Learning and Instruction 75 (2021) 101496

Contents lists available at ScienceDirect

Learning and Instruction

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

Enhancing spatial skills through mechanical problem solving

Felipe Munoz-Rubke [a,b,c,*], Russell Will [d], Zachary Hawes [e], Karin H. James [d]

[a] Instituto de Psicología, Universidad Austral de Chile, Puerto Montt, Chile
[b] Cognitive Science Program, Indiana University, Bloomington, IN, USA
[c] Program in Neuroscience, Indiana University, Bloomington, IN, USA
[d] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
[e] Department of Applied Psychology & Human Development, Ontario Institute for Studies in Education, Ontario, Canada

ARTICLE INFO

Keywords:
Spatial skills
Spatial training
Mechanical problem solving
Mechanical reasoning
STEM

ABSTRACT

Higher spatial skills are associated with increased interest, performance, and creativity in STEM fields (Science, Technology, Engineering, Mathematics). However, evidence for causal relations between spatial skills and STEM performance remains scarce. In this study, we test the extent to which mechanical problem solving, a spatially demanding STEM activity, facilitates spatial performance. Participants (N = 180) were randomly assigned to one of four training conditions: mechanical reasoning with a hands-on component; mechanical reasoning without a hands-on component; an active control condition involving spatial training with cross-sectioning; and an active control group involving a reading exercise. All participants were tested immediately before, after, and one-week following training. Both mechanical conditions were associated with enhanced spatial visualization performance, an effect that was similar for both conditions and remained stable across immediate and delayed post-tests. These findings suggest that mechanical problem solving is a potentially viable approach to enhancing spatial thinking.

For a while I gave myself up entirely to the intense enjoyment of picturing machines and devising new forms. . . . The pieces of apparatus I conceived were to me absolutely real and tangible in every detail, even to the minutest marks and signs of wear. I delighted in imagining the motors constantly running. . . . In less than 2 months I evolved virtually all the types of motors and modifications of the system which are now identified with my name. (Tesla, 1919/1995, p. 65)

1. Introduction

This quote provides a glimpse into the mind of Nikola Tesla, the famed mechanical engineer and inventor. Central to Tesla's inventive creativity, including the invention of the induction motor, was his capacity to "picture" and operate on objects in his mind as if they were real (Tesla, 1995; von Károlyi, 2013). This type of thinking bears striking resemblance to what is often referred to as spatial thinking; the ability to generate, recall, transform, and manipulate visual-spatial information (Lohman, 1996). Critically, Tesla is not alone in the importance he placed on spatial thinking. Many other scientific advancements were said to have relied heavily on spatial thinking, including Einstein's theory of relativity, the discovery of the structure of DNA, and mapping the spread of disease (Newcombe, 2010, 2016).

In addition to anecdotal evidence, there is strong empirical support for the importance of spatial thinking in scientific discovery and innovation (Kell, Lubinski, Benbow, & Steiger, 2013). For example, spatial skills have been found to uniquely predict creativity and technical innovation in the workplace (Kell et al., 2013). Especially strong relations have been revealed between spatial thinking and STEM performance (Science, Technology, Engineering, and Mathematics; Wai, Lubinski, & Benbow, 2009). Evidence from large-scale longitudinal studies ($N = 400{,}000$) have found spatial skills to strongly predict which students enter, and succeed in STEM disciplines, even after taking verbal and quantitative reasoning into account (Lubinski, 2010; Shea, Lubinski, & Benbow, 2001; Wai et al., 2009).

This evidence, coupled with the growing need to increase the number of qualified STEM professionals (Fayer, Lacey, & Watson, 2017), has led to increased efforts to better establish and elucidate the underlying mechanisms that potentially link spatial thinking and STEM performance. In this paper, we ask whether and to what extent engaging in a STEM-relevant skill, mechanical problem solving, transfers to spatial visualization performance. More specifically, we examine the effect that

* Corresponding author. Los Pinos s/n Balneario Pelluco, Puerto Montt, Chile.
E-mail address: felipe.munoz@uach.cl (F. Munoz-Rubke).

https://doi.org/10.1016/j.learninstruc.2021.101496
Received 1 December 2019; Received in revised form 14 May 2021; Accepted 19 May 2021
Available online 8 June 2021

a brief mechanical training intervention has on participants' spatial visualization performance compared to two active control conditions.

1.1. Malleability of spatial thinking

Given the close link between spatial skills and STEM performance, an important question concerns the extent to which spatial skills are malleable and transferable. That is, can spatial thinking be improved through training, and if so, does training generalize to related domains such as science and mathematics? Findings from a recent meta-analysis indicate that spatial thinking not only can be improved through training, but that these effects also appear durable and transferrable to other spatial tasks not directly trained (Uttal et al., 2013). In fact, improvements in both trained and untrained measures were nearly identical, with effect sizes approaching 0.5 standard deviations. These findings suggest that spatial thinking is a highly malleable cognitive construct, with potentially significant and far reaching implications for STEM education.

However, the evidence regarding the extent to which spatial training generalizes to STEM performance has been met with mixed and largely inconclusive results. For example, while there is some evidence that spatial training generalizes to mathematics performance (e.g., see Hawes, Moss, Caswell, Naqvi, & MacKinnon, 2017; Cheng & Mix, 2014; Gilligan, Flouri, & Farran, 2017), there is also evidence suggesting the contrary (e.g., see Hawes, Moss, Caswell, & Poliszczuk, 2015; Cornu, Schiltz, Pazouki, & Martin, 2019; Xu & LeFevre, 2016). Further complicating the issue are the relatively few tightly controlled and adequately powered studies. By and large, researchers have used underpowered quasi-experimental (non-randomized) study designs to investigate transfer effects (e.g., see Hawes et al., 2017). Thus, more tightly controlled experiments are needed to test for causal relations between spatial thinking and STEM performance. The current study fills this gap by employing a randomized controlled study, involving two active control groups: A mental cross-sectioning training group to test for the specificity of transfer (Active Control 1) and a reading control group to control for test-retest effects (Active Control 2).

1.2. Approaches to spatial training

Investigations of causal relations between spatial thinking and STEM performance have assumed a directionality of effects. To date, most spatial interventions have approached training either through repeated exposure to spatial tests (e.g., repeated practice with mental rotation items; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008) or spatially demanding computer/video games (e.g., Tetris; Sims & Mayer, 2002). This approach assumes a directionality in relations in which more "basic" and narrowly defined spatial skills are hypothesized to underlie more advanced STEM-related performance and disciplinary practice. It is common practice to train participants on a single psychometrically defined spatial skill, such as mental rotation, and then test for transfer on more broadly defined and domain-relevant skills, such as mathematics or engineering (e.g., see Hawes et al., 2015; Cheng & Mix, 2014; Gilligan et al., 2017; Sorby, 1999). However, it is also possible that spatially demanding activities, of the type inherent in many STEM activities, such as mechanical engineering or architecture, might also be important sources for improving spatial thinking.

The purpose of the present study was to further examine this idea through a brief intervention that involved training participants on a series of 3D mechanical reasoning puzzles. We were interested in whether exposure to domain-relevant skills (i.e., mechanical problem solving) might facilitate performance on untrained measures of spatial visualization. As noted above, this approach is novel in that it reverses the directionality of transfer typically assumed. Rather than train participants on abstract objects that are not explicitly connected to a specific STEM domain/skill, we test the possibility that spatial thinking can be improved through directly interacting with a domain-relevant STEM activity, mechanical reasoning. Because this approach is better aligned with current and emerging STEM curricula and educational practices (e.g., makerspaces), it avoids the opportunity costs potentially associated with isolated spatial training approaches. For example, the time and effort used to train participants on a single spatial skill, such as mental rotation (Hawes et al., 2015), potentially takes away time and effort that could be spent training spatial thinking through more naturalistic, contextualized, and educationally relevant STEM activities (e.g., making activities as described in Ramey, Stevens, & Uttal, 2020).

