Brod, Garvin; Breitwieser, Jasmin; Hasselhorn, Marcus; Bunge, Silvia

Being proven wrong elicits learning in children - but only in those with higher executive function skills

Developmental science 23 (2020) 3, 14 S., 10.1111/desc.12916

Received: 8 June 2018 | Revised: 1 September 2019 | Accepted: 11 October 2019

DOI: 10.1111/desc.12916

PAPER

Developmental Science WILEY

Being proven wrong elicits learning in children – but only in those with higher executive function skills

Garvin Brod[1,2,3] | Jasmin Breitwieser[1,2] | Marcus Hasselhorn[1,2,3] | Silvia A. Bunge[4]

[1]Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

[2]DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany

[3]Department of Psychology, Goethe University, Frankfurt, Germany

[4]Department of Psychology & Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA

Correspondence
Garvin Brod, German Institute for International Educational Research, Frankfurt am Main, Germany.
Email: garvin.brod@dipf.de

Funding information
Jacobs Foundation

Abstract

This study investigated whether prompting children to generate predictions about an outcome facilitates activation of prior knowledge and improves belief revision. 51 children aged 9–12 were tested on two experimental tasks in which generating a prediction was compared to closely matched control conditions, as well as on a test of executive functions (EF). In Experiment 1, we showed that children exhibited a pupillary surprise response to events that they had predicted incorrectly, hypothesized to reflect the transient release of noradrenaline in response to cognitive conflict. However, children's surprise response was not associated with better belief revision, in contrast to a previous study involving adults. Experiment 2 revealed that, while generating predictions helped children activate their prior knowledge, only those with better inhibitory control skills learned from incorrectly predicted outcomes. Together, these results suggest that good inhibitory control skills are needed for learning through cognitive conflict. Thus, generating predictions benefits learning – but only among children with sufficient EF capacities to harness surprise for revising their beliefs.

KEYWORDS

belief revision, cognitive conflict, executive functions, surprise, violation of expectation

1 | INTRODUCTION

A well-known assumption of constructivist theories of learning is that new content has to be connected to prior knowledge in order to promote meaningful learning. Indeed, a plethora of research has found that activating prior knowledge in learners strongly improves their comprehension and memory for the novel content (e.g., Alexander, 1996; Bransford & Johnson, 1972). Developmental research has likewise found this statement to be generally true in children of all ages (Bjorklund, 1987; Chi, 1978; Stahl & Feigenson, 2015; for an overview see Brod, Werkle-Bergner, & Shing, 2013). However, the ability to make deliberate and strategic use of one's prior knowledge also follows a developmental trajectory that does not reach its peak before late adolescence (Bjorklund, Muir-Broaddus, & Schneider, 1990; Brod, Lindenberger, & Shing, 2017; Hasselhorn, 1990). It is thus imperative to better understand how children can best activate their prior knowledge and connect it to new content.

One instructional strategy that has been widely adopted is the induction of cognitive conflict, wherein learners are confronted with inconsistencies between new information and their prior concepts (Carey, 1985; Piaget, 1985; Posner, Strike, Hewson, & Gertzog, 1982). Ideally, the learners notice the conflict, experience a disequilibrium, and try to resolve the conflict by changing their existing concepts. One strategy that has been proposed to achieve this is to let learners generate predictions before presenting them with the correct information (Liew & Treagust, 1995; White & Gunstone, 1992).

We hypothesize that prediction generation is a promising instructional technique because it invokes the activation of prior knowledge, and, when a prediction is falsified, elicits cognitive conflict. Indeed, it has recently been demonstrated in university students that a specific effect of making predictions is that incorrectly predicted outcomes induce surprise (Brod, Hasselhorn, & Bunge, 2018), which is the initial reaction to a perceived conflict between new information and activated knowledge (Mandler, 1990). Learners' surprise about a particular outcome that refuted their beliefs was positively related to the correction of this (false) belief – that is, successful belief revision (Brod et al., 2018).

Is the elicitation of surprise also beneficial for belief revision in children? On the one hand, children's inefficiency in making deliberate and strategic use of their prior knowledge suggests that any form of instruction that invokes the activation of prior knowledge in children should yield strong beneficial effects. On the other hand, to achieve belief revision, the induced conflict also needs to be resolved.

A key hypothesis of the current study was that good executive functioning would be needed for this final step, particularly when the prior belief was strong. Executive functions (EFs) encompass a set of domain-general abilities including updating and monitoring of information, inhibition of prepotent responses, and switching between tasks or mental sets (Diamond, 2013; Miyake et al., 2000; but see Miyake & Friedman, 2012). The slow developmental trajectory of EFs (Diamond, 2013; Luna, Garver, Urban, Lazar, & Sweeney, 2004) has important consequences for children's ability to acquire complex knowledge structures. With increasing EF capacity, children can handle rules of increasing complexity and switch between them if the context suggests so (Marcovitch & Zelazo, 2009; Zelazo, 2004). Recent research has demonstrated that children's EF abilities are correlated with their construction of a vitalist theory of biology (Bascandziev, Tardiff, Zaitchik, & Carey, 2018; Zaitchik, Iqbal, & Carey, 2014). Furthermore, impairments in EFs interfere with both the construction and the expression of conceptual knowledge (Gropen, Clark-Chiarelli, Hoisington, & Ehrlich, 2011; Johnson & Carey, 1998; Zaitchik & Solomon, 2008). In sum, there is emerging evidence suggesting that EFs are important for learning of complex material.

