frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 10 February 2025
DOI 10.3389/fdpys.2025.1496651

OPEN ACCESS

EDITED BY
Kim P. Roberts,
Wilfrid Laurier University, Canada

REVIEWED BY
Vanessa R. Simmering,
University of Kansas, United States
Anastasia Datsogianni,
University of Nicosia, Cyprus
Diana Selmeczy,
University of Colorado Colorado Springs,
United States

*CORRESPONDENCE
Igor Bascandziev
igb078@mail.harvard.edu

RECEIVED 15 September 2024
ACCEPTED 15 January 2025
PUBLISHED 10 February 2025

CITATION
Bascandziev I, Abutto A, Walker CM and Bonawitz E (2025) Mind over matter: consistency monitoring and domain-specific learning. *Front. Dev. Psychol.* 3:1496651.
doi: 10.3389/fdpys.2025.1496651

Mind over matter: consistency monitoring and domain-specific learning

Igor Bascandziev[1]*, Adani Abutto[2], Caren M. Walker[3] and Elizabeth Bonawitz[1]

[1]Graduate School of Education, Harvard University, Cambridge, MA, United States, [2]Department of Psychology, Stanford University, Stanford, CA, United States, [3]Department of Psychology, University of California, San Diego, San Diego, CA, United States

Introduction: Children's naïve understanding of the physical world is permeated with inconsistencies among beliefs. For example, young children who believe that air does not occupy space also believe that balloons are filled up with air. Here, we asked if an ability to explicitly notice inconsistencies among statements is associated with a more mature understanding of the physical world.

Method: We tested 100 children who received a Physics Interview, a battery of Executive Functioning measures, a Cognitive Reflection measure, and a Consistency Monitoring measure.

Results and discussion: We found that Consistency Monitoring is associated with Physics Understanding, even when controlling for Age, Executive Functioning, and Cognitive Reflection. This finding highlights the importance of explicit consistency monitoring skills in the accumulation and expression of domain-specific understanding of the physical world, and it suggests future avenues for development and research of educational interventions that take into account the role of consistency monitoring skills in science learning.

KEYWORDS

consistency monitoring, physics understanding, cognitive reflection, executive functioning (EF), naïve theories and misconceptions

Introduction

Domain-specific theories help bind together our systems of beliefs and provide broad explanatory accounts of the data we observe. However, our theories about the world are fallible: they are sometimes wrong, sometimes internally inconsistent, and often incomplete. This fallibility is particularly pronounced in children's and adults' naïve theories. Indeed, one of the biggest obstacles that science teachers face in the classroom is "not what children lack, but what they have," namely children's naïve theories constructed in early childhood (Carey, 2000). As an example of internal inconsistency in children's understanding of the physical world, consider the common naïve belief that "air is nothing" or that "air does not take up any space" held by 6- and 7-year-olds (Carey, 2009). Contrast that with the belief—also held by many 6- and 7-year-olds—that "there is no air in outer space" or that "we use air to fill up balloons." The belief that "air does not occupy space" is inconsistent with the belief that "air can fill up a balloon." Importantly, children do not easily notice this inconsistency at an explicit level (by explicit, we mean available to verbal report; Limón and Carretero, 1997), and they do not easily change their belief that "air does not take up any space" upon seeing anomalous data, namely balloons being filled up with air (c.f., Posner et al., 1982). The existence of such inconsistent beliefs raises the question of how important the general ability to explicitly notice inconsistencies is for the process of theory revision and theory construction. Indeed, explicitly noticing inconsistencies in one's own understanding—as contrasted with implicit measures of feelings of uncertainty or slowing down—might facilitate the process of generating and assimilating more accurate models of reality (Limón and Carretero, 1997).

In the present study, we investigate the relationship between children's *domain-general* ability to explicitly notice inconsistencies among statements and their *domain-specific* progress toward generating more accurate models of reality. By domain-general, we mean cognitive resources and skills that are not specific to any given domain. This includes executive functioning, cognitive reflection, and consistency monitoring skills, as well as other broad cognitive abilities. By domain-specific, we mean areas (e.g., physics) with specific developmental trajectories, in which learners acquire the identity of the entities that belong to the domain (e.g., matter or energy in the case of physics), as well as the specific causal mechanisms they then use to predict and explain phenomena in that domain (e.g., mechanical forces in the case of physics; Carey, 1995; Wellman and Gelman, 1992).

More specifically, we investigated the interplay between domain-general skills like consistency monitoring, executive functioning, and cognitive reflection, and children's domain-specific understanding of the physical world. Notably, these domain-general skills are closely related to a broader category of metacognitive monitoring: e.g., consistency monitoring is often invoked when discussing metacognitive comprehension monitoring and error detection processes (Fleur et al., 2021) and executive functioning is also often implicated in metacognitive regulation (Fernandez-Duque et al., 2000). Additionally, although the initial cognitive reflection (CRT) task (Frederick, 2005) has mostly been discussed in the theoretical framework of intuitive and fast cognitive System 1 vs. a more reflective and slow cognitive System 2 type of reasoning (Kahneman, 2011), the slower and more reflective type of reasoning implies a metacognitive component (Shtulman and Young, 2023). Despite this similarity, however, in the present study we treat them as potentially overlapping but still different sets of skills.

As far as young children's domain-specific intuitive theories about the physical world are concerned, they are different from those of adults and from the current scientific understanding (Carey, 2009; Piaget and Inhelder, 1974). In some respects, children's intuitive beliefs and concepts resemble those of Aristotelian physics. For example, young children's concept of weight, similar to Aristotle's concept, is that weight is an accidental property of matter (akin to odor; Jammer, 1961). On this view, some things weigh something, and some things weigh nothing at all, just like some things have odor and some things do not. In line with this, children typically claim that a big pile of rice weighs something and that small pieces like a grain of rice or a small piece of Styrofoam weigh nothing at all (Smith, 2007; Smith et al., 1985, 1992, 1997). Furthermore, young children do not think of occupying space as a necessary property of matter, and they think that some small objects do not take up any space at all (Bascandziev and Carey, 2022). Another aspect of children's intuitive theory of matter is that it does not differentiate between non-material (but physically real) entities and material entities. For example, young children often say that non-material things like shadows, sound, and electricity are material (DeVries, 1986; Carey, 1991, 2009; Piaget, 1960). In short, from a modern scientific point of view, young children's intuitive understanding of the physical world is permeated with various misconceptions about the nature of matter and its properties.

Other related theoretical concepts that also undergo developmental change are the concept of density, children's understanding of the principles of water displacement, and their understanding of the law of conservation of weight. Young children, up to age 10 or older, have not yet constructed an adult-like concept of *density* that is differentiated from the concept *weight* (we use "weight" instead of "mass" because the concept *mass* is also not yet differentiated from *weight* at this age; see Carey, 2009 for review). This is evident across many tasks where children exhibit weight intrusions when making density judgments, for example, judging a large aluminum block to belong to the steel metals family because the absolute weight of a large aluminum piece is bigger than the weight of a small piece of steel. Children also exhibit density intrusions when making weight judgments, for example, judging that a small steel block would cause a bridge to collapse, while also correctly judging that a larger and heavier wooden block would not cause the same bridge to collapse (Smith, 2007; Smith et al., 1985, 1997, 1992; Smith and Unger, 1997; Snir et al., 1993). Similarly, children exhibit misconceptions when reasoning about the principles of water displacement. The typical error that many 6- to 8-year-olds make is that heavier objects, rather than objects with bigger volume, displace more water. When given an example of two objects with an identical volume but different weights, or two objects of different volumes where the smaller one is heavier, children tend to ignore the volume of the object and claim that the heavier one will displace more water (Colantonio et al., 2022, 2023; Theobald et al., 2024). Finally, it has been widely documented that young children up to age six or older fail to appreciate the law of conservation of weight (Piaget and Inhelder, 1974). Taken together, these findings show that many interrelated concepts—that together constitute children's understanding of the physical world—undergo change in early and middle childhood.

Inconsistencies among beliefs and domain-specific learning

A common feature of (children's) naïve theories is that they are globally inconsistent (diSessa, 1988; Friedman and Forbus, 2011; Friedman et al., 2018). For example, a child who believes that a grain of rice weighs nothing at all, also knows that a pile of rice weighs something, and that the sum of an infinite number of zeros equals to zero (Bascandziev and Carey, 2022). Thus, the belief that a grain of rice weighs zero units of weight and the belief that a pile of rice weighs some non-zero number of units of weight cannot be true at the same time. The same line of argument applies to reasoning about space. The belief that a piece of material takes up no space at all and the belief that a larger piece of the same material takes up some non-zero amount of space cannot be true at the same time, given that the sum of zeros equals zero. As mentioned above, children's reasoning about the material/non-material distinction also runs into inconsistencies. The statement that air does not take up space, because air is nothing, is incompatible with the statement that air is used to fill up balloons.

