Cognitive Development 45 (2018) 113–124

Contents lists available at ScienceDirect

Cognitive Development

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

Where will it go? How children and adults reason about force and motion

Justin Harris[a,*], Nathan R. George[b], Kathy Hirsh-Pasek[c], Nora S. Newcombe[c]

[a] Museum of Science, Boston, United States
[b] Derner School of Psychology, Adelphi University, United States
[c] Department of Psychology, Temple University, United States

ARTICLE INFO

Keywords:
Developmental comparison
Force and motion
Sex differences

ABSTRACT

Even infants can recognize physically impossible patterns of motion, seem to expect correct trajectories, and as they develop motor skills, move as necessary to achieve a goal. Yet in adulthood, the majority of people perform poorly when asked to make explicit predictions about motion in the same problems, and are influenced by irrelevant surface features. To characterize the changes that occur during development and the nature of individual differences, we developed a new assessment of force and motion that is age appropriate for both 6-year-old children and adults. Participants at both ages were generally able to reason at above-chance levels about motion in one dimension, although adults showed superior performance. On problems involving motion in two dimensions, adult males did better than boys, but adult females were equivalent to girls. These data provide the basis for a reinvigorated investigation of the factors supporting the development of the ability to think about force and motion.

1. Introduction

A stick strikes a hockey puck sliding across the ice. Where will it go? From a physics professor's standpoint, this problem is a simple one, similar to problems in most college-level, introductory physics courses. However, students often struggle to achieve the coherent conceptual understanding of force and motion that these problems are meant to measure (diSessa, 2013; Vosniadou & Skopeliti, 2014). Conceptual confusion is evident even in students who appear to be performing well in physics courses (Hestenes, 1985a, 1985b; ; Viennot, 1979). Many high-performing students generate correct *scalar* mathematical solutions by using a strategy of matching variables in functions, but do not show conceptual mastery when probed more deeply (Halloun & Hestenes, 1985b; Hammer, 1989; Larkin, McDermott, Simon, & Simon, 1980; Van Heuvelen, 1991). Mastery of the underlying physics is inhibited by the vector nature of core concepts such as velocity, acceleration, and force. Accurately solving problems regarding motion in two dimensions requires successful vector addition and students generally lack a precise understanding of vector quantities and vector mathematics (Flores, Kanim, & Kautz, 2004; Nguyen & Meltzer, 2003).

Adults' difficulties with basic concepts of force and motion may appear puzzling, however, when considered in the light of developmental research with very young children. Infants often show surprisingly good performance in recognizing causal interactions and in attending to specific details of the forces involved in motion events, such as expecting larger forces to result in more motion (Cohen & Oakes, 1993; Göksun, George, Hirsh-Pasek, & Golinkoff, 2013; Hood, 1998; Kotovsky & Baillargeon, 2000; Leslie & Keeble, 1987; Oakes & Cohen, 1990). In addition, by the time they have left toddlerhood, children are able to walk, run, push, pull,

* Corresponding author at: Museum of Science, Boston, 1 Science Park, Boston, MA, 02114, United States.
E-mail address: jharris@mos.org (J. Harris).

https://doi.org/10.1016/j.cogdev.2018.01.002
Received 28 November 2016; Received in revised form 14 January 2018; Accepted 16 January 2018
Available online 06 February 2018

throw, and catch to achieve a variety of desired outcomes in a way that involves taking account of complex concepts such as friction.

Ideally, apparent adult difficulties could be reconciled with the seemingly advanced abilities of children by directly comparing their performance on force and motion problems. However, there is wide variety in the methodologies used with different age groups. Studies of early childhood often use looking-time and other recognition paradigms, while studies with adults typically use explicit prediction paradigms. When adults do not have to make explicit predictions, they do much better. For example, Kaiser, Proffitt, Whelan, & Hecht (1992) presented adults with a simple pendulum problem – if the string on the pendulum is cut while the bob is at its apex how should it fall? When participants viewed five videos to *recognize* which was correct, they did well. When shown a static image of five potential paths and asked to *predict* which was correct, the same participants did poorly, often selecting a physically-impossible drawing (i.e., one depicting motion that would not occur under the specified conditions). Thus, adults typically succeed in the types of recognition tasks used with children. Similarly, both children and adults show success in action but errors when asked to reflect. For example, Krist, Fieberg, and Wilkening (1993) asked adults and children to identify how fast a ball needed to be launched off a platform to hit a target, as the height of the platform and distance of the target changed. When child and adult participants were asked to *predict* the required speed, performance was poor, at both ages. However, when the same participants were instead asked to *act*, to throw the ball with sufficient speed, performance was better.

We can also compare children and adults using experimental techniques more comparable to those used with adults, involving explicit prediction. In this case, children struggle with reasoning correctly, just as adults do (Hood, 1995; Kaiser, McCloskey, & Proffitt, 1986; Kaiser, Proffitt, & McCloskey, 1985; Kim & Spelke, 1999). For example, both children and adults make incorrect predictions about scenarios involving multiple components of motion (i.e., velocity or acceleration in different dimensions), often partially or completely ignoring one of the components of motion (diSessa, 1982; Göksun et al., 2013; Halloun & Hestenes, 1985a). Thus, much like adults succeeded when tested in a manner comparable to methods relied upon with children, children struggle when tasks require prediction. It is worth noting that dimensionality affects the difficulty of prediction problems, but not action or perception problems. This difference can be attributed to the fact that, for a single dimension the quantities can be reduced to scalars with different signs indicating the direction, but vector addition is required when the motion spans more than one dimension. Overall, the perception-action system appears to be much "smarter" than the cognitive system, for reasons that have yet to be well specified.

One conclusion from this body of research might be that there are no age-related differences in understanding force and motion. But this conclusion would be premature. The variation in paradigms used with children and adults has obscured the goal of determining whether children and adults differ at all in their understanding. Describing any natural age-related change and elucidating the factors associated with conceptual advance might allow the design of better educational support to address the areas where understanding lags.

In exploring age-related change, we also need to address individual differences. Although performance is poor overall on prediction problems, some people gain conceptual mastery of force and motion. While it has been argued that physics may just be too hard for most people, in a categorically different way than other difficult topics (Sobel, 2009), many researchers in the physics community reject this idea (Lasry, Finkelstein, & Mazur, 2009). They suggest that specific aspects of individual cognitive profiles interact with the specific experiences that people encounter in informal and formal settings to determine their progress on particular physics problems. One example might be that spatial skills are involved because force, acceleration, and velocity are vector quantities. Consider again a stick striking a hockey puck sliding across the ice. The first step in solving this problem is noticing how the direction of the force of the strike compares to the direction of the hockey puck's motion. If they are in the same direction (i.e., zero degrees difference) the magnitude of the puck's velocity will increase, but its direction will not change. If they are in the opposite direction (i.e., 180 ° difference), the magnitude of the puck's velocity will decrease and might even become negative (i.e., its direction would be reversed). Between these two simple extremes, the puck's acceleration may result in a change of both the magnitude and direction of the puck's velocity. The underlying features of these problems suggest spatial thinking might be a potential explanatory factor for individual differences in whether physically-possible predictions eventually develop. Furthermore, variations in attention to the spatial aspects of these problems has been suggested as a key difference between unsuccessful novices and experts (Larkin, 1981, 1982, 1983; Van Heuvelen, 1991) and as an important predictor of initial success across all STEM disciplines (Uttal & Cohen, 2012). Spatial thinking has also been shown to explain individual differences on force and motion problems specifically (Kozhevnikov, Motes, & Hegarty, 2007).