But why should EFs matter for learning from cognitive conflict? For belief revision to take place, learners must first detect the conflict between their prior knowledge and the new information and then update their existing beliefs. A large body of cognitive neuroscience research suggests that error/conflict monitoring and behavior updating are separable processes. Specifically, a conflict or error signal is generated by one brain system and transmitted to another system that mediates the updating of thought and behavior (Botvinick, Carter, Braver, Barch, & Cohen, 2001; Carter & Van Veen, 2007; Kerns et al., 2004). Critically, these brain systems, and the interaction between them, develop throughout childhood (Supekar et al., 2010; Waxer & Morton, 2011; Zielinski, Gennatas, Zhou, & Seeley, 2010), and there is evidence that these developmental changes underlie increases in behavioral updating (Crone, Zanolie, Van Leijenhorst, Westenberg, & Rombouts, 2008; Fiske & Holmboe, 2019; Morton & Munakata, 2009). Thus, we posit that belief revision requires strong EF skills. Since EF development is slow and highly variable, this could mean that children – especially those with lower EF – struggle to learn through cognitive conflict.

Highlights

- Violated predictions are considered key drivers for learning, but it is unclear whether children can harness the resulting surprise signal for revising incorrect beliefs.
- The current study tested this assumption in children aged 9–11 using a prediction–feedback learning task along with recordings of pupillary responses.
- Children exhibited a strong and specific pupillary surprise response to incorrectly predicted events, but only some of the children revised their initial beliefs.
- We found that children's ability to learn from incorrectly predicted events was related to their inhibitory control capacity.

We tested these ideas in a sample of 9- to 12-years-olds who performed an EF task as well as two experimental memory tasks: a geography knowledge task and an episodic memory task. Both memory tasks were designed to assess belief revision in a simplistic, well-controlled, and repeatable scenario. The geography knowledge task has recently been established in adults as an effective paradigm for measuring the pupillary surprise reaction (Brod et al., 2018). The pupillary surprise reaction is an objective index of the release of norepinephrine in the brainstem's locus coeruleus, which leads to a short burst of arousal that is reflected in a transient dilation of the pupil (see Kloosterman et al., 2015; Preuschoff, 't Hart, & Einhäuser, 2011; for a general overview of pupillometry in the study of cognition, see Laeng, Sirois, & Gredebäck, 2012; Sirois & Brisson, 2014). Brod and colleagues (2018) found that actively generating a prediction enhances the pupillary surprise reaction to expectancy-violating events. This finding was taken to reflect enhanced cognitive conflict as a result of explicitly committing to an (incorrect) outcome.

The present study investigates whether children show a pupillary surprise response to expectancy-violating events, and whether the extent of subsequent belief revision is related to individual differences in EFs. Late childhood was chosen because it encompasses a time in which basic EF skills can be expected to be in place (see Chatham, Frank, & Munakata, 2009), but with substantial individual differences among children. The EF task enabled us to assess the importance of two central EF components – inhibition and switching – for learning from cognitive conflict.

Memory performance in the two tasks was compared between two within-subject conditions: a *prediction* condition, as well as a well-matched post hoc evaluation condition henceforth called the *postdiction* condition. For the episodic memory task, these conditions, which required activation of prior knowledge, were contrasted with a *baseline* condition that did not, thereby enabling us to test the general effect of prior knowledge activation on children's memory performance. Furthermore, the episodic memory task enabled us to directly examine memory performance for conflict-inducing (i.e. incorrectly predicted) events. If participants' average memory performance for these events correlated with one or both EF components, this finding would be the first demonstration of a relation between EFs and learning through the induction of cognitive conflict.

2 | METHODS

2.1 | Participants

Fifty-one children (mean age 9.9 years; 24 female, 27 male) who were German native speakers participated in the study. All but one of these children were 9–11 years old; the remaining participant was 12 years old. Children attended grades 4–6 of Frankfurt metropolitan area public schools. Children were recruited through bulletins and flyers that were handed out at sports clubs and via email lists of parents. The sole inclusion criterion, which was stressed in the advertisements, was that the children had to have at least some interest in soccer. This approach to recruitment yielded a strong gender imbalance (27 boys vs. 2 girls). Therefore, in response to manuscript reviews, we tested an additional 22 girls in a second wave of data collection performed about 1 year later.

Children and their parents gave written informed consent prior to testing. The two experimental tasks took place in two separate 60-min sessions on different days. Children were given small gifts at the end of each session. Parents were paid €5 per session to cover their travel costs. Ethics approval was obtained from the ethics committee of the DIPF | Leibniz Institute for Research and Information in Education.

2.2 | Study design & testing procedures

2.2.1 | Overview

Three computerized experimental tasks were performed in two separate sessions of about 60 min each. The EF task (ca. 10 min) and the geography task (ca. 40 min) were performed during the first session, in that order, and the soccer task (ca. 50 min) during the second session. One child did not perform the first session due to a computer malfunction, and therefore did not provide data for the EF task or the geography task.

The geography and soccer tasks were similar in structure. Both tasks included an initial knowledge assessment, a computerized eye-tracking task (the study phase), and a final assessment of knowledge or memory. They also both included the following two conditions: a prediction condition in which participants were prompted to make a prediction about the outcome before the answer was revealed, and a postdiction condition in which the outcome was presented first, after which the participants were asked to judge which outcome they would have predicted. Thus, prior knowledge had to be activated either before or after seeing the actual outcome, depending on the condition. The two conditions were performed within-subjects in separate blocks, and differed only in the presentation order of the stimuli; participants had to state their expectations either before or after seeing the actual outcome (see Figure 1).