A similar type of conflict arises in children's reasoning about water displacement. The belief that heavier objects displace more water is inconsistent with the belief that two material entities—such as water and blocks—cannot occupy the same space at the same time—a belief that young children and even infants seem to have (Carey, 2009; Spelke et al., 1992). Consider the following example as an illustration. Imagine two blocks that are a perfect fit for their respective containers. When the blocks are in the containers, there is no space left in the containers. In other words, nothing else but the blocks can fit in the containers, because two material objects cannot occupy the same space at the same time. If the two blocks are pushed in two containers that have an equal amount of water, the blocks will displace the same amount of water, namely all of the water that is inside the containers. How heavy the blocks are has no bearing on how much water they will displace. Despite this, however, children routinely claim that heavier blocks displace more water.

The existence of inconsistent beliefs within the same individual has been well documented in the psychological literature. Those range from coexistence of supernatural and scientific explanations (Legare et al., 2012; Legare and Shtulman, 2018) to coexistence of naïve and scientific explanations (Bascandziev, 2022, 2024; Shtulman and Harrington, 2016; Shtulman and Legare, 2020; Shtulman and Lombrozo, 2016). For example, although most scientifically naïve adults have acquired a vitalist theory of biology according to which plants are living things but the sun is not, they (and even university professors) are slower and less accurate to confirm that plants are alive than to confirm that animals are alive under speeded conditions (Goldberg and Thompson-Schill, 2009). This suggests that these individuals continue to implicitly harbor conflicting scientific and naïve beliefs. Similarly, healthy elderly with weakened executive functioning sometimes will say that the "sun is alive because it's moving" under normal (i.e., not speeded) conditions, although their biological theory and explanations seem to remain intact (Tardiff et al., 2017). This also suggests that naïve and scientific representations that are in conflict with each other are held by the same individual.

Importantly, various studies have investigated the effects of a conflict between the observed evidence and a model of the world (e.g., Bascandziev, 2024; Limón, 2001; Posner et al., 1982; Theobald et al., 2024; see Potvin, 2023 for review). Indeed, children and even infants seem to learn from data that are in conflict with their model of the world (Bonawitz et al., 2012; Legare et al., 2010; Schulz et al., 2008; Stahl and Feigenson, 2015). For example, children who erroneously believe a block should balance at its geometric center (rather than center of mass) are more likely to explore a block balancing at a non-geometric point and correctly revise their beliefs following exploration of this "anomalous" data (Bonawitz et al., 2012). At some level, the children in these looking-time and behavioral tasks are registering a conflict and acting on it. On the other hand, the conflict between anomalous data and the learner's model of the world is not always explicitly noticed (i.e., accessible to verbal report), and it does not always lead to learning (Chinn and Brewer, 1993; Dreyfus et al., 1990; Kuhn, 1989; Limón and Carretero, 1997).

The findings that conflicting evidence sometimes generates behaviors that support learning and other times goes unnoticed raises two important, inter-related questions. First, given that anomalous data and inconsistent beliefs often go unnoticed, which domain-general cognitive skills are involved in consistency monitoring? Second, how are those consistency monitoring skills related to the acquisition and accumulation of domain-specific knowledge?

Present study: consistency monitoring vis a vis other domain-general cognitive skills and domain-specific knowledge

What kind of cognitive capacity is required for one to explicitly notice inconsistencies among beliefs or statements? There are several plausible, not mutually exclusive answers that we investigate in the present study. The first is that executive functions are foundational to consistency monitoring skills. Executive functioning is a set of skills including updating or working memory, cognitive flexibility or set shifting, and inhibitory control (Miyake et al., 2000). These skills are implicated in self-regulation, planning, metacognitive control, comprehension and conflict monitoring (Botvinick et al., 2004; Hofmann et al., 2012; Roebers et al., 2019; Neuenschwander et al., 2012). Indeed, the term executive functioning is often used interchangeably with terms such as error monitoring or conflict monitoring (Checa et al., 2014). Thus, it is plausible to say that the capacity to notice inconsistencies among beliefs and statements, especially in the context of domain-specific learning, is functionally dependent on executive functioning.

A second possibility is that consistency monitoring skills overlap with the ability for cognitive reflection. Cognitive reflection is somewhat independent of executive functioning, and it is defined as an ability to override an initial intuitive response with an analytic or reflective response (Frederick, 2005; Stanovich et al., 2016). By definition, engaging in reflective reasoning means reasoning about a particular issue, monitoring for (intuitive) errors (i.e., consistency monitoring), inhibiting intuitive incorrect responses, and generating analytic (accurate) responses (Shtulman and Young, 2023). Thus, it is plausible that the capacity to explicitly notice inconsistencies among beliefs and statements overlaps greatly with the capacity for reflective reasoning.

A third option is that the ability to explicitly notice inconsistencies among beliefs or statements is sufficiently independent from both executive functioning and reflective reasoning, and it is implicated in domain-specific knowledge acquisition. We review the three options in more detail below.

The relationship between domain-general executive functioning and domain-specific knowledge

There is a large literature showing that executive functioning is related to domain-specific knowledge acquisition and the expression of the acquired knowledge (see Carey et al., 2015 for review). That is, executive functioning has been implicated in the acquisition and expression of mathematical knowledge (Bull and Lee, 2014), understanding of the psychological world (Carlson and Moses, 2001; Devine and Hughes, 2014), understanding of the biological world (Bascandziev et al., 2018; Tardiff et al., 2020), and understanding of the physical world (Colantonio et al., 2024; Thibault and Potvin, 2018). Across all these domains, executive functioning may be involved in the online processing and the expression of an already acquired understanding, or it may be involved in the construction of such understanding (Carey et al., 2015). However, it is unclear what *specific role* executive functioning is playing in the construction of domain-specific understanding. Although some proposals have speculated that executive functioning may be playing a role in conflict monitoring and in the process of resolving noticed inconsistencies (Bascandziev et al., 2018), the hypothesis that consistency monitoring predicts domain-specific knowledge (concurrently, independently, or as part of executive functioning) has not been tested directly.

The relationship between domain-general cognitive reflection and domain-specific knowledge

Similarly, there is a growing literature showing that cognitive reflection is related to domain-specific knowledge (Shtulman and Young, 2023). Cognitive reflection has been shown to be associated with mathematical understanding, understanding of the physical and biological worlds (Young and Shtulman, 2020a,b), and a wide range of skills that are important for scientific reasoning and science learning (Don et al., 2016; Gervais, 2015; Pennycook and Rand, 2019; Shtulman and McCallum, 2014; Stanovich et al., 2016). Importantly, cognitive reflection has been shown to predict domain-specific performance over and above executive functioning (Young and Shtulman, 2020a). It has been argued that the main role of cognitive reflection in the expression of domain-specific understanding is the ability to override intuitive ideas and responses while engaging in reflective reasoning processes (Shtulman and Young, 2023). However, whether engaging in cognitive reflection also means engaging in consistency monitoring is not clear.

Consistency monitoring as an independent predictor of domain-specific knowledge

In the present study, we are testing the hypothesis that consistency monitoring is associated with young children's understanding of the physical world. In order to test this hypothesis, we developed an individual differences measure that, at face value, measures explicit consistency monitoring directly. The measure was developed by adapting tasks from Markman's (1977, 1979) pioneering work on comprehension monitoring. The tasks included in this measure involve short texts that have numerous inconsistencies or straightforward contradictions. The texts used in the present study were about animals and animal behavior, which is a domain unrelated to the children's developing physics understanding. In this task, after hearing the texts/stories, children are asked whether the story makes sense, whether there is anything confusing about the story, and whether the story is true or not. We reasoned that children who notice the inconsistencies and contradictions in the text would answer that the story did not make sense, that it is confusing, and that the story as a whole is not true. We predicted that children's performance on this task is going to be related to their domain-specific understanding of the physical world.

We did not have any specific predictions about the predictive power of the consistency monitoring measure over and above executive functioning and cognitive reflection. As reviewed above, one possibility is that executive functioning, cognitive reflection, or both might be foundational to explicit consistency monitoring. If so, then we should find that the consistency monitoring measure is unrelated to children's understanding of the physical world after controlling for executive functioning and cognitive reflection. Another possibility, however, is that the consistency monitoring measure is sufficiently independent from executive functioning and cognitive reflection and therefore predictive of children's physics understanding over and above executive functioning and cognitive reflection.

In summary, the first and main hypothesis tested in the present study is that children's ability for consistency monitoring is associated with the domain-specific understanding of the physical world. To test this hypothesis, we tested the prediction that, when controlling for age, the newly developed consistency monitoring measure (i.e., the Inconsistent Stories task) will be correlated with children's performance on the Physics Interview.

Second, we tested the hypothesis that children's executive functioning and cognitive reflection abilities are associated with the domain-specific understanding of the physical world. Although this hypothesis has been tested in other domains (e.g., Bascandziev et al., 2018; Carlson and Moses, 2001; Colantonio et al., 2024; Devine and Hughes, 2014; Tardiff et al., 2020; Thibault and Potvin, 2018; Zaitchik et al., 2014), to our knowledge, it has not been tested in the domain of physics. We predicted that, when controlling for age, the executive functioning and cognitive reflection measure will be correlated with children's performance on the Physics Interview.