Appropriate examination of the development of physics ideas, of individual differences, and of the factors related to differential development of physically-possible ideas requires a conceptual test that provides quantitative results. This test must be presented in a format that is age-appropriate for a wide age-range, but does not rely on either recognition or action, given that the perception-action system appears to undergo little developmental change. Here we present a novel digital test of understanding of force and motion (called the Hedgehog Game) that meets these criteria, allowing the investigation of the development of naïve ideas of force and motion from the age at which children first enter school and begin formally instruction through adulthood. Participants' performance is compared across two age groups to evaluate change over time. We also test whether measures of spatial thinking might explain individual differences that emerge in this task.

2. Method

2.1. Participants

Fifty-eight adults between the ages of 18–65 (M = 23.21 years, 28 male) from a major northeastern U.S. city participated in this study. Participants were recruited either 1) through undergraduate courses and participated for course credit or 2) when accompanying a child to participate in an unrelated study. Six adults were excluded for having received college physics instruction within the previous year or having taken more than one college physics course at any time. The final sample consisted of 52 physics-naïve adults (M = 22.75 years, 27 male).

Sixty-three typically-developing children between 66 and 78 months old (5.5–6.5 years; M = 70.63 months, 33 male) were recruited from the suburbs of a major northeastern U.S. city. They were predominately Caucasian and from upper-middle class families. The data from three children were discarded because they did not complete the study. The final sample consisted of 60 children (M = 70.63 months, 31 male).

2.2. Materials

2.2.1. The Hedgehog Game

The assessment of naïve ideas of force and motion represented forces using cartoon hedgehogs, referred to as the Hedgehog Game. The structure of the game was inspired by similar work by Göksun et al. (2013) in which children and adults were presented with a live-action task requiring them to predict the end location of a small ball after it was acted upon by one or two forces (a ramp and hairdryer) in various configurations. Here, we present a digital, multiple-choice version of this task, with increased variation in factors such as the relative size and timing of the forces involved. Ideas about force and motion were measured by eliciting predictions about motion from forces and inferences about the forces involved in a motion event, two formats we believed would both tap into the cognitive system only (i.e., no motion was depicted for test items; inferences were made based on a final potion with no path shown from the start position). Given large variation in results associated with different test formats in previous research, introducing multiple measurements assures a more stringent test of the game's internal validity. Finding similar results across these formats would suggest that additional concerns about the influence of task demands would not be warranted.

2.2.1.1. Introduction. In the game, cartoon hedgehogs jointly blew a ball around a gridded board. The forces represented by the hedgehogs could vary in direction, magnitude, and spatial relation (to both each other and the ball). Participants were first introduced to the concept that the direction of the force (up, down, left, right, and diagonally) would influence the direction of the motion. Then they were shown that the magnitude would influence motion. Magnitude was represented by both size of the hedgehog and color (red hedgehogs were small and exerted less force, brown hedgehogs were large and exerted more force). The distances were shown, but not specified verbally. The significance of spatial relation to the ball was demonstrated by showing that the distance traveled by the ball was inversely related to the hedgehog's distance from the ball (i.e., a hedgehog standing far away from the ball could only act on it weakly). Participants were told that the hedgehogs cannot move or turn around, to avoid any misinterpretation of non-demonstrated events. Practice items were administered to ensure the participants understood each of these aspects (i.e., direction, hedgehog size, and distance from ball) and concurrently to explain the two trial types (i.e. prediction and inference) that would be encountered.

The test began with single-force questions (Fig. 1). Single-force trials were followed by one-dimension trials in which two forces act in the same direction and then one-dimension trials where the forces act in opposition (i.e., the forces act at 180° to each other; Fig. 2). The remaining problems introduced a second dimension and orthogonal forces (i.e. the forces act at 90° to each other; Fig. 3). The problem set addressed a range of contexts to measure ideas about how multiple components of motion combine (see Table 1 for ideas tested at each level). Differences in difficulty were created by varying the context (i.e., orientation, size, and timing) of the forces.

All answers were given in a multiple-choice format. Every two-force level included a context for interpreting the motion – the square in which the ball would land if only one hedgehog blew the ball was highlighted. Each trial began with an explanation of the context for interpreting answers ("This square shows where this hedgehog would blow the ball if he was all alone."). For prediction trials, distracter items were designed to be meaningful, albeit incorrect, combinations of the two forces. The combinations reflected correct responses of other orientations (i.e., 0°, 180°, or 90°) and a common naïve idea– dominance (Halloun & Hestenes, 1985a) – that when two forces combine one will overwhelm the other to completely determine the motion. For inference trials, the distracter

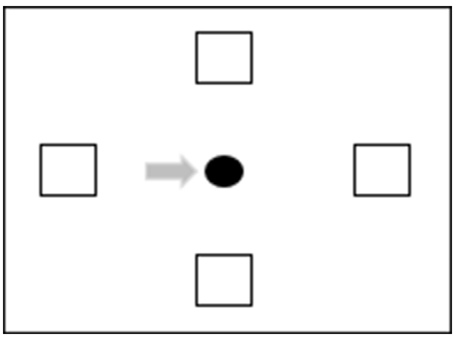

Fig. 1. One-force array (prediction).

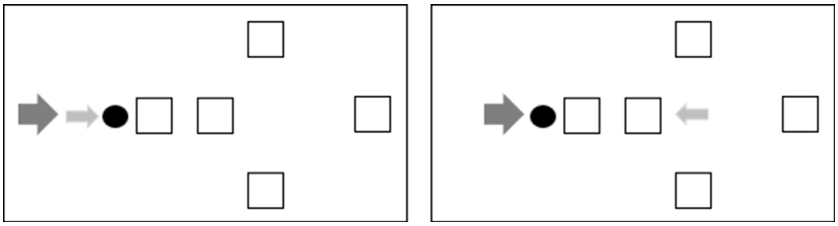

Fig. 2. One-dimension two-force array (prediction). Forces can act in the same direction (left) or directly opposing direction (right).