2.3 | Geography task

2.3.1 | Design

Children were asked, on each of a series of trials, to consider which of two countries had a larger population. The dependent measure was the change in hierarchy knowledge of the population size of European countries. Changes in knowledge were assessed by asking participants to rank order a list of 12 countries by their number of inhabitants, both before and after completion of a block of trials (Figure 1; see Procedures section for additional details). This procedure of pre-test, study phase, and post-test for a list of countries was carried out first for a block of one condition (prediction or postdiction) and then – using a new list of countries – for a block of the other condition. Assignment of the two lists of countries to condition, as well as the ordering of the blocks, were counterbalanced across participants. This design enabled us to compare the improvement in hierarchy knowledge between the prediction and postdiction condition.

2.3.2 | Testing procedures

The testing session started and ended with the knowledge test, in which the participants were asked to rank order two decks of 12 European countries (represented by the flag and the name below) by number of inhabitants, starting with the country that they thought had the most inhabitants (for details regarding the stimulus lists, see Brod et al., 2018). After participants were finished sorting the first deck, the cards were removed and participants repeated the procedure with the second deck of cards. Before the computerized task blocks were administered, participants were given time to familiarize themselves with the flags, and they were made aware of the fact that no country names would be shown in the task. Participants were told that during the following study phase they would see a pair of flags from the list of 12 countries' flags, along with the correct population sizes. They were not told to memorize those numbers, but were instead informed that they would be asked to sort the cards again after the computerized study phase was finished.

Each of the two blocks started with four practice trials to familiarize participants with the task (prediction or postdiction). Next, participants saw 40 unique pairs of countries (see Figure 1 for an overview of the trial sequence and timing). To facilitate learning of the hierarchy, participants saw only six of the 12 countries during

GEOGRAPHY TASK
Pretest
1 2 3 4 5 6 7 8 9 10 11 12
By name
B DK FIN GB IRL L LV RO RS E CZ H
By inhabitants
GB E RO B CZ H LV RS DK FIN IRL L
By participant
GB H IRL CZ RO FIN B LV E DK RS L
Study Phase
Prediction condition
Baseline Phase
Response Phase
Pupil Baseline
Results Phase
x + x
? + ?
x + x
5.7 + 11.3
2 s
4.25 s
.75 s
3.5 s
or
Postdiction condition
Baseline Phase
Pupil Baseline
Results Phase
Response Phase
x + x
x + x
5.7 + 11.3
? + ?
2 s
.75 s
3.5 s
4.25 s
Posttest
1 2 3 4 5 6 7 8 9 10 11 12
By inhabitants
GB E RO B CZ H LV RS DK FIN IRL L
By participant
GB E RO B DK CZ H LV RS IRL FIN L
SOCCER TASK
Study Phase
Prediction condition
Baseline Phase
Response Phase
Pupil Baseline
Results Phase
x + x
? + ?
x + x
2 + 0
2 s
4.25 s
.5 s
3.5 s
or
Postdiction condition
Baseline Phase
Pupil Baseline
Results Phase
Response Phase
x + x
x + x
2 + 0
? + ?
2 s
.5 s
3.5 s
4.25 s
Baseline condition
Baseline Phase
Results Phase
x + x
2 + 0
2 s
8.25 s
Test Phase
+
x + x
.5 s
Response dependent

FIGURE 1 Schematic overview of the task design. Geography task design: In the pre- and posttest, participants were presented with flags of European countries (ordered alphabetically, upper row) and had to try to rank order them by number of inhabitants (middle row: correct rank order; lower row: exemplary rank order by one participant). In between pre- and posttest, they underwent the computerized study phase, in which they performed one of two conditions (counterbalanced). In the prediction condition, participants had to make a prediction first and then saw the correct population sizes (in millions), whereas in the postdiction condition, they first saw the population sizes and then had to make a post hoc statement regarding which results they would have predicted. Participants were only able to respond when the question marks appeared on the screen, using the same five-point scale for both conditions (far left: clearly the left country, left: probably the left country, middle: don't know, right: probably the right country, far right: clearly the right country). For illustrative purposes, the background is shown in white and the print in black. For the real experiment, the background was gray and the print was white, so as to reduce luminance contrasts. Flags © dikobrazik/Fotolia. Soccer task design: The study phase of the soccer task was highly similar in structure to the one of the geography task. However, the soccer task contained an additional baseline condition in which they were instructed to memorize the outcome and to indicate which of the two logos is bigger (cover task). In addition, for the prediction and postdiction conditions, the labels of the five-point scale were adapted to the scores: Far left: >1 goal difference victory for the left team, left: 1 goal victory for the left team, middle: draw, right: 1 goal victory for the right team, far right: >1 goal victory for the right team. For the baseline condition, the children used a simplified three-point scale: Left: left logo is bigger, middle: same size, right: right logo is bigger. During the subsequent test phase, all club pairs were presented again. Here the children had to recall the correct outcome of the match as presented during the study phase, again using the five-point scale

the first half of each block; only during the second half of each block did they see all of the countries. In the prediction block, participants were instructed to predict which country of each pair had the greater population size, and to do so while the question marks appeared on the screen (i.e. 'Response Phase', see Figure 1). Participants used a button box to indicate on a five-point scale whether they thought the answer was clearly or probably the country on the left (buttons 1–2), whether they did not know (button 3), or whether they thought it was probably or clearly the country on the right (buttons 4–5). The same scale was used in the postdiction condition, in which participants were instructed to make a post hoc evaluation (What would you have expected?). After each block was completed, participants sorted the 12 countries again, following the same instructions as during the first knowledge test.