Third, we investigated several research questions for which we did not have specific predictions. We asked whether children's ability for explicit consistency monitoring is associated with physics performance over and above executive functioning, over and above cognitive reflection, and over and above both executive functioning and cognitive reflection. Learning the answers to these research questions is important because it will shed light on whether the relationship between explicit consistency monitoring and physics understanding is independent or dependent on executive functioning and cognitive reflection.

Method

Participants

A total of 100 children were recruited (M_{Age} = 84.79 months; SD = 12.21; range = 57 to 114 months). All 100 children participated in the first assessment session of the study in which they received the Physics Interview. Due to attrition, a total of 85 children participated in the second assessment session of the study (about a month later) in which they received the battery of domain general tasks (i.e., executive functions tasks, the Cognitive Reflection Task—Developmental, and the consistency monitoring measure (Inconsistent Stories Task). The average age of the 85 participants was 84.39 months (SD = 12.5; range = 57 to 114 months) at the start of the study. The convenience sample was drawn from elementary schools in the Boston metro area, which comprise a predominantly white, Non-Hispanic, middle-class population. The school district from which the majority of the sample was drawn is composed of 71.8% White, 13.7% Asian, 6.7% Hispanic, 1.9% African American, 5.8% Multi Race, Non-Hispanic, and 0.1% Non-Hawaiian Pacific Islander families.

Procedure

The data presented here were collected as a part of a larger study that included pre-training assessment (Physics Interview), four teaching sessions about the material world ∼1 week apart, and post-training assessment (that included the same Physics Interview and domain-general tasks designed to measure children's executive functions, cognitive reflection, and consistency monitoring).[1] The present study reports only a portion of the data collected for the larger study, namely children's performance on the Physics Interview at pre-training only and their performance on the domain-general tasks that were administered at post-training. We investigated children's performance on the Physics Interview at pre-training (as opposed to post-training) because the four training sessions administered after pre-training targeted children's physics understanding, which were manipulated in three different conditions in the larger study [Thought Experiments, Real Experiments, and Baseline (no training) condition], which systematically influenced children's physics understanding at post-training. However, the training that targeted physics concepts exclusively was not expected to have any influence on children's executive functioning, cognitive reflection, or consistency monitoring. Indeed, a one-way ANOVA comparison of the three groups across these three domain-general variables showed that there were no significant differences on those measures as a function of a group. The data on pre- to post-training improvement are presented elsewhere. The two assessment sessions (i.e., the pre-training Physics Interview and the post-training domain-general measures) presented here were ∼6 weeks apart.[2] The prediction tested in the present study, namely that children's performance at pre-training will be related to consistency monitoring, executive functioning, and reflective reasoning is included in the pre-registration among the other predictions that are tested and presented elsewhere. In addition, the pre-registration describes the sample size, and the coding procedures for the physics interview. All assessments involved one-on-one testing, and they were conducted in a quiet classroom or a quiet corner in a hallway in the children's schools.

Physics interview

The Physics Interview was designed to cover concepts that are the target of early elementary STEM education (NGSS Lead States, 2013) and have been researched extensively in prior literature (e.g., see Carey, 2009 for review). Moreover, the concepts targeted in the Physics Interview undergo a prolonged acquisition period because of their complexity and also undergo dramatic change in early and middle childhood (Carey, 2009). For example, at age 6, children have not yet differentiated material from physically real but non-material entities (e.g., shadows, electricity, or sound) and they do not think that gasses are material (Carey, 2009). Furthermore, children at this age have not yet constructed a concept of *density* that is differentiated from the concepts *weight* and *size* (Smith et al., 1985), they have incorrect beliefs about the principles of water displacement (Theobald and Brod, 2021), and they still make conservation errors (e.g., that the weight of an object would change if its shape changes; Piaget and Inhelder, 1974). Below, we give a short overview of the Physics Interview questions that targeted these concepts.[3]

Material non-material distinction, weight, and occupying space

In order to tap into children's understanding of the distinction between material and non-material entities, as well as their understanding of weight and occupying space as necessary properties of matter, the interview included a series of questions on this topic. After providing an introduction, a few examples, and a child-friendly locution about what we mean by the word material (i.e., something material is something made of stuff; see Carey, 2009 for examples where the "made of stuff" locution was used to question children about their understanding of the material non-material distinction), the experimenter first asked whether a list of 10 different entities were material or not (e.g., "Is air made of stuff?"). These questions were designed to examine the material/non-material distinction (DeVries, 1986; Carey, 1991). Next, the experimenter asked if the same list of 10 entities occupy space, and whether they weigh anything at all (note that children at this age do not differentiate weight from mass; e.g., "Does air take up space?" and "Does air weigh anything at all?"). Then, children were asked to agree or disagree with an argument made by a different child according to which if you put the *shadow* of an elephant on your hand, you will not be able to lift your hand because elephants are so big. These questions were designed to examine children's understanding that shadows are not material and do not weigh anything at all. The next set of questions also considered children's understanding of occupying space and having weight as necessary properties of matter, but this time by using actual visible, but very small pieces of matter, namely a tiny ball made of playdough, a grain of sand, and a tiny piece of sponge. For each piece, the experimenter placed the entity on the table in front of the child and asked if the piece takes up a lot of space, a tiny bit of space, or no space at all, and whether it weighs a lot, a tiny bit, or nothing at all (Bascandziev and Carey, 2022).

Density, water displacement, and conservation

In order to tap into children's understanding of density and the differentiation of the concepts *weight, volume, and density,* we

1 The raw data, interviews, and coding schemes are available at https://osf.io/ua3rb/.

2 We pre-registered the predictions for the larger study at https://aspredicted.org/DJG_YWR.

3 The full interview and the coding scheme are available at https://osf.io/ua3rb/.

administered a set of questions adapted from Smith et al. (1985). For example, children were shown two blocks that are different sizes but weigh the exact same amount, and they were asked if the two blocks could be made of the same material. Similarly, children were shown blocks of the same size that weigh different amounts, and they were asked if those two blocks could be made of the same material.

Next, children received items designed to test their understanding of the principles of water displacement. Children were shown two containers with an equal amount of water and two balls made out of different materials, with different (or same sizes), and with different weights. Children were asked to imagine pushing the two balls all the way to the bottom of the container, and they were asked if one of the two balls will push up more water or if the two balls would push up an equal amount of water (Theobald and Brod, 2021).

Finally, children received a question about the conservation of weight (Piaget and Inhelder, 1974) and questions borrowed from The Inquiry Project curriculum about units of weight measurement and about conservation of weight (TERC, 2011). As specified in our pre-registration, we scored the judgments that children made in the Physics Interview by assigning a score of 1 to correct judgments and a score of 0 to incorrect judgments. All 100 children in the sample completed the Physics Interview. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was ICC = 0.99, $p < 0.001$.

Consistency monitoring (inconsistent stories task)

Inconsistent stories task

The Inconsistent Stories Task (IST) has been adapted from Markman's (1979) work on children's metacognitive unawareness of their comprehension failures. To our knowledge, this is the first adaptation of this task as an individual difference measure. Before reading the stories, the experimenter told the children that he would read them a couple of stories, and that their job is to pay close attention, because they would be answering a few questions afterwards. The task included a total of two stories (adapted from Markman, 1979), which were about animals and included inconsistencies. For example, in one story children heard that snakes find insects by listening to them and that snakes do not have ears and they cannot hear insects. After hearing each story, children received three questions: i) "Did the story make sense?" ii) "Do you think the story is true?" iii) "Was there anything confusing about the story?" If children said that the story makes sense, that it is true, and that there is nothing confusing about the story, they received 0 points on the respective question. If children said that the story did not make sense, that the story is not true, and that the story was confusing, they received 1 point on each respective question. If children answered any of the three questions correctly, they were asked a follow up question to explain why they thought the story did not make sense, why it was not true, or why it was confusing. A total of 84 children completed the Inconsistent Stories task. Given our two stories with three questions each, the possible range of scores on this task was between 0 and 6. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was ICC = 0.99, $p < 0.001$.[4]

Executive functions (backward digit span, verbal fluency, and day-night)

To measure the three different aspects of executive functions, namely working memory, set-shifting, and inhibitory control (Miyake et al., 2000), we administered three different executive function tasks: backward digit span, verbal fluency, and the day-night task.

Backward digit span

The backward digit span is a working memory task. Children were told that the experimenter would read them some numbers, and that their job is to remember the numbers and tell the experiment what the numbers were, but in backwards order. The experimenter gave a few examples and said: "if I say 1, 2, you should say 2, 1; if I say 3, 4, you should say 4, 3. Okay?" Next, children were able to complete a few practice trials with 2-digit numbers during which trials they were given corrective feedback if they made any errors. The test trials began with a block of two 2-digit numbers, then a block of two 3-digit numbers, all the way to a block of two 7-digit numbers. The testing was discontinued after children made two consecutive errors within a block, but not before the block with 5 digits was reached. The score that each child received was equal to the cardinal value of the largest set that the child repeated correctly. For example, if the child correctly repeated at least one string with 2 digits and repeated incorrectly both strings with 3 digits, then the child received a score of 2. If the child correctly repeated at least one string with 3 digits and made two consecutive errors within the block with 4 digits, then the child received a score of 3. The range of possible scores is between 2 and 7. A total of 82 children completed the Backward Digit task. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was ICC = 0.99, $p < 0.001$.