1.3. Theoretical underpinnings of the current study

1.3.1. Descriptive accounts of relations between mechanical reasoning and spatial visualization

There are several reasons why mechanical problem solving may enhance individuals' performance in spatial visualization tasks, including measures of mental paper folding and mental rotation. Both of these measures have been strongly linked to spatial visualization, though the consensus seems to be stronger for paper folding (Carroll, 1993; Lohman, 1988; McGee, 1979; Michael, Guilford, Fruchter, & Zimmerman, 1957) than for mental rotation (see Linn & Petersen, 1985). Ontologically, present day definitions and tests of spatial abilities originate from the development and use of three-dimensional mechanical reasoning tasks of the early 20th century (Hegarty & Waller, 2005; Smith, 1964). These early mechanical tests involved a combination of both physical and mental manipulation and provided the groundwork and inspiration for paper-and-pencil assessments of spatial abilities (see Smith, 1964). A look at the early definitions of spatial abilities reflect their mechanical origins. For example, Thurstone (1950) defined spatial visualization as "*the ability to imagine the movement or internal displacement among the parts of a configuration that one is thinking about*" (p. 518). Against this background, perhaps it should be of little surprise that descriptive accounts of the cognitive processing involved in mechanical problem solving are conceptually similar – if not isomorphic – to those ascribed to spatial visualization (Harris, Hirsh-Pasek, & Newcombe, 2013; Hegarty, 2004). For example, the ability to imagine solutions to problems that require complicated, multi-step manipulations and/or simulations of spatially presented information is common to both spatial visualization and mechanical reasoning tasks (Hegarty & Sims, 1994). Indeed, spatial visualization has been posited as a primary means for how people reason about mechanical systems (Hegarty, 1991, 2004);an approach that involves evaluating the physical structure of the system and then mentally simulating in piecewise fashion the fundamental relations of the component parts, eventually allowing one to make sense of the system as a whole. To summarize, spatial visualization and mechanical reasoning are operationalized in similar ways and appear to rely on similar multi-step reasoning and problem-solving strategies.

1.3.2. Empirical accounts of relations between mechanical reasoning and spatial visualization

Critically, there is also empirical support for shared cognitive processing of mechanical reasoning and spatial visualization. Prior research has revealed strong positive correlations between spatial visualization skills and mechanical problem solving (Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997; Kozhevnikov, Motes, & Hegarty, 2007). For example, Hegarty and Sims (1994) found mechanical reasoning to correlate with mental rotation at $r = 0.60$ and with mental paper folding at $r = 0.76$. Moreover, there is a long history of factor analytic studies demonstrating that both constructs load on the same factor (Hamilton, Nussbaum, Kupermintz, Kerkhoven, & Snow, 1995; Humphreys, Lubinski, & Yao, 1993; Smith, 1964); a finding that perhaps speaks to the fact that spatial visualization assessments were borne out of mechanical reasoning tasks (Smith, 1964). Taken together, the available evidence suggests that spatial visualization and mechanical reasoning may recruit highly similar cognitive resources. Thus, it is possible that experience with one task may facilitate performance in the other. In the present study, we set out to test this hypothesis and examine the extent to which mechanical problem-solving transfers to spatial visualization performance. Based on the evidence reviewed above, we had reasons to expect near transfer from mechanical reasoning to spatial visualization tasks.

1.4. Enhancing spatial visualization through active vs. passive mechanical reasoning

Prior research indicates that spatial visualization performance, particularly mental rotation, can be improved following spatial training approaches that involve an active 'hands-on' feedback component (e.g., see Adams, Stull, & Hegarty, 2014; Wiedenbauer & Jansen-Osmann, 2008; Wiedenbauer, Schmid, & Jansen-Osmann, 2007). As theorized by Wohlschläger and Wohlschläger (1998), mental imagery and the physical instantiations of one's spatial imaginings are functionally dependent on one another and experience in one should affect the other. Research into the neural mechanisms underlying mental rotation further implicate the motor system (Wraga, Thompson, Alpert, & Kosslyn, 2003; Zacks, 2008). According to Wexler, Kosslyn, and Berthoz (1998), "*visuomotor anticipation is the engine that drives mental rotation*" (p. 79). These findings indicate the potential benefits of having participants physically interact with the mechanical puzzles and test the accuracy of their visualized solutions. Accordingly, the effects of mechanical reasoning on participants' spatial visualization performance should be stronger under training conditions in which physical manipulation is present than when absent.

To test this possibility, the present study included two forms of mechanical training; a hands-on active approach vs. a hands-off passive approach. More specifically, we included two mechanical training conditions to distinguish whether the potential gains in spatial skills were related to the visualization and/or feedback phase of training. The two mechanical reasoning groups trained on an identical set of 3D mechanical puzzles. In both mechanical conditions, participants were required to visualize (or mentally simulate) solutions to a given mechanical problem. Critically, the groups differed in the feedback phase of training. Whereas the *active mechanical condition* executed their planned solutions to the problems by directly interacting with the puzzles, participants in the *passive mechanical condition* observed video recordings of the correct solutions to the problems. Because both groups differed only in the feedback phase (and not in their need to engage in spatial visualization), we reasoned that any differences in spatial performance may be attributed to differences afforded through the active vs. passive component of training. In the absence of group differences, any gains in participants' spatial visualization may be attributed to the spatial visualization phase. That is, if spatial performance is observed to be similar between mechanical conditions, this suggests that the feedback phase (active vs. passive) has little influence and would provide stronger support for the spatial visualization phase of training.

1.5. Aims and hypotheses of the present study

The general purpose of the present study was to test the effects of mechanical training on spatial visualization performance. We hypothesized that compared to two active control groups (i.e., mental cross-sectioning and reading), participants assigned to either mechanical training condition would demonstrate gains in spatial visualization performance. This prediction was based on the shared-processing account and the need to engage in multi-step spatial visualization processes across both mechanical conditions.

A more specific aim of the present study was to further test the effects of active vs. passive mechanical training on participants' spatial visualization performance. Based on prior research indicating the benefits of active participation, we hypothesized that the active mechanical condition would confer even greater benefits to participants' spatial performance than the passive mechanical condition.

To further test the specificity of the mechanical training, we included an active control condition which involved training participants on mental cross-sectioning problems (see C. A. Cohen & Hegarty, 2014). This condition was included to provide a more stringent comparison than the reading condition, allowing us to directly compare the potentially differential effects of two different forms of visualization training. Although we had reason to believe that mental cross-sectioning would facilitate mental cross-sectioning performance (replicating the findings of C. A. Cohen and Hegarty, 2014), we were less certain that this form of training would be as effective as mechanical training in facilitating spatial visualization performance (as assessed through mental rotation and mental paper folding measures). Specifically, we hypothesized that mechanical training would more effectively facilitate spatial visualization performance than mental cross-sectioning due to the greater need to engage in multi-step spatial visualization processes (e.g., Hegarty & Sims, 1994; Smith, 1964). In summary, based on the shared-processing account as well as the potential added benefit of physical manipulation, we predicted differential training effects on participants' spatial visualization across all four conditions.

2. Methods

2.1. Participants

One hundred and eighty native English speakers ($M = 19.8$ years, $SD = 2.98$, range = 18–52 years, 99 females) participated in this experiment. Participants were randomly assigned to one of four conditions: Active Mechanical (n = 45, 31 females), Passive Mechanical (n = 45, 23 females), Active Control 1 (cross-sectioning) (n = 45, 21 females), and Active Control 2 (reading) (n = 45, 24 females). According to a power analysis, this was the number of participants required to achieve a power of .80 ($\alpha = 0.05$) in detecting a medium effect size ($f = 0.25$) when using a one-way ANCOVA (number of groups = 4, numerator df = 3) (J. Cohen, 1988; Faul, Erdfelder, Lang, & Buchner, 2007).

Participants were recruited through the Department of Psychological and Brain Sciences Course Credit Subject Pool and by posters located around the Indiana University campus. Participants received course credit or financial compensation for their participation, had normal or corrected to normal vision and reported no history of neurological disorders. Informed consent was obtained before the experiment, in accordance with the Indiana University Institutional Review Board approved protocol.

2.2. Materials

2.2.1. Measures and testing procedures

Participants completed the same four measures before (pre-test), immediately after (post-test), and a week after training (delayed post-test). These measures included two well-known tests of spatial visualization (Mental Paper Folding and Mental Rotation), one recently developed test of cross-sectioning ability (Santa Barbara Solids; C. A. Cohen & Hegarty, 2012), and a Visual Search task. We selected these assessments because they measure distinct skills. Moreover, mental paper folding and mental rotation involve multi-step visualization, cognitive processes previously shown to be similar to those involved with mechanical problem solving (e.g., Hegarty & Sims, 1994; Smith, 1964). The visual search task, on the other hand, does not involve spatial visualization. Instead, the visual search task is a measure of visuospatial perception. Including this measure allowed us to test the specificity of transfer between mechanical training and spatial visualization. If mechanical reasoning provides a means to train spatial visualization skills, then we should expect the largest gains to occur on measures of spatial visualization.