Upon completion of the final block, participants were given a brief questionnaire in which they had to indicate on a scale from 1–6 (1 = clearly prediction, 6 = clearly postdiction) which of the two conditions they thought was more fun, and in which they thought they had learned more.

2.4 | Soccer task

2.4.1 | Design

In the soccer task, children were asked to memorize the result of a soccer match between two teams of Germany's first division (see Figure 1 for a graphical depiction of the study phase). The general structure of the soccer task was similar to the geography task, although slight variations in the task design were necessary based on the demands of each task. In addition to the prediction and postdiction conditions, in which participants had to predict or postdict the outcome of a soccer match, the task additionally contained a baseline condition. In this condition, children were asked to decide which of the two soccer team logos was bigger. This task was intended as a cover task that did not involve semantic elaboration but ensured that children paid attention to the screen. The order of the conditions, and the assignment of matches to conditions, were counterbalanced across participants. There were two blocks per condition, which were performed successively. Each block consisted of a study phase followed by a test phase. Participants saw 25 unique pairs of soccer teams in each block's study phase. In the test phase, participants saw all 25 pairings again and had to state the actual results of the match.

Match results were taken from real matches that took place during the 2014/15 season, 2–3 years prior to the testing session. We reasoned that taking real results and telling children that the results were real should enhance the relevance of participants' prior knowledge (as noted previously, recruitment materials indicated that participants should be at least somewhat interested in soccer). However, since two teams always play twice against each other in the course of a season and the matches dated back two seasons, even children with high soccer knowledge could not know the actual results beforehand. This assumption was confirmed with a questionnaire administered after the experiment.

2.4.2 | Testing procedures

First, children were instructed to rank order the 18 teams of the 2014–15 season of Germany's premier soccer division by their final standing. This task served as a baseline measure of prior knowledge, and ensured familiarity with the stimulus material. Children were then given time to familiarize themselves with the club logos, and were told that they would now see real results of matches from the 2014–15 season, which they should memorize for a subsequent memory test. No details were given regarding the specifics of the later memory test.

During the study phase, children were instructed to predict or make a post hoc evaluation regarding the likely outcome of the match, or, in the baseline condition, to indicate which of the two club logos was bigger. Participants again had to respond on a five-point scale using a button box (see Figure 1). The study phase of each condition started with four practice trials. For the test phase, which followed shortly thereafter, participants were told that they would now see all match pairs again and that they should try to recall the actual result of the match, using the same five-point scale as during the study phase.

2.5 | EF task: Hearts and flowers task

2.5.1 | Design

As our EF task, we used the so-called Hearts &Flowers Task (HFT) developed by Diamond and colleagues (Diamond, Barnett, Thomas, & Munro, 2007; Wright & Diamond, 2014). It includes three blocks, congruent, incongruent, and mixed, all of which require sustained attention and maintenance of task rules in working memory. The incongruent and mixed blocks additionally require inhibitory control, as it is necessary to override a prepotent response tendency, and the mixed block additionally requires switching between mental sets, as it is necessary to switch between rules. We also collected pilot data of 16 children on a card sorting test; these data are not reported here due to the low number of participants for individual differences analyses.

2.5.2 | Testing procedures

The HFT was performed at a desktop computer during the beginning of the first testing session, that is, prior to the geography task. On each trial, a red heart or flower appeared on the right or left side of the screen for 1,500 ms, and the children had to press a button with their left or right index finger during the presentation of the stimulus or during the following 500-ms fixation-cross display (i.e. response window = 2,000 ms). The children were given instructions and practice trials prior to each block to ensure that they understood the task.

In the congruent block, a heart was presented on every trial, and the children were instructed to press the button on the same side on which the heart appeared. In the incongruent block, a flower was presented on every trial, and the children were instructed to press the button on the side opposite to the one on which the flower appeared. In the mixed block, heart and flower trials were intermixed, and the children had to continue to follow the rules learned previously, switching between the heart condition (respond on the same side) and the flower condition (respond on the opposite side). The task parameters were identical to those used by Brod, Bunge, and Shing (2017) (the task, including the exact stimuli and stimulus lists, can be found at https://osf.io/c8gbj/). The blocks, which each included 20 trials, were performed in a fixed order, with congruent trials followed by incongruent and then mixed trials. The HFT is typically performed in that order, and previous research has shown that condition differences in response times are not explained by the order of the blocks (Wright & Diamond, 2014).

2.6 | Stimulus presentation & eye-tracking data acquisition

Subjects were seated about 68 cm from the screen in a dimly lit room. The eye-tracking apparatus (EyeLink 1000, SR Research) was located below the computer screen and recorded continuously throughout both experiments at a frequency of 500 Hz. Eye-tracking was performed to record changes in participants' pupil size in response to the presentation of the correct outcome (i.e., during the ‘Results Phase’, see Figure 1). The key measure was the difference in the pupillary response between outcomes that match versus violate expectancies. This difference can be interpreted as a measure of the amount of surprise experienced by a participant, and can be compared between the prediction and postdiction conditions.