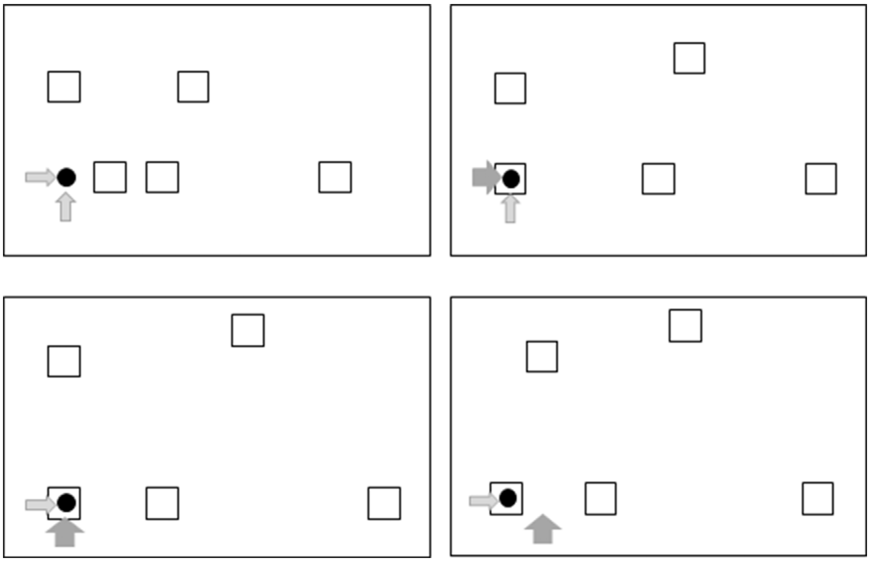

Fig. 3. Two-dimension two-force array (prediction). Forces can be equal (upper left) or, when separate, the big force (upper right) or the little force (both lower) may be the "context" force. Forces may act simultaneously (both upper and lower left) or sequentially (lower right).

Table 1
Dimensionality and concepts of games levels.

Forces	Dimensionality	Concept
1 force	1 dimension	Motion occurs in the direction of the force
2 forces, same direction	1 dimension	Vector superposition (addition)
2 forces, opposite direction	1 dimension	Vector superposition (subtraction)
2 forces, orthogonal, simultaneous	2 dimensions	Vector superposition (addition – diagonal resultant). The resultant is now in a different direction (diagonal), requiring more careful attention to be paid to the directional component of the vector.
2 forces, orthogonal, sequential	2 dimensions	Vector superposition (addition – diagonal resultant) and inertia. The addition of inertia makes this a physically different problem, though the vector addition is the same. This problem requires some physics knowledge, though this is knowledge that could be gained through observation.
Same force		A basic case, should not bias people to focus on either force.
Big force first		The errors reflect the idea that people may fail to integrate the forces and instead focus on one alone. Test-takers may apply the dominance principle to the first force because it is larger or the second force because it is more recent.
Small force first		With the big force second, both strength and recency should lead people to apply to the dominance principle to the second force, if it is applied at all.

items were also built around the dominance principle – that one force must be directly aimed at the destination. Other distracters had a reversed orientation from the correct and dominance-consistent answers. The design of the distracter items allows for incorrect responses to be evaluated as being meaningful, by exploring whether errors were randomly distributed amongst the distracters or if they consistently followed the dominance principle. Dominance as a physically-impossible idea has been demonstrated when the forces are unequal in size (the biggest force wins) and when they act sequentially (the most recently applied force wins), so the test includes trials featuring both factors separately and in conjunction. Simultaneous trials always preceded sequential trials.

2.2.1.2. Prediction trials. The problem sets are first presented in terms of prediction; given the forces, the test-takers determined where the ball will stop up (Figs. 1–3). This response format is common on tests. This highlighted square that showed where one of

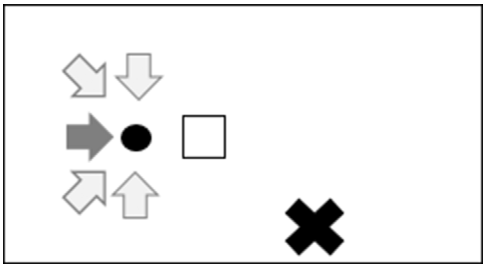

Fig. 4. Two-dimension two-force inference array (same force, simultaneous).

the forces would individually blow the ball was a potential answer choice for the prediction trials.

2.2.1.3. Inference trials. Inference, was included as an alternate, cognitive measurement of force and motion ideas (Fig. 4). For inference trials, the end point for the ball is specified and test-takers are required to determine where a force should be added to make the ball reach the specified target. Answer choices were represented as four faded hedgehogs that indicated the size, color, and direction of the hedgehog that could stand at that choice. For inference trials, participants were only asked to place a single force; one force was already in place for two-force trials.

2.2.1.4. Design. Prediction trials always preceded inference trials. Pilot data suggested that this presentation order helps participants understand the convention that two hedgehogs would both blow on a ball, even if one stands between the other and the ball (Fig. 2).

Participants were randomly assigned to one of two sets of problems at each of the nine levels. The sets differed on direction (horizontal or vertical for the "context" hedgehog) to ensure this factor would not impact performance. Each set contained three problems.

The test was untimed and each participant could move forward at his or her own pace. For each question, participants were required to confirm their answer by selecting the "Next" button and were unable to change their answer after progressing to the next problem.

2.2.1.5. Physics note. The explanations and the graphics in the game are designed to convey the idea that the hedgehogs exert an instantaneous force on the ball as soon as the ball passes in front of them, with an accompanying instantaneous acceleration. However, the program actually models the motion, assigning a specific initial velocity to the ball. The velocity is decremented as the ball moves to simulate the negative acceleration due to friction. The program does not model forces directly, but does simulate movement similarly to that which people encounter in their everyday life by modeling the resulting acceleration and associated velocity. The one slightly unrealistic element is that hedgehogs exert a force on the ball even if there is another hedgehog standing between the hedgehog and the ball. This feature was required so that two-force, same direction trials could be presented clearly.

The correct solution to the two force problems cannot be found by performing vector addition on the *distance traveled* when each force acts in isolation. Rather, vector addition should be performed on the instantaneously imparted *velocities*, because the total distance traveled by a ball before coming to a stop is a function of its initial velocity. A ball with a higher initial velocity will roll for a longer amount of time before its velocity is reduced to zero, resulting in an increased distance traveled. When an object is acted on simultaneously by two forces oriented at 90° (or less) to each other the initial velocity will be higher than if the same object was acted upon by either of the two forces individually (e.g., $1 + 1 = \sqrt{2}$ in the perpendicular case, $\sqrt{2} > 1$). This leads to the somewhat counterintuitive truth that the ball will go further in both dimensions than performing vector addition on the distance traveled would indicate. For example, if the forces were arranged so a brown hedgehog blew the ball left (5 squares) and then red hedgehog was placed at the end of its trajectory so as to blow the ball only once it had come to a complete stop, then the ball would have rolled 5 squares to the left and 3 squares up. However, because the ball will be moving under two perpendicular components of motion at once it moves slightly further. In the Hedgehog Game, the ball will not only move further, but it will end up in a different square (i.e., 6 squares to the left and 4 squares up). Answer choices that correspond to this quantitative error were not included.

