procedure: novel).

Challenge Preference Measure

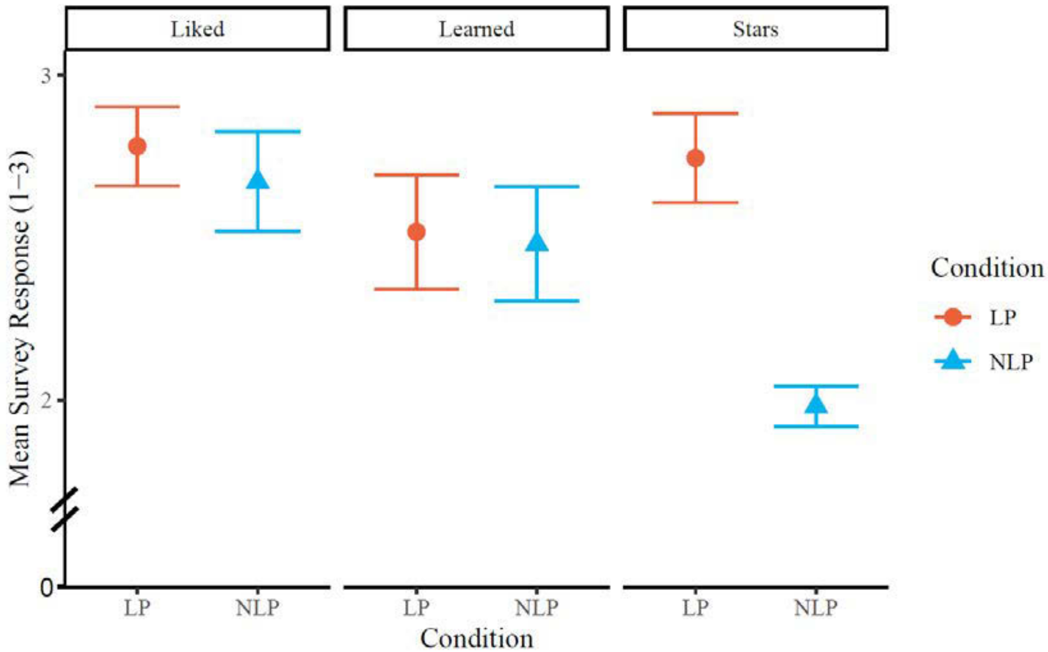
To analyze whether individual differences in challenge preference impacted choices of future tasks, a Point-Biserial Spearman Correlation test was run using self-challenge as an independent predictor of fraction of similar future task selections. These results did not suggest a significant correlation between the self-challenge measure and frequency of similar future task selections, $r_{pb}(109) = -.08$, $p = .438$. Additional Point-Biserial Spearman Correlation tests were run splitting type of similarity into character-based and procedure-based, again finding no correlation between self-challenge and future task similarity choices, $r_{pb}(109) = -.02$, $p = .850$ and $r_{pb}(109) = -.04$, $p = .694$, respectively.

Survey Data Analysis

The initial game survey responses were averaged for each condition, and the results are presented in Figure 7. The question regarding how many stars participants thought that they received throughout the game served as the primary manipulation check. A chi-square test of participants' responses to the manipulation check revealed that they were aware of their performance depending on their condition, suggesting that the learning progress manipulation was successful, $X^2(2, N = 110) = 9.94$, $p < .001$. Contrastingly, in analyzing their subjective perceptions of learning, a chi-square test of independence found that the participant's learning progress condition did not significantly impact how much they felt that they learned about the concept, $X^2(2, N = 110) = 0.25$, $p = .882$. Finally, an exploratory analysis found that condition

Figure 7

Mean Survey Responses by Condition



Note. Mean responses to survey questions by condition. For the "Liked" and "Learned" questions: 3 = a lot, 2 = a little, 1 = not at all. For the "Stars" question: 3 = more each day, 2 = same each day, 1 = less each day. Error bars were constructed using a 95% confidence interval.

did not significantly impact how much participants liked the game, $X^2(2, N = 110) = 1.47$, $p = .479$.

Discussion

The aim of this study was to investigate how learning progress interacts with future task similarity preferences in children. I examined these concepts with the hypothesis that children who experience learning progress will prefer similar future tasks, whereas children who do not experience learning progress will prefer novel future tasks. I found that learning progress and individual differences in challenge preference did not impact task similarity preferences in 4- to 6-year-old children. In other words, children who experienced learning progress or a lack thereof did not differ in whether they pursued similar or novel future tasks. This finding is directly contrary to both of my hypotheses, as participants in both conditions chose similar and novel future tasks at or near chance value.

