Thinking Skills and Creativity 58 (2025) 101936

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

Understanding the link between spatial skills and mechanical reasoning

Felipe Munoz-Rubke[a,*], Zachary Hawes[b], Ramón D. Castillo[c], Karla Avendaño[a], Valentina Zamorano[a], Javier Soto-Mardones[a]

[a] *Instituto de Psicología, Universidad Austral de Chile, Puerto Montt, Chile*
[b] *Department of Applied Psychology & Human Development, Ontario Institute for Studies in Education, Ontario, Canada*
[c] *Centro de Investigación en Ciencias Cognitivas, Facultad de Psicología, Universidad de Talca, Talca, Chile*

ARTICLE INFO

Keywords:
Spatial skills
Mechanical reasoning
STEM

ABSTRACT

The link between spatial thinking and the understanding of physical events has long been of interest in cognitive science. Previous research suggests that spatial cognition plays a pivotal role in comprehending mechanical systems, given that physical events unfold in space. However, it remains unclear whether all subtypes of spatial ability contribute equally to mechanical reasoning. To investigate this, we administered six tasks to participants: four spatial assessments organized within a 2 (intrinsic vs. extrinsic) × 2 (static vs. dynamic) factorial framework, and two mechanical reasoning measures (the DAT-5 Mechanical Reasoning test and a Gears-and-Belts task). While we initially hypothesized that intrinsic and dynamic spatial skills would be most predictive of mechanical reasoning, our findings challenged this prediction. Contrary to our expectations, we found that mechanical reasoning was best characterized by the use of more extrinsic (rather than intrinsic) and more dynamic (rather than static) spatial processing, at least at an overall level. Additionally, our results revealed that the predictive power of spatial skills is highly dependent on the task, as this association was stronger for a broad assessment of varied mechanical problems (DAT-5) than for a narrow task focused on a specific system (Gears-and-Belts). This finding suggests that future investigations should employ a wider variety of physical problems to better understand the interconnections between specific spatial abilities and distinct forms of mechanical reasoning.

1. Introduction

There are many ways to solve mechanical problems, such as deciding the safest way to cut down a tree that will threaten a house during a storm or choosing the right type of nail to fix a broken bookshelf. These different strategies are often used interchangeably in real-world situations. One particularly useful approach when lacking specialized knowledge about a problem is to use mental simulations to envision the possible outcomes of our actions. This is because mental simulations allow us to test different scenarios and imagine how they might play out before deciding what to do (Hegarty, 2004). For example, we can predict how a tree will fall after we cut it down, accounting for the many possible physical factors that may interfere with its trajectory while it is falling. This mental simulation pathway to mechanical problem solving (Ullman et al., 2017), though limited in scope due to its dependence on our internal

* Corresponding author.
E-mail address: felipe.munoz@uach.cl (F. Munoz-Rubke).

https://doi.org/10.1016/j.tsc.2025.101936
Received 31 October 2023; Received in revised form 20 July 2025; Accepted 25 July 2025
Available online 25 July 2025

cognitive abilities (Ludwin-Peery et al., 2021), can be highly effective in predicting the behavior of simple systems.

Studies show that people mentally simulate mechanical systems, a kind of physical system, in a piecemeal fashion, rather than holistically (Hegarty, 2004). This means that they do not try to understand how a mechanical system works all at once, but instead focus on individual components and how they interact with each other. When predicting the movement of a component, people start at the assumed point of origin and move forward, following the connections between components, until they reach the target of interest. The farther away the target is from the point of origin, the longer it takes to make a prediction and the more likely it is that the prediction will be wrong. Eye-tracking data supports this account, showing that people first focus on the overall layout of a mechanical system, and then explore the sequential movement of components along the causal chain Hegarty (1992).

Mechanical reasoning is naturally linked to spatial reasoning because mechanics is the science of the forces that underlie motion, and motion happens in space (Hegarty, 2004). Consistently, evidence shows that people with better spatial visualization abilities are better at solving kinematics problems. They are better at predicting the motion of objects, understanding graphs that show information about the movement of objects, and translating the motion of objects across frames of reference (Kozhevnikov et al., 2007). Marked differences have also been found between high and low-spatial ability individuals in motion verification tasks (Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997). These differences are more pronounced when participants predict the behavior of systems with a higher number of components.

More recently, we have also learned that engaging in the solution of challenging mechanical problems (Munoz-Rubke et al., 2018) can lead to gains in spatial visualization ability that can be observed even a week after the intervention (Munoz-Rubke et al., 2021). However, the overall evidence indicates that mechanical reasoning is not a purely spatial process. Instead, it appears to be a composite skill that involves not only mental simulation but also a series of non-spatial cognitive tools known as beliefs and heuristics to make sense of physical interactions (Hegarty, 2004; Mitko & Fischer, 2020; Schwartz & Black, 1996a, 1996b).

Among these non-spatial cognitive tools, beliefs can be understood as intuitive theories about how the world works, and heuristics as the mental shortcuts derived from them. These heuristics are often highly effective, allowing for rapid and generally accurate judgments. For example, a reliable heuristic for pulley systems is that an open, uncrossed belt causes both pulleys to rotate in the same direction. While efficient, these same shortcuts can be misleading when applied to less intuitive systems. This blend of mental simulation and abstract rules within mechanical reasoning raises an important question about the specific nature and role of the spatial component.

Therefore, one of the most urgent matters within this line of research is to distinguish whether all types of spatial skills are similarly associated with our capacity for mechanical reasoning. To address this, in the current study we further explored the connection between four different types of spatial ability, following a 2×2 theoretical classification (Uttal et al., 2013), and the capacity to predict the behavior of mechanical systems.

1.1. *The mental simulation pathway to mechanical problem solving*

We know how people predict the behavior of mechanical systems through mental simulation (Hegarty, 2004). That is, systems are decomposed into components, and such elements are mentally animated in a piecewise fashion, starting at the assumed point of origin and continuing until the desired target is reached (Hegarty, 1992). We also know that the capacity to mentally simulate how a mechanical system works is more closely linked to spatial ability than verbal ability (Hegarty & Sims, 1994). Furthermore, in a dual-task study aimed at discovering whether mechanical reasoning makes use of the visuospatial sketchpad, mechanical reasoning interfered with performance during a visuospatial task but not as much with performance during a verbal task (Sims & Hegarty, 1997).

Additional support in favor of the existence of a mental simulation pathway to mechanical problem solving comes from the work of (Schwartz & Black, 1996a). These authors supplied evidence in favor of the use of mental simulation for the prediction of the joint movement of two interlocking gears, since the time that it took participants to predict whether the interlocking gears would mesh was proportional to the angle of rotation. Additionally, they also demonstrated that mental simulation is linked to, but not reducible to, spatial abilities.

In recent years, further research has expanded our understanding of the mental simulation of physical events. The *mental physics-engine hypothesis* suggests that we make sense of the physical world by running internal simulations, which are akin to the software instantiated in current video games (Ullman et al., 2017). This approach went beyond previous research on mechanical reasoning by focusing on the understanding of general physical scenes.

According to this approach, our capacity for intuitive physics is based on runnable mental models. These models are built as incomplete internal reconstructions of physical scenes that retain relevant elements while discarding inconsequential information. Although they account for important properties such as mass, friction, and gravity, they are not detailed or comprehensive representations. Instead, these models are incomplete and approximate (Ullman et al., 2017). The primary purpose of runnable mental models is not to create a faithful representation of the external world but to enable accurate predictions of relevant physical occurrences. This simplification process helps us make fast, practical decisions in real-world situations.

Runnable mental models are approximations, which we interact with through a series of hacks and shortcuts (Lake et al., 2017; Ullman et al., 2017). Once the internal reconstruction is created, it can be mentally explored and manipulated to generate predictions. These predictions can then be compared with observations, leading to the formulation of adjusted beliefs.

The mental simulation process entails the formation of repeated simulations of a physical scene, which, as a whole, approximate the ground truth model described by Newtonian mechanics. This means that the mental simulations from which we derive our predictions can be understood as noisy physics simulations (Battaglia et al., 2013; Ullman et al., 2017). The multiple simulations created during the observation of a single physical scene can be organized within a distribution of simulations, which explain the noisy estimates people make when predicting physical scenes (Battaglia et al., 2013; Sanborn et al., 2013; Smith & Vul, 2013).

The resulting mental model does not keep track of every single object present in a scene, as those elements that are not actively being acted upon remain dormant, until the moment they become relevant to the unfolding physical event. It is only when an object is affected by the action of another object, via collision, for instance, that it becomes activated and actively tracked by the physics-engine.

This engine is not only based on spatial information, as it contains rules and beliefs that help resolve the interactions between objects and their future states. One of the most powerful attributes of this perspective is that it can be applied to a variety of physical problems, even those in which individuals have no prior experience. It is not computationally intensive and can explain our ability to make somewhat accurate predictions in tasks we have not seen before. Additionally, the physics-engine in the brain hypothesis has a developmental inspiration as its general-purpose procedures are consistent with results from developmental research (Baillargeon, 2004; Spelke, 1990).