The items of each assessment were divided into three groups, so that they could be presented before, immediately after, and a week after the training session. Due to this, no test items were shown more than once to a participant. This was done so that participants would not respond to any stimuli by relying on memory. The order of presentation of these item groups was counterbalanced across participants to avoid order effects. The order of presentation of the items within each group was fully randomized. Raw scores were used for all analyses.

2.2.2. Description of measures

Mental Paper Folding Test (French, Ekstrom, & Price, 1963): Participants were shown a diagram of a piece of paper being folded multiple times and then a hole being punched through a specific location. Participants then indicated how the unfolded sheet of paper would look like by selecting among five alternatives. Previous investigations indicate that the Paper Folding test measures spatial visualization (Carroll, 1993; Linn & Petersen, 1985; Lohman, 1988; McGee, 1979; Michael et al., 1957), as it requires the multistep manipulation of spatial information. This test contained the first 18 items out of the original list of 20 items, which were equally divided among the pre-, post-, and delayed post-test assessments.

Mental Rotation Test (Peters et al., 1995; Vandenberg & Kuse, 1978): Participants were presented with two-dimensional representations of three-dimensional cube figures and then asked to identify which two of four figures were identical to the target figure. The correct alternatives corresponded to a rotated version of the target figure. Although Linn & Petersen suggest that this test evaluates mental rotation ability, other studies indicate that this test is also strongly related to spatial visualization (Lohman, 1988; Vandenberg & Kuse, 1978). The test used in the current study contained 24 questions from the Peters et al. (1995) adaptation of the Vandenberg Mental Rotation Test, equally divided among the pre-, post-, and delayed post-test assessments.

Santa Barbara Solids Test (C. A. Cohen & Hegarty, 2012): Participants were presented with images of geometric solids containing a flat plane cutting through them. Subsequently, they indicated which two-dimensional shape would be obtained if the geometric solid were to be sectioned by the cutting plane by selecting among four alternatives. According to C. A. Cohen and Hegarty (2012), this test is a measure of cross-sectioning ability and is only partially related to spatial visualization. This test contained 30 questions, equally divided among the pre-, post-, and delayed post-test assessments.

Visual Search Task (Saarinen, 1994): Participants were required to identify a target pattern among distracting visual information. Participants looked for a local pattern within a global shape. In a background that contained randomly assigned letters “L” and “T”, participants first looked for a circle of “Ls” (global shape), and then indicated whether this circle contained one letter “T” (local pattern). This task did not measure spatial visualization, but instead served as a control task by measuring our participants’ visuospatial perception. This task contained 30 stimuli, equally divided among the pre-, post-, and delayed post-test assessments.

2.2.3. Reliability

The Intraclass Correlation Coefficient (ICC) for test-retest reliability of each assessment was calculated by considering the average performance of participants in the Active Control 2/Reading group across all three time points (pre-test, post-test, delayed post-test). We only considered these participants when calculating the ICCs, as they took part in the only group that did not receive an intervention targeted at modifying their spatial skills. The ICC for the Paper Folding test was 0.54 (fair reliability), for the Vandenberg Mental Rotation test was 0.76 (good reliability), for the Santa Barbara Solids test was 0.36 (poor reliability), and for the Visual Search task was 0.50 (fair reliability).

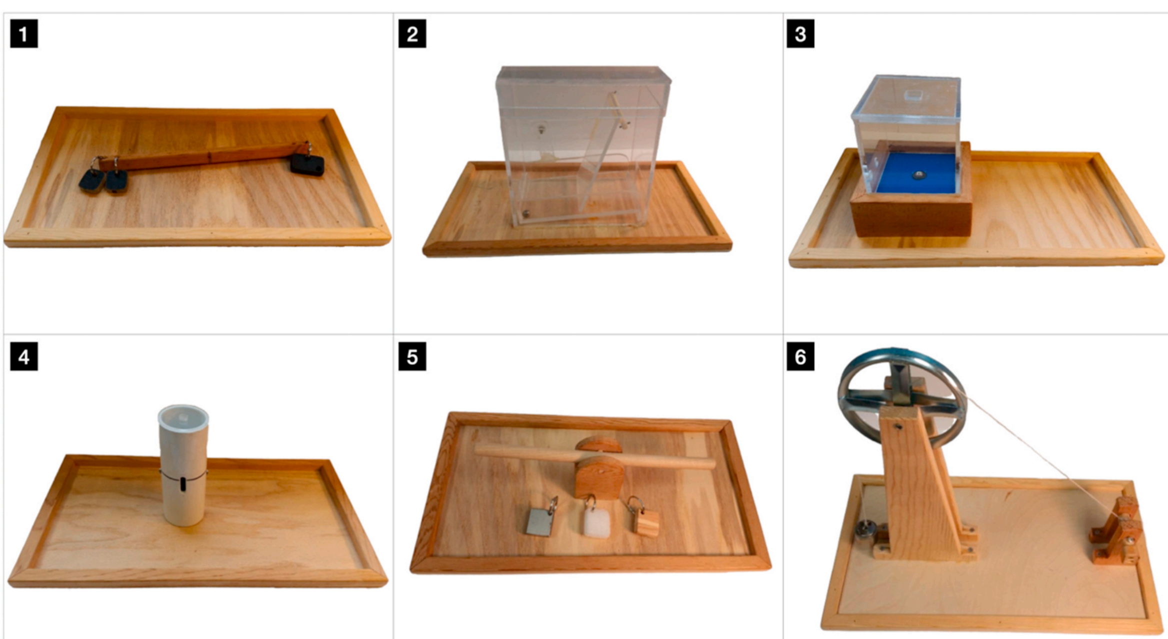

Fig. 1. This figure shows the mechanical puzzles used in the present experiment. (1) *Scale Balance*: There is a wooden bar with 3 weights hung asymmetrically on it, such that one side is heavier than the other. The goal is to balance the bar so that it is parallel to the wooden tray that holds the task. (2) *Box Flap*: One marble is located at the bottom of a transparent plastic structure. The goal is to remove the marble from the box. The plastic structure has 2 holes, but only one of them can be used to successfully solve the problem. (3) *Marble Push*: A marble is located within a hollow rubber circle that is positioned on the floor of a transparent plastic cube. Underneath the rubber circle there is a PVC structure connecting the floor of the transparent structure with the wooden tray. The goal is to get the marble out of the box. (4) *PVC Pipe:* There is a tube that contains an internal metal platform. The metal platform is located halfway up the tube and has a marble atop it. The goal is to remove the marble from the tube. (5) *Scale Prop*: There is a wooden scale and 3 different weights. The first weight is made of metal, the second is made of wood, and the third is made of styrofoam. The goal is to balance this scale parallel to the wooden tray that holds the task. (6) *Weight Wheel:* There is a large, elevated wheel that contains a string within its crevice. The string is attached to a reel on one side of the wheel and a weight on the other side, making up a pulley mechanism. The string is attached to the reel through a hole in the center and contains an arm that can be rotated to spin the string. The goal is to make the weight remain elevated at the base of the wheel. For further details, please see.(Munoz-Rubke et al., 2018)

2.2.4. Mechanical puzzles

Both mechanical conditions involved training with an identical set of 3D mechanical puzzles (Munoz-Rubke, Olson, Will, & James, 2018). In total, participants were presented with six puzzles: Scale Balance, Box Flap, Marble Push, PVC Pipe, Scale Prop, and Weight Wheel (see Fig. 1). These names were not divulged to the participants. The Scale Balance puzzle was included as a practice item to familiarize participants with the training procedure.

The mechanical puzzles were made of wood, metal, and/or plastic and were individually positioned on top of a wooden tray with dimensions 45 cm × 25 cm x 1.5 cm. Each puzzle was presented alongside three tools, one of which was required to solve it. The tools were constructed out of wood, metal, and/or plastic and their sizes ranged from 4 to 25 cm along their largest axis (see Fig. 2). Results from a previous investigation (Munoz-Rubke et al. 2018), showed that the mechanical puzzles had different levels of difficulty. Specifically, Marble Push has been found to be the most complicated mechanical puzzle (median = 11.6%, 95% CI = [6.7%, 18.3%]), followed by Weight Wheel (median = 28.7%, 95% CI = [21.7%, 36%]), PVC Pipe (median = 45.9%, 95% CI = [33.7%, 60%]), Box Flap (median = 47.2%, 95% CI = [35.5%, 63.4%], Scale Prop (median = 47.3%, 95% CI = [35.6%, 61.8%], and Scale Balance (median = 73.6%, 95% CI = [56.24%, 93.9%]).