2.2.2. Spatial battery

A spatial battery was administered to both age groups, focusing on two spatial abilities: mental folding and mental rotation. Mental folding has previously found to be related to the types of problems used in the Hedgehog Game (Kozhevnikov et al., 2007). However, mental folding typically fails to show sex differences (Linn & Petersen, 1995; Voyer, Voyer, & Bryden, 1995), while sex differences are often observed in this type of problem (Coletta, Phillips, & Steinert, 2011; Docktor & Heller, 2008; Kost, Pollock, & Finklestein, 2009; Lorenzo, Crouch, & Mazur, 2006). The addition of a spatial measure that does show sex difference would be required to fully investigate this relation. Mental rotation was selected as a similar, but distinct, measure of spatial thinking to mental folding, which is well-known for its sex differences (Harris, Hirsh-Pasek, & Newcombe, 2013).

2.2.2.1. Mental folding. Adult mental folding was measured using the Educational Testing Service's Paper Folding Test (PFT; Ekstrom, French, Harman, & Dermen, 1976). Images depicted a piece of paper being folded multiple times and then a hole being punched through it. Adults were required to select the image that shows how the paper would look when it was unfolded, from amongst five answer choices. It has two pages, with 10 items each. Participants were given 3 min for each page. The manual reports reliability between 0.75 and 0.84, across three samples (Ekstrom et al., 1976).

The University of Chicago Children's Mental Folding Test (CMFT) was used to measure mental folding in children. This measure is valid for 5- through 9-year-olds (Butts, Foley, Schiffman, & Levine, 2017), consisting of two practice items and 11 test items. The examiner folded a piece of a paper and then punched a hole in. The child selected the answer from amongst the choices that depicts how the paper will look when it is unfolded. The test was untimed.

2.2.2.2. Mental rotation. Adult mental rotation was measured using Vandenberg and Kuse's (1978) Mental Rotation Task-A (MRT-A). A target 3D object, made of multiple small blocks, was presented. Participants identified which two of four potential images showed the same object at a different orientation. There were two pages, with 12 items each. Each participant was given 3 min per page. This test has a (KR-20) reliability of 0.88.

Mental rotation in children was measured through the animal portion of Neuburger, Jansen, Heil, and Quaiser-Pohl's (2011) mental rotation task (MRT). This task consisted of two completed sample items demonstrating the task, two sample items the child must complete correctly, and 16 test items. This task was the same as the adult version, but with 2D images of animals instead of blocks. The instructions were provided to us in German and translated by a research assistant on the project. This test was untimed.

2.2.3. Other cognitive domains

2.2.3.1. Expressive vocabulary. Vocabulary was measured as a control for general intelligence (*g*), using the WJ-III Achievement Picture Vocabulary (PV) subtest (Woodcock, McGrew, & Mather, 2001). This test was used for both groups. This measure was untimed and free-response. Participants were shown drawings of objects and asked to identify them. A minimum of 12 items and a maximum of 44 items can be administered. Children begin at an earlier test item than adults, but follow the same basal and ceiling rules. This test has a median reliability of 0.81, across the age of intended use (24 months–90 years and older).

2.2.3.2. Number sense. Two measures of number sense were administered, because early mathematics has often been found to be related to spatial skills (Mix et al., 2016; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017). There is also a minor quantitative component to the Hedgehog Game (i.e. the large force blow the ball 5 squares while the small force blows the ball 3 squares), which, while not stated explicitly, may benefit children with higher number sense. This skill was therefore also assessed in children, but not adults. First, linearity of number sense was measured through a number line task (NL). Children were asked to place a hatch mark to represent where the numbers 3, 25, 86, 6, 2, and 67 (in that order) fell on a set of 0–100 number lines (Siegler & Opfer, 2003). This task was untimed. Scores for this test were percentage of variance (R^2) in hatch mark placement explained by a linear number representation (as in Gunderson, Ramirez, Beilock, & Levine, 2012). Second, children's ability to estimate was measured through a symbolic approximation (SA) task adapted from Gunderson et al. (2012). Two quantities were compared to a third, all shown symbolically as numerals. This version only included approximate addition. Two practice items and 10 test items were presented. Children were shown that one character received two quantities of cookies, which were then hidden. They were then shown a second character receiving a different of amount of cookies and asked to determine who has more. This task was untimed.

2.3. Procedure

After obtaining consent, participants were brought to a quiet room for testing. Testing always began with mental folding. For adults this was followed by mental rotation. For children the order of the mental rotation, number line, and symbolic approximation tasks were counterbalanced. For both groups, the vocabulary measure followed these tests and testing concluded with the Hedgehog Game. The Hedgehog Game was delivered on a touch screen tablet. The tablet was always placed horizontally on a table, to avoid any implication that gravity was involved.

The entire testing procedure lasted between 40–80 min for children and 40–60 min for adults. Children were allowed to take breaks between tests at their request. Breaks were encouraged between tests if a child appeared to be losing focus. However, breaks were rarely requested and lasted no longer than 5 min.

3. Results

Much of the data was non-normal, with some high-performing outliers but also the floor effects that would be expected with challenging problems such as the two-dimension trials of the Hedgehog Game. In order to address this issue, analyses which rely on the normality of the distribution were conducted using SPSS Version 25's bootstrap module. The advantage of a bootstrapping approach is that it requires no assumptions about the shape of the distribution. These assumptions are avoided by using a Monte Carlo approach to estimate the population parameters directly from the sample. To do so, we generate a large number of bootstrap samples, or samples that are drawn from the original data set randomly, with replacement, and of the same size as the initial sample. The statistic of interest is calculated for each bootstrap sample and parameter estimates are generated based on the samples (see Efron & Tibshirani, 1986 for an introduction to the approach). Bootstrapping analyses are computationally intensive, but very large numbers of bootstrap samples can be generated in a few hours with modern processors. The analyses presented herein where all conducted with 10,000 bootstrap samples drawn from the data, except where noted otherwise due to low variance in the initial sample (resulting in bootstrap samples with zero variance). For ease of interpretation, the statistic of interest is presented with confidence intervals (bias-corrected and accelerated) generated from the bootstrap analysis, rather than with a *p*-value.