Verbal fluency

The Verbal Fluency task is designed to measure cognitive flexibility or set shifting (Munakata et al., 2012; Troyer et al., 1997). This task has been used as a measure of endogenous set shifting with children of similar ages as the ones in the sample of the present study (Bascandziev et al., 2018; Shtulman et al., 2023; Snyder and Munakata, 2010; Young and Shtulman, 2020a,b). Children were given two tasks: Animal Naming and Food Naming. For each task, the experimenter told children that they should name as many animals (or foods) as they can in 1 min. After the experimenter ensured that the child understood the task, children began naming animals (or foods) until the time was up. To be successful on

4 The coding schemes and the raw data for all tasks described above are available at https://osf.io/ua3rb/. Furthermore, the larger study, including predictions made for the results presented here, have been pre-registered at: https://aspredicted.org/DJG_YWR.

this task requires the use of an abstract superordinate concept *animal* or *food* and then search a vast lexical database of individual instances of animals and foods. A strategy that finds subcategories of animals (or foods; e.g., farm animals, jungle animals, sea animals, etc.), monitoring when the subcategory is exhausted, and flexibly switching to a different subcategory, which includes endogenous set shifting, leads to higher scores on this task (Snyder and Munakata, 2010). Children's scores reflect the number of unique animals and foods that they named in 1 min. Repetitions (with some exceptions)[5] and incorrect responses (e.g., drinks instead of foods) were excluded. Children's verbal fluency score is a simple sum of the Animal and Food naming. A total of 84 children completed the Verbal Fluency task. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was (ICC = 0.98, $p < 0.001$).

Day-night task

The Day-Night task is designed to measure children's inhibitory control ability. We modeled our task after Gerstadt et al. (1994) classic study in which they administered this task to 3 ½ to 7-year-olds. Children were shown two pictures, one of a day sky and one of a night sky, and they were told that when they see a picture of a day sky, they are supposed to say "night" as fast as possible, and when they see a picture of a night sky, they are supposed to say "day" as fast as possible. Next, children received four practice trials with corrective feedback and ample time to respond. The test trials consisted of 10 trials during which one picture was presented at a time for ∼ 1to 2 s. In order to succeed, children needed to inhibit the prepotent response, namely, to say the word that describes the picture, and then produce the opposite word. Children's scores reflect the number of correct responses on the 10 trials. Incorrect responses, false starts (e.g., "da.. night"), switched answers (e.g., "day, no... night"), and missed trials were given 0 points. The range of possible scores is from 0 to 10. A total of 82 children completed the Day Night task. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was ICC = 0.96.

Cognitive reflection (cognitive reflection task)

Cognitive reflection task—developmental (CRT-D)

The Cognitive Reflection Task—Developmental is designed to measure children's reflective reasoning, which involves recognizing and rejecting an intuitive but incorrect response, and then providing a counterintuitive but correct response (Young and Shtulman, 2020a). An example of an item is: "If you are running a race and you pass the person in second place, what place are you in?" The intuitive lure is to say 1st place, which is an incorrect answer. The correct answer is that the person who passed the person in 2nd place will be in 2nd place. Before administering the task, the experimenter told the children that they will hear some really tricky questions, that they should listen carefully, and that they should try their best to give a correct answer. The task included a total of five questions. Each answer was scored as correct (1 point) or incorrect (0 points), thus producing a possible range between 0 and 5 points. A total of 84 children completed the Cognitive Reflection task. To check the interrater agreement, 30% of the data were coded by two independent coders. The interrater agreement was ICC = 1; i.e., perfect agreement.

Results

We first present the descriptive statistics of the Physics Understanding measure as well the domain-general measures. Next, we present the bivariate correlations between the variable of interest (Physics Understanding) and the domain-general measures. Finally, we present a series of regression analyses, which focuses on the relationship between the Physics Understanding outcome variable and the control and domain-general predictor variables.

Descriptive statistics: physics understanding

Table 1 presents the descriptive statistics of the six sub composites designed to measure different aspects of children's understanding of the material world. The interview's questions targeted several concepts, including children's understanding of the material/non-material distinction (e.g., that air is material, but shadows are not), the understanding that material things occupy space, and the understanding that material things have weight. In addition, the interview tapped into children's understanding of the concept density and how it is differentiated from weight, the understanding of water displacement principles, and the understanding of conservation of weight.

Inspection of Table 1 shows that there was sufficient variability among children on the six sub composites and children were showing neither floor, nor ceiling effects on any of the six sub composite variables. As far as individual variability is concerned, some children were performing near the bottom of the possible range and some children were performing at the very top of the possible range on each of the six sub composites. As a group, children's average scores indicated that they have not yet acquired an adult-like theory of matter. The errors that children typically made on the Physics interview were consistent with prior findings in the literature (see Carey, 2009 for review). For example, on average, children denied that gasses such as air and steam are material, but they frequently said that non-material but physically real entities, such as electricity, are material. Similarly, on average, children denied that material things such as air and steam occupy space and have weight (Carey, 2009), they believed that heavier objects (rather than objects with bigger volume) displace more water (Theobald and Brod, 2021), that weight could change if one changes the shape of the object (Piaget and Inhelder, 1974), or that two blocks of the same size but different weights could be made of the same material (Smith et al., 1985).

The internal consistency of the six sub composites, where each sub composite was treated as an item, was acceptable (Cronbach's

5 See coding scheme at https://osf.io/ua3rb.

TABLE 1 Descriptive statistics of the six sub composites of the Physics Interview ($n = 100$).

	Material/Non-material distinction	Space	Weight	Density	Water displacement	Conservation
Mean	6.19	7.82	9.68	1.99	1.69	1.54
SD	1.76	2.18	2.67	0.75	1.33	0.96
Range	2–9	4–12	3.5–15	1–3	0–4	0–3
Possible range	0–10	0–13	0–15	0–3	0–4	0–3

Alpha = 0.7). This allowed for the construction of a Physics Understanding composite variable. To construct a composite variable, each of the six sub composites was first standardized,[6] and then the six composites were averaged into a single variable, which was again standardized with a mean of 0 and a standard deviation of 1. Thus, a score of 0 on the Physics Understanding outcome variable represents an average performance.

Descriptive statistics: domain general measures

Table 2 presents the descriptive statistics of the domain general measures, including the consistency monitoring measure [measured with the Inconsistent Stories Task (IST)], the three executive function measures: working memory (measured with the Backward Digit Span task), set shifting ability (measured with the Verbal Fluency task), and inhibitory control (measured with the Day Night task), as well as the cognitive reflective measure [measured with the Cognitive Reflection Task - Developmental (CRT-D) task].

Inspection of Table 2 shows that there is sufficient variability across all domain-general measures. The range of scores that children achieved on all tasks was near or equivalent to the possible range. In other words, some children were performing at the bottom of the possible range and some children were performing at the very top of the possible range of scores. As far as the average group performance is concerned, the performance on the Day Night task was near the ceiling, suggesting that the task was relatively easy for children at this age. Conversely, the average group performance on the Cognitive Reflection Task—Developmental was near floor, suggesting that children were providing incorrect responses most of the time on this particular task. Importantly, however, although the average performance was high on the Day Night task and low on the CRT-D task, there was sufficient variability around those average scores. The average group performance on the remaining tasks was neither near floor nor near ceiling.

In order to reduce the number of predictor variables, on theoretical grounds, we created a composite Executive Function (EF) variable by standardizing and averaging the Backward Digit Span, Verbal Fluency, and Day Night tasks. The composite EF variable was then standardized. The composite EF variable was based on the data from 81 children. We also include a supplementary analysis posted on OSF (see link above) where each of the EF variables is treated as an independent construct. The supplementary analysis presents results that are consistent with the results presented here. Finally, on theoretical grounds, we kept the Inconsistent Stories and Cognitive Reflection tasks as separate predictors. Both variables were also standardized. The variable Age was also standardized, which means that Age of 0 represents an average age of the sample.

Correlations among the outcome, predictor, and control variables

Figure 1 presents the bivariate correlations among the outcome Physics Understanding variable, the domain general predictor variables, and Age. In addition, Figure 1 represents the distributions of each variable as well as the scatterplots of the bivariate relationships. Inspection of the correlation coefficients reveals that the outcome variable of interest, Physics Understanding, is significantly correlated with all domain-general predictor variables and also with Age. The strength of the correlation coefficients between Physics Understanding and the domain general predictor variables ranged between moderate to high. Inspection of the correlation coefficients among the domain-general variables shows that all three constructs, namely Consistency Monitoring, Executive Functioning, and Cognitive Reflection are correlated with each other, and they are all correlated with Age.

In addition, we investigated the intercorrelations between the three predictor variables, namely Executive Functioning, Cognitive Reflection, and Inconsistent Stories while controlling for Age. We found that controlling for Age, the correlation between EF and CRT-D was $r_{(78)} = 0.16$, $p = 0.15$, the correlation between EF and Inconsistent Stories was $r_{(78)} = 0.07$, $p = 0.55$, and the correlation between CRT-D and Inconsistent Stories was $r_{(78)} = 0.17$, $p = 0.14$. This finding suggests that the measures of executive functioning, cognitive reflection, and consistency monitoring tap into unique constructs that are independent from each other.