The results of a study that compared the physics-engine model to a deterministic ground truth model (based on Newtonian mechanics) across five different physical scenes showed that participants' judgements were consistent with the physics-engine model (Battaglia et al., 2013). Further investigations have confirmed that there is a strong correspondence between people's judgments and the results generated by the physics-engine model (Hamrick et al., 2016).

1.2. 2 × 2 classification of spatial skills

Spatial skills relate to the capacity to think about spatial processes. This includes the ability to think, while assuming different frames of reference, about the spatial distribution present between objects and their parts, as well as about the interactions and displacements that objects experience while moving through space. This broad description encompasses cognitive skills that have been grouped differently by distinct groups of researchers. Due to this, there is no consensus on how to classify the different types of spatial skills.

Lohman (1979) identified three main spatial factors: visualization, spatial orientation, and spatial relations (presented as speeded rotation by Lohman in 1988). Visualization is a broad category that includes complex spatial processes. Lohman does not delve into the mental operations grouped under this factor, as the main grouping characteristic is that the spatial tasks loading on this factor are complex and cognitively demanding (Lohman, 1979, 1988). In contrast, spatial orientation refers to the ability to discover how objects will be seen from different perspectives, and spatial relations involves the ability to solve mental rotation problems through different means and as quickly as possible.

Another classification offered by Linn and Petersen (1985) distributed spatial skills into spatial perception, mental rotation, and spatial visualization. Spatial perception abilities are used to understand the spatial relation present between stimuli and our own bodies. Mental rotation refers to the capacity to mentally simulate the rotation of objects along two or three dimensions. It has been debated whether mental rotation involves analog or analytic processing strategies, though the evidence points to both as valid routes to problem solving. Spatial visualization abilities serve the purpose of performing mental transformations on the spatial representations, such that answers to multi-step sequences can be found.

A more recent account presents a theoretical classification of spatial skills according to two dimensions: intrinsic vs extrinsic and static vs dynamic (Uttal et al., 2013). The first contrast, intrinsic vs. extrinsic, is concerned with whether the spatial process happens at the level of individual objects or at the level of the arrangement present among objects. The intrinsic level refers to spatial information that allows for object identification and discrimination. For example, if you are looking at a set of tools, your intrinsic spatial skills play a role in identifying the hammer and distinguishing it from other objects in the environment, such as pliers or screwdrivers. The extrinsic level, in contrast, focuses on the relationship that can be established among objects. Therefore, the concern is not on differentiating objects, but on being able to understand the relative position that a certain object occupies with respect to other objects in space. For example, if you are looking at a hammer and a screwdriver, your extrinsic spatial skills play a role in understanding that the hammer is located to the left of the screwdriver (Newcombe & Shipley, 2015; Uttal et al., 2013).

The second contrast, static vs. dynamic, refers to whether the spatial representation can be used as it is or if it needs to be modified to be operative. At the static level no transformation is necessary, as the spatial representation can be utilized in its current state. For example, if you are looking at a city map, your static spatial skills play a role in finding your relative position with respect to different landmarks that appear on the map. The dynamic level, in contrast, requires the spatial representation to undergo some kind of transformation to be useful. As such, if you are looking at a city map, your dynamic spatial skills play a role in distinguishing the routes that you and your friends would have to take to arrive at the same restaurant, starting from different initial positions (in this example, it is assumed that the identification of routes is performed through internal mental effort, and not based on the mere manual rotation of the map, as in this latter case you would be using your static spatial skills in a sequential manner). Another example of a task that requires the use of dynamic spatial skills is predicting the direction of the final gear in a mechanical system. This process typically requires mentally animating the transfer of motion from one gear to the next. The interaction between these two contrasts generates a 2 × 2 matrix of spatial skills: intrinsic static, intrinsic dynamic, extrinsic static, and extrinsic dynamic (Newcombe & Shipley, 2015; Uttal et al., 2013).

This 2 × 2 model is based on evidence from cognitive psychology and cognitive neuroscience (Newcombe & Shipley, 2015). Chatterjee (2008), interested in the neural substrates connecting language and spatial thought, also differentiates along the two aforementioned dimensions. Previous studies have also suggested a partial dissociation between intrinsic and extrinsic spatial skills (Hegarty et al., 2006), as well as a marked distinction between intrinsic dynamic and extrinsic dynamic skills (Kozhevnikov & Hegarty, 2001). Research has also pointed out that there could be a significant difference between static and dynamic spatial skills (Kozhevnikov et al., 2005)

1.3. The link between mechanical reasoning and spatial skills

Given that physical events develop in space, we could be tempted to deduce that our understanding of physical situations is the result of our capacity for spatial thinking. The evidence, however, does not support this claim. Physical predictions can be formed by both simulation and rule-based strategies (Hegarty, 2004; Schwartz & Black, 1996b; Smith et al., 2023), and the construction of physical simulations themselves rely, at least in part, on the use of heuristics and beliefs (Ullman et al., 2017).

The connection between the understanding of physical events and spatial skills has been shown by previous studies (Hegarty & Sims, 1994; Kozhevnikov et al., 2002, 2007; Kozhevnikov & Thornton, 2006; Mitko & Fischer, 2020; Mitko et al., 2024). Hegarty and Sims (1994) showed that performance in paper folding (intrinsic dynamic), mental rotation (intrinsic dynamic), and spatial orientation (extrinsic dynamic) were positively associated with performance on a mechanical comprehension test. In this study, spatial skills predicted performance in mechanical trials, where participants judged sentences describing a physical system's motion, but not in non-mechanical trials, where they assessed sentence descriptions of diagrams of a physical system. Furthermore, individuals with higher spatial ability were more capable of verifying kinematic sentences involving events at different stages of the mechanism (beginning, middle, end) than their counterparts with lower spatial skills. This second group made significantly more mistakes when verifying information at the end of the causal chain.

Additionally, performance in paper folding (intrinsic dynamic) and form board (intrinsic dynamic) was also positively correlated with achievement on a kinematics questionnaire (Kozhevnikov et al., 2002), a result similar to that found by the same researchers in a subsequent study (Kozhevnikov et al., 2007). In this later study, the researchers created a composite spatial visualization ability score by including the performance in paper folding, form board, card rotation, and cube comparison, all measurements of intrinsic dynamic spatial skills that were positively correlated with overall accuracy in kinematic problems (Kozhevnikov et al., 2007).

In a subsequent study, to quantify the link between the intuitive understanding of physical events and spatial skills, researchers presented participants with an intuitive physics task, three spatial skills evaluations, and a verbal working memory task (Mitko & Fischer, 2020). In the physical task, participants predicted the direction in which an unstable tower would fall. The first spatial evaluation was the paper folding test, the second was a mental rotation task, and the final assessment was a spatial working memory task. According to the 2×2 classification of spatial skills, both the paper folding test and the mental rotation task involve intrinsic dynamic spatial skills, whereas the spatial working memory task involves intrinsic static spatial skills. Results suggested a significant correlation between both intrinsic dynamic spatial skill measures and performance on the unstable towers task, but not for the intrinsic static spatial skill. In fact, neither the spatial nor verbal working memory task showed a statistically significant correlation with the unstable towers task. In their interpretation of the results, the authors concluded that despite the notable connection between our capacity to predict physical events and spatial skills, the former capacity cannot be fully accounted for by the latter. In a subsequent investigation by the same research team, Mitko et al. (2024) found that performance on a test of intuitive physics was associated with performance in a mental rotation task, but that the former construct could not be fully explained by the latter.

Even though the link between mechanical reasoning and spatial abilities has been well substantiated by previous work (Hegarty & Sims, 1994; Kozhevnikov et al., 2002, 2007; Kozhevnikov & Thornton, 2006; Mitko & Fischer, 2020; Mitko et al., 2024), we lack a more nuanced understanding of the specific contributions of the distinct types of spatial abilities on our capacity to make sense of the physical world. This led us to investigate the extent to which the four distinct subtypes of spatial ability, as defined by the 2×2 framework of Uttal et al. (2013), differentially contribute to mechanical reasoning.

To pursue this goal, we assessed mechanical reasoning using two different instruments, the DAT-5 Mechanical Reasoning subtest

Table 1
Summary of previous investigations linking spatial skills with mechanical reasoning.

Investigation	Spatial Skills Associated with Mechanical Reasoning	Spatial Skills Not Associated with Mechanical Reasoning
Hegarty and Sims (1994)	Paper folding (intrinsic dynamic), mental rotation (intrinsic dynamic), and spatial orientation (extrinsic dynamic) were positively associated with performance on a mechanical comprehension test	
Kozhevnikov et al. (2002)	Paper folding (intrinsic dynamic) and form board (intrinsic dynamic) were positively correlated with achievement on a kinematics questionnaire	
Kozhevnikov and Thornton (2006)	Before an educational intervention, paper folding scores (intrinsic dynamic) were positively associated with performance in mechanical problems	After an educational intervention, paper folding scores (intrinsic dynamic) were not associated with performance in mechanical problems
Kozhevnikov et al. (2007)	Paper folding (intrinsic dynamic), form board (intrinsic dynamic), card rotation (intrinsic dynamic), and cube comparison (intrinsic dynamic) were used to create a composite spatial visualization ability score that was positively correlated with overall accuracy in kinematic problems	
Mitko and Fischer (2020)	Paper folding (intrinsic dynamic) and mental rotation (intrinsic dynamic) were positively associated with performance in the unstable towers task	Spatial working memory (intrinsic static) was not associated with performance in the unstable towers task
Mitko et al. (2024)	Mental rotation (intrinsic dynamic) was positively associated with performance on a test of intuitive physics	

Note. The table summarizes findings from key studies on the relationship between spatial skills and mechanical reasoning. The classification of spatial skills shown in parentheses is based on the theoretical framework proposed by Uttal et al. (2013), which categorizes skills along intrinsic vs. extrinsic and static vs. dynamic dimensions.