In a secondary analysis, I found that children preferred tasks of character-based similarity significantly more than tasks of procedure-based similarity across conditions. Additionally, a replication of Arend et al.'s (1979) measure of individual differences in challenge preference did not predict children's choices of future tasks. In an analysis of the survey data, I found that children were aware of their performance over time, in alignment with previous research (Leonard et al., 2020; Leonard et al., 2022). However, I also found that participants did not differ in their subjective perceptions of learning, providing interesting insight into how children understand and internalize performance feedback. Possible interpretations of the primary and secondary analyses are discussed in the following sections.

Learning Progress and Future Task Selections

The most obvious interpretation of these results is that experience of learning progress does not impact task similarity preferences in children of this age range. As there is no direct prior research on children's preference for task feature similarity in the context of learning progress, it may just be the case that task feature similarity preferences are not mediated by whether children experience progress in their learning. In the context of the learning environment, this would mean that the similarity of learning tasks to each other does not need to be individually catered to a child's objective performance on a learning outcome. Additionally, the finding that participants preferred tasks of character-based similarity significantly more than procedure-based similarity across conditions indicates that instructors should attempt to incorporate similar superficial task features throughout learning outcomes to promote engagement. Though it may simply be the case that learning progress does not impact future task similarity preference, it is worthwhile to consider facets of the experiment itself that may have moderated these findings.

The first question to consider is why learning progress did not impact task selections in either condition. One explanation for the lack of between-group differences in task selections as compared to previous literature is that the present study utilized a computer-based design, whereas recent learning progress literature with children has used more tactile measures (e.g., Leonard et al., 2020; Leonard et al., 2022). While Leonard et al. (2022) found an effect of learning progress on task selection across both computer-based and tactile modalities, their computer task involved a touch-screen game in which children were able to engage freely with the game. Contrastingly, participants in the present study were instructed to point to the shapes that they wanted to select, and experimenters manually interacted with the game for the participant. This separation of agency in engaging with the game itself may have skewed participants' engagement with the game. While there is extensive research into how specific types of educational games (e.g., immersive, puzzle-based, multiplayer; Xi & Hamari, 2019) promote autonomy in children's learning, little research has examined whether physically engaging with the game itself impacts engagement. As such, it is possible that participant's lack of physical autonomy in the game may have mitigated the effect of the learning progress feedback on task similarity preference.

Another key difference between the present study and previous learning progress literature is the manner in which performance feedback was presented throughout the game. Recall that Leonard et al. (2022) used a pulley with a magnetic egg attached to it, and experimentally manipulated when the egg would fall from the pulley in the process of attempting to raise the egg over a branch of a toy tree. Each time the egg fell, the experimenter would place a marker to indicate the height that the participant was able to pull the egg on that trial. In this way, performance feedback was built into the mechanism of the game itself and reinforced by continuous summary feedback of performance over time, which was available to the participant throughout the full game. While the present study attempted to emulate this with the inclusion of stars, positive/negative trial-by-trial feedback, and summary feedback at the end of each block, participants were not able to see their full-game performance continuously throughout the game in the way that participants in Leonard et al.'s (2022) study could. This difference could have influenced how children internalized their performance over the course of the game itself. Future research should examine whether the inclusion of a progress bar or some other metric of continuous feedback improves subjective perceptions of learning.

Internalization of Performance Feedback

Stepping back from specific feedback metrics, it is interesting to consider the possible interpretations of the finding that children were reasonably aware of the rate of change of their performance across the game itself but did not report a subjective experience of learning in alignment with this feedback. In other words, how is it that participants were aware that they received incrementally increasing feedback in the learning progress condition and stagnant feedback in the no learning progress condition, but that this awareness did not impact how much they felt that they learned across conditions? Little is known about how children or adults conceptualize learning progress (Oudeyer et al., 2016), making it all the more fascinating that children felt as though they learned both in the presence and absence of learning progress feedback.