Predicting children's physics understanding

To test the hypotheses and research questions outlined above, we performed a series of regression analyses. Table 3 presents the regression coefficients and the associated statistics of six different models (Models A through F). The outcome variable in all models

6 This allowed each of the six sub composites to have an equal weighting when averaged together. Creating a composite variable from the raw scores, however, does not change the results presented in this study.

TABLE 2 Descriptive statistics of the domain-general predictors ($n = 82–84$).

	Inconsistent stories	Backward digit	Verbal fluency	Day-night	Cognitive reflection
Mean	2.73	3.43	28.62	8.45	0.99
SD	1.88	0.98	9.47	2.04	1.22
Range	0–6	2–5	0–50	1–10	0–5
Possible Range	0–6	2–7	0–n/a	0–10	0–5

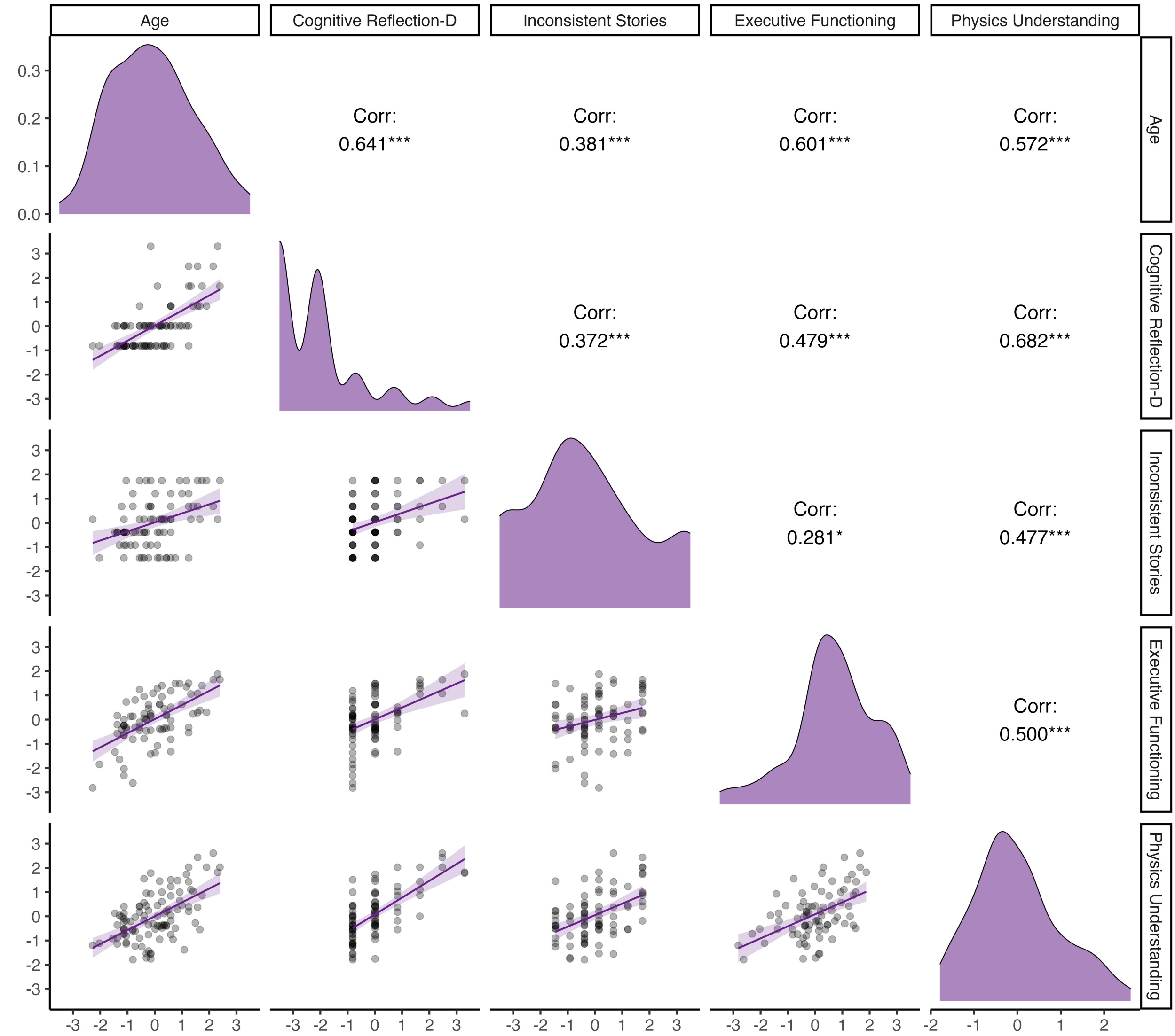

FIGURE 1
Bivariate correlations between Physics Understanding and the domain general predictor variables ($n = 81–100$). * < 0.05, ** < 0.01, *** < 0.001. The distributions of all variables are on the diagonal. Below the diagonal are the bivariate scatterplots. All plotted variables represent z-scored measures.

is children's Physics Understanding and the goal of each model was to test the predictive value of each of the domain general variables while controlling for other variables.

We first asked whether Consistency Monitoring, Executive Function, and Cognitive Reflection would continue to be significantly associated with Physics Understanding after controlling for Age. The first Model A tests the main prediction that Inconsistent Stories is going to be correlated with children's performance on the Physics Interview even after controlling for Age. The regression analysis confirmed this prediction, and it showed that controlling for Age, the predicted Physics Understanding score is 0.29 higher for every 1-unit difference in Inconsistent Stories. Conversely, Model B shows that controlling for Age, Executive Function is no longer a significant predictor of Physics Understanding. Finally, Model C shows that controlling for Age, Cognitive Reflection remains a significant predictor of

TABLE 3 Comparison of regression models predicting children's physics understanding.

Predictor	Model A	Model B	Model C	Model D	Model E	Model F
Intercept	**0.07**	**0.1**	**0.09**	**0.09**	**0.08**	**0.09**
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
	0.87	1.33	1.13	1.02	1	1.11
Age	**0.49**	**0.46**	**0.31**	**0.38**	**0.26**	**0.19**
	(0.09)	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)
	5.56***	4.34***	3.10**	3.56***	2.59*	1.75
Inconsistent stories	**0.29**			**0.26**	**0.22**	**0.21**
	(0.09) 3.20**			(0.09) 2.81**	(0.09) 2.61*	(0.09) 2.41*
Executive function		**0.21**		**0.19**		**0.13**
		(0.11)		(0.11)		(0.10)
		1.94		1.82		1.37
Cognitive reflection			**0.49**		**0.43**	**0.40**
			(0.10)		(0.10)	(0.11)
			4.82***		4.16***	3.83***
R^2	0.44	0.40	0.52	0.45	0.52	0.54
F	32.00	25.54	25.54	21.17	30.98	22.36
(df)	(2, 81)	(2, 78)	(2, 78)	(3, 77)	(3, 79)	(4, 76)
P	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

* < 0.05, ** < 0.01, *** < 0.001.
Cell entries are estimated regression coefficients in bold, (standard errors in parentheses), and t-statistics.

Physics Understanding. Taken together, these results show that once Age is controlled for, the variability in Executive Function is no longer predictive of the variability in Physics Understanding. Conversely, the relationship between Physics Understanding on the one hand and Inconsistent Stories and Cognitive Reflection on the other hand is significant even after statistically controlling for Age. There were no significant interactions between Cognitive Reflection and Age and between Inconsistent Stories and Age.

Next, we addressed the research question of whether Inconsistent Stories predicts Physics Understanding over and above Executive Functioning and Cognitive Reflection. Model D shows that after controlling for Age and Executive Function, Inconsistent Stories remains a significant predictor of Physics Understanding. Similarly, Model E shows that after controlling for Age and Cognitive Reflection, Inconsistent Stories remains a significant predictor of Physics Understanding. Finally, Model F shows that after controlling for Age, Executive Functioning, and Cognitive Reflection, Inconsistent Stories remains a significant predictor of Physics Understanding. Importantly, Model F shows that controlling for Age and Executive Function, both Cognitive Reflection and Inconsistent Stories predict unique variance in Physics Understanding over and above the other variables. That is, controlling for Age, Executive Function, and Cognitive Reflection, the predicted score of Physics Understanding is 0.21 higher for 1-unit difference in Inconsistent Stories. Similarly, controlling for Age, Executive Function, and Inconsistent Stories, the predicted score of Physics Understanding is 0.40 higher for every 1-unit difference in Cognitive Reflection. Figure 2 shows the fitted lines (i.e., based on Model F) of prototypical children who scored at the 25th percentile (bottom line) and the 75th percentile (top line) on CRT-D where the slopes of the lines represent the relationship between Physics Understanding and performance on Inconsistent Stories, controlling for Age and Executive Functioning.