(Cordero & Corral, 2006), which is a well-established measure, and a custom Gears-and-Belts task created in our lab. For the four spatial quadrants, we measured intrinsic static skills with a Visual Working Memory task (Makovski et al., 2008), intrinsic dynamic skills with a Paper Folding test (Ekstrom et al., 1976), extrinsic static skills with a Proportional Reasoning task (Boyer & Levine, 2012; Boyer et al., 2008), and extrinsic dynamic skills with a Spatial Orientation task (Friedman et al., 2019).

To guide our inquiry, we pre-registered two primary hypotheses based on the existing literature (see Table 1). Considering the strong evidence in favor of intrinsic dynamic spatial skills found in previous research (Hegarty & Sims, 1994; Kozhevnikov et al., 2002, 2007; Kozhevnikov & Thornton, 2006; Mitko & Fischer, 2020), we considered it more plausible that higher scores in mechanical reasoning would be associated with higher scores in intrinsic rather than extrinsic spatial skills. Similarly, we favored a higher role for dynamic over static spatial skills due to the core nature of mental simulation.

Furthermore, our data analysis was designed to determine which of these four spatial subtypes were the most powerful predictors of mechanical reasoning. To achieve this, our approach moved beyond simple correlations. We used a multiple regression model because it forces the predictors to statistically compete with one another. This method evaluated the relationship between any one predictor and the outcome only after accounting for the influence of all other predictors in the model. This process allows us to isolate the unique contribution of each subtype, providing a much clearer perspective on their relative importance in mechanical reasoning.

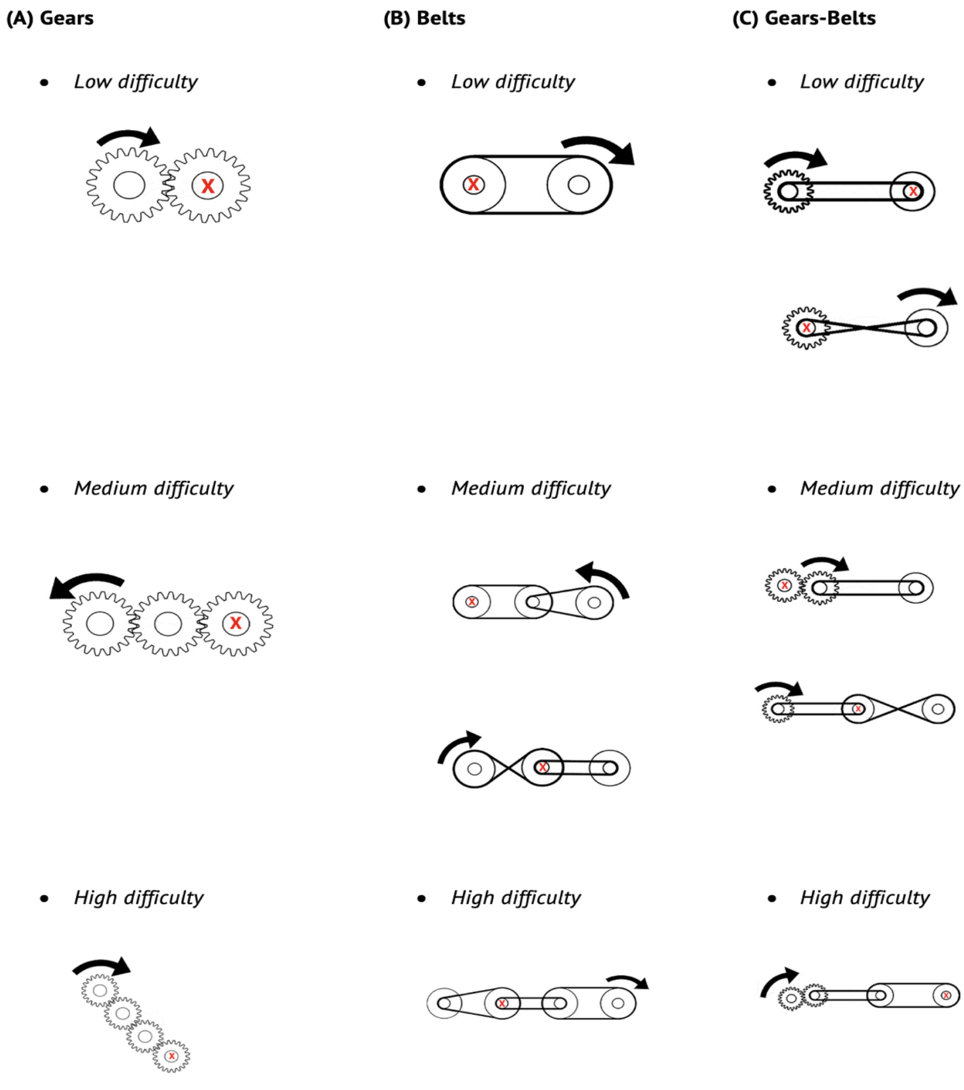

Fig. 1. Sample stimuli from the gears-and-belts task.
Note. The figure displays sample items for the three types of stimuli: (A) Gears, (B) Belts, and (C) Gears-Belts, categorized by three levels of difficulty. Difficulty was defined by the number of elements in the system (two = low, three = medium, and four = high). For each item, participants determined the direction of motion of a target element (marked with a red X) based on the movement of a source element (indicated by the arrow).

2. Methods

2.1. Participants

Three hundred participants (M = 23.11 years, SD = 3.93, range = 18–48 years, 183 females) took part in this study. Seventeen participants were removed from the analysis due to incomplete data resulting from a software malfunction.

Most participants were Universidad Austral de Chile students from both the Puerto Montt and Valdivia campuses (95.67 % of the total sample), though any native Spanish speaking adult was allowed to participate. Participants received economic compensation for their participation (3300 CLP, equivalent to approximately 5 USD at the time of data collection).

Informed consent was obtained before the experiment, in accordance with the Universidad Austral de Chile Institutional Review Board approved protocol.

2.2. Materials

We presented participants with six different tasks: two that evaluated mechanical reasoning and four that assessed spatial abilities.

2.2.1. DAT-5 mechanical reasoning test

The 5th version of the Differential Aptitude Test (Cordero & Corral, 2006) evaluates seven competences: verbal reasoning, numerical reasoning, abstract reasoning, mechanical reasoning, spatial relations, spelling, perceptual speed, and accuracy. Of these, we only presented participants with level 1 mechanical reasoning items. This is the easiest of the two available levels, as it is designed for students aged 12 to 16, whereas the more advanced Level 2 is targeted at older students (16–18) and adults (Cordero & Corral, 2006).

This mechanical assessment included 60 questions designed to measure the ability to understand the basic principles of how different machines work and to predict outcomes in scenarios involving interactions between forces and objects. For instance, in a practice trial, participants were shown two pictures of a baseball bat hitting a ball. The question posed was which ball would reach a higher elevation after being hit. In the upper picture, the bat hits the lower part of the ball, while in the lower picture, the bat hits the upper part. Participants were required to choose one of three alternatives. In this example, A indicates that the ball will reach a higher elevation in the upper picture, B indicates that the ball will reach a higher elevation in the lower picture, and C indicates that there will be no difference between the two scenarios. In this case, the correct answer is A.

All 60 questions included three possible response alternatives. The maximum possible score was 60 points. Participants were given 15 min to respond to as many questions as possible. The items appeared on the computer screen in a random order, and participants were required to provide an answer to move onto the next question.

Previous data suggest that Cronbach's alpha for the first level of the DAT-5 mechanical reasoning test ranges between 0.85 and 0.90, indicating high inter-item reliability (Cordero & Corral, 2006). From the data collected in the current study, we calculated a Cronbach's alpha of 0.86 and a McDonald's Omega of 0.88.

2.2.2. Gears-and-Belts task

The Gears-and-Belts task was created in our lab. We created 10 gear stimuli, 10 belt stimuli, and 10 gear-belt stimuli organized along three levels of difficulty. For all categories of stimuli, low difficulty was characterized by the presence of two elements in the diagram, medium difficulty by the presence of three elements, and high difficulty by the presence of four elements (see Fig. 1).