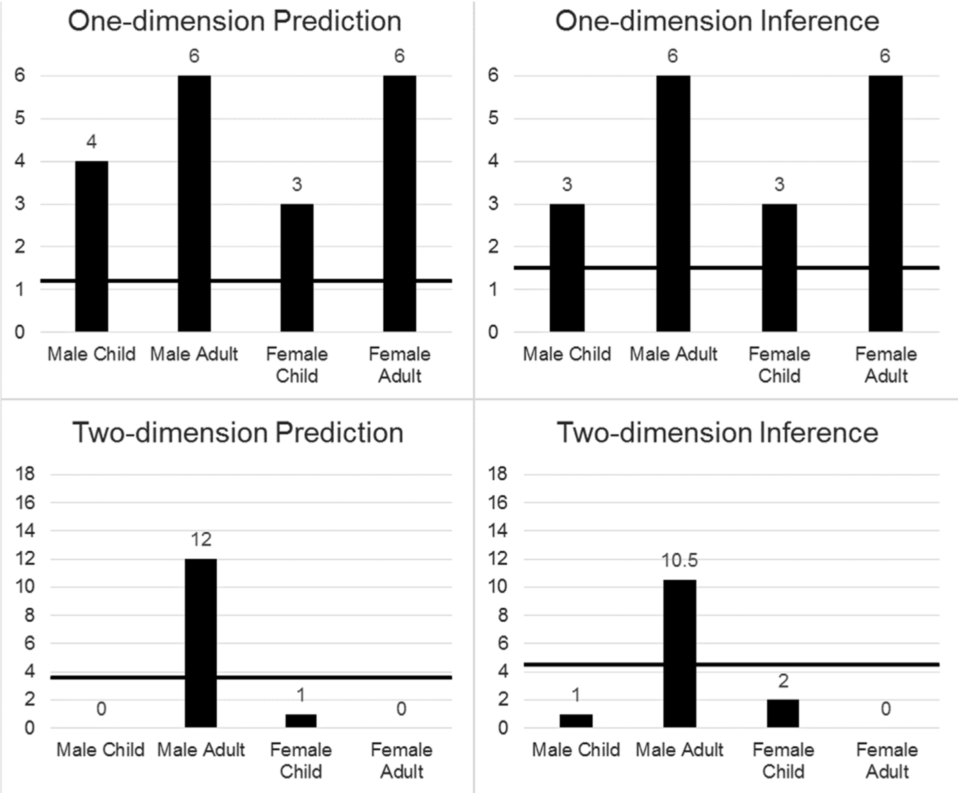

Fig. 5. Median score group by age and sex. The line represents chance and the graphs are scaled from the minimum possible score to the maximum possible score.

3.1. Hedgehog Game

Analyses were conducted separately for one-dimension and two-dimension trials because of the additional demands of two-dimension vector addition. Prediction items and inference items were also examined separately. Sex differences have been found in adults on qualitative physics measures (Coletta et al., 2011; Docktor & Heller, 2008; Kost et al., 2009; Lorenzo et al., 2006), so sex was included as a factor in addition to age.

Fig. 5 shows the performance of each group broken down by trial type and Table 2 shows the bootstrapped comparisons. For one-dimension trials, adults performed better than children on both prediction and inference trials. However, all groups showed above-

Table 2
Performance by age and sex.

Parameter	B	Std. Error	BCa 95% CI lower limit	BCa 95% CI upper limit
Prediction – One dimension				
Age	−1.29[a]	0.44	−2.16	−0.42
Sex	0.38	0.46	−0.54	1.30
Age * Sex	−1.00	0.63	−2.26	0.253
Inference – One dimension				
Age	−2.63[a]	0.35	−3.33	−1.94
Sex	−0.51	0.37	−1.24	0.22
Age * Sex	0.03	0.50	−0.97	1.02
Prediction – Two dimensions				
Age	−7.87[a]	1.42	−10.69	−5.06
Sex	−7.58[a]	1.50	−10.55	−4.61
Age * Sex	7.40[a]	2.05	3.35	11.45
Inference – Two dimensions				
Age	−8.12[a]	1.31	−10.70	−5.53
Sex	−6.35[a]	1.38	−9.08	−3.62
Age * Sex	7.42[a]	1.88	3.69	11.15

[a] Effect was statistically significant.

Table 3
Comparison to chance for two dimension scores.

Trial Type	BCa 95% CI lower limit	BCa 95% CI upper limit
Prediction[b]		
Female children	1.83	4.72
Male children	1.61	5.42
Female adults	1.92	5.56
Male adults[a]	9.00	13.48
Inference[c]		
Female children	2.00	4.34
Male children	1.19	3.00
Female adults	2.04	5.72
Male adults[a]	7.56	12.63

[a] Effect was statistically significant.
[b] Chance is 3.6.
[c] Chance is 4.5.

chance performance (Table 3). No sex-related effects were detected. In contrast, an age by sex interaction was detected for two-dimension trials, for both prediction and inference trials. When forces acted in two dimensions, adult males were the only group to perform above chance (Table 3).

It is also possible to explore the nature of the errors in two-dimension trials, specifically whether or not incorrect responses are randomly distributed among the distracters or if they conform to Halloun and Hestenes's (1985a) "dominance principle." It is important to note that among the prediction trials two trial types have more than one distracter that could indicate adherence to the dominance principle. For the trials in which the forces are equal in timing and size, allowing either force to "win" must be taken as an indicator of the dominance principle. For the trials in which the bigger force acts first, the two potential indicators of dominance are in conflict, again requiring that allowing either force to "win" be taken as an indicator of the dominance principle.

Since each trial type was measured three times for each participant the proportion of dominance-consistent errors was compared to chance, rather than a chi-square analysis. All groups, regardless of age or sex, showed more responses in line with the dominance principle than chance levels (Table 4). Differential performance with respect to age and sex on selection of dominance-consistent responses was also explored. There was a main effect for both age and sex for prediction trials, but only for age on inference trials (Table 5).

3.2. Individual differences

Sex differences are reported for mental rotation but not mental folding in adults (Harris et al., 2013; Linn & Petersen, 1995; Voyer et al., 1995). To determine whether analyses of potential factors to explain individual differences could be collapsed across sex, the additional measures were tested for sex differences, in both age groups (Table 6). Male children scored higher on the symbolic approximation task and bordered on significance for number line, but otherwise no sex differences were detected. A lack of sex differences on mental rotation was unexpected as previous research shows a small, but significant male advantage (Neuburger et al., 2011). Analyses of potential factors to explain individual differences in children's performance on the Hedgehog Game were collapsed across sex, with the exception of correlations with the number sense measures. Adult males scored higher on mental rotation,

Table 4
Comparison to chance for proportion of dominance errors.