Discussion

The present study investigated the relationship between explicit consistency monitoring skills and domain-specific knowledge. We found that even after controlling for Age, Executive Functioning, and Cognitive Reflection, children's performance on the Physics Interview was associated with Consistency Monitoring (i.e., with the Inconsistent Stories task). This result suggests that the consistency monitoring skill, as measured by the Inconsistent Stories task, is an independent predictor of children's performance on the Physics Interview. In other words, the ability to notice inconsistent statements in text is different from the suite of executive functioning abilities, as well as the ability for cognitive reflection, and it independently predicts variance on measures of physics understanding.

A novel aspect of the present study is the inclusion of the new individual differences measure, namely the Inconsistent Stories task, designed to measure children's consistency monitoring skill. To our knowledge, this is the first study to investigate the individual differences in *explicit* consistency monitoring and how those individual differences relate to other domain-general skills

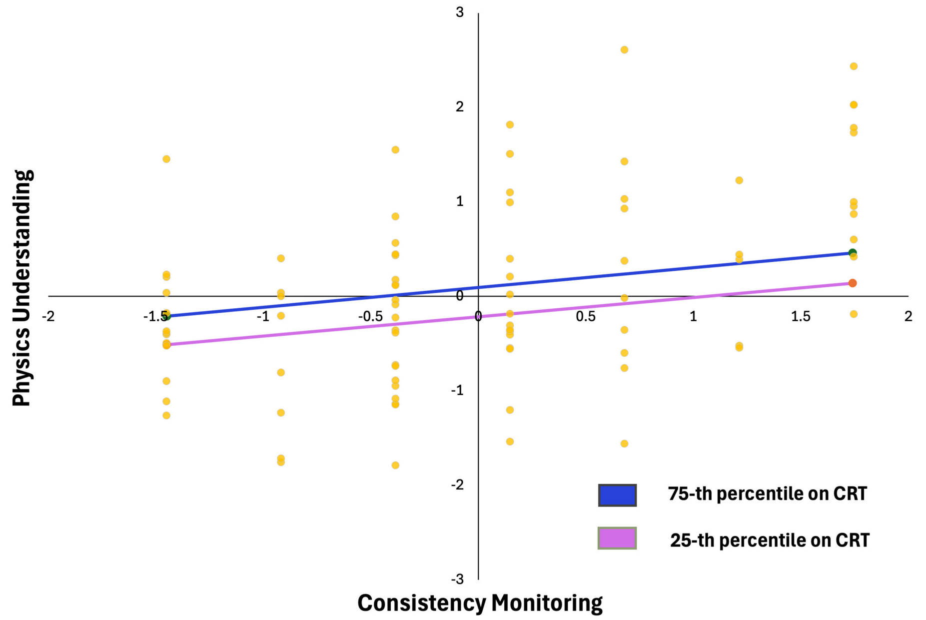

FIGURE 2
Consistency monitoring is positively correlated with Physics Understanding when controlling for Age, Executive Functioning, and Cognitive Reflection. Fitted lines of prototypical children who scored at the 25th percentile (bottom line) and the 75th percentile (top line) on Cognitive Reflection—Developmental. The slope of the line represents the relationship between Consistency Monitoring and Physics Understanding on the basis of Model F (see Table 3). The scatter plot represents the bivariate relationship between Consistency Monitoring and Physics Understanding.

such as executive functioning and cognitive reflection, as well as how consistency monitoring relates to the domain-specific physics understanding. As outlined in the Introduction, we reasoned that explicit consistency monitoring might be sufficiently independent from executive functioning and cognitive reflection and that it may be independently related to physics understanding. We found that although the bivariate correlations between Inconsistent Stories, Executive Functioning, and CRT-D were significant, they were not statistically significant after controlling for Age, suggesting that these measures tap into different types of abilities. Collectively, these findings provide a basis for advocating further exploration of the *explicit consistency monitoring* construct and for the expansion and improvement of the tasks that measure it.

In what ways is the consistency monitoring construct different from executive functioning and cognitive reflection, and why does it predict physics understanding over and above executive functioning and cognitive reflection? At face value, consistency monitoring, as measured by the Inconsistent Stories task, requires participants to encode the information, hold it in working memory, inhibit intuitive interpretations, draw relevant *long-range inferences* from that information (e.g., that not having light at the bottom of the ocean means that one cannot see the color of other animals), and then *compare the long-range inferences for consistency*. Executive functioning and cognitive reflection seem to tap into abilities that overlap to some extent. The Cognitive Reflection Task specifically, also contains a lure that elicits a fluent, first to mind kind of intuitive response that neither the executive functioning tasks nor the consistency monitoring task seem to have. In a similar vein, neither the executive functioning tasks nor the cognitive reflection task seem to tap into the individual's propensity to reason about the *consequences* of having certain beliefs (i.e., drawing *long-range inferences* that follow from those beliefs) and comparing them for consistency. Indeed, these processes are also important for scientific and domain-specific reasoning. For example, the consequence of having a belief that air is nothing and that it occupies no space, is that air cannot inflate balloons, it cannot fill up one's lungs, and that there is no difference between the "air" on Earth and in outer space. By noticing the inconsistencies that follow from holding the belief that air is nothing, one is in a better position to learn that air is something. Future research should test these possibilities more directly.

While acknowledging that explicitly noticing inconsistencies does not automatically lead to learning, nor does it mean that it automatically puts the learner in a better position to learn (e.g., Chi, 2013; Chinn and Brewer, 1993; Festinger, 1957), we list several reasons why noticing inconsistencies might sometimes help learners to engage in theory revision and theory construction. The first possibility is that noticing inconsistencies in one's understanding may play a motivational role. That is, if one notices a tension among one's beliefs, then that may motivate the process of seeking new explanations in the service of resolving the tension (e.g., Loewenstein, 1994). The quest for new explanations may include seeking information in the physical world (e.g., conducting new observations and experiments), in the social world (e.g., asking questions of knowledgeable others), as well as in one's mind (e.g., conducting thought experiments; see Bascandziev and Carey, 2022; Bascandziev and Harris, 2019; Bascandziev, 2022, 2024 for examples of how thought experiments can help learning). Another related possibility is that drawing long-range inferences, noticing inconsistencies, and attempting to resolve them may have cognitive benefits for the learner. By definition, engaging in such processes implies that the learner engages in deep processing of the material and making more connections among the relevant pieces of information, which is known to benefit memory and learning (Craik and Lockhart, 1972). Finally, noticing inconsistencies in one's understanding may pay dividends when one is encountering new explanations in informal or formal educational settings. For example, a child who has noticed an inconsistency between the belief that a grain of rice weighs nothing at all and that a pile of rice weighs something may find it easier to encode and assimilate the information that all material bodies, no matter how small they are, weigh something (Bascandziev and Carey, 2022).

The present study found an association between explicit consistency monitoring and domain-specific learning. The emphasis is on explicit, because the kind of consistency monitoring investigated in the present study, as measured by the Inconsistent Stories task, should be differentiated from many forms of implicit consistency monitoring. For example, implicit uncertainty could be measured by physiological indexes such as pupil dilation (Preuschoff et al., 2011), theta activation (Begus and Bonawitz, 2020, 2024); reaction times (Roebers et al., 2019), search behavior (Andreuccioli et al., 2024), or exploration measures (Lapidow et al., 2022; Wang et al., 2021). Moreover, studies have reported a link between implicit consistency monitoring measures and domain-specific learning. For example, several studies have reported a link between surprise, as measured by pupillometry, and domain-specific learning (Brod et al., 2018; Colantonio et al., 2023; Theobald and Brod, 2021; Theobald et al., 2024). However, the link between implicit and explicit measures of consistency monitoring is not clear. In other words, it is not clear whether the implicit forms of consistency monitoring give rise to an explicit (i.e., accessible to verbal report) consistency monitoring, and if not, then what additional steps are needed to attain an explicit representation of an inconsistency. Furthermore, it is not clear whether the reported association between implicit consistency monitoring and domain-specific learning and the association between explicit consistency monitoring and domain-specific learning are akin to each other. It is quite possible that different mechanisms underlie each association. In sum, the present study reports an association between explicit consistency monitoring and domain-specific knowledge. Future studies should explore the relationship between implicit and explicit measures of consistency monitoring, as well as how each type of consistency monitoring contributes to domain-specific learning.

In addition to finding that consistency monitoring is related to physics understanding, the present study also showed that cognitive reflection is related to physics understanding and failed to show any relationship between executive functioning and physics understanding (after controlling for age). Whereas, prior work has found a relationship between domain-general skills and the specific domains of biology (Bascandziev et al., 2018; Tardiff et al., 2020; Zaitchik et al., 2014), intuitive psychology (Carlson and Moses, 2001; Devine and Hughes, 2014), mathematics, and physics in adults (Bull and Lee, 2014; Colantonio et al., 2024; Thibault and Potvin, 2018), there have been no studies to our knowledge that have investigated young children's domain-general skills and their understanding of the physical world. This is important because each domain is different and the construction of knowledge within each domain may entail different domain-specific learning mechanisms (Wellman and Gelman, 1992), and by extension, it may also recruit different domain-general skills. Indeed, past research has shown that even learning different types of knowledge within a single domain (e.g., learning factual vs. conceptual knowledge), is associated with different types of domain-general skills (e.g., Bascandziev et al., 2018). As a case in point, the present study showed that young children's accumulated knowledge about the physical world is related both to their ability for cognitive reflection and their consistency monitoring, but not with their executive functioning (when controlling for Age). This suggests possible differences between this domain of physics and other domains (e.g., biology or mathematics) for which past studies have found strong associations with executive functioning. Future research could more systematically compare different domains and how conceptual learning in those domains relates to a wide range of domain-general skills, to better understand how different systems of knowledge might be built via different domain general supports.