In the case of gear stimuli, three low-difficulty items were presented first, followed by three medium-difficulty items, and concluding with four high-difficulty items. In the case of belt stimuli, three low-difficulty items were presented first, followed by four medium-difficulty items, and concluding with three high-difficulty items. Finally, for gear-belt stimuli, three low-difficulty items were presented first, followed by five medium-difficulty items, and concluding with two high-difficulty items. All participants were presented with the same sequence of stimuli, starting with gears, followed by belts, and finishing with gear-belt items.

In all trials, participants were presented with the direction of movement of one source element in the diagram and had to determine whether a target element - physically connected to the source - would move in a clockwise or counterclockwise direction. Participants pressed "a" if they believed the target would move clockwise or "k" if they thought it would move counterclockwise.

Participants were required to provide an answer before moving on to the next item. Participants were given a maximum of 12 min to complete the task. For each correct item, one point was awarded. The maximum possible score was 30 points. From the data collected in the current study, we calculated a Cronbach's alpha of 0.85 and a McDonald's Omega of 0.87, suggesting high inter-item reliability.

2.2.3. Visual working memory task (Intrinsic static spatial ability)

The visual working memory task used in the current study followed the work of Makovski et al. (2008). We included 20 items, divided into two categories of 10 items each: whole-report and partial-report. In both whole-report and partial-report trials, participants were initially presented with a 2 × 2 spatial arrangement containing four figures for 1000 ms (circle, square, right triangle, non-right triangle, hexagon, heart, moon, or rhombus). Following that, we presented a fixation cross for 1600 ms. In the whole-report trials, participants saw another 2 × 2 arrangement displaying four figures, while in the partial-report trials they saw only one figure positioned within a unique location in the 2 × 2 spatial arrangement (see Fig. 2).

Participants were asked to indicate whether the first and second spatial arrangements were "similar" or "different." Following explicit instructions provided before the task, a "similar" response was only correct if the figures in the second arrangement were

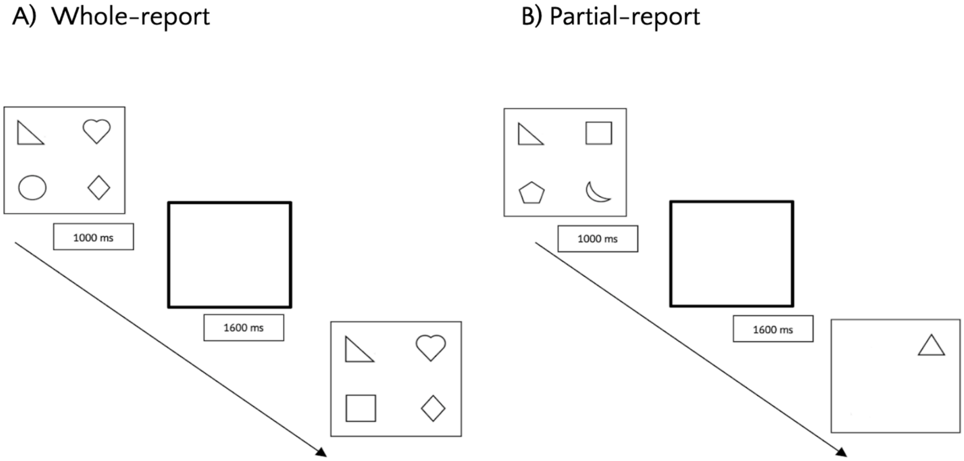

Fig. 2. Schematic of the Visual Working Memory task.
Note. The task design was based on the paradigm from Makovski et al. (2008). In both conditions, an initial array of four shapes was displayed for 1000 ms, followed by a 1600 ms retention interval. Panel (A) depicts a whole-report trial, where a full array was presented for comparison. Panel (B) depicts a partial-report trial, where only a single shape was presented for comparison. In both cases, participants judged whether the second display was identical to the first in terms of shape and location.

identical in shape and, at the same time, occupied the same spatial locations as in the first. Conversely, a "different" response was correct if there was any change in either the shape or the location of any figure. These definitions were reinforced with several visual and practical examples during the instruction phase.

Participants had to provide an answer to move on to the next item. For each correct response, one point was awarded. Three of the original 20 trials were removed from the analysis because a reliability analysis indicated that their removal improved the scale's internal consistency. This resulted in a maximum possible score of 17 points. For the remaining 17 items, we obtained a Cronbach's alpha of 0.58 and a McDonald's Omega of 0.63, suggesting moderate inter-item reliability.

2.2.4. Paper folding test (Intrinsic dynamic spatial ability)

We used the paper folding test VZ-2 (Ekstrom et al., 1976). In this test, participants observed a sequence showing how a piece of paper is folded, punctured, and then unfolded (French et al., 1963). In each trial, participants indicated how the unfolded sheet of paper would look at the end of the folding and puncturing sequence by selecting among five alternatives (see Fig. 3). Previous investigations indicate that the paper folding test measures spatial visualization (Linn & Petersen, 1985; Lohman, 1988), as it requires the multistep manipulation of spatial information.

This test contained 20 items, and one point was given for each correct answer. From the data collected in the current study, we calculated a Cronbach's alpha of 0.83 and a McDonald's Omega of 0.86, indicating a high inter-item reliability.

2.2.5. Proportional reasoning task (Extrinsic static spatial ability)

In the proportional reasoning task, based on the work of Boyer and colleagues (Boyer & Levine, 2012; Boyer et al., 2008), participants indicated which of the five bars was visually proportional to a reference bar.

The proportional reasoning task contained 36 stimuli, divided into an equal amount of low, medium, and high-difficulty items (see Fig. 4). Low-difficulty items included one distracting stimulus, while medium and high-difficulty items included two and three distracting stimuli, respectively. Distracting stimuli deviated from the correct answer by a proportion between 10 and 20 percent. In contrast, incorrect but non-distracting stimuli deviated from the correct answer by 30 % or more.

Items were presented randomly. Participants responded by clicking on the stimuli that they considered to be proportional to the

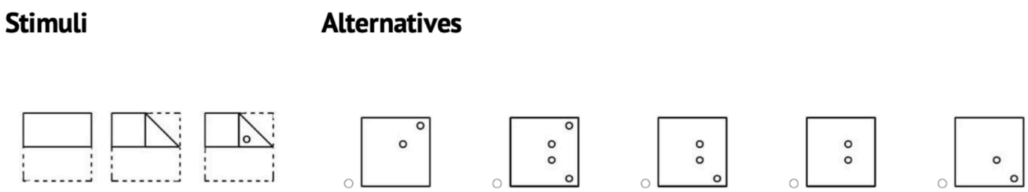

Fig. 3. Example item from the paper folding test.
Note. The figure shows a sample item from the paper folding test (Ekstrom et al., 1976). The "Stimuli" panel on the left depicts a sequence in which a piece of paper is folded and punctured. The "Alternatives" panel on the right presents five possible outcomes. The task is to mentally unfold the paper and select the alternative that correctly shows the resulting pattern of holes.

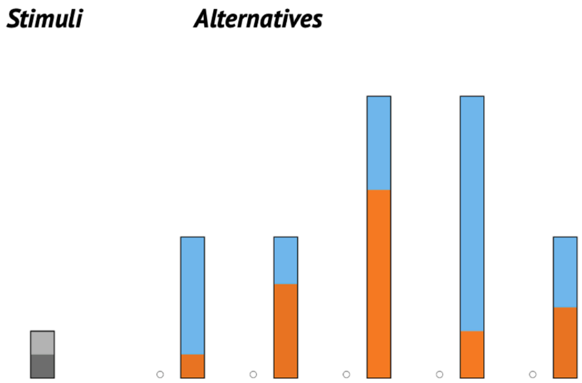

Fig. 4. Example item from the proportional reasoning task.
Note. The task was adapted from the work of Boyer and colleagues (Boyer & Levine, 2012; Boyer et al., 2008). In a typical trial, a reference bar ("Stimuli") was presented on the left. The participant's task was to select which of the five bars in the "Alternatives" panel was visually proportional to the reference bar.

reference. Participants were given 15 s to provide an answer. If no response was provided, the computer presented a new item. One point was awarded for each correct answer. The maximum possible score was 36 points. From the data collected in the current study, we calculated a Cronbach's alpha of 0.84 and a McDonald's Omega of 0.86, indicating a high inter-item reliability.

2.2.6. Spatial orientation task (Extrinsic dynamic spatial ability)

In the spatial orientation task, we used the items developed by Friedman et al. (2019). Participants were asked to imagine themselves at a specific location, facing an object. They then had to determine the position of a secondary object from this initial perspective.

Participants responded by writing down the angular distance in degrees between their current perspective and the location of the secondary object (see Fig. 5). The task comprised 12 items, and participants were given 10 min to complete it. The maximum possible score was 12 points. From the data collected in the current study, we calculated a Cronbach's alpha of 0.79 and a McDonald's Omega of 0.84, indicating high inter-item reliability.

2.3. Procedures

Due to data collection taking place during the COVID-19 pandemic, participants responded using their personal computers. All tasks were programmed using lab.js and presented to participants through OpenLab (Henninger et al., 2022).

The researcher shared links to the six tasks individually, and the order in which these tasks were presented was fully randomized. Throughout the instruction phase of each task, a researcher was readily available to address any questions raised by the participants. Each session was carefully planned with an anticipated duration ranging from 75 to 90 min.