One possible interpretation of this finding is that the participants felt more familiar with the game by the end of the three blocks and internally correlated this familiarity to their learning. It is well established that children of this age range tend to overestimate the scope and breadth of their prior knowledge about a recently learned or familiarized novel concept (see Xia et al., 2023 for a review). As such, it may be the case that children in this study used their recent familiarity with the game itself to overestimate their learning throughout the game without necessarily attending to the objective performance feedback itself. Alternatively, participants may also have just viewed the act of receiving stars in any capacity as indicative of some level of learning, regardless of the presence or absence of growth throughout the game. That being said, this still begs the question as to why they are not viewing the objective performance feedback as indicative of learning progress.

One potential explanation for the misalignment between objective and subjective reports of performance is that the feedback was experimentally manipulated, making it impossible for children to experience genuine learning. That being said, the finding that children reported their subjective perception of learning in the game between a little and a lot in both conditions demonstrates that they did feel as though they were learning in both conditions, even without the ability to do so. In other words, children still felt that they learned even without the possibility of genuine learning, so the fact that the feedback was fixed likely was not the main reason for the observed distinction between participants' objective and subjective reports. Additionally, experimentally manipulated feedback is standard practice in cognitive developmental research and has been used reliably in the specific context of learning progress research with children (e.g., Leonard et al., 2020; Leonard et al., 2022). With this, we can reasonably reject the notion that the lack of potential to learn led to the misalignment in performance reports.

It is also worth considering whether the metacognitive load of asking participants how much they learned over the course of a three-block game is too substantial for children of this age range. While it is generally accepted in the field that theory of mind begins to develop around the ages 3-4 years old, metacognitive abilities – and more specifically, implicit metacognitive processing – does not arise until 5 years of age (see Whitebread & Neale, 2020 for a review). Though the present study included 40 4-year-olds, there were no significant age effects in reports of subjective learning. As such, while it could still be true that the metacognitive load of the question itself could have been too much for 4- to 6-year-old children, the 4-year-old participants did not skew this finding. More research is necessary to understand children's ability to integrate performance feedback into implicit metacognitive processing of learning progress.

Assuming children of this age range are able to reflect on their own knowledge, it is interesting to consider whether this specific type of performance feedback is intrinsically indicative of learning to them. The present study used stars for both trial-by-trial and summary feedback, which was intended to indicate their performance over the course of the game. This decision was made in line with Leonard et al.'s (2020) finding that children are more sensitive to the rate of change of their performance than their total amount of correct or incorrect guesses, as well as the fact that stars are commonly used to indicate positive feedback. In the present study, children were aware of their rate of change in performance based on the stars they received, but this did not translate to their subjective perception of learning progress. This calls into question the efficacy of performance metrics such as stars in inducing an intrinsic feeling of learning progress in children. Considered in the educational context, stars and stickers are often used to reward children for good behavior. It is worthwhile to consider how children internalize this feedback, and whether they are able to understand that these performance metrics are intended to indicate learning and improvement. Future research should examine how children understand performance metrics and how learning can best be indicated to children.

In essence, these findings suggest that children do not prefer similar or novel tasks significantly more depending on their performance on a previous task. However, children generally prefer superficially similar tasks significantly more than procedurally similar tasks. Individual differences in challenge preference did not impact task feature similarity preference. More research should be conducted on differential internalization of various performance metrics in the context of learning progress.

Limitations

One of the aforementioned limitations of the present study is the fact that performance was experimentally manipulated, making it such that participants could not experience genuine learning. This decision was made to control for individual differences in intellectual acuity and ensure equal distributions of children who either did or did not experience learning progress. While experimentally manipulated feedback is common practice in the field, it now appears that this may not be indicative of learning progress to children. As such, future studies should attempt to incorporate genuinely learnable novel concepts in the context of learning progress. Another potential limitation is that children in the no learning progress condition overall received more stars than those in the learning progress condition (18 total for NLP: 12 total for LP). This decision was made in alignment with the aforementioned Leonard et al. (2020) finding that children attend more to the rate of change of their performance than to their accumulative performance feedback. This potential confound could be mitigated in future studies by using a different proportion of correct guesses each day such that each condition accrues the same number of stars by the end of the game. Finally, not allowing participants to engage with the game themselves may have impacted their perceived autonomy, therein impacting their perceptions of learning. Related studies should incorporate a more interactive experimental design mechanic (e.g., touch-screen computer or tablet-based format) to promote a sense of agency in engaging with the game.