2.3. Experimental procedures

Following random assignment to a group, participants completed the four pre-training assessments (Mental Paper Folding, Mental Rotation, Santa Barbara Solids, Visual Search). Then, each participant received no more than 15 min of training. The pre-test, training, and immediate post-tests were all conducted in a laboratory setting. For all computer tasks, including the assessments and the two active control conditions, a 2013 Apple iMac (16:9 aspect ratio, 2560 × 1440 native resolution) was used. All spatial measures were administered using Qualtrics™. Although no participants took part in their respective condition for more than 15 min (excluding the initial instructions and practice trials), we used a hard deadline of 15 min for the Active Control 2 (Reading) condition.

While post-training assessments were administered immediately after the training was completed, delayed post-training assessments were emailed to participants exactly one week after their involvement in the experiment. Participants were required to complete the delayed post-training assessment on the same day they received such email. Four participants were excluded from the study due to lack of compliance with this rule.

2.3.1. Active and passive mechanical training

Each puzzle was presented one at a time in a completely randomized order. Participants were first presented with a photograph depicting each mechanical puzzle, as well as with the three accompanying tools used to solve each puzzle (e.g., wrench, rod, paint mixer). A trained examiner then communicated the goal of each problem through verbal instructions. Only then the examiner introduced the actual 3D puzzle, placing it on the desk directly in front of the participant. Participants were then given up to 60 s to plan/visualize their solutions to the problem. They were told to work as quickly as possible, but to prioritize finding the correct solution. They were also instructed not to touch the puzzles or tools during this planning phase. However, they were permitted to rotate the tray on which the puzzle was positioned to gain a complete perspective of the puzzles.

Each plan comprised selecting one tool out of the three alternatives, as well as defining a specific course of action. Following this, participants in the active mechanical condition proceeded to manually execute their plans, while participants in the passive mechanical condition observed a video in which an experimenter demonstrated how to solve each mechanical problem. Note that this was the sole difference between conditions.

2.3.2. Active control 1 (cross-sectioning training)

Both the task and procedures for the cross-sectioning condition were based on the work of C. A. Cohen and Hegarty (2014), which we tried to replicate as faithfully as possible. In this training condition, participants were presented with pictures of 3D geometric figures on a computer screen. Each figure was presented as a static 3D solid (e.g., cube, cylinder, pyramid) intersected by a flat plane at a vertical, horizontal, or

Fig. 2. This figure shows the tools that were available to participants while solving each mechanical puzzle. The first row shows the tools associated with (1) Scale Balance, (2) Box Flap, and (3) Marble Push. The second row shows the tools associated with (4) PVC Pipe, (5) Scale Prop, and (6) Weight Wheel. For a detailed description of each tool, please see (Munoz-Rubke et al., 2018).

diagonal crosscut. Participants were told that they could move the flat plane through the shape, while they attempted to draw the intersection of the geometric figure and the cutting plane. Once satisfied with their drawings of the predicted cross-section, participants were presented with the correct drawing of the cross-section. Participants then assessed whether both drawings were similar or not. Participants marked with a '+' sign if both drawings were the same, or with a '−' sign if the drawings were different. A total of 15 geometric figures were presented and the experimenter cycled back to any figures that were drawn incorrectly. The order in which the geometric figures were presented to participants was fully randomized.

2.3.3. Active control 2 (reading)

Participants had to read an autobiographical novel written by one of the Founding Fathers of the United States. Participants read the text for 15 min, on the same computer that they used to take the assessments.

2.4. Statistical procedures

We used logistic multilevel models and ANCOVAs to analyze our data. While we used the logistic multilevel models to evaluate changes in performance in each assessment by training condition and across the three time points, we used the ANCOVAs to compare performances among conditions at the post-test and delayed post-test stages, using average performance at pre-test as a covariate.

Each of the logistic multilevel models was used to evaluate whether participants' solution accuracy in each of the four assessments changed as a function of the training condition (Condition: Active Mechanical, Passive Mechanical, Active Control 1, and Active Control 2) and time (Time: Pre-test, Post-test, Delayed post-test). For each of these four models, Participants were included as a random-effect variable, and Condition and Time as fixed-effects. Each of these models included our participants' responses to each assessment as the response variable. We decided to use logistic models because for each test item, a 0 was assigned when participants did not select the correct answer, and a 1 when they selected the correct response. We choose to report all main effects of the logistic multilevel models by means of an Analysis of Deviance table with Type III Wald chi-square tests for ease of interpretation (McCullagh & Nelder, 1989). For each of these models, we also ran multiple comparisons to compare the performance of the training groups across the three time points. All multiple comparisons were FDR corrected.

After estimating changes in performance within each training condition, we compared the gains obtained at the post-test and delayed post-test stages among groups. We ran ANCOVA models, one for each assessment at post-test and one for each assessment at delayed post-test. In each of these models, Condition was the main predictor, average performance at pre-test was the covariate, and average performance at either the post-test or delayed post-test was the response variable. For each of these models, we also ran multiple comparisons to compare the groups pairwise. All multiple comparisons were FDR corrected.

All statistical procedures were implemented in the R programming language. The logistic multilevel models were estimated using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and the multiple comparisons with FDR corrections were estimated using the lsmeans package (Lenth, 2016).

3. Results

3.1. Descriptive statistics for both mechanical training conditions

Participants in the Active Mechanical condition took an average of 34.61 s ($SD = 12.61$) in planning each of their solutions and an average of 32.51 s ($SD = 8.06$) in executing each of their solutions. Participants in the Passive Mechanical condition took an average of 36.11 s ($SD = 11.81$) in planning each of their solutions. Participants in the passive group did not have to execute their solutions to the mechanical problems.

Both groups showed similar accuracy rates in mechanical problem solving, as participants in the active condition solved 40.4% of the mechanical puzzles ($SD = 21.5$) and participants in the passive condition solved 43.5% of the mechanical puzzles ($SD = 17.7$).

3.2. Main results

3.2.1. Mental Paper Folding Test

We first evaluated whether solution accuracy in the Mental Paper Folding Test changed as a function of Condition and Time. An Analysis of Deviance showed a main effect of Time ($\chi^2(2) = 40.77$, $p < .001$) and an interaction between Condition and Time ($\chi^2(6) = 13.68$, $p = .033$). No main effect of Condition was observed ($\chi^2(3) = 5.75$, $p = .124$).

FDR corrected multiple comparisons showed that the Active Mechanical group showed higher odds of solution accuracy during the post-test than during the pre-test ($OR = 1.95$, $p < .001$). Similarly, this group showed higher odds of solution accuracy during the delayed post-test than during the pre-test ($OR = 2.40$, $p < .001$). In contrast, no statistically significant difference was found between the group's post-test and delayed post-test performances ($OR = 0.81$, $p = .263$). These results indicate that the Active Mechanical group increased their performance in the Mental Paper Folding Test following training, and that this increase remained largely intact for at least one week (we will refer to this pattern as *'immediate and sustained gains'*).

The Passive Mechanical group also showed *immediate and sustained gains*, as both its post-test ($OR = 2.03$, $p < .001$) and delayed post-test ($OR = 2.10$, $p < .001$) performances were linked to higher odds of solution accuracy than its pre-test performance. Additionally, the post-test and delayed post-test performances in this condition did not statistically differ from each other ($OR = 0.97$, $p = .854$).

For both control groups, no differences were found across time points. The Active Control 1 (cross-sectioning) group did not show statistically significant differences between the post-test and the pre-test ($OR = 1.36$, $p = .245$), the delayed post-test and the pre-test ($OR = 1.25$, $p = .289$), and the post-test and the delayed post-test ($OR = 1.08$, $p = .660$). Nearly identical patterns were observed for the Active Control 2 (reading) group, as there were no statistically significant differences between the post-test and the pre-test ($OR = 1.22$, $p = .575$), the delayed post-test and the pre-test ($OR = 1.16$, $p = .575$), and the post-test and the delayed post-test ($OR = 1.05$, $p = .792$).

An ANCOVA model that estimated changes in average performance in the Mental Paper Folding test during the post-test, using Condition as the main predictor and Pre-test (average performance during the pre-test) as a covariate, showed a main effect of Condition ($F(3,172) = 3.28$, $p = .022$, $\eta^2 = 0.06$), a main effect of Pre-test ($F(1,172) = 7.75$, $p = .006$, $\eta^2 = 0.04$), and no interaction between Condition and Pre-test ($F(3,172) = 2.30$, $p = .080$, $\eta^2 = 0.04$). Subsequent FDR corrected pairwise comparisons showed that the Active Mechanical group showed higher average performances than both the Active Control 1 ($p = .031$) and the Active Control 2 ($p = .031$) groups. Similarly, the Passive Mechanical group also showed higher average performances than both the Active Control 1 ($p = .045$) and the Active Control 2 ($p = .031$). No further contrasts were statistically significant (all $ps > .792$).