2.4. Data analysis

Our data analysis procedures were pre-registered before viewing the data[1] and were implemented in R (R Core Team, 2020). The analysis was conducted in three main stages. These were a preliminary psychometric analysis to validate our instruments, a primary analysis to address our research hypothesis, and a supplementary exploratory analysis.

To evaluate the internal consistency of our six measurement instruments, we calculated Cronbach's alpha and McDonald's omega for each. Although we did not pre-register this step, it was considered important to evaluate the quality of our data. Since Cronbach's alpha can sometimes be too conservative and even unreliable, some researchers suggest using McDonald's omega as a better alternative (Hayes & Coutts, 2020). For dichotomous performance data (correct vs. incorrect responses, as in the case of all assessments used in this study except the Spatial Orientation task), McDonald's omega has also been suggested as a better estimator of inter-item reliability than Cronbach's alpha, which becomes the KR-20 coefficient when working with this type of variable (Béland & Falk, 2022). Results from the reliability analysis are presented in the Materials section.

To explore the construct validity of our measures, we conducted a pre-registered Exploratory Factor Analysis (EFA) that included all six vectors. We expected this statistical model to account for a minimum of 40 % of the variance present in the data. Furthermore, we expected both mechanical reasoning tasks and the intrinsic/dynamic spatial skills to load on the same factor. We also expected another factor to include all four spatial skills tasks.

After establishing the psychometric properties of our measures, we proceeded with the main analyses as specified in our pre-

[1] https://archive.org/details/osf-registrations-njseh-v1

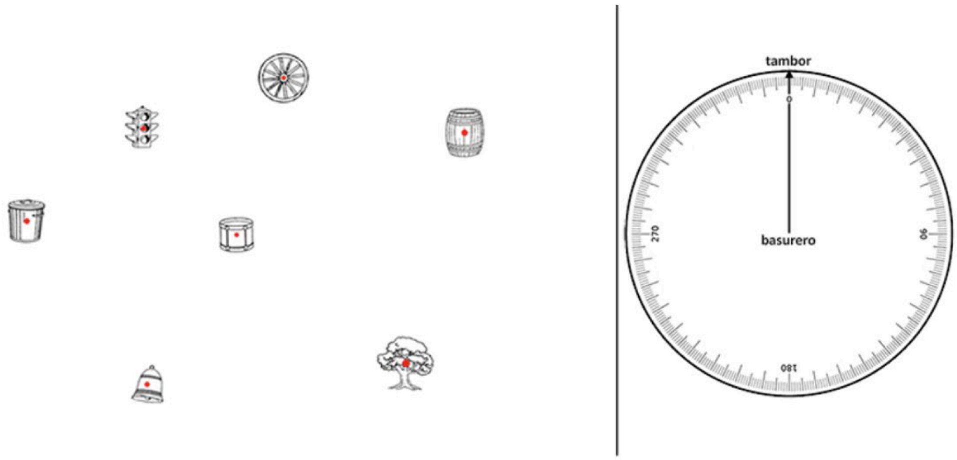

Fig. 5. Example item from the spatial orientation task.
Note. The task items were those developed by Friedman et al. (2019). The left panel shows a map of the objects' locations, and the right panel displays the response interface. For each item, participants had to imagine standing at one object (e.g., basurero = trash can), facing directly towards a second object (e.g., tambor = drum), and from that perspective, use the compass to indicate the direction of a third, target object (e.g., semáforo = traffic light).

registration. First, we reported the descriptive statistics for each variable (M, Q1, Q2, Q3, Q4, SD, etc.). Then, to examine the relationships between our constructs of interest, we created a matrix of bivariate correlations that included all four spatial skills variables on one side and both mechanical reasoning variables on the other. Next, to test our primary hypotheses about the unique predictive power of each spatial subtype, we ran two multiple regressions. One model included the DAT-5 mechanical reasoning scores as the response variable and all four spatial assessments as predictors. The second model included the same predictor variables but used the Gears-and-Belts task as the response variable. This regression approach was chosen because it forces the predictors to statistically compete with one another, allowing us to isolate the unique contribution of each subtype after accounting for the influence of all other predictors in the model.

Finally, we ran two additional, non-preregistered multiple regression models to include gender and age as possible predictors of performance in the mechanical reasoning assessments.

3. Results

First, we report the results according to the analysis plan proposed during the pre-registration process. Subsequently, we present some exploratory analyses that were deemed of interest after analyzing the data.

3.1. Preliminary analysis

To explore the construct validity of our measures, we first conducted an Exploratory Factor Analysis (EFA) using a maximum likelihood method of extraction and a varimax rotation. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.84, exceeding the recommended value of 0.60, and Bartlett's test of sphericity was statistically significant ($\chi^2(15) = 412.81, p < .001$), indicating that the correlations between variables were sufficiently large for a factor analysis.

In our pre-registered plan, we expected to account for a minimum of 40 % of the variance present in the data with this statistical model. We anticipated that the following three variables would load on the same factor: the DAT-5 mechanical reasoning task, the Gears-and-Belts task, and the Intrinsic Dynamic task. We also expected another factor to include all four subtypes of spatial skills.

Data from 282 participants were considered for this EFA, which, according to the pre-registration plan, considered the extraction of two factors. The first factor explained 27 % of the variance, while the second explained 18 % of the variance. Together, these two factors explained a total of 45 % of the variance.

Results showed that the variables DAT-5 mechanical reasoning, Intrinsic Dynamic, Extrinsic Static, and Extrinsic Dynamic had high loadings on the first factor, with loadings of 0.64, 0.58, 0.61, and 0.57, respectively. The variables Gears-and-Belts and Intrinsic Static had relatively lower loadings of 0.28 and 0.31, respectively, on the first factor. The second factor included the variable Gears-and-Belts with a high loading of 0.80, and the variables DAT-5 mechanical reasoning, Intrinsic Dynamic, Extrinsic Static, and Extrinsic Dynamic with lower loadings of 0.45, 0.30, 0.22, and 0.30, respectively. The variable Intrinsic Static was not associated with the second factor.

3.2. Main analysis

Descriptive statistics for each variable are reported in Table 2. All variables except Extrinsic Dynamic could range from a minimum score of 0 to a maximum score of 1, indicating that all responses were correct. In the case of Extrinsic Dynamic, the scores represent the degrees of deviation from the correct answer and could range from a minimum of −359.99 to a maximum of 0.

Prior to conducting the correlations, we assessed the data for normality using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The results indicated that all variables deviated from a normal distribution ($p<.05$). To ensure our findings were sound, we calculated both parametric (Pearson) and non-parametric (Spearman) correlations. The correlation coefficients from both methods were highly similar (e.g., the correlation between DAT-5 and Gears-and-Belts was $r = 0.54$ with Pearson and rs=0.57 with Spearman). Given this similarity and the robustness of the Pearson test, we report the Pearson correlations for ease of interpretation. The full normality test results and the Spearman correlation matrix are available in the supplementary materials (Tables S1 and S2, respectively).

Bivariate correlations showed a large positive correlation between the DAT-5 mechanical reasoning test and the Gears-and-Belts task ($r = 0.54$, 95 % CI = [0.45, 0.62], $p < .001$), as well as positive correlations for both mechanical assessments with all types of spatial ability (see Table 3). The DAT-5 mechanical scores showed a large positive association with Intrinsic Dynamic, Extrinsic Static, and Extrinsic Dynamic spatial abilities (all with $r = 0.50$, $p < .001$) and a small positive association with Intrinsic Static skills ($r = 0.23$, 95 % CI = [0.11, 0.34], $p < .001$). The Gears-and-Belts task showed a moderate positive association with Intrinsic Dynamic, Extrinsic Static, and Extrinsic Dynamic spatial abilities (all with $r > 0.35$, $p < .001$), while it showed a small positive association with Intrinsic Static skills ($r = 0.15$, 95 % CI = [0.03, 0.26], $p < .001$). All p-values were FDR corrected for multiple comparisons.

As shown in Table 4, the multiple regression model that included the DAT-5 mechanical reasoning scores as the response variable and the four types of spatial skills as the predictors explained a substantial portion of variance ($F(4, 277) = 48.72$, $p < .001$, $R^2 = 0.41$, adj. $R^2 = 0.40$). A similar multiple regression model that included the Gears-and-Belts scores as the response variable and the four types of spatial skills as the predictors explained a moderate portion of variance ($F(4, 277) = 21.65$, $p < .001$, $R^2 = 0.24$, adj. $R^2 = 0.23$). With the exception of the Intrinsic Static spatial skills (both $p > .208$), all variables showed a positive and statistically significant association with the dependent variable after controlling for the effects of the other predictors (all $p < .015$). Intrinsic Dynamic skills, assessed by the paper folding task, Extrinsic Static skills, assessed by the proportional reasoning task, and Extrinsic Dynamic skills, assessed by the spatial orientation task, were predictors of performance in both mechanical reasoning tasks (see Fig. 6).