Future Directions

Moving forward, the primary question that still remains is how humans conceptualize learning progress and use this feedback to inform future decision making. This is an ongoing question in the field that warrants further study. Relatedly, future research should examine how various forms of performance metrics differentially induce subjective perceptions of learning progress in children. For instance, understanding whether continuous within-game feedback produces stronger perceptions of learning than discrete post-game summary feedback, or whether either type of feedback is fundamentally representative of learning to children, would inform how future studies present learning progress feedback to children. The latter highlights children's implicit metacognitive processing abilities concerning their own learning progress as another area for further research. In all, we have to first understand *if* children are able to comprehend when they have progressed in their learning before being able to ask *how* children conceptualize learning progress.

References

Abdelghani, R., Oudeyer, P.-Y., Law, E., de Vulpillières, C., & Sauzéon, H. (2022). Conversational agents for fostering curiosity-driven learning in children. *International Journal of Human-Computer Studies*, *167*, 102887. https://doi.org/10.1016/j.ijhcs.2022.102887

Adams, N., Little, T. D., & Ryan, R. M. (2017). Self-Determination Theory. In M. L. Wehmeyer, K. A. Shogren, T. D. Little, & S. J. Lopez (Eds.), *Development of Self-Determination Through the Life-Course* (pp. 47–54). Springer Netherlands. https://doi.org/10.1007/978-94-024-1042-6_4

Arend, R., Gove, F. L., & Sroufe, L. A. (1979). Continuity of Individual Adaptation from Infancy to Kindergarten: A Predictive Study of Ego-Resiliency and Curiosity in Preschoolers. *Child Development*, *50*(4), 950–959. https://doi.org/10.2307/1129319

Baranes, A. F., Oudeyer, P.-Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnins.2014.00317

Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *British Journal of Psychology*, 41(1-2), 68-80.

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

Chen, Z. (1996). Children's Analogical Problem Solving: The Effects of Superficial, Structural, and Procedural Similarity. *Academic Press, Inc*. https://doi.org/10.1006/jecp.1996.0037

Chong, S. W., & Reinders, H. (2022). Autonomy of English language learners: A scoping review of research and practice. *Language Teaching Research*, 13621688221075812. https://doi.org/10.1177/13621688221075812

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47(1), 1-12. doi:10.3758/s13428-014-0458-y.

Eun, B. (2019). The zone of proximal development as an overarching concept: A framework for synthesizing Vygotsky's theories. *Educational Philosophy and Theory*, *51*(1), 18–30. https://doi.org/10.1080/00131857.2017.1421941

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), Article 12. https://doi.org/10.1038/s41583-018-0078-0

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, *231*, 105325. https://doi.org/10.1016/j.cognition.2022.105325

Kane, S. P. (2019). Sample size calculator. Retrieved April 1, 2023, from https://clincalc.com/stats/samplesize.aspx

Laversanne-Finot, A., Péré, A., & Oudeyer, P.-Y. (2021). Intrinsically Motivated Exploration of Learned Goal Spaces. *Frontiers in Neurorobotics*, *14*. https://www.frontiersin.org/articles/10.3389/fnbot.2020.555271

Leonard, J. A., Cordrey, S. R., Liu, H. Z., & Mackey, A. P. (2022). Young children calibrate effort based on the trajectory of their performance. *Developmental Psychology*. https://doi.org/10.1037/dev0001467

Leonard, J. A., Sandler, J., Nerenberg, A., Rubio, A., Schulz, L. E., & Mackey, A. P. (2020). Preschoolers are sensitive to their performance over time. *Proceedings of the 42st Annual*

Conference of the Cognitive Science Society. Cognitive Science Society. https://cognitivesciencesociety.org/cogsci20/papers/0312/0312.pdf

Lim, Y., & Park, H. (2022). Who have fallen behind? The educational reform toward differentiated learning opportunities and growing educational inequality in South Korea. *International Journal of Educational Development*, *92*, 102599. https://doi.org/10.1016/j.ijedudev.2022.102599

Metcalfe, J., Schwartz, B. L., & Eich, T. S. (2020). Epistemic Curiosity and the Region of Proximal Learning. *Current Opinion in Behavioral Sciences*, *35*, 40–47. https://doi.org/10.1016/j.cobeha.2020.06.007