Trial Type	BCa 95% CI lower limit	BCa 95% CI upper limit
Prediction[a]		
Female children	0.34	0.51
Male children	0.42	0.62
Female adults	0.46	0.64
Male adults	0.62	0.87
Inference[b]		
Female children	0.41	0.68
Male children	0.43	0.73
Female adults[c,e]	0.99	1.00
Male adults[d,e]	0.97	1.00

[a] Factoring in the two trial types for which two distracters match the dominance principle, chance is 0.33.
[b] Chance is 0.33.
[c] Scores based on 6402 samples.
[d] Scores based on 8796 samples.
[e] Results could not be computed for some samples so CIs were computed by percentile method, not BCa.

Table 5
Proportion of dominance errors by age and sex.

Parameter	B	Std. Error	BCa 95% CI lower limit	BCa 95% CI upper limit
Prediction				
Age	−0.24[a]	0.08	−0.39	−0.09
Sex	−0.21[a]	0.07	−0.36	−0.06
Age * Sex	0.11	0.10	−0.09	0.31
Inference				
Age	−0.41[a]	0.08	−0.56	−0.25
Sex	0.01	0.01	−0.01	0.03
Age * Sex	−0.05	0.10	−0.24	0.15

[a] Effect was statistically significant.

Table 6
Sex differences.

Measure	Female Biased-correct Mean	Male Biased-correct Mean	Mean Difference BCa 95% CI lower limit	Mean Difference BCa 95% CI upper limit	Biased-corrected *d*
Children					
Folding	3.90	3.58	−0.64	1.27	0.16
Rotation	5.06	5.83	−2.93	1.44	0.18
Vocab	19.18	20.48	−2.79	0.09	0.44
NL	0.61	0.74	−0.28	0.00	0.49
SA	7.54	8.65	−1.85	−0.34	0.73[a]
Adults					
Folding	9.96	11.49	−3.33	0.30	0.42
Rotation	7.40	12.32	−7.17	−2.59	1.03[a]
Vocab	32.94	32.37	−1.58	2.79	0.14

[a] Effect was statistically significant.

but there was no sex difference for paper folding or vocabulary, as expected. All adult analyses of individual differences were tested separately for males and females.

Correlations were conducted on child scores with CMFT, MRT, SA, and number line scores, controlling for *g* using WJIII-PV score. Correlations were done with respect to score on one-dimension trials only since children were not above chance two-dimension trials. Mental folding was the only consistent correlate, with relations to success on both prediction and inference trials, $r_{pred} = 0.43$, $p_{pred} < 0.01$, $r_{inf} = 0.34$, $p_{inf} < 0.01$. Symbolic approximation was correlated with success for male children on prediction trials, $r_{pred} = 0.41$, $p_{pred} < 0.03$, but not inference trials, $r_{inf} = -0.01$, $p_{inf} = 0.95$ and not correlated with success for female children on either trial type, $ps > 0.05$. Neither mental rotation nor number line performance were predictive for either trial type for either sex, $ps > 0.05$.

Correlations were conducted on adult scores on the Hedgehog Game with MRT-A and PFT scores, controlling for *g* using WJIII-PV score. Adult males were analyzed using total score for each trial type, while adult females were analyzed using their scores on the one-dimension trials only since they were not above chance for two-dimension trials. None of the correlations reached significance, $ps > 0.05$.

4. Discussion

These data provide evidence concerning the development of ideas about force and motion in one and two dimensions, using a test that is age-appropriate across a wide range, which taps into the cognitive system rather than the simpler tasks of recognition or action, and that has invariant surface features. This assessment allowed us to ask whether there is change in how people reason about a non-trivial force and motion problem across development. We found some surprising patterns.

For one-dimension problems, children have a good foundational understanding even as they are first entering into formal schooling. However, even for these problems on which children generally perform well, adults do better than children, with the majority performing at ceiling. This pattern suggests some foundation on which to build instruction, but also shows that there is age-related improvement, due either to instruction or to simple everyday experience.

A strikingly different pattern was seen for two-dimension problems – only adult males showed above-chance performance. Despite similarly low performance for both male and female children, the results for adult men provide evidence that more accurate representations can develop without extensive physics instruction. However, it is unclear why a sub-set of men developed accurate representations while other participants did not. The analysis of errors suggests that the less accurate representations were at least of a kind, regardless of age or sex; they followed the dominance principle reported by Halloun and Hestenes (1985a). While children showed a wider variety of responses, they resembled adults.

How does change occur on two-dimension problems? And why does it seem to progress towards physically-possible representations for males but not females? One possibility we explored concerned the role of spatial thinking and sex-related differences in spatial skill. However, this hypothesis was not supported by the present data set; spatial thinking was not found to predict performance on the Hedgehog Game among adults. This result was surprising because very similar problems were found to be related to spatial thinking in previous work (Kozhevnikov et al., 2007). Mental folding did predict performance for children, but that test typically does not show sex differences (Harris et al., 2013) so it cannot fully explain the results. It is possible that adults in this study may have been relying on heuristics, rather than spatial visualization, due to a sense of familiarity with the problems presented (Schwartz & Black, 1996). The Hedgehog Game may have appeared familiar due to its modeling of resistance forces, which were excluded in the problems used in Kozhevnikov et al. (2007) study but exist in the real world, or due to the inclusion of simpler problems at the beginning of the game. Perhaps spatial thinking could be useful for adults even in familiar problems, if they are explicitly encouraged to use visualization as a problem-solving approach or instructed in how to bring spatial skills to bear on these problems. Along those lines, Joh, Jaswal, and Keen (2011) demonstrated that asking 3-year-olds to imagine the path of a ball in a problem that typically resulted in physically-impossible responses significantly increased performance. Adults might similarly benefit from such a simple intervention.

Another approach to sex differences on the difficult problems might be to suggest that stereotypes, anxiety, and stereotype threat are at work, perhaps coupled with differential course-taking. Of course, appealing to course-taking runs into the problem that prior research has suggested that formal instruction does not always support the development of conceptual expertise, and the current study did not involve participants who had taken much formal physics, though it is possible there were some differences in high school. Stereotype threat and anxiety could be investigated, as they have been found to be important in other areas of learning, notably mathematics (e.g., Shih, Pittinsky, & Ambady, 1999; Lyons & Beilock, 2012). However, a recent meta-analysis revealed only weak effects of stereotype threat (Doyle & Voyer, 2016).