3.3. Exploratory analysis

For exploratory purposes, we ran additional multiple regression models on both mechanical reasoning tasks, including predictors for spatial skills, gender, and age. As seen in Table 5, the results are very similar to those reported in Table 4. After accounting for the effects of the remaining predictors, Intrinsic Dynamic, Extrinsic Static, and Extrinsic Dynamic spatial skills were associated with the DAT-5 mechanical reasoning test and the Gears-and-Belts task. Additionally, gender was also predictive of both response variables, as males showed overall higher scores in both mechanical reasoning evaluations after controlling for the effects of the other predictor variables (both $p < .045$).

Comparing the adjusted R^2 values from both models (Table 4 vs. Table 5) shows only a modest increase in explained variance from 0.41 to 0.45 for the DAT-5 and from 0.23 to 0.24 for the Gears-and-Belts task. This suggests that while gender was a statistically significant predictor on its own, it did not explain a large amount of additional variance beyond what was already captured by the spatial skills. No statistically significant effect was found for either Intrinsic Static skills or age.

4. Discussion

Our results show a strong connection between spatial skills and mechanical reasoning, a finding that aligns with the broader literature on this topic (see Table 1). However, the central contribution of our study is demonstrating that not all spatial abilities are equally relevant. We found a consistent pattern across all analyses indicating that intrinsic static spatial skills have a distinctly weaker association with mechanical reasoning compared to their counterparts. Furthermore, our findings reveal that the strength of this

Table 2
Descriptive statistics.

Variable	M	SD	min	Q1	Q2	Q3	max
DAT-5 mechanical reasoning	0.67	0.14	0.28	0.58	0.68	0.78	0.98
Gears-and-Belts	0.87	0.14	0.03	0.80	0.93	0.97	1.00
Intrinsic Static (Visual Working Memory)	0.76	0.14	0.18	0.68	0.76	0.88	1.00
Intrinsic Dynamic (Paper Folding)	0.61	0.17	0.00	0.51	0.64	0.75	0.95
Extrinsic Static (Proportional Scaling)	0.52	0.17	0.00	0.39	0.56	0.64	0.89
Extrinsic Dynamic (Spatial Orientation)	−45.25	37.86	−199,25	−62.58	−33.17	−16.62	−2.20

Note. Descriptive statistics for mechanical reasoning and spatial skills variables. *M* = Mean; SD = Standard Deviation; min = Min score; Q1 = First Quartile; Q2 = Median; Q3 = Third Quartile; max = Max score. With the exception of Extrinsic Dynamic, all variables are scored on a 0 to 1 scale, where 1 indicates perfect accuracy. Scores for Extrinsic Dynamic (Spatial Orientation) represent degrees of deviation from the correct answer; therefore, a score of 0 is the maximum achievable score.

Table 3
Bivariate correlations among tasks.

TEST / TASK	Intrinsic Static (Visual Working Memory)	Extrinsic Dynamic (Spatial Orientation)	Extrinsic Static (Proportional Scaling)	Intrinsic Dynamic (Paper Folding)	Gears-and-Belts
DAT-5 mechanical reasoning	0.23 ***	0.50 ***	0.50 ***	0.50 ***	0.54 ***
Gears-and-Belts	0.15 *	0.38 ***	0.35 ***	0.40 ***	
Intrinsic Dynamic (Paper Folding)	0.23 ***	0.41 ***	0.43 ***		
Extrinsic Static (Proportional Scaling)	0.19 **	0.41 ***			
Extrinsic Dynamic (Spatial Orientation)	0.20 ***				

Note. The table displays bivariate Pearson's r correlations among all study variables. All reported p-values were adjusted for multiple comparisons using the False Discovery Rate (FDR) correction.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4
Multiple regression model for the DAT-5 mechanical reasoning test and the Gears-and-Belts task.

	DAT-5 mechanical reasoning				GB			
Predictors	β	*SE* β	CI	*p*	β	*SE* β	CI	*p*
(Intercept)	−0.00	0.05	[−0.09, 0.09]	<0.001	−0.00	0.05	[−0.10, 0.10]	<0.001
Intrinsic Static (Visual Working Memory)	0.06	0.05	[−0.03, 0.15]	0.208	0.02	0.05	[−0.09, 0.13]	0.709
Intrinsic Dynamic (Paper Folding)	0.27	0.05	[0.16, 0.37]	<0.001	0.24	0.06	[0.12, 0.36]	<0.001
Extrinsic Static (Proportional Scaling)	0.26	0.05	[0.15, 0.36]	<0.001	0.15	0.06	[0.03, 0.27]	0.015
Extrinsic Dynamic (Spatial Orientation)	0.28	0.05	[0.17, 0.38]	<0.001	0.23	0.06	[0.11, 0.35]	<0.001
Observations	282				282			
R^2 / R^2 adjusted	0.42 / 0.41				0.24 / 0.23			

Note. Results from two multiple regression models predicting performance on the DAT-5 mechanical reasoning test and the Gears-and-Belts (GB) task from the four spatial skills variables. β = standardized regression coefficient; SE β = standard error of the coefficient; CI = 95 % confidence interval; R^2 = *R*-squared. Total observations *N* = 282.

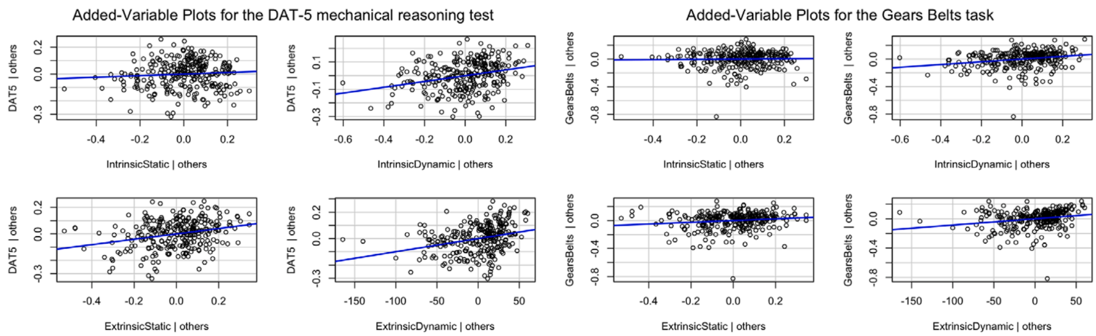

Fig. 6. Association between spatial skills and mechanical reasoning.
Note. The set of plots on the left shows the association between each spatial skill predictor and the DAT-5 mechanical reasoning scores, after controlling for the effects of the other predictors. The set on the right shows the same associations for the Gears-and-Belts task.

connection is significantly affected by the mechanical reasoning task itself; the association was much stronger for a broad assessment (the DAT-5 Mechanical Reasoning test) than for a narrow task focused on a specific system (the Gears-and-Belts task).

While all skills showed a positive bivariate correlation with mechanical reasoning performance, the predictive power of the intrinsic static skills became statistically non-significant when forced to compete with the other spatial skills in our multiple regression models. This was true for both the broad DAT-5 assessment and the specific Gears-and-Belts task.

The Exploratory Factor Analysis (EFA) provided converging evidence for this conclusion. Different from what we predicted, the EFA revealed a primary factor that grouped the DAT-5 mechanical test with the three most predictive spatial skills (intrinsic dynamic, extrinsic static, and extrinsic dynamic), while intrinsic static skills failed to load strongly onto it. Taken together, these findings strongly suggest that intrinsic static ability is less critical for the kind of mechanical reasoning assessed in this study.

Table 5
Multiple regression model for the DAT-5 mechanical reasoning test and the Gears-and-Belts task, including predictors for gender and age.

	DAT-5 mechanical reasoning				GB			
Predictors	B	*SE* β	CI	*P*	β	*SE* β	CI	*p*
(Intercept)	−0.18	0.06	[−0.29, −0.07]	<0.001	−0.09	0.07	[−0.22, 0.05]	<0.001
Intrinsic Static (Visual Working Memory)	0.06	0.05	[−0.03, 0.15]	0.193	0.02	0.05	[−0.09, 0.13]	0.727
Intrinsic Dynamic (Paper Folding)	0.26	0.05	[0.16, 0.36]	<0.001	0.24	0.06	[0.12, 0.36]	<0.001
Extrinsic Static (Proportional Scaling)	0.20	0.05	[0.10, 0.30]	<0.001	0.12	0.06	[0.00, 0.25]	0.048
Extrinsic Dynamic (Spatial Orientation)	0.25	0.05	[0.15, 0.35]	<0.001	0.22	0.06	[0.10, 0.34]	<0.001
Gender (Male)	0.46	0.09	[0.27, 0.64]	<0.001	0.23	0.11	[0.01, 0.45]	0.045
Age	0.04	0.05	[−0.05, 0.13]	0.338	0.01	0.05	[−0.10, 0.12]	0.854
Observations	282				282			
R^2 / R^2 adjusted	0.47 / 0.45				0.25 / 0.24			

Note. Results from two exploratory multiple regression models predicting performance on the DAT-5 mechanical reasoning test and the Gears-and-Belts (GB) task, with spatial skills, gender, and age as predictors. β = standardized regression coefficient; SE β = standard error of the coefficient; CI = 95 % confidence interval; R^2 = *R*-squared. Total observations $N = 282$.