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Chapter 11 - Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In B. Studer & S. Knecht (Eds.), *Progress in Brain Research* (Vol. 229, pp. 257–284). Elsevier. https://doi.org/10.1016/bs.pbr.2016.05.005

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*. https://doi.org/10.3389/neuro.12.006.2007

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265–286. https://doi.org/10.1109/TEVC.2006.890271

Okada, R. (2023). Effects of Perceived Autonomy Support on Academic Achievement and Motivation Among Higher Education Students: A Meta-Analysis. *Japanese Psychological Research*, *65*(3), 230–242. https://doi.org/10.1111/jpr.12380

Patall, E. A. (2013). Constructing motivation through choice, interest, and interestingness. *Journal of Educational Psychology*, *105*(2), 522–534. https://doi.org/10.1037/a0030307

Poli, F., Meyer, M., Mars, R. B., & Hunnius, S. (2022). Contributions of expected learning progress and perceptual novelty to curiosity-driven exploration. *Cognition*, *225*, 105119. https://doi.org/10.1016/j.cognition.2022.105119

Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist*, *56*(1), 54–77. https://doi.org/10.1080/00461520.2020.1862657

Reeve, J., Vansteenkiste, M., Assor, A., Ahmad, I., Cheon, S. H., Jang, H., Kaplan, H., Moss, J. D., Olaussen, B. S., & Wang, C. K. J. (2014). The beliefs that underlie autonomy-supportive and controlling teaching: A multinational investigation. *Motivation and Emotion*, *38*(1), 93–110. https://doi.org/10.1007/s11031-013-9367-0

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Smiley, P. A., & Dweck, C. S. (1994). Individual Differences in Achievement Goals among Young Children. *Child Development*, *65*(6), 1723–1743. https://doi.org/10.2307/1131290

Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's Understanding of Knowledge Acquisition: The Tendency for Children to Report That They Have Always Known What

They Have Just Learned. *Child Development*, *65*(6), 1581–1604. https://doi.org/10.1111/j.1467-8624.1994.tb00837.x

Ten, A., Gottlieb, J., & Oudeyer, P.-Y. (2021). Intrinsic Rewards in Human Curiosity-Driven Exploration: An Empirical Study. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/13b6p5ms

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, *26*(4), 1377–1387. https://doi.org/10.3758/s13423-019-01598-6

Wangwongwiroj, T., & Bumrabphan, K. (2021). Self-Determination Theory: Statistical Correlations Between Motivational Regulations and Basic Psychological Needs. *International Journal of Higher Education Pedagogies*, *2*(2), Article 2. https://doi.org/10.33422/ijhep.v2i2.39

Waterschoot, J., Vansteenkiste, M., & Soenens, B. (2019). The effects of experimentally induced choice on elementary school children's intrinsic motivation: The moderating role of indecisiveness and teacher–student relatedness. *Journal of Experimental Child Psychology*, *188*, 104692. https://doi.org/10.1016/j.jecp.2019.104692

Westera, W. (2019). Why and How Serious Games can Become Far More Effective: Accommodating Productive Learning Experiences, Learner Motivation and the Monitoring of Learning Gains. *Journal of Educational Technology & Society*, *22*(1), 59–69.

Whitebread, D., & Neale, D. (2020). Metacognition in early child development. *Translational Issues in Psychological Science*, *6*(1), 8–14. https://doi.org/10.1037/tps0000223

Xi, N., & Hamari, J. (2019). Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management*, *46*, 210–221. https://doi.org/10.1016/j.ijinfomgt.2018.12.002

Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2023). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development*, *n/a*(n/a). https://doi.org/10.1111/cdev.14042

Appendix A

Script – *{Note: Everything in a box like* this *is an audio recording}*

"We're going to play a game on the computer, are you ready to start?"

[Verbal Assent]

"Great, let's get started!" * MAKE SURE THE VOLUME IS TURNED UP! *

Shape Task

> In this game, you're going to find out which shapes [Monster/Robot] likes to eat. There are some shapes that [Monster/Robot] likes to eat and there are other shapes he doesn't like to eat. When you feed [Monster/Robot] a shape he likes, he will say 'Mmm' and give you a star, like this: [MMM DEMO]. When you feed [Monster/Robot] a shape he doesn't like, he will say 'Yuck' and not give you a star, like this: [YUCK DEMO]. [Monster/Robot] eats all day until his bedtime. When [Monster/Robot] goes to bed, we'll see how many stars you got, like this: [EXAMPLE]. Your job is to feed [Monster/Robot] as many of the shapes he likes as you can and get as many stars as you can each day before [Monster/Robot]'s bedtime. Try to feed him all of his favorite shapes! Do you want to hear the instructions again or are you ready to start?