Another ANCOVA model that estimated changes in average performance during the delayed post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Condition ($F(3,172) = 6.79$, $p < .001$, $\eta^2 = 0.10$), no effect of Pre-test ($F(1,172) = 0.28$, $p = .594$, $\eta^2 = 0.00$) and no interaction between Condition and Pre-test ($F(3,172) = 1.76$, $p = .156$, $\eta^2 = 0.03$). Subsequent FDR corrected pairwise comparisons showed that the Active Mechanical group showed higher average performances than both the Active Control 1 ($p = .001$) and the Active Control 2 ($p = .003$) groups. The Passive Mechanical group showed a higher average performance than the Active Control 1 ($p = .040$), but not than the Active Control 2 ($p = .080$) condition. No further contrasts were statistically significant (all *ps* > .191) (see Fig. 3A).

3.2.2. Vandenberg Mental Rotation Test

The model that evaluated whether solution accuracy in the Mental Rotation test changed as a function of Condition and Time showed a main effect of Time ($\chi^2(2) = 43.84, p < .001$) and an interaction between Condition and Time ($\chi^2(6) = 12.94, p = .044$). No main effect of Condition was observed ($\chi^2(3) = 0.02, p = .999$).

FDR corrected multiple comparisons showed that both the Active and Passive Mechanical groups showed *immediate and sustained gains*. In the Active group, both its post-test ($OR = 1.84, p < .001$) and delayed post-test ($OR = 1.53, p = .008$) performances were linked to higher odds of solution accuracy than its pre-test performance. No statistically significant difference was found between this group's post-test and delayed post-test performances ($OR = 1.20, p = .227$). Similarly, in the Passive group both its post-test ($OR = 1.72, p < .001$) and delayed post-test ($OR = 2.03, p < .001$) performances were linked to higher odds of solution accuracy than its pre-test performance. No statistically significant difference was found between this group's post-test and delayed post-test performances ($OR = 0.85, p = .290$).

The Active Control 1 (cross-sectioning) group showed higher odds of solution accuracy during the post-test than during the pre-test ($OR = 1.60, p = .006$). However, they also showed higher odds of solution accuracy during the post-test than during the delayed post-test ($OR = 1.48, p = .015$) and no statistically significant difference was found between the delayed post-test and the pre-test performances ($OR = 1.08, p = .594$). This suggest that what was gained immediately following the intervention was lost one week later. The Active Control 2 (reading) group did not show statistically significant differences between the post-test and the pre-test ($OR = 1.43, p = .057$), the delayed post-test and the pre-test ($OR = 1.14, p = .403$), and the post-test and the delayed post-test ($OR = 1.26, p = .195$).

An ANCOVA model that estimated changes in average performance in the Mental Rotation test during the post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Pre-test ($F(1,172) = 6.13, p = .014, \eta^2 = 0.03$), no main effect of Condition ($F(3,172) = 0.31, p = .815, \eta^2 = 0.00$), and no interaction between Condition and Pre-test ($F(3,172) = 2.02, p = .113, \eta^2 = 0.03$).

Another ANCOVA model that estimated changes in average performance during the delayed post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Pre-test ($F(1,172) = 16.14, p < .001, \eta^2 = 0.08$) and an interaction between Condition and Pre-test ($F(3,172) = 5.17, p = .002, \eta^2 = 0.07$), but not a main effect of Condition ($F(3,172) = 2.51, p = .060, \eta^2 = 0.05$). Subsequent FDR corrected pairwise comparisons showed that the Passive Mechanical group showed higher average performances than both the Active Control 1 ($p = .027$) and the Active Control 2 ($p = .028$) groups. No further contrasts were statistically significant (all *ps* > .225) (see Fig. 3B).

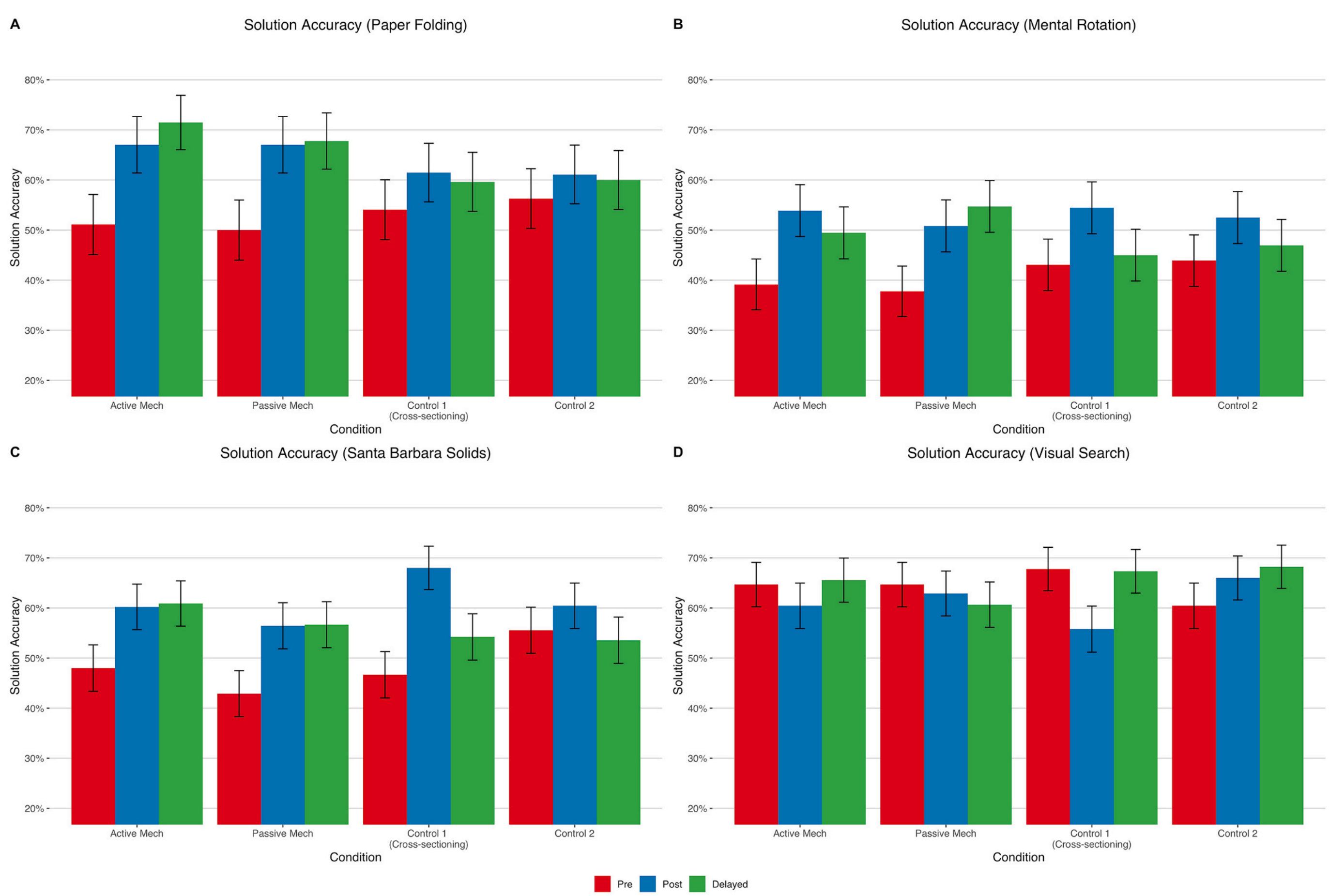

Fig. 3. This figure shows the results for each training condition across all time points of data collection. (A) In the Paper Folding test, both mechanical groups showed immediate and sustained gains. Compared to at least one of the control groups, both mechanical conditions demonstrated gains immediately following the training, as well as one week later. (B) In the Mental Rotation test, both mechanical groups showed immediate and sustained gains. During the delayed post-test, one-week later, the passive mechanical condition demonstrated significant gains in mental rotation compared to both control groups. (C) In the Santa Barbara Solids test, both mechanical groups showed immediate and sustained gains. (D) In the Visual Search task, the mechanical groups did not show differences in performance across all three time points.