These findings, particularly the weaker role of intrinsic static skills, align well with previous research. Mitko and Fischer (2020), for instance, also found that while intrinsic dynamic skills were correlated with performance on their mechanical task, a measure of intrinsic static skills (spatial working memory) was not. However, our conclusions should be interpreted with caution. The spatial working memory task used to assess this ability in our study exhibited the lowest internal consistency of all our measures. This moderate reliability means its weak association with mechanical reasoning could be partially attributable to measurement error, not only a true lack of relationship. Therefore, future studies using instruments with higher reliability are needed to definitively clarify the role of intrinsic static skills.

The classification of spatial skills used in our framework comprises two dimensions: intrinsic vs. extrinsic and dynamic vs. static (Uttal et al., 2013). Initially, we hypothesized that mechanical reasoning would rely more heavily on intrinsic processing than extrinsic processing. This prediction was grounded in numerous studies demonstrating a strong correlation between mechanical reasoning and performance in spatial visualization tasks (Hegarty & Sims, 1994; Kozhevnikov et al., 2007; Kozhevnikov & Thornton, 2006; Mitko & Fischer, 2020; Mitko et al., 2024), which are typically characterized as intrinsic spatial skills. However, our results did not fully align with this prediction. Although the data showed that performance in the paper folding test (intrinsic dynamic) was indeed predictive of performance in both assessments of mechanical reasoning, it also indicated that both extrinsic static and extrinsic dynamic spatial skills are also highly correlated with mechanical reasoning.

This unexpected finding could be explained by the nature of mechanical reasoning tasks, which often involve the interaction of multiple objects, an aspect more aligned with extrinsic processing. For instance, some items on the DAT-5 mechanical reasoning test require participants to predict the behavior of interacting objects. Earlier research had already hinted at a connection between extrinsic spatial skills and mechanical reasoning (Hegarty & Sims, 1994), particularly through an assessment of spatial orientation. Taken as a whole, the evidence suggests that while both intrinsic and extrinsic spatial skills are connected to performance in mechanical problems, mechanical reasoning could be characterized by the use of more extrinsic (rather than intrinsic) spatial processing.

Within the static vs. dynamic dimension, previous research has favored the linkage between dynamic spatial skills and mechanical reasoning. Indeed, multiple studies support the idea that dynamic transformation of spatial representations play a potential role in predicting the future behavior or state of mechanical systems (Hegarty, 2004). Some of the strongest evidence in favor of the role of dynamic processing comes from the work of Hegarty and colleagues. In their studies, these researchers have demonstrated that the human mind can predict the behavior of mechanical systems by evaluating the displacement of the underlying components in a piecemeal fashion and following the causal chain of events. Both behavioral and eye-tracking data support the role of mental simulation in solving these problems (Hegarty, 1992). Our results are consistent with these studies in showing a preferential association between dynamic spatial skills and mechanical reasoning. However, they also point to a potential role for static spatial skills, particularly when interacting with extrinsic information processing. After all, our proportional reasoning task (extrinsic static) showed an association with performance in both mechanical reasoning evaluations, even after accounting for the effect of both dynamic spatial skills.

Overall, our results are consistent with those of Hegarty and Sims (1994), who found a connection between measurements of intrinsic and extrinsic dynamic skills with performance in a mechanical comprehension test. They also align with the findings of Kozhevnikov et al. (2002), Kozhevnikov and Thornton (2006), Kozhevnikov et al. (2007), Mitko and Fischer (2020), and Mitko et al. (2024), who described a relationship between intrinsic dynamic spatial skills and mechanical reasoning.

Another interesting finding of the current study is the different amount of variance explained in the multiple regression models. As shown in Table 4, both multiple regression models included the same predictors. The only difference between them was whether the response variable was the DAT-5 mechanical reasoning scores or the Gears-and-Belts scores. Interestingly, the statistically significant predictors in both models are the same: intrinsic dynamic, extrinsic static, and extrinsic dynamic spatial abilities. However, the amount of variance explained by these predictors differs significantly between the two models. In the DAT-5 mechanical reasoning model, the proportion of variance explained is 0.42 (adj. $R^2 = 0.41$). In contrast, the Gears-and-Belts model accounts for only 0.24 of the variance (adj. $R^2 = 0.23$). This indicates that the spatial skills predictors are much better at forecasting performance on a mechanical reasoning evaluation involving various physical scenarios, such as the DAT-5 mechanical reasoning test, than on tasks focused solely on the motions of systems made of gears and belts. This finding suggests that future investigations should focus on a variety of physical problems. Understanding these connections could lead to more targeted educational interventions and assessments, ultimately improving our comprehension of how different spatial skills contribute to mechanical reasoning across diverse contexts. In an exploratory analysis, we added predictors for both gender and age to the aforementioned multiple regression models. As seen in Table 5, the amount of variance explained increased for both models, as males tended to score higher than females on both mechanical reasoning assessments.

4.1. Theoretical contributions

The strong association we found between various spatial abilities and mechanical reasoning aligns with the view that individuals, at least in some instances, predict mechanical behavior through mental simulation that incorporates a spatial component (Hegarty, 2004). Our results contribute to prior knowledge by indicating that some spatial skills seem more important for this process than others.

The proposal of Schwartz and Black (1996a) that simulation is linked to but not reducible to spatial abilities resonates with our own results. Indeed, the fact that the four spatial measures in our study did not account for all performance variance underscores that other factors are at play. Our findings are also broadly consistent with the work of Mitko and Fischer (2020), who found that intrinsic dynamic spatial abilities (paper folding, mental rotation) showed small-to-medium correlations with their unstable towers task, while an intrinsic static skill (spatial working memory) showed no link. A point of convergence is that both their study and ours identified visuospatial working memory as a poor predictor of performance in mechanical tasks. This consistent finding across different studies is particularly telling. It may suggest that mechanical reasoning depends more on the dynamic transformation of mental representations than on the simple static maintenance of spatial information that is the primary function of spatial working memory.

A closer look at the tasks used in our study and in the work of Mitko and Fischer (2020) reveals a pattern regarding the strength of the association between spatial skills and mechanical reasoning. A point of convergence is that Mitko and Fischer used a single, specific task, the unstable towers, and similarly, when we analyze our more specific Gears-and-Belts task, we also find a weaker association with spatial abilities. In contrast, this association becomes much stronger when we look at the DAT-5, an assessment encompassing a wide diversity of physical problems. This suggests that the diversity of problems within a mechanical reasoning test could be a critical factor. Broader assessments may consistently engage spatial processing across varied scenarios, while more specific tasks might allow participants to develop and apply task-specific heuristics, thus reducing the apparent contribution of the spatial ability. Despite these details, both studies affirm the same fundamental conclusion that there is a foundational link between spatial abilities and mechanical reasoning, yet mechanical reasoning cannot be reduced to mere spatial processing.

To more fully account for the complexities in our data, it is helpful to turn to the physics-engine hypothesis (Ullman et al., 2017). This framework proposes that our capacity for intuitive physics is based on runnable mental models built as incomplete and approximate reconstructions of physical scenes. These mental models necessarily make use of spatial processes, as the physical interactions they represent unfold in space and time. However, as these authors propose, the models cannot be explained by spatial ability alone because they also incorporate cognitive shortcuts to generate predictions. Our findings lend empirical support to this theoretical point and provide a layer of nuance. While the physics-engine hypothesis describes a multi-stage process that would likely draw on a variety of spatial skills rather than a single, monolithic ability, our results begin to delineate the nature of this complexity. The differential predictive power of the four spatial abilities offers a specific profile of the most relevant spatial processes by showing that intrinsic static skills seem to be less critical than their counterparts. Furthermore, the remaining unexplained variance in our models reinforces the view that these mental models may also incorporate non-spatial, rule-based components.

The current debate on physical reasoning often centers on whether understanding relies more on mental simulation or on the application of heuristics and rules (Ludwin-Peery et al., 2021; Smith et al., 2023; Sosa et al., 2025). The choice between these strategies is likely not fixed. Instead, it may be moderated by an interaction between an individual's expertise and the problem's context. For instance, an individual with prior knowledge of gear systems understands that adjacent gears rotate in opposite directions. When presented with a simple two-gear problem, they can apply this rule directly, bypassing the need for mental simulation and thereby not necessarily engaging the specific spatial abilities found to be relevant in the present study (Sosa et al., 2025).

The critical influence of context becomes evident in more complex scenarios. Consider a system with a thousand gears. Applying the simple adjacent-gear rule sequentially would be profoundly inefficient, and a full mental simulation of such a large system would be cognitively intractable. An expert, however, might possess a higher level of abstraction based on a more sophisticated rule, for example, identifying that all even-numbered gears turn one way and all odd-numbered gears turn the opposite way, allowing for a rapid and accurate solution (Schwartz & Black, 1996b). This progression from a simple, concrete rule to a more abstract principle suggests a potential developmental pathway in physical reasoning, from effortful mental simulation to the formation of basic rules based on experience, and finally to the development of more abstract and efficient principles. Ultimately, this highlights that a complete model must account for the dynamic interplay between an individual's knowledge and the specific demands of the problem. Exploring this interaction in greater detail presents a promising avenue for future research.