(*If necessary at the start of each "day"*: "It's time to feed [Monster/Robot] all of his favorite shapes!")

"Now I have some questions for you."

1. "How much did you like the game? Did you like it a lot, a little or not at all?"
2. "How much did you learn about which shapes Monster likes to eat? Did you learn a lot, a little or not at all?"
    a. "Why do you say [a lot/ a little/ not at all]?"
3. "How many stars did you get throughout the game? Did you get more stars each day, less stars each day, or about the same each day?"

Choosing new game

"Now, we're going to choose a new game to play. We're going to look at different games, and for each set of games you'll choose the one you want to play the most. Here are the options:"

> [P] In this game, you're going to find out which shapes Alien likes to eat.
>
> [S] In this game, you're going to learn how to play with [Monster/Robot]'s ball.
>
> [N] In this game, you're going to learn how to play with Alien's ball.

"Which game do you want to play?"

(If necessary: "If you had to choose one game, which game would you choose?"*)*

"Why did you choose that game?" [type their answer]

"Alright, here are the next three options:"

> [P] In this game, you're going to find out which shapes Monkey likes to eat.
>
> [S] In this game, you're going to help [Monster/Robot] find the right path home.
>
> [N] In this game, you're going to help Monkey find the right path home.

"Which game do you want to play?"

(If necessary: "If you had to choose one game, which game would you choose?"*)*

"Why did you choose that game?" [type their answer]

"Alright, here are the last three options:"

> [P] In this game, you're going to find out which shapes [Monster/Robot] likes to eat.
>
> [S] In this game, you're going to find out where Monster is hiding.
>
> [N] In this game, you're going to find out where Robot is hiding.

"Which game do you want to play?"

(If necessary: "If you had to choose one game, which game would you choose?"*)*

"Why did you choose that game?" [type their answer]

"Now let's play that game." [they play the future task for 24 trials]

"Now I have some questions for you."

1. "How much did you like the game? Did you like it a lot, a little or not at all?"
2. "How much did you learn about [which shapes [M/R] likes to eat/ where [M/R] was hiding]? Did you learn a lot, a little or not at all?"
    a. "Why do you say you [learned a lot/ learned a little/ didn't learn at all]?"
3. "How many right guesses did you make throughout the game? Did you make more, less, or the same amount of right guesses throughout the game?"
4.

Tower Task:

"In this game, you are going to build a tower using these blocks. On this board, you will point to how tall you think that you can make the tower. You will play the game a few times, and then we'll be all done!"

"How high do you think you can build the tower?"

(*If needed:* "Can you point to it on the chart?")

[record their estimate, have them build the tower, record their actual performance/ why it fell]

"You built that tower [##] blocks high. How high do you think you can build the next tower?"

(*If needed:* "Can you point to it on the chart?")

[record their estimate, have them build the tower, record their actual performance/ why it fell]

IF…

They stop: wait until obvious disengagement, then "Is that as high as you can build this tower?"

Become increasingly disinterested: "Just a few more!"

________________________________ [Repeat for 6 total towers]________________________________

"Alright we're all done! Thank you so much for playing with me!"

Appendix B

Tower Task Sheet

Tower 1 Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________	**Tower 2** Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________
Tower 3 Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________	**Tower 4** Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________
Tower 5 Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________	**Tower 6** Estimate: _________ Actual: _________ End: Fell Stopped Other: _________________________

Tower Task Guidelines

Count the height of the tower when it was **last stable** (i.e., the height *before* the block that made it fall)

{ex. If the tower was stable at 19, they placed the 20th block and it fell → height = 19}

If they…

Place the next block and the tower falls ………………………………………. height – last block

ex. If they placed the 20th block and the tower fell… height = 19

Place two blocks and the tower falls ………………………………………. height – last 2 blocks

ex. If they placed the 20th and 21st blocks before it could stabilize… height = 19

Knock over their own tower after it was already stable……………………… height before it fell

ex. If it is stable at 20 and then they knock it over… height = 20

If ANY block falls……………………………………………………………….. height before that block fell

ex. If they have 19 blocks and only the 20th falls… height = 19