One limitation of the present study is that it did not include a language measure. This is a limitation because both the Inconsistent Stories task and the Cognitive Reflection Task are language dependent, so it is possible that the effects observed in the present study are driven by children's language abilities rather than their consistency monitoring or cognitive reflection ability. We think that this possibility is unlikely. First, many studies that have investigated the relationship between executive functioning and domain-specific understanding have found that the effect of executive functioning continues to be significant even after controlling for language measures, suggesting that it is not the language comprehension component of the tasks that drives the effect (Tardiff et al., 2020; Carlson and Moses, 2001; Zaitchik et al., 2014). Moreover, one study in a different science domain (biological reasoning) found a double dissociation between language measures and executive functioning on the one end and domain-specific learning on the other. Whereas, executive functioning was predictive of improvement on domain-specific causal-explanatory learning, it was not predictive of factual learning. Conversely, whereas receptive vocabulary was predictive of factual learning, it was not predictive of domain-specific causal explanatory learning (Bascandziev et al., 2018). Taken together, these findings suggest that the role of the domain-general cognitive abilities such as executive functioning in domain-specific learning goes beyond language abilities. Future research should test the prediction that the roles of consistency monitoring and cognitive reflection in domain-specific learning is also disassociated from language abilities.

In conclusion, the present study investigated the relationship between young children's progress in the domain of physics and several domain-general predictors, including consistency monitoring, executive functioning, and cognitive reflection. We found a relationship between explicit consistency monitoring and physics understanding when controlling for age, executive functioning, and cognitive reflection. This finding highlights the importance of a domain-general skill implicated in the accumulation and expression of domain-specific understanding, and it points to important avenues for future research and educational interventions.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving humans were approved by Harvard University Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

IB: Writing – original draft, Writing – review & editing. AA: Writing – review & editing. CW: Writing – review & editing. EB: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Caplan Foundation for Early Childhood to IB, EB, and CW and by the National Science Foundation (Division of Research on Learning; 2301180) to EB, IB, CW, and Patrick Shafto.

Acknowledgments

We are thankful to the elementary schools in the Westwood Public Schools District in Westwood, MA, the Cambridge Montessori School in Cambridge, MA, and the Atrium School in Watertown, MA. Special thanks to the families and children who participated in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Andreuccioli, L., Mazor, S., Begus, K., Denison, S., Bonawitz, E., and Walker, C. M. (2024). "Young children adapt their search behavior for necessary versus merely possible outcomes," in *Proceedings of the 46th Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society).

Bascandziev, I. (2022). "Representational pluralism in the service of learning: the case of thought experiments," in *Multidisciplinary Perspectives on Representational Pluralism in Human Cognition*, eds. M. Bélanger, P. Potvin, S. Horst, S. Shtulman, and E. Mortimer (New York, NY: Routledge).

Bascandziev, I. (2024). Thought experiments as an error detection and correction tool. *Cogn. Sci.* 48:e13401. doi: 10.1111/cogs.13401

Bascandziev, I., and Carey, S. (2022). "Young children learn equally from real and thought experiments," in *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, eds. J. Culbertson, A. Perfors, H., Rabagliati, and V. Ramenzoni (Toronto, OA: Cognitive Science Society).

Bascandziev, I., and Harris, P. L. (2019). "Can children benefit from thought experiments?," in *The Scientific Imagination*, eds. P. Godfrey-Smith and A. Levy (New York, NY: Oxford University Press).

Bascandziev, I., Tardiff, N., Zaitchik, D., and Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cogn. Psychol.* 104, 1–28. doi: 10.1016/j.cogpsych.2018.03.002

Begus, K., and Bonawitz, E. (2020). The rhythm of learning: theta oscillations as an index of active learning in infancy. *Dev. Cogn. Neurosci.* 45:100810. doi: 10.1016/j.dcn.2020.100810

Begus, K., and Bonawitz, E. (2024). Infants evaluate informativeness of evidence and predict causal events as revealed in theta oscillations and predictive looking. *Commun. Psychol.* 2:77. doi: 10.1038/s44271-024-00131-3

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

Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* 8, 539–546. doi: 10.1016/j.tics.2004.10.003

Brod, G., Hasselhorn, M., and Bunge, S. A. (2018). When generating a prediction boosts learning: the element of surprise. *Learn. Instr.* 55, 22–31. doi: 10.1016/j.learninstruc.2018.01.013

Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059

Carey, S. (1991). "Knowledge acquisition: Enrichment or conceptual change?," in *The Epigenesis of Mind: Essays on Biology and Cognition*, eds. S. Carey and R. Gelman (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 257–291.

Carey, S. (1995). "On the origins of causal understanding," in *Causal Cognition*, eds. D. Sperber, D. Premack, and A. J. Premack (Oxford: Clarendon Press), 268–308.

Carey, S. (2000). Science education as conceptual change. *J. Appl. Dev. Psychol.* 21, 13–19. doi: 10.1016/S0193-3973(99)00046-5

Carey, S. (2009). *The Origin of Concepts.* New York, NY: Oxford University Press.

Carey, S., Zaitchik, D., and Bascandziev, I. (2015). Theories of development: in dialog with jean piaget. *Dev. Rev.* 38, 36–54. doi: 10.1016/j.dr.2015.07.003

Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72, 1032–1053. doi: 10.1111/1467-8624.00333

Checa, P., Castellanos, M. C., Abundis-Gutiérrez, A., and Rosario Rueda, M. (2014). Development of neural mechanisms of conflict and error processing during childhood: implications for self-regulation. *Front. Psychol.* 5:326. doi: 10.3389/fpsyg.2014.00326

Chi, M. T. (2013). "Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes," in *International Handbook of Research on Conceptual Change, 2nd Edn.*, ed. S. Vosniadou (New York, NY: Psychology Press), 49–70.

Chinn, C. A., and Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Rev. Educ. Res.* 63, 1–49. doi: 10.3102/00346543063001001

Colantonio, J., Bascandziev, I., Theobald, M., Brod, G., and Bonawitz, E. (2022). Priors, progressions, and predictions: theory-based Bayesian models of children's revising beliefs of water displacement. *IEEE Trans. Cogn. Dev. Syst.* 15, 1487–1500. doi: 10.1109/TCDS.2022.3220963

Colantonio, J., Bascandziev, I., Theobald, M., Brod, G., and Bonawitz, E. (2023). Seeing the error in my "bayes": a quantified degree of belief change correlates with children's pupillary surprise responses following explicit predictions. *Entropy* 25:211. doi: 10.3390/e25020211

Colantonio, J. A., Bascandziev, I., Theobald, M., Brod, G., and Bonawitz, E. (2024). Predicting learning: understanding the role of executive functions in children's belief revision using bayesian models. *Top. Cogn. Sci.* doi: 10.1111/tops.12749

Craik, F. I., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learning Verb. Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Devine, R. T., and Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Dev.* 85, 1777–1794. doi: 10.1111/cdev.12237

DeVries, R. (1986). Children's conceptions of shadow phenomena. *Genet. Soc. Gen. Psychol. Monogr.* 112, 479–530.

diSessa, A. A. (1988). "Knowledge in Pieces," in *Constructivism in the Computer Age*, eds G. Forman and P. B. Pufall (Hillsdale, NJ: Erlbaum), 49–70.

Don, H. J., Goldwater, M. B., Otto, A. R., and Livesey, E. J. (2016). Rule abstraction, model-based choice, and cognitive reflection. *Psychon. Bull. Rev.* 23, 1615–1623. doi: 10.3758/s13423-016-1012-y

Dreyfus, A., Jungwirth, E., and Eliovitch, R. (1990). Applying the "cognitive conflict" strategy for conceptual change: some implications, difficulties, and problems. *Sci. Educ.* 74, 555–569. doi: 10.1002/sce.3730740506

Fernandez-Duque, D., Baird, J. A., and Posner, M. I. (2000). Executive attention and metacognitive regulation. *Conscious. Cogn.* 9, 288–307. doi: 10.1006/ccog.2000.0447

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fleur, D. S., Bredeweg, B., and van den Bos, W. (2021). Metacognition: ideas and insights from neuro- and educational sciences. *NPJ Sci. Learn.* 6:13. doi: 10.1038/s41539-021-00089-5

Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732

Friedman, S. E., and Forbus, K. D. (2011). "Repairing incorrect knowledge with model formulation and metareasoning," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence* (Barcelona: AAAI Press). doi: 10.5591/978-1-57735-516-8/IJCAI11-154

Friedman, S. E., Forbus, K. D., and Sherin, B. (2018). Representing, running, and revising mental models: a computational model. *Cogn. Sci.* 42, 1110–1145. doi: 10.1111/cogs.12574

Gerstadt, C. L., Hong, Y. J., and Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition* 53, 129–153 doi: 10.1016/0010-0277(94)90068-X