3.2.3. Santa Barbara Solids Test

A model that evaluated whether solution accuracy in the Santa Barbara Solids test changed as a function of Condition and Time indicated a main effect of Time ($\chi^2(2) = 64.05$, $p < .001$) and an interaction between Condition and Time ($\chi^2(6) = 27.22$, $p < .001$). No main effect of Condition was observed ($\chi^2(3) = 4.97$, $p = .174$).

FDR corrected multiple comparisons indicated that both mechanical groups showed *immediate and sustained gains*. In the Active group both its post-test ($OR = 1.66$, $p < .001$) and delayed post-test ($OR = 1.70$, $p < .001$) performances were linked to higher odds of solution accuracy than its pre-test performance. No statistically significant difference was found between this group's post-test and delayed post-test performances ($OR = 0.97$, $p = .836$). Similarly, in the Passive Mechanical group both its post-test ($OR = 1.75$, $p < .001$) and delayed post-test ($OR = 1.76$, $p < .001$) performances were linked to higher odds of solution accuracy than its pre-test performance. No statistically significant difference was found between this group's post-test and delayed post-test performances ($OR = 1.01$, $p = .946$).

The Active Control 1 (cross-sectioning) group showed higher odds of solution accuracy in the post-test than in the pre-test ($OR = 2.48$, $p < .001$) and in the delayed post-test than in the pre-test ($OR = 1.36$, $p = .022$). However, unlike both mechanical groups, this group's post-test performance was linked to higher odds of solution accuracy than its delayed post-test performance ($OR = 1.82$, $p < .001$). This suggest that part of what was gained right after the intervention was lost a week after.

The Active Control 2 (reading) group did not show statistically significant differences between the post-test and the pre-test ($OR = 1.23$, $p = .199$), the delayed post-test and the pre-test ($OR = 0.92$, $p = .542$), and the post-test and the delayed post-test ($OR = 1.33$, $p = .104$).

An ANCOVA model that estimated changes in average performance in the Santa Barbara Solids test during the post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Condition ($F(3,172) = 3.73$, $p = .012$, $\eta^2 = 0.06$), no main effect of Pre-test ($F(1,172) = 3.13$, $p = .078$, $\eta^2 = 0.02$), and no interaction between Condition and Pre-test ($F(3,172) = 2.43$, $p = .067$, $\eta^2 = 0.04$). Subsequent FDR corrected pairwise comparisons showed that the Active Control 1 group showed a higher average performance than the Passive Mechanical group ($p = .011$). No further contrasts were statistically significant (all *ps* $> .078$).

Another ANCOVA model that estimated changes in average performance during the delayed post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Pre-test ($F(1,172) = 21.67$, $p < .001$, $\eta^2 = 0.11$), no main effect of Condition ($F(3,172) = 1.39$, $p = .247$, $\eta^2 = 0.03$), and no interaction between Condition and Pre-test ($F(3,172) = 0.18$, $p = .912$, $\eta^2 = 0.00$) (see Fig. 3C).

3.2.4. Visual search task

We evaluated whether solution accuracy in the Visual Search task changed as a function of Condition and Time. An Analysis of Deviance showed a main effect of Time ($\chi^2(2) = 7.21$, $p = .027$) and an interaction between Condition and Time ($\chi^2(6) = 21.32$, $p = .002$). No main effect of Condition was observed ($\chi^2(3) = 1.44$, $p = .697$).

FDR corrected multiple comparisons showed that for both mechanical groups, no differences were observed across all three time points. The Active Mechanical condition did not show statistically significant differences between the post-test and the pre-test ($OR = 0.83$, $p = .286$), the delayed post-test and the pre-test ($OR = 0.84$, $p = .780$), and the post-test and the delayed post-test ($OR = 0.80$, $p = .286$). Similarly, the Passive group did not show statistically significant differences between the post-test and the pre-test ($OR = 0.93$, $p = .579$), the delayed post-test and the pre-test ($OR = 1.03$, $p = .579$), and the post-test and the delayed post-test ($OR = 1.10$, $p = .579$).

The Active Control 1 (cross-sectioning) group showed lower odds of solution accuracy during the post-test than during the pre-test ($OR = 0.60$, $p < .001$) and the delayed post-test ($OR = 0.61$, $p < .001$). However, we did not find statistically significant differences between the delayed post-test and the pre-test ($OR = 0.98$, $p = .887$), suggesting that participants showed similar levels of performance at the beginning and at the end of the investigation. The Active Control 2 (reading) group showed higher odds of solution accuracy in the delayed post-test than in the pre-test ($OR = 1.40$, $p = .451$). No statistically significant differences were found between the pre-test and the post-test ($OR = 1.27$, $p = .126$) and between the post-test and the delayed post-test ($OR = 0.90$, $p = .478$).

An ANCOVA model that estimated changes in average performance in the Visual Search task during the post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Condition ($F(3,172) = 5.04$, $p = .002$, $\eta^2 = 0.08$), no main effect of Pre-test ($F(1,172) = 0.03$, $p = .842$, $\eta^2 = 0.00$), and no interaction between Condition and Pre-test ($F(3,172) = 0.95$, $p = .417$, $\eta^2 = 0.02$). Subsequent FDR corrected pairwise comparisons indicated that both the Passive Mechanical ($p = .020$) and the Active Control 2 ($p = .001$) groups showed higher average performance than the Active Control 1 group. No further contrasts were statistically significant (all *ps* $> .080$).

A further ANCOVA model that estimated changes in average performance during the delayed post-test, using Condition as the main predictor and Pre-test as a covariate, showed a main effect of Condition ($F(3,172) = 4.44$, $p = .005$, $\eta^2 = 0.07$), a main effect of Pre-test ($F(1,172) = 10.07$, $p = .002$, $\eta^2 = 0.05$), and an interaction between Condition and Pre-test ($F(3,172) = 4.49$, $p = .005$, $\eta^2 = 0.06$). Subsequent FDR corrected pairwise comparisons indicated that both the Active Control 1 ($p = .005$) and the Active Control 2 ($p < .001$) groups showed higher average performance than the Passive Mechanical group. No further contrasts were statistically significant (all *ps* $> .063$) (Fig. 3D).

4. Discussion

Nearly a century of research has revealed close connections between mechanical reasoning and spatial visualization skills (Smith, 1964). However, investigations into the nature of this relationship have been limited to self-report, as alluded to in the opening quote by Nikola Tesla, and correlational study designs. The purpose of this study was to fill this gap in the literature and experimentally probe the potentially causal relationship between mechanical reasoning and spatial visualization. Overall, our results indicated that both active and passive mechanical interventions were associated with improved spatial visualization performance. Analyses based on within-group change revealed that both groups demonstrated immediate and sustained gains in measures of mental paper folding, mental rotation, and mental cross-sectioning. Compared to the control groups, both mechanical groups demonstrated gains in mental paper folding immediately following training, as well as one week later. There was no evidence of training-induced gains in mental rotation immediately following training. However, at the delayed post-test, one-week later, the passive mechanical condition demonstrated significant gains in mental rotation compared to both control groups. The active and passive mechanical conditions demonstrated similar patterns of performance across time points in all four measures (i.e., *immediate and sustained gains* in measures of mental rotation, mental paper folding, and cross-sectioning, as well as no differences in performance in the visual search task). Taken together, these findings provide initial evidence and proof of concept that mechanical reasoning, a highly spatial but domain-relevant STEM skill, transfers to untrained measures of spatial visualization.

4.1. Active vs. passive mechanical training

Contrary to our predictions, there was no evidence that the active mechanical group was the most effective in improving spatial visualization performance. Although prior research has demonstrated the positive effects of hands-on spatial training (e.g., see Adams et al., 2014;

Wiedenbauer & Jansen-Osmann, 2008; Wiedenbauer et al., 2007), our study failed to show any advantage of solving the mechanical puzzles by hand versus passively viewing someone else solve the problems. Our results indicated that both mechanical groups demonstrated highly similar performance patterns in all four measures enlisted and across all three time points. Indeed, both mechanical groups showed *immediate and sustained gains* in measures of spatial visualization, mental cross-sectioning, and no major changes in the visual search control task. As both mechanical conditions were largely equivalent in terms of their overall effectiveness, we conclude that the spatial visualization phase, in which participants from both groups had to derive solutions to the mechanical puzzles, is the most probable reason for the observed effects (a hypothesis we revisit further below).