4.2. Future work

We identify three main avenues for future research within this topic. The first avenue involves examining the specific characteristics of distinct mechanical problems to determine whether they require more intrinsic or extrinsic spatial processing, or more static or dynamic spatial processing. This could further clarify the connection between these constructs, building on the findings of this investigation and previous studies in the scientific literature.

A second avenue for future research should address the fundamental relationship between the mental simulation and rule-based pathways, including whether they can operate as truly independent strategies. Experimentally isolating these processes would allow us to test a key hypothesis, that the contribution of spatial skills is exclusive to the mental simulation pathway. Beyond this, it is also critical to investigate the nature of the transition between these modes of thinking about the world. We propose that this transition is primarily directional, with individuals moving from an initial reliance on effortful simulation toward the formation and application of more efficient rules as they gain expertise. Conversely, the reverse pathway from established rule use back to simulation is likely weak or non-existent. Understanding both the isolated functions of these pathways and the transition between them is critical, as recent research suggests that mechanical reasoning involves the flexible use of both (Smith et al., 2023; Sosa et al., 2025).

The third avenue of research involves evaluating the potential causal effects of mechanical reasoning training on the improvement of performance in spatial skills, and vice versa. Munoz-Rubke et al. (2021) provided data suggesting that short-term involvement in mechanical reasoning training can lead to gains in spatial visualization lasting up to a week, an effect not observed with control training. Kozhevnikov and Thornton (2006) also showed that using software for real-time experimental data graphing during physics lectures led to gains in spatial visualization abilities. These results provide a solid foundation for exploring potential causality, but much more work remains to be done. Additionally, there is also a need to investigate whether spatial skills training can lead to improvements in mechanical reasoning. Meta-analyses have suggested that spatial skills are malleable (Uttal et al., 2013), but the effect of spatial skills training on mechanical reasoning capacity remains uncharted territory.

4.3. Limitations

While the current results deepen our understanding of the connection between mechanical reasoning and spatial skills, this study has some limitations. One limitation is that the linkage between mechanical reasoning and spatial skills was explored at an overall level, without delving into the specific connections that might exist between subtypes of spatial ability and different types of mechanical problems. It is reasonable to expect that particular subsets of physical problems could be more strongly connected to certain subtypes of spatial skills.

Another limitation of the current study is that we used single evaluations to target each subtype of spatial ability. Although previous work suggests that each specific evaluation used in this study is representative of the interactions created by the axes of intrinsic vs. extrinsic and static vs. dynamic spatial abilities, our results would have been stronger had we utilized multiple measures to cover each quadrant. Due to time constraints, it was difficult to include more evaluations in the current study. However, future investigations could incorporate multiple measures for each subtype of spatial ability, which would provide a more robust understanding of these connections. This approach could be particularly important for clarifying the connection between intrinsic static spatial skills and mechanical reasoning, given that the task measuring this subtype had the lowest inter-item reliability in our set. By using multiple measures, future research could address this limitation and further validate the findings reported here.

5. Conclusion

Taken together, these results deepen our understanding of the connection between mechanical reasoning and spatial skills. They suggest that mechanical problem solving is associated with intrinsic dynamic, extrinsic static, and extrinsic dynamic spatial processing. This means that, at least at an overall level, mechanical problem solving could be related to more extrinsic (rather than intrinsic) and more dynamic (rather than static) spatial processing. Our findings are consistent with previous research, such as Hegarty (2004), Mitko and Fischer (2020) and Schwartz and Black (1996a, 1996b), which also indicate that mechanical reasoning cannot be fully explained by spatial processing, despite the strong association between these constructs.

Future research should aim to examine the specific characteristics of distinct mechanical problems to determine whether they require more intrinsic or extrinsic processing, or more static or dynamic spatial processing. Additionally, isolating the mental simulation pathway from the rule-based pathway to mechanical problem solving could clarify whether the contribution of spatial skills is exclusive to the former. Lastly, exploring the potential causal effects of training in both mechanical reasoning and spatial skills could provide deeper insights and lead to more effective educational strategies and tools that enhance these cognitive abilities.

CRediT authorship contribution statement

Felipe Munoz-Rubke: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zachary Hawes:** Writing – original draft, Methodology, Conceptualization. **Ramón D. Castillo:** Writing – original draft, Conceptualization. **Karla Avendaño:** Writing – original draft, Investigation, Data curation. **Valentina Zamorano:** Writing – original draft, Investigation, Data curation. **Javier Soto-Mardones:** Writing – original draft, Software, Investigation, Formal analysis, Data curation.

Acknowledgments

FMR would like to thank the Chilean National Agency for Research and Development (ANID) for its support through the grantANID FONDECYT INICIACION 11190742. RDC would like to thank ANID for its support through the grant ANID FONDECYT EXPLORACION 13240168, and the Programa de Investigación Asociativa (PIA) en Ciencias Cognitivas (RU-158-2019) funded by the Universidad de Talca.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2025.101936.

Data availability

Data will be made available on request.

References

Baillargeon, R. (2004). Infants' Physical world. *Current Directions in Psychological Science, 13*(3), 89–94.
Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America, 110*(45), 18327–18332.
Béland, S., & Falk, C. F. (2022). A comparison of modern and popular approaches to calculating reliability for dichotomously scored items. *Applied Psychological Measurement, 46*(4), 321–337.
Boyer, T. W., & Levine, S. C. (2012). Child proportional scaling: Is 1/3=2/6=3/9=4/12? *Journal of Experimental Child Psychology, 111*(3), 516–533.
Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*(5), 1478–1490.
Chatterjee, A. (2008). The neural organization of spatial thought and language. *Seminars in Speech and Language, 29*(3), 226–238. quiz C6.
Cordero, A., & Corral, S. (2006). *DAT–5 tests de aptitudes diferenciales* (Version 2). *TEA Ediciones*.
Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational testing service.
French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for Kit of Reference Tests for Cognitive Factors*. Educational Testing Service.
Friedman, A., Kohler, B., Gunalp, P., Boone, A. P., & Hegarty, M. (2019). A computerized spatial orientation test. *Behavior Research Methods, 22*, 327. –314.
Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition, 157*, 61–76.
Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24.
Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18*(5), 1084–1102.
Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*(2), 151–176.
Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411–430.
Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and Individual Differences, 9*(1), 19–42.
Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences, 8*(6), 280–285.
Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.Js: A free, open, online study builder. *Behavior Research Methods, 54*(2), 556–573.
Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745–756.
Kozhevnikov, M., Hegarty, M., & Mayer, R. (2002). Spatial abilities in problem solving in kinematics. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 155–171). London: Springer.
Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition, 33*(4), 710–726.
Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*(4), 549–579.
Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education and Technology, 15*(1), 111–132.
Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *The Behavioral and Brain Sciences, 40*, e253.
Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479–1498.
Lohman, D.F. (1979). *Spatial ability: A review and reanalysis of the correlational literature (Technical report, N. 8)*. Stanford University, School of Education: Aptitude Research Project. https://apps.dtic.mil/sti/citations/ADA075972.
Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In *Advances in the psychology of human intelligence, 4* pp. 181–248). Psychology Press.
Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on simulation approaches in intuitive physics. *Cognitive Psychology, 127*, Article 101396.
Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*(2), 369–380.
Mitko, A., & Fischer, J. (2020). When it all falls down: The relationship between intuitive physics and spatial cognition. *Cognitive Research: Principles and Implications, 5*(1), 24.
Mitko, A., Navarro-Cebrián, A., Cormiea, S., & Fischer, J. (2024). A dedicated mental resource for intuitive physics. *IScience, 27*(1), Article 108607.
Munoz-Rubke, F., Olson, D., Will, R., & James, K. H. (2018). Functional fixedness in tool use: Learning modality, limitations and individual differences. *Acta Psychologica, 190*, 11–26.
Munoz-Rubke, F., Will, R., Hawes, Z., & James, K. H. (2021). Enhancing spatial skills through mechanical problem solving. *Learning and Instruction, 75*, 1–12.
Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial Thinking: New typology, new assessments. *Studying Visual and Spatial Reasoning for Design Creativity*, 179–192.
R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological Review, 120*(2), 411–437.
Schwartz, D. L., & Black, J. B. (1996a). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology, 30*(2), 154–219.

Schwartz, D. L., & Black, J. B. (1996b). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*(4), 457–497.
Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition, 25*(3), 321–332.
Smith, K.A., Battaglia, P., & Tenenbaum, J. (2023). *Integrating heuristic and simulation-based reasoning in intuitive physics*. https://doi.org/10.31234/osf.io/bckes.
Smith, K. A., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science, 5*(1), 185–199.
Sosa, F. A., Gershman, S. J., & Ullman, T. D. (2025). Blending simulation and abstraction for physical reasoning. *Cognition, 254*(105995), Article 105995.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*(1), 29–56.
Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind Games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences, 21*(9), 649–665.
Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.