A third possible way to approach the problem of age-related change in men but not women is to suggest that men have more experience with two-force situations in everyday life, perhaps through sex-typed play, sports or hobbies, either directly through participation or indirectly through observation. Indeed, even if we accept an explanation like stereotype threat for sex differences, we still need to explain why men who have not taken much formal physics do better than young boys. There are recent in-principle demonstrations of the idea that people can accumulate memories of experiences with physical phenomena in the real world from which they can ultimately extract enough signal to reach valid conclusions (Sanborn, Mansinghka, & Griffiths, 2013; Ullman, Stuhlmuller, Goodman, & Tenenbaum, 2014) − what Sanborn et al. call a noisy Newton model. Furthermore, action experiences can improve students' conceptual understanding of some principles in physics, such as angular momentum and torque (Kontra, Lyons, Fischer, & Beilock, 2015). Of course, we would need to specify what the relevant experiences would be and to quantify how densely they occur in the everyday lives of girls and boys. A more direct strategy would be to conduct experiments providing both women and men with massed experience with the key perceptual phenomena and examine if they draw correct inferences without formal instruction, or if there are simple instructional manipulations that would lead to correct inferences. We could also add action experiences, perhaps literally with hockey pucks, or, more practically with gesture, which has supportive effects on learning in mathematics and other sciences (e.g., Atit, Gagnier, & Shipley, 2015; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014). Of course, there are potential pitfalls to this approach. Simple exposure to correct displays on one-dimension motion under gravity (i.e., falling-object problems) leads to correct predictions by novices in the short run, but fMRI data suggests that their answers involved fact retrieval rather than reasoning about the motion, which would not support long-term learning or transfer to problems involving different surface features (Foisy et al., 2016).

Development may also involve reconciling conflicting implicit and explicit knowledge that co-exist. This idea was already clear in research by Kaiser et al. (1992) and Krist et al. (1993), but has gained increased recent support from neuroscience research. Reiner and Gera (2016) showed that even when people make physically-possible predictions their ERP responses to viewing such physically-impossible events do not differ from people who made predictions that match the videos. There is also fMRI evidence that even experts may be inhibiting naïve ideas when providing physically-possible predictions, suggesting that these ideas often continue to co-exist (Foisy, Potvin, Riopel, & Masson, 2015). Instruction could usefully build on this idea. One potential intervention would be to attempt to unify students' ideas through a process of self-correction − iterating between eliciting students' explicitly held ideas and then asking them to rate how accurate those prediction appeared when demonstrated. Indeed, making a series of predictive sketches and then seeing immediately how they turn out has been shown to propel change in penetrative thinking in geoscience as compared to sketching descriptively or making predictions without sketching (Gagnier, Atit, Ormand, & Shipley, 2017). Similar success has been observed in physics as well. For example, Frederiksen and White's (1998) ThinkerTools platform allows learners to explore virtual microworlds with computer support for building and expressing new ideas, resulting in increased levels of physically-possible predictions.

It is noteworthy that, despite a distinct format and required thought process, the data from prediction and inference trials show the same pattern of results, with only one exception. While men did show a higher proportion of dominance responses than women in prediction trials, there was no such sex differences on inference trials. This disparity likely has its roots in the differing properties of the distracters across these problems. In inference problems, the remaining answer choices (beyond correct and dominance) would result in the ball traveling distinctly away from the intended goal. Both male and female adults were likely able to easily dismiss these possibilities. The non-dominance related distracter items in prediction problems, however, could not be so easily eliminated based on direction. Thus, despite this one difference, the overall similarity across prediction and inference problems suggests that the influence of task demands on differential performance, such as those found between tasks measuring the perception-action system and those measuring the cognitive system, have been successfully avoided.

This study clearly documented a puzzling pattern of age-related but sex-conditioned growth that demands explanation. Effective science education will be much more within our grasp if we know whether to concentrate on encouraging students to bring spatial skills to bear, adding manipulations to reduce stereotype threat and anxiety, including perceptual and embodied experiences, building on current improvements over traditional lectures, or some combination thereof. Furthermore, understanding why conceptual advance sometimes occurs, but far from invariably, is essential to building effective theory of cognitive development.

We presented an objective and understandable assessment technique to evaluate how change occurs and how best to support it. Questions remain with respect to the validity of the Hedgehog Game across development. First, it is necessary to confirm that this test is appropriate for physics learners between the two age groups that were tested. Such work would also allow for a detailed description of the developmental trajectory of ideas about force and motion between the time when students first enter primary schooling through the end of secondary education. Second, it is unclear if this test would be appropriate for children under the age of 5. Given that our results showed evidence that adult-like, physically-impossible predictions and inferences were already present in the children we tested, it is important to determine when these ideas first begin to form.

Acknowledgements

We wish to thank the Temple University Infant and Child Lab staff and interns, the Temple University Research in Spatial Cognition (RISC) group, Tilbe Göksun, Joanna Schiffman, Nicole Hallinen, and Dan Lyons. This work could not have been completed without the help of the children, their parents, and the adults who participated. This research was supported in part by a grant from the National Science Foundation to the Spatial Intelligence and Learning Center SBE-0541957 and SBE-1041707.

References

Atit, K., Gagnier, K., & Shipley, T. F. (2015). Student gestures aid penetrative thinking. *Journal of Geoscience Education, 63*(1), 66–72. http://dx.doi.org/10.5408/14-008.1.

Butts, J. R., Foley, A. F., Schiffman, J., & Levine, S. (2017). How an understanding of layers unfolds: a new mental folding task for young children. *Poster presented at the biennial meeting of the Cognitive Development Society*.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29*(3), 421–433. http://dx.doi.org/10.1037/0012-1649.29.3.421.

Coletta, V. P., Phillips, J. A., & Steinert, J. (2011). FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects. In N. Rebello, P. Engelhardt, & C. Singh (Eds.). *2011 Physics Education Research Conference* (pp. 23–26). New York: American Institute of Physics.

diSessa, A. A. (1982). Unlearning Aristotelian physics: a study of knowledge-based learning. *Cognitive Science, 6*, 37–75. http://dx.doi.org/10.1207/s15516709cog0601_2.

diSessa, A. A. (2013). A bird's-eye view of the pieces vs coherence controversy (from the pieces side of the fence). In S. Vosniadou (Ed.). *International handbook of research on conceptual change* (pp. 35–60). (2nd ed.). New York: Routledge.

Docktor, J., & Heller, K. (2008). Gender differences in both force concept inventory and introductory physics performance. *AIP (American Institute of Physics) Conference Proceedings, 1064*, 15–18.

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47*, 103–116. http://dx.doi.org/10.1016/j.lindif.2015.12.018.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1*(1), 54–75. http://www.jstor.org/stable/2245500.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Flores, S., Kanim, S. E., & Kautz, C. H. (2004). Student use of vectors in introductory mechanics. *American Journal of Physics, 72*, 460–468. http://dx.doi.org/10.1119/1.1648686.

Foisy, L.-M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience, 4*(1), 26–36. http://dx.doi.org/10.1016/j.tine.2015.03.001.

Foisy, L.-M. B., Potvin, P., Riopel, M., Allaire-Duquete, G., Nenciovici, L., & Masson, S. (2016). *Novices' neural correlates of error-correction in mechanics.* Toronto: International Mind, Brain, Education Society September.

Frederiksen, J. R., & White, B. Y. (1998). Teaching and learning generic modeling and reasoning skills. *Interactive Learning Environment, 5*(1), 33–51. http://dx.doi.org/10.1080/1049482980050103.