Since the pupil is highly reactive to changes in luminance, and is also modulated by eye movements, the design of the study phase was carefully tailored to the measurement of changes in pupil sizeby (a) matching stimulus luminance across each trial, (b) including a ‘Pupil Baseline’ phase right before the ‘Results Phase’ during which the pupillary surprise response was measured, and (c) presenting all stimuli close to the center of the screen to render saccades unnecessary. Stimuli were presented using PsychoPy v1.8, which was developed for psychophysics experiments and offers high control over visual presentation and timing (Peirce, 2007).

2.7 | Data analyses

2.7.1 | Performance data analysis

Data were analyzed using R (R Core Team, 2014). The α level was set at .05 throughout the analyses. For both tasks, expectancy-inconsistent trials were defined as a scale difference between expected and actual result of two or greater, which means that the actual result is also qualitatively different from the expected one. This difference score has been shown to be highly correlated with participants' experienced surprise, as assessed by surprise ratings (Reisenzein, 2000).

For the geography task, hierarchy knowledge was assessed by calculating the mean absolute difference between the estimated rank position and the true rank position. Thus, smaller differences represent greater knowledge. Improvement in hierarchy knowledge was defined as pretest–posttest change in the mean absolute difference. A within-subject *t* test was calculated to test for condition differences in change in hierarchy knowledge. Questionnaire data were evaluated using a one sample *t* test comparing participants' responses to the mean of the scale (3.5).

For the soccer task, prior knowledge was assessed using the initial knowledge test. It was calculated in the same way as in the geography task (i.e. average absolute difference between the estimated ranks and the true final rank positions in the 2014–15 season of Germany's premier soccer division). Memory performance was evaluated via two repeated-measures ANOVAs, each with the percentage of correctly retrieved results (i.e. correct differences) as the dependent variable. The first ANOVA compared mean accuracy between the three conditions (prediction, postdiction, baseline). The second ANOVA examined memory accuracy as a function of expectancy (consistent, inconsistent) and condition (prediction, postdiction); the baseline condition was excluded since it contained no expectancy measure. Three participants were excluded due to below chance level performance (<18.5% correct), leaving 48 participants for the analyses.

For the HFT, mean accuracy and reaction time (RT) were calculated for each of the three blocks (congruent, incongruent, mixed) and subjected to separate one-way ANOVAs. We focused our analyses on inverse efficiency scores, which combine accuracy and RTs into one score and are calculated by dividing RTs by accuracy. To obtain separate scores of children's inhibition and switching performance while controlling for processing speed and sustained attention, we calculated relative efficiency difference scores ([y−x]/y) between blocks for each child. Similar to the classic Stroop effect, the inhibition score reflects the relative difference in efficiency for correctly answered trials in the incongruent and congruent block ($M = -0.33$, $SD = 0.17$, range = −0.87 to −0.02), and, thus, how much more difficult the incongruent blocks are as a result of the need to inhibit the prepotent response. As in studies of task-switching, the switching score reflects the relative difference in efficiency between congruent trials in the mixed block and congruent trials in the congruent block ($M = -1.14$, $SD = 0.53$, range = −3.02 to −0.30), and, thus, how much more difficult congruent trials are in the mixed block, due to the added switching demands. Notably, both difference scores share the same minuend and denominator (denoted by *y* in the equation) – efficiency in the congruent block –which leads to normalized scores that can be directly compared. The correlation between the two scores was moderately strong ($r = .52$, $t(48) = 4.08$, $p < .001$, one-tailed).

To assess the importance of inhibition and shifting for learning from cognitive conflict, we performed a multiple regression analysis with the inhibition and shifting scores as predictors, and the difference in memory performance between expectancy-consistent and expectancy-violating outcomes in the prediction condition of the soccer task as the dependent variable. We used simple difference scores instead of relative difference scores for memory performance because differences in memory accuracy (% correct) are straightforward to interpret and compare between conditions. To test the generalizability of the observed relation between the inhibition score and the difference in memory performance, we also compared models with versus without controlling for prior knowledge or gender to test for gender differences that go beyond differences in prior knowledge.

2.7.2 | Eye-tracking data analysis for the geography task

Pupil data were analyzed in R using itrackR (https://github.com/jashubbard/itrackR) along with self-developed analysis scripts. Building on prior research in adults with these tasks (see Brod et al., 2018), we focused on the data recorded during the study phase of the geography task. As in the previous adult study that established this paradigm, pupillary data were also collected in the soccer task, but were ultimately not analyzed. Pupil data are most reliable if averaged over a large number of trials, and the probabilistic nature of the task did not yield many strong expectancy violations. Since our study was run to investigate whether the findings revealed in adults hold in children as well, we did not analyze the pupillary data of the soccer task.

The main goal of the pupillometry analyses was to determine whether children were surprised by outcomes that they had predicted incorrectly. We calculated the difference between the average pupil diameter after seeing an expectancy-inconsistent outcome to the one after seeing an expectancy-consistent outcome. Our time interval of interest was the first two seconds of the 'Results Phase' – that is, right after children saw the correct outcome. Data from four children could not be used for the pupillometry analyses because button presses were not recorded, which precluded the sorting into expectancy-consistent and expectancy-inconsistent outcomes. This resulted in 45 children whose data were usable for the pupillometry analyses.

As first steps in the analyses, eye-tracking data and behavioral data were merged. Second, periods of blinking were removed and interpolated using cubic spline interpolation (see Hepach & Westermann, 2016), which fits a third-order polynomial function to the missing data interval. Third, pupil data were aligned relative to the onset of the 'Results Phase'. Then, pupil data were normalized by subtracting the diameter at each time point from the average diameter 200 ms before the onset of the 'Results Phase' until 100 ms after the onset and dividing by it. This calculation results in a baseline-corrected percent signal change measure. With this normalization, any nonspecific effect that lasts longer than an individual trial (e.g. arousal, fatigue) cannot confound the results.