Gervais, W. M. (2015). Override the controversy: analytic thinking predicts endorsement of evolution. *Cognition* 142, 312–321. doi: 10.1016/j.cognition.2015.05.011

Goldberg, R. F., and Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychol. Sci.* 20, 480–487 doi: 10.1111/j.1467-9280.2009.02320.x

Hofmann, W., Schmeichel, B. J., and Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends Cogn. Sci.* 16, 174–180. doi: 10.1016/j.tics.2012.01.006

Jammer, M. (1961). *The Concepts of Mass in Classical and Modern Physics*. Cambridge, MA: Harvard University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychol. Rev.* 96:674. doi: 10.1037/0033-295X.96.4.674

Lapidow, E., Killeen, I., and Walker, C. M. (2022). Learning to recognize uncertainty vs. recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Dev. Sci.* 25:e13178doi: 10.1111/desc.13178

Legare, C. H., Evans, E. M., Rosengren, K. S., and Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Dev.* 83, 779–793. doi: 10.1111/j.1467-8624.2012.01743.x

Legare, C. H., Gelman, S. A., and Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Dev.* 81, 929–944. doi: 10.1111/j.1467-8624.2010.01443.x

Legare, C. H., and Shtulman, A. (2018). "Explanatory pluralism across cultures and development," in *Interdisciplinary Approaches to Metacognitive Diversity,* eds. J. Proust and M. Fortier (Oxford: Oxford University Press). doi: 10.1093/oso/9780198789710.003.0019

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learn. Instr.* 11, 357–380. doi: 10.1016/S0959-4752(00)00037-2

Limón, M., and Carretero, M. (1997). Conceptual change and anomalous data: a case study in the domain of natural sciences. *Eur. J. Psychol. Educ.* 12, 213–230. doi: 10.1007/BF03173085

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* 116:75. doi: 10.1037/0033-2909.116.1.75

Markman, E. M. (1977). Realizing that you don't understand: a preliminary investigation. *Child Dev.* 48, 986–992. doi: 10.2307/1128350

Markman, E. M. (1979). Realizing that you don't understand: elementary school children's awareness of inconsistencies. *Child Dev.* 50, 643–655. doi: 10.2307/1128929

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Munakata, Y., Snyder, H. R., and Chatham, C. H. (2012). Developing cognitive control: three key transitions. *Curr. Dir. Psychol. Sci.* 21, 71–77. doi: 10.1177/0963721412436807

Neuenschwander, R., Röthlisberger, M., Cimeli, P., and Roebers, C. M. (2012). How do different aspects of self-regulation predict successful adaptation to school? *J. Exp. Child Psychol.* 113, 353–371. doi: 10.1016/j.jecp,0.2012.07.004

NGSS Lead States (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Pennycook, G., and Rand, D. G. (2019). Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188, 39–50. doi: 10.1016/j.cognition.2018.06.011

Piaget, J. (1960). *The Child's Conception of the World*. Totowa, NJ: Littlefield.

Piaget, J., and Inhelder, B. (1974). *The Child's Construction of Quantities: Conservation and Atomism*. London: Routledge and Kegan Paul.

Posner, G., Strike, K., Hewson, P. W., and Gertzog, W. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Sci. Educ.* 66, 211–227. doi: 10.1002/sce.3730660207

Potvin, P. (2023). Response of science learners to contradicting information: a review of research. *Stud. Sci. Educ.* 59, 67–108. doi: 10.1080/03057267.2021.2004006

Preuschoff, K., 't Hart, B. M., and Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N. S., and van Loon, M. H. (2019). The role of children's metacognitive experiences for cue utilization and monitoring accuracy: a longitudinal study. *Dev. Psychol.* 55, 2077–2089. doi: 10.1037/dev0000 776

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., and Jenkins, A. C. (2008). Going beyond the evidence: abstract laws and preschoolers' responses to anomalous data. *Cognition* 109, 211–223. doi: 10.1016/j.cognition.2008.07.017

Shtulman, A., Harrington, C., Hetzel, C., Kim, J. Palumbo, C., and Rountree-Shtulman, T. (2023). Could It? Should It? Cognitive reflection facilitates children's reasoning about possibility and permissibility. *J. Exp. Child Psychol.* 235:105727.

Shtulman, A., and Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Top. Cogn. Sci.* 8, 118–137. doi: 10.1111/tops.12174

Shtulman, A., and Legare, C. H. (2020). Competing explanations of competing explanations: accounting for conflict between scientific and folk explanations. *Top. Cogn. Sci.* 12, 1337–1362. doi: 10.1111/tops.12483

Shtulman, A., and Lombrozo, T. (2016). "Bundles of contradiction: a coexistence view of conceptual change," in *Core Knowledge and Conceptual Change,* eds. D. Barner and A. S. Baron (Oxford University Press), 49–67.

Shtulman, A., and McCallum, K. (2014). "Cognitive reflection predicts science understanding," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society,* eds. P. Bello, M. Guarini, M. McShane, and B. Scassellati (Austin, TX: Cognitive Science Society), 2937–2942.

Shtulman, A., and Young, A. G. (2023). The development of cognitive reflection. *Child Dev. Perspect.* 17, 59–66. doi: 10.1111/cdep.12476

Smith, C., Carey, S., and Wiser, M. (1985). On differentiation: a case study of the development of the concepts of size, weight, and density. *Cognition* 21, 177–237. doi: 10.1016/0010-0277(85)90025-3

Smith, C., Snir, J., and Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: the case of weight-density differentiation. *Cogn. Instr.* 9, 221–283. doi: 10.1207/s1532690xci0903_3

Smith, C., and Unger, C. (1997). What's in dots-per-box? conceptual bootstrapping with stripped-down visual analogs. *J. Learn. Sci.* 6, 143–181. doi: 10.1207/s15327809jls0602_1

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories: using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cogn. Instr.* 25, 337–398. doi: 10.1080/07370000701632363

Smith, C. L., Maclin, D., Grosslight, L., and Davis, H. (1997). Teaching for understanding: a comparison of two approaches to teaching students about matter and density. *Cogn. Instr.* 15, 317–393. doi: 10.1207/s1532690xci1503_2

Snir, J., Smith, C., and Grosslight, L. (1993). Conceptually enhanced simulations: a computer tool for science teaching. *J Sci. Educ. Technol.* 2, 373–388. doi: 10.1007/BF00694526

Snyder, H. R., and Munakata, Y. (2010). Becoming self-directed: abstract representations support endogenous flexibility in children. *Cognition* 116, 155–167. doi: 10.1016/j.cognition.2010.04.007

Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychol. Rev.* 99:605. doi: 10.1037/0033-295X.99.4.605

Stahl, A. E., and Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science* 348, 91–94. doi: 10.1126/science.aaa3799

Stanovich, K. E., West, R. F., and Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. MIT Press.

Tardiff, N., Bascandziev, I., Carey, S., and Zaitchik, D. (2020). Specifying the domain-general resources that contribute to conceptual construction: evidence from the child's acquisition of vitalist biology. *Cognition* 195:104090. doi: 10.1016/j.cognition.2019.104090

Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., and Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: a case study of vitalist biology. *Cogn. Psychol.* 95, 145–163. doi: 10.1016/j.cogpsych.2017.04.004

TERC (2011). *The Inquiry Project: Seeing the World through a Scientist's Eye*. Available at: https://inquiryproject.terc.edu/curriculum/index.html (accessed August 1, 2024).

Theobald, M., and Brod, G. (2021). Tackling scientific misconceptions: the element of surprise. *Child Dev.* 92, 2128–2141. doi: 10.1111/cdev.13582

Theobald, M., Colantonio, J., Bascandziev, I., Bonawitz, E., and Brod, G. (2024). Do reflection prompts promote children's conflict monitoring and revision of misconceptions? *Child Dev.* 95, e253–e269. doi: 10.1111/cdev.14081

Thibault, F., and Potvin, P. (2018). Executive function as a predictor of physics-related conceptual change. *Neuroeducation* 5, 119–126. doi: 10.24046/neuroed.20180502.119

Troyer, A. K., Moscovitch, M., and Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11, 138–146. doi: 10.1037/0894-4105.11.1.138

Wang, J., Yang, Y., Macias, C., and Bonawitz, E. (2021). Children with more uncertainty in their intuitive theories seek domain-relevant information. *Psychol. Sci.* 32, 1147–1156. doi: 10.1177/0956797621994230

Wellman, H. M., and Gelman, S. A. (1992). Cognitive development: foundational theories of core domains. *Annu. Rev. Psychol.* 43, 337–375. doi: 10.1146/annurev.ps.43.020192.002005

Young, A. G., and Shtulman, A. (2020a). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychol. Sci.* 31, 1396–1408. doi: 10.1177/0956797620954449

Young, A. G., and Shtulman, A. (2020b). How children's cognitive reflection shapes their science understanding. *Front Psychol.* 11:1247. doi: 10.3389/fpsyg.2020.01247

Zaitchik, D., Iqbal, Y., and Carey, S. (2014). The effect of executive function on biological reasoning in young children: an individual differences study. *Child Dev.* 85, 160–175. doi: 10.1111/cdev.12145