Göksun, T., George, N. R., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and motion: How young children understand causal events. *Child Development, 84*(4), 1285–1295. http://dx.doi.org/10.1111/cdev.12035.

Gagnier, K. M., Atit, K., Ormand, C., & Shipley, T. F. (2017). Comprehending 3D diagrams: Sketching to support spatial reasoning. *Topics in Cognitive Science, 9*(4), 883–901. http://dx.doi.org/10.1111/tops.12233.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*. http://dx.doi.org/10.1037/a0027433 Advance online publication.

Halloun, I. A., & Hestenes, D. (1985a). Common sense concepts about motion. *American Journal of Physics, 53*(11), 1056–1065. http://dx.doi.org/10.1119/1.14031.

Halloun, I. A., & Hestenes, D. (1985b). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043–1055. http://dx.doi.org/10.1119/1.14030.

Hammer, D. (1989). Two approaches to learning physics. *The Physics Teacher, 27*(9), 664–670. http://dx.doi.org/10.1119/1.2342910.

Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing, 14*, 105–115. http://dx.doi.org/10.1007/s10339-013-0544-6.

Hood, B. M. (1995). Gravity rules 2- to 4-year olds? *Cognitive Development, 10*, 577–598. http://dx.doi.org/10.1016/0885-2014(95)90027-6.

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science, 1*(1), 59–63. http://dx.doi.org/10.1111/1467-7687.00013.

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

Kaiser, M. K., Proffitt, D. R., & McCloskey, M. (1985). The development of beliefs about falling objects. *Perception & Psychophysics, 38*(6), 533–539. http://dx.doi.org/10.3758/BF03207062.

Kaiser, M. K., McCloskey, M., & Proffitt, D. R. (1986). Development of intuitive theories of motion: Curvilinear motion in the absence of external forces. *Developmental Psychology, 22*(1), 67–71. http://dx.doi.org/10.1037/0012-1649.22.1.67.

Kaiser, M. K., Proffitt, D. R., Whelan, S. M., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 669. http://dx.doi.org/10.1037/0096-1523.18.3.669.

Kim, I., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and intertia on object motion. *Developmental Science, 2*(3), 339–362. http://dx.doi.

org/10.1111/1467-7687.00080.
Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science, 26*(6), 737–749. http://dx.doi.org/10.1177/0956797615569355.
Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Unpacking gender differences in students' perceived experiences in introductory physics. *AIP (American Institute of Physics) Conference Proceedings, 1179*, 177–180.
Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5-month-old infants. *Developmental Science, 3*(3), 344–359. http://dx.doi.org/10.1111/1467-7687.00129.
Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*(4), 549–579. http://dx.doi.org/10.1080/15326900701399897.
Krist, H., Fieberg, E. L., & Wilkening, F. (1993). Intuitive physics in action and judgment: The development of knowledge about projectile motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 952–966. http://dx.doi.org/10.1037/0278-7393.19.4.952.
Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335–1342. http://dx.doi.org/10.1126/science.208.4450.1335.
Larkin, J. (1981). Cognition of learning physics. *American Journal of Physics, 49*(6), 534–541. http://dx.doi.org/10.1119/1.12667.
Larkin, J. H. (1982). *Spatial reasoning in solving physics problems. C.I.P. Paper #434.* Pittsburgh, PA: Department of Psychology, Carnegie-Mellon University.
Larkin, J. H. (1983). Problem representation in physics. In D. Gentner (Ed.). *Mental models*. Hillsdale, NJ: Erlbaum.
Lasry, N., Finkelstein, N., & Mazur, E. (2009). Are most people too dumb for physics? *The Physics Teacher, 47*, 418–422. http://dx.doi.org/10.1119/1.3225498.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265–288. http://dx.doi.org/10.1016/S0010-0277(87)80006-9.
Linn, M. C., & Petersen, A. C. (1995). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479–1498. http://dx.doi.org/10.2307/1130467.
Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics, 74*, 118. http://dx.doi.org/10.1119/1.2162549.
Lyons, I. M., & Beilock, S. L. (2012). When math hurts: math anxiety predicts pain network activation in anticipation of doing math. *Public Library of Science, 7*(10), e48076. http://dx.doi.org/10.1371/journal.pone.0048076.
Mix, K. S., Levine, S. C., Cheng, Y. L., Young, C., Hambrick, D. Z., Ping, R., et al. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145*(9), 1206–1227. http://dx.doi.org/10.1037/xge0000182.
Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2011). Gender differences in pre-adolescents' mental rotation performance: Do they depend on grade and stimulus type? *Personality and Individual Differences, 50*, 1238–1242. http://dx.doi.org/10.1016/j.paid.2011.02.017.
Nguyen, N., & Meltzer, D. E. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics, 71*, 630–638. http://dx.doi.org/10.1119/1.1571831.
Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction using the hands to learn math. *Psychological Science, 25*(4), 903–910. http://dx.doi.org/10.1177/0956797613518351.
Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5*(2), 193–207. http://dx.doi.org/10.1016/0885-2014(90)90026-P.
Reiner, M., & Gera, L. (2016). *The brain recognizes misconceptions that explicit cognition does not: Explicit vs. implicit physics knowledge of the falling pendulum.* Toronto: International Mind, Brain, Education Society September.
Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120*(2), 411. http://dx.doi.org/10.1037/a0031912.
Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*, 457–497. http://dx.doi.org/10.1207/s15516709cog2004_1.
Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80–83. http://dx.doi.org/10.1111/1467-9280.00111.
Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–250. http://dx.doi.org/10.1111/1467-9280.02438.
Sobel, M. (2009). Physics for the non-scientist: A middle way. *The Physics Teacher, 47*, 346–349. http://dx.doi.org/10.1119/1.3204113.
Ullman, T., Stuhlmuller, A., Goodman, N., & Tenenbaum, J. (2014). Learning physics from dynamical scenes. *Proceedings of the 23rd annual conference of the Cognitive Science Society,* 1–6.
Uttal, D. H., & Cohen, S. A. (2012). Spatial thinking and STEM education: When, why, and how? In B. Ross (Vol. Ed.), *Psychology of learning and motivation: Vol. 57*, (pp. 147–181). San Diego, CA: Elsevier, Inc.
Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics, 59*(10), 891–897. http://dx.doi.org/10.1119/1.16667.
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604. http://dx.doi.org/10.2466/pms.1978.47.2.599.
Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and mathematical skills across the preschool years [Monograph]. *Monographs of the Society for Research in Child Development, 82*(1), 1–150.
Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education, 1*(2), 205–221. http://dx.doi.org/10.1080/0140528790010209.
Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education, 23*(7), 1427–1445. http://dx.doi.org/10.1007/s11191-013-9640-3.
Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250–270.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement.* Itasca, IL: Riverside Publishing.