The average percentage change in pupil diameter was calculated per participant across the first two seconds of the 'Results Phase', because the surprise response can be interpreted as the initial consequence of a perceived discrepancy that acts as an interruption mechanism for ongoing cognitive processes (Mandler, 1990; Meyer, Reisenzein, & Schützwohl, 1997). The 2-s duration was chosen based on data of younger adults in the same paradigm (Brod et al., 2018) that indicated that the pupillary response displays a peak roughly 500 ms after result presentation and further unfolds for 1–2 s. This is a typical pupillary trajectory, whereby the response is sluggish and curbed to a frequency range below 4 Hz (Kloosterman et al., 2015; Loewenfeld & Lowenstein, 1993). Average percentage change was calculated separately for outcomes that were consistent versus inconsistent with expectancies, and separately for the prediction and postdiction condition.

To measure the surprise response, we calculated the average percentage change in pupil diameter for expectancy-consistent outcomes and subtracted it from the change in pupil diameter for expectancy-inconsistent outcomes, separately for prediction and postdiction. One participant's data were excluded from this analysis because of extreme values (>3 SDs from the mean). We performed *t* tests to determine statistical significance of the pupillary surprise response in each condition and to test for condition differences.

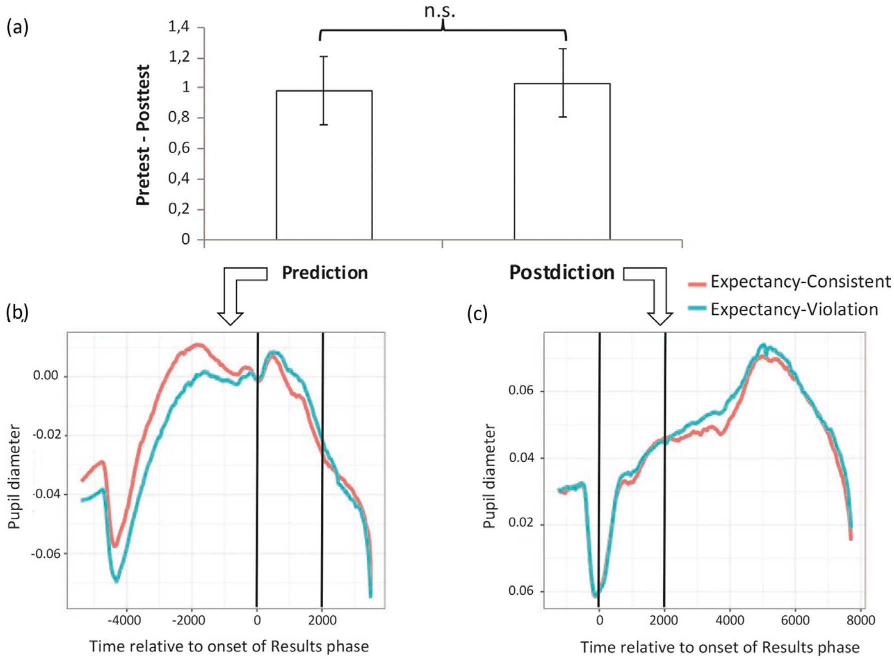

FIGURE 2 Geography task results. Panel (a) shows a strong increase in hierarchy knowledge in both the prediction and postdiction condition, with no significant differences between conditions. Error bars represent within-subject standard error. Panels b and c show the full time series of the pupillary response in the prediction (b) and postdiction condition (c), separately for expectancy-consistent and expectancy-violating outcomes. Black lines indicate the time window of interest, which was the first two seconds of the 'Results Phase'. The full time series is plotted for illustrative purposes only; inferential statistics were performed for the average percentage change in pupil diameter during the time window of interest. Percentage change was calculated relative to a baseline period 200 ms before to 100 ms after the onset of the time window of interest. Results showed that the pupil size was enhanced for expectancy-violating as compared to expectancy-consistent events in the prediction condition only, indicating that generating a prediction enabled children to experience surprise about events that violated expectancies

3 | RESULTS

3.1 | Geography task

We first examined pretest–posttest changes in hierarchy knowledge, which was defined as the mean absolute difference between the estimated and the correct rank position of the countries. Children improved their hierarchy knowledge from pretest (prediction: 2.78 ± 0.89 [*M* ± *SD*]; postdiction: 2.81 ± 0.77) to posttest (prediction: 1.79 ± 1.15; postdiction: 1.78 ± 0.97) in both conditions (difference from pretest to posttest for prediction: 0.98 ± 1.15, range −1.0–4.0, $t(49) = 6.04$, $p < .001$; postdiction: 1.03 ± 0.92, range −0.5–3.17, $t(49) = 7.97$, $p < .001$). As is apparent in Figure 2, however, there were no differences in knowledge improvement between prediction and postdiction conditions ($t(49) = -0.26$, $p = .60$).

We then looked at children's pupillary surprise response, defined as the mean difference in pupil diameter between expectancy-consistent and expectancy-inconsistent events during the first 2 s of the Results Phase. Children exhibited a significant pupillary surprise response in the prediction condition ($t(44) = 2.79$, $p = .004$), but not in the postdiction condition ($t(44) = 0.28$, $p = .782$); this difference between conditions was significant ($t(44) = 1.88$, $p = .033$). The observed pupillary surprise response in the prediction condition did not correlate with knowledge improvement, however ($r = -.07$, $t(43) = -0.46$, $p = .65$), in contrast to the study involving adults (Brod et al., 2018).