Despite highly similar performance patterns, a difference between mechanical conditions emerged in the mental rotation test. Our results indicated that the passive mechanical group, but not the active mechanical group, demonstrated greater mental rotation performance at the delayed post-test compared to the control groups. One explanation for this finding is that the passive group may have benefited from the more immediate feedback received during training. While participants in the passive group were exposed to immediate feedback following their spatial visualization phase, those in the active group received feedback through executing their own solutions. It is possible that the opportunity to physically manipulate the puzzles provided too much of a time lag/buffer between the visualization phase of the training and the solution phase. Said differently, the immediate feedback afforded to the passive group may have provided a more optimal condition to learn about the accuracy of one's visualized solutions to each problem. For participants in the active condition, the opportunity to physically act on their visualizations may have been especially harmful in the case of incorrect solutions. Moreover, whereas the passive group were informed of the correct solutions for each puzzle, the active group were only informed of the correct solution if they happened to solve the puzzle themselves, through active manipulation. Thus, these differences in both the immediacy of feedback, but also insights gleaned from seeing the correct solution, may have driven the differences in results between mechanical conditions. Follow-up studies are needed to test these possibilities.

4.2. Accounts of improved spatial thinking after mechanical training

There are several reasons why our mechanical reasoning training may have conferred benefits to participants' spatial performance. According to the shared-processing account, transfer of learning from one context (mechanical reasoning) to another (spatial reasoning) may have occurred due to recruitment of common mental processes. Indeed, as mentioned in the Introduction, there is evidence to suggest that mechanical and spatial reasoning are highly correlated with one another. There is a long history of factor analytic studies showing that various measures of mechanical reasoning and spatial abilities tend to load on the same factor; a finding consistent with the shared-processing account (Hamilton et al., 1995; Humphreys et al., 1993; Smith, 1964). One hypothesis, and the one that we aimed to test in the current study, states that mechanical and spatial reasoning are linked insofar as they both rely on spatial visualization. Thus, it is possible that one of the reasons why participants' spatial visualization skills were stronger following mechanical training was due to the shared need to engage in multi-step spatial visualization processes. According to this view, the mechanical reasoning condition may have provided participants with an extended opportunity to practice spatial visualization. In turn, this extra practice solving spatial visualization problems, albeit through solving mechanical reasoning problems, may have driven the post-tests improvements observed. That both mechanical reasoning conditions were approximately equal in their effectiveness is potentially due to the common need to engage in spatial visualization across both conditions. Our finding of gains in measures of spatial visualization, but not in our control measure of spatial processing devoid of spatial visualization (i.e., a visual search task), lends further support to this hypothesis. Moreover, the need to engage in multi-step spatial visualization processes in the mechanical conditions (e.g., mental rotation, translation, inferred spatial relations, etc.), may explain why the mechanical reasoning condition led to better outcomes than the mental cross-sectioning condition, which arguably involves a more singular spatial process. This explanation also offers insight into why the training effects may have been stronger for mental paper folding compared to mental rotation. That is, compared to mental rotation, mental paper folding places more demands on multi-step spatial visualization processes (e.g., see Harris et al., 2013). Taken together, the degree of transfer observed may depend on the extent to which the outcome measures required multi-step spatial visualization. Although more research is necessary to test this possibility, our results suggest that mechanical reasoning may be an effective means for training multi-step spatial visualization processes.

The mechanical reasoning conditions may have also facilitated spatial performance by providing participants with new insights and strategies for solving spatial visualization problems. It has been well-established that individuals differ in the strategies they use to solve spatial reasoning problems, including measures of mental rotation and mental paper folding (Bethell-Fox & Shepard, 1988; Harris et al., 2013; Schultz, 1991). For example, a distinction exists between individuals who generally employ a more analytic, verbally mediated strategy for solving spatial visualization problems compared to individuals who employ a more holistic, spatially mediated strategy (Bethell-Fox & Shepard, 1988; Schultz, 1991). Critically, higher spatial performance is associated with the latter ; (Glück, Machat, Jirasko, & Rollett, 2002; Janssen & Geiser, 2010). Thus, it is possible that participation in the mechanical training conditions may have prompted individuals to engage in more holistic, spatially mediated approaches to spatial problem solving. The adoption of such spatial strategies may have contributed to increased spatial visualization performance after training. To test this possibility in follow-up studies, it will be important to collect information about the strategy participants use prior to, during, and following training.

To conclude, the present findings add to a large body of research that suggest that spatial thinking is a highly malleable construct (Uttal et al., 2013). Our findings are unique, however, in that they provide evidence that mechanical problem solving, a highly spatial but domain-relevant skill, transfers to untrained measures of spatial visualization. We interpret this finding as evidence for the shared-processing account. In addition, we posit that the mechanical reasoning conditions may have also yielded spatial gains by means of encouraging participants to adopt more effective strategies for solving spatial visualization problems. Future research should aim to further probe these, and other plausible accounts, in order to better understand the underlying mechanisms that support rapid, generalizable, and durable improvements in spatial thinking.

4.3. Implications for spatial training and STEM education

Our results provide preliminary support that directly engaging with STEM content may be a viable and effective means for training spatial skills. This approach contrasts with the predominant approach to spatial training, which involves repeated exposure to abstract and decontextualized spatial stimuli (e.g., 3D cube figures). In addition, a directionality of effects is typically assumed, in which spatial reasoning is seen as a foundational cognitive ability on which other skills, namely STEM-related learning, are based (e.g., see Hawes et al., 2015). However, as demonstrated in the present paper, it is also possible that STEM-related activities help spatial thinking. This finding raises some important implications for both theories of learning and the planning/design of classroom-based interventions and instruction. For example, in terms of theories of learning, our study challenges assumptions implicit in hierarchical models of learning. That is, the belief that higher-order learning depends on lower-level and more basic cognitive capacities and concepts. Although there are clear cases in which this model of learning is most accurate (e.g., one must learn basic arithmetic before advanced algebra), there is also evidence that engaging with 'higher-level' tasks might also work towards strengthening 'lower-level' capacities (e.g., see Lyons, Bugden, Zheng, De Jesus, & Ansari, 2018). The present findings are but one counter example of hierarchical learning. However, because we did not test the effect of spatial training on mechanical reasoning, it was not possible to directly compare the strength of directional and reciprocal effects. The shared-processing account makes some important predictions in this regard (see Lourenco, Cheung, & Aulet, 2018). First, the degree of transfer will depend on the degree of recruitment and reliance on shared underlying mechanisms (e.g., the need to engage in spatial visualization). Second, transfer effects are expected to be bidirectional and relatively symmetrical in their effects on one another. Based on the present findings, as well as prior research demonstrating strong relations between spatial visualization and mechanical reasoning, we hypothesize that training in one domain will generalize to the other. Moving forward, it will be important to test this prediction, as the findings will further elucidate the mechanisms of transfer and the nature of the relationship between spatial and mechanical reasoning.

A better understanding of the direction of transfer is important when considering classroom-based approaches to spatial training and learning. There are risks associated with isolated approaches to spatial training (i.e., training spatial skills in isolation from their disciplinary use), including risks of the training not working and, in turn, the allocation of time and resources that could have been spent engaging in spatially demanding STEM content. As argued by Ramey et al. (2020) even if isolated spatial training is effective at improving spatial skills, "*something is being lost by denying students the opportunity to reason with the actual tools with which they might be expected to reason spatially in professional practice*" (p. 469). Thus, although tightly controlled training approaches have been instrumental in establishing that spatial thinking is malleable (Uttal et al., 2013), it is unclear whether this same approach should be endorsed as a means of improving STEM performance. The findings from our study, albeit still very much lab-based, suggest the potential benefits of having individuals interact with 3D mechanical problems. A promising avenue for follow-up research involves examining the extent to which the benefits observed in the present study translate to more authentic mechanical reasoning educational tasks. For example, the rapid rise of 'makerspaces' across K-12 education represent one such opportunity (Bevan, 2017; Giannakos, Divitini, & Iversen, 2017). It has recently been shown that 'making activities' afford a rich context in which to engage in and further develop a variety of spatial reasoning capacities (Ramey et al., 2020). Whether or not makerspace activities that involve mechanical problem solving transfer to spatial reasoning skills will be an important question to address in future research.

4.4. Limitations

The current study had several limitations. First, the delayed post-test was completed in the participants' homes. Thus, it is impossible to rule out the possibility that the delayed post-tests were completed without external assistance. However, given random assignment to condition, it is likely that the occurrence of such instances would also be randomly distributed across groups. Nonetheless, future research should aim to create equal testing conditions across all time points of data collection. Second, time spent on the computer was not equivalent across groups; the two mechanical conditions provided a break from the computer while the two active control conditions received no computer break. Thus, it is possible that the computer break may have influenced post-test performance. However, given that the cross-sectioning group also demonstrated improvements in their spatial performance, this possibility seems unlikely. To eliminate such a confound it will be important in future research to include an active control group that also receives a break from the computer during the training phase of the study. Third, we did not measure participant motivation or expectancy effects. It remains plausible that the two mechanical reasoning conditions provided a more motivating training condition, potentially increasing one's interest and investment in achieving success at post-test. Fourth, because the Santa Barbara Solids test showed poor reliability, and both the Paper Folding test and Visual Search task showed fair reliability, the current results must be interpreted with caution. Future research will be needed to further explore the causal connection between mechanical training and spatial visualization performance. Fifth, the power analysis that was used to estimate the sample size did not account for the interactions included in the ANCOVA models.[1] Moving forward, an even larger sample size is recommended to implement the study design employed here. For example, a larger sample size would better ensure more equivalent baseline performance across groups. In the present study, despite randomization, participants assigned to the mechanical conditions obtained slightly lower scores than the control conditions in several measures. However, it is also important to note that this minor difference in baseline performance was adequately accounted for in our statistical models. Moreover, our findings of sustained gains by the mechanical groups suggest that any initial differences in performance were overcome by the training. Sixth, our results suggest a potential benefit of receiving passive feedback compared to active feedback. However, it is clear that more research is needed to further substantiate these results. For example, it is unclear why slightly more favorable outcomes were observed in the paper folding measure compared to mental rotation. It is also unclear why the benefit afforded to the passive condition would be present at the delayed post-test but not the immediate post-test. Finally, the small number of training trials prevented the opportunity to examine individual differences in training-related improvements and its association with gains in spatial thinking. Moving forward, the inclusion of a more extensive training intervention will provide an opportunity to address important follow-up questions to the current study, including an examination of individual differences at baseline, rates of learning, and the impact these variables have on one's spatial performance.

5. Conclusion

Our findings indicate that mechanical reasoning may facilitate spatial visualization performance. This study adds to a large body of research that suggest spatial thinking is a highly malleable construct. Our findings are unique, however, in that they suggest that directly engaging with STEM content may be a viable and effective means for training spatial skills. Moving forward, more research is needed to further substantiate the current findings, as the implications of such work are of critical importance in the planning and design of classroom-based spatial learning.

CRediT authorship contribution statement

Felipe Munoz-Rubke: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Russell Will:** Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Zachary Hawes:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Karin H. James:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

[1] Note that our initial plan of analysis changed as a result of the review process. The recommended analytical approach, reported here, requires a larger sample size than we had originally calculated.

Acknowledgements and Funding

Felipe Munoz-Rubke would like to thank Fulbright and CONICYT for financial support provided through an academic scholarship. Portions of this research was supported by the Indiana University Emerging Areas of Research grant: Brains, Babies, and Machines.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2021.101496.

References

Adams, D. M., Stull, A. T., & Hegarty, M. (2014). Effects of mental and manual rotation training on mental and manual rotation performance. *Spatial Cognition and Computation, 14*(3), 169–198.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 12–23.
Bevan, B. (2017). The promise and the promises of Making in science education. *Studies in Science Education, 53*(1), 75–103.
Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society, 15*(1), 2–11.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Routledge.
Cohen, C. A., & Hegarty, M. (2012). Inferring cross sections of 3D objects: A new spatial thinking test. *Learning and Individual Differences, 22*(6), 868–874.
Cohen, C. A., & Hegarty, M. (2014). Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences, 33*, 63–71.
Cornu, V., Schiltz, C., Pazouki, T., & Martin, R. (2019). Training early visuo-spatial abilities: A controlled classroom-based intervention study. *Applied Developmental Science, 23*(1), 1–21.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.
Fayer, S., Lacey, A., & Watson, A. (2017). *Science, technology, engineering, and mathematics (STEM) occupations: Past, present, and future: Spotlight on statistics*. Washington, DC: U.S. Bureau of Labor Statistics.
French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.
Giannakos, M. N., Divitini, M., & Iversen, O. S. (2017). Entertainment, engagement, and education: Foundations and developments in digital and physical spaces to support learning through making. *Entertainment Computing, 21*, 77–81.
Gilligan, K. A., Flouri, E., & Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology, 163*, 107–125.
Glück, J., Machat, R., Jirasko, M., & Rollett, B. (2002). Training-related changes in solution strategy in a spatial test. *Learning and Individual Differences, 13*(1), 1–22.
Hamilton, L. S., Nussbaum, E. M., Kupermintz, H., Kerkhoven, J. I. M., & Snow, R. E. (1995). Enhancing the validity and usefulness of large-scale educational assessments: II. NELS:88 science achievement. *American Educational Research Journal, 32*(3), 555–581.
Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing, 14*(2), 105–115.
Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction, 35*(3), 236–264. https://doi.org/10.1080/07370008.2017.1323902
Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001
Hegarty, M. (1991). Knowledge and processes in mechanical problem solving. In R. Sternberg, & P. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 253–285).
Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences, 8*(6), 280–285.
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., & Waller, D. A. (2005). Individual differences in spatial abilities. In *The cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge University Press.
Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78*(2), 250–261.
Janssen, A. B., & Geiser, C. (2010). On the relationship between solution strategies in two mental rotation tasks. *Learning and Individual Differences, 20*(5), 473–478.
von Károlyi, C. (2013). From Tesla to Tetris: Mental rotation, vocation, and gifted education. *Roeper Review, 35*(4), 231–240.
Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*(9), 1831–1836.
Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*(4), 549–579.
Lenth, R. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, Articles, 69*(1), 1–33.
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. (1988). Spatial abilities as traits, processes, and knowledge. In *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Psychology Press.
Lohman, D. F. (1996). Spatial ability and G. In I. Dennis, & P. Tapsfield (Eds.), *Human abilities: Their nature and assessment* (pp. 97–116). Hilldale, NJ: Lawrence Erlbaum Associates.
Lourenco, S. F., Cheung, C.-N., & Aulet, L. S. (2018). Is visuospatial reasoning related to early mathematical development? A critical review. In *Heterogeneity of function in numerical cognition* (pp. 177–210). Elsevier.
Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49*(4), 344–351.
Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., & Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. *Developmental Psychology, 54*(3), 440–457.
McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). CRC Press.
McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*(5), 889–918.
Michael, W. B., Guilford, J. P., Fruchter, B., & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement, 17*, 185–199.
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. https://doi.org/10.1016/j.actpsy.2018.06.006
Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator, 34*(2), 29.
Newcombe, N. S. (2016). Thinking spatially in the science classroom. *Current Opinion in Behavioral Sciences, 10*, 1–6.
Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58.
Ramey, K. E., Stevens, R., & Uttal, D. H. (2020). In-FUSE-ing STEAM learning with spatial reasoning: Distributed spatial sensemaking in school-based making activities. *Journal of Educational Psychology, 112*(3), 466–493.
Saarinen, J. (1994). Visual search for global and local stimulus features. *Perception, 23*(2), 237–243.
Schultz, K. (1991). The contribution of solution strategy to spatial performance. *Canadian Journal of Psychology, 45*(4), 474–491.
Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604–614.
Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*(1), 97–115.
Smith, I. M. (1964). *Spatial ability: Its educational and social significance*. University of London Press.
Sorby, S. A. (1999). Developing 3-D spatial visualization skills. *Engineering Design Graphics Journal*. Retrieved from http://www.edgj.org/index.php/EDGJ/article/view/126/122.
Tesla, N. (1995). My inventions: The autobiography of Nikola Tesla. In B. Johnston (Ed.), *Barnes & noble (original work published 1919)*.
Thurstone, L. L. (1950). Some primary abilities in visual thinking. *Proceedings of the American Philosophical Society, 94*(6), 517–521.
Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual & Motor Skills, 47*(2), 599–604.
Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817–835.
Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*(1), 77–94.
Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction, 18*(1), 30–41.
Wiedenbauer, G., Schmid, J., & Jansen-Osmann, P. (2007). Manual training of mental rotation. *European Journal of Cognitive Psychology, 19*(1), 17–36.
Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 397–412.
Wraga, M., Thompson, W. L., Alpert, N. M., & Kosslyn, S. M. (2003). Implicit transfer of motor strategies in mental rotation. *Brain and Cognition, 52*(2), 135–143.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15*(4), 763–771.

Xu, C., & LeFevre, J.-A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology, 52*(6), 854–866.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19.