On the post-test questionnaire, most children reported that the prediction condition was more fun than the postdiction condition (mean rating on 6-point scale = 2.18, $t(49) = -5.68$, $p < .001$), and that it was slightly better at promoting learning (mean rating = 2.98, $t(49) = -2.15$, $p = .04$).

In sum, these results suggest that although children experience surprise about events that they have predicted incorrectly, this surprise does not lead to better learning, as indicated both by the lack of performance differences between the prediction and postdiction condition and the lack of a correlation between the pupillary surprise response and improvements in hierarchy knowledge. To understand these results, which differ from those observed previously with this paradigm in adults (Brod et al., 2018), we needed to look more closely at how expectancy-inconsistent events may or may not lead to learning in children. The soccer task allowed us to do so by providing separate measures of episodic memory for expectancy-consistent and expectancy-inconsistent events.

3.2 | Soccer task

We first compared differences in memory accuracy between the three conditions (prediction: 0.32 ± 0.08; postdiction: 0.33 ± 0.09;

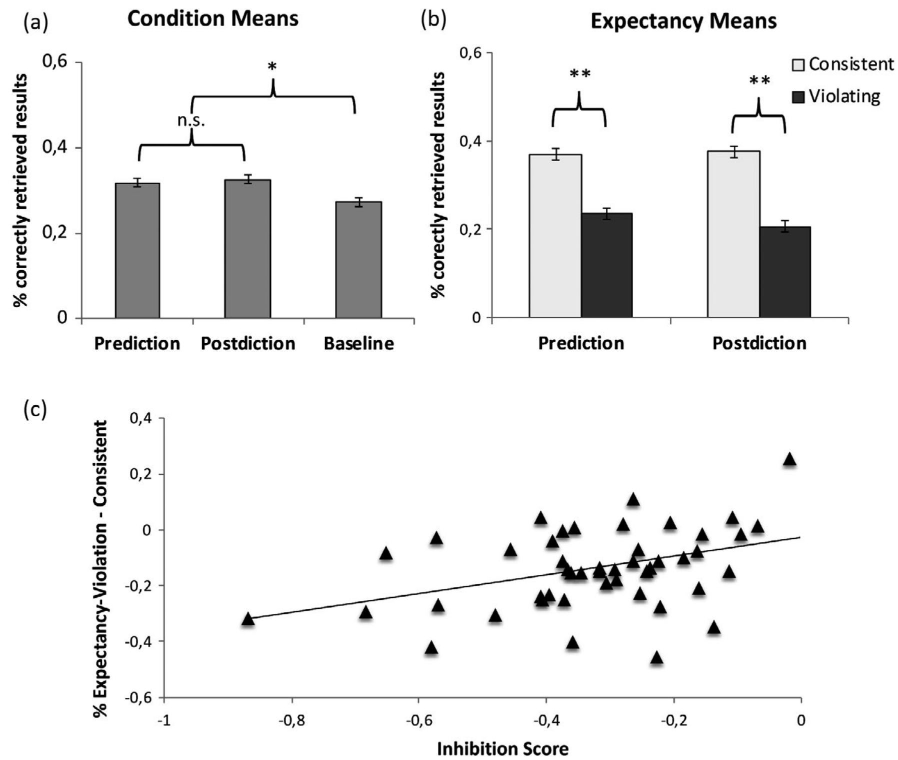

FIGURE 3 Soccer task results. (a) Children's overall memory accuracy in the baseline condition was significantly worse than in the prediction and postdiction condition, but there were no differences between the latter two. (b) For both the prediction and postdiction condition, memory for expectancy-violating events was strongly reduced compared to memory for expectancy-consistent events, with no interaction. (c) Inhibition, as measured by the hearts & flowers task, was moderately strongly correlated with the decrease in memory accuracy for expectancy-violating events in the prediction condition (r = .40). For a and b, error bars are within-subject standard errors. *$p < .05$, **$p < .01$

baseline: 0.27 ± 0.09; see Figure 3a). A within-subjects ANOVA revealed a main effect of condition (F(2,94) = 8.09, $p < .001$). Follow-up contrasts (Bonferroni–Holm corrected) indicated that children's memory accuracy in the prediction and postdiction conditions were significantly better than in the baseline condition (p = .03 and .01, respectively). However, accuracy did not differ between the prediction and postdiction conditions (p = .61). Next, we focused on the prediction and postdiction conditions and included expectancy in the model (prediction: consistent: 0.38 ± 0.09, inconsistent: 0.21 ± 0.13; postdiction: consistent: 0.37 ± 0.11, inconsistent: 0.23 ± 0.10; see Figure 3b). This model revealed a significant effect of congruency (F(1,47) = 80.59, $p < .001$), no effect of condition (F(1,47) = 0.64, p = .43), and no interaction (F(1,47) = 1.80, p = .19). Thus, memory for expectancy-inconsistent events was strongly reduced compared to memory for expectancy-consistent events, independent of condition.

We additionally tested whether prior knowledge of the relative rankings of the soccer teams were correlated withthis reduction in memory performance for expectancy-inconsistent events. We found that higher prior knowledge was associated with reduced memory for expectancy-inconsistent events in both the prediction condition (r = .27, t(46) = 1.90, p = .03, one-tailed) and postdiction condition (r = .33, t(46) = 2.38, p = .001, one-tailed), indicating that prior knowledge was an obstacle for remembering expectancy-inconsistent events.

In sum, we observed an overall memory enhancement for the two elaboration conditions relative to the baseline condition in the soccer task. However, in line with the results of the geography task, no differences in memory performance were observed between prediction and postdiction. In both conditions, children exhibited significantly worse memory for events that were inconsistent with their expectancies. This reduction in memory was magnified for children with high prior knowledge. Next, we explored one potential explanation as to why children struggle to incorporate information that stands in conflict with their prior knowledge: limited EF skills – that is, a limitation in their ability to inhibit attention to their prior knowledge and to incorporate new, conflicting information. The episodic nature of the soccer task allowed us to test whether children's memory for expectancy-inconsistent events was positively correlated with their EF skill level.

3.3 | EF task: Descriptives & correlations with soccer task

Although accuracy in the HFT was close to ceiling for all three blocks (Congruent: 98.7 ± 3.0%; Incongruent: 97.4 ± 4.3%; Mixed: 93.8 ± 7.0%), a within-subjects ANOVA revealed a main effect of block (F(2,98) = 14.53, $p < .001$), indicating a difficulty gradient in the expected direction: congruent < incongruent < mixed. RTs for correctly answered trials showed an even more pronounced difficulty gradient (Congruent: 393 ms ± 71; Incongruent: 509 ms ± 79; Mixed: 764 ms ± 109) (F(2,98) = 414.21, $p < .001$). For the subsequent correlational analyses, we combined accuracy and RTs into a reversed efficiency score (RT/accuracy), which displayed a similar difficulty gradient (Congruent: 399 ± 73; Incongruent: 525 ± 94; Mixed: 821 ± 145) (F(2,98) = 313.12, $p < .001$).

We calculated relative efficiency difference scores for each participant in an effort to isolate inhibitory and switching components while controlling for processing speed and sustained attention. Based on the way these scores are computed (see Methods), higher values reflect better performance. We sought to test whether children's inhibition and/or switching scores were correlated with learning from conflict, as measured by the relative drop in memory performance for expectancy-inconsistent (i.e. conflict-inducing) outcomes as compared to expectancy-inconsistent outcomes (see Methods). Similar to the inhibitory and switching scores, higher scores reflect a smaller performance decrement. This measure provides a measure of learning through cognitive conflict that is unbiased by overall memory performance, since memory performance for expectancy-consistent outcomes is accounted for. The drop in memory performance did not differ between the prediction and postdiction conditions (prediction: −0.13 ± 0.14, range −0.45–0.26, postdiction: −0.17 ± 0.16, range −0.43–0.30; ($t(46) = 1.27$, $p = .210$).

Testing for correlations between EFs and learning from conflict, we found that higher inhibition scores were associated with a smaller drop in memory for expectancy-inconsistent versus -consistent events in the prediction condition ($r = .40$, $t(45) = 2.89$, $p = .003$, one-tailed, see Figure 3c), but not in the postdiction condition ($r = .17$, $t(45) = 1.16$, $p = .128$, one-tailed). The switching score did not correlate with the drop in memory for expectancy-inconsistent events, neither in the prediction ($r = .14$, $t(45) = 0.92$, $p = .185$, one-tailed) nor in the postdiction condition ($r = -.14$, $t(45) = -0.96$, $p = .344$). Accordingly, in a multiple regression analysis that simultaneously included the inhibition and shifting scores, the memory drop for expectancy-violating events compared with expectancy-consistent events was predicted by the inhibition score only (see Table 1). Comparing models with versus without controlling for prior knowledge or gender yielded a significant effect of prior knowledge when gender was controlled for (i.e. a better model fit: $\chi^2(1) = 4.16$, $p = .048$), but no significant effect of gender when prior knowledge was controlled for ($\chi^2(1) = 0.38$, $p = .540$). Together, these results suggest that the reported results do generalize across girls and boys. To conclude, results of the individual differences analyses suggest that children's EFs, in particular the inhibitory component, specifically predicted their ability to remember expectancy-inconsistent events.

4 | DISCUSSION

This study investigated whether prompting children to generate predictions about an outcome helps them to activate their prior knowledge and improves their learning. Results of the soccer task indicated that the prediction condition lead to better learning than a baseline condition in which no knowledge activation was required. However, across both tasks, this benefit of prediction was not observed when prediction generation was compared to another condition that required knowledge activation. This failure to benefit from generating predictions is striking because the pupillometry data suggested that children experienced surprise about events that they had predicted incorrectly. Thus, although incorrectly predicted outcomes evoked surprise in children, most of them did not leverage their surprise to revise their beliefs. We explored one potential reason for why many children struggled to incorporate information that conflicts with their prior knowledge: lower executive functioning. Regression results revealed that children's ability to remember expectancy-inconsistent events was strongly related to their inhibitory control skills.

TABLE 1 Results of the multiple regression analysis

	B	SE B	t	p
Intercept	−0.140	0.021	−6.51	<.001**
Inhibition	0.383	0.141	2.72	.009**
Shifting	−0.021	0.045	−0.46	.650
Inhibition × Shifting	0.032	0.014	0.23	.818

Note: Dependent variable: percentage correct expectancy-violating – expectancy-consistent trials in prediction condition.
$^{**}p < .01$.

The first important finding from this study is that asking children to generate a prediction about an outcome leads to better learning of the outcome than just asking them to try to memorize it for a later test. In contrast to a recent finding in university students (Brod et al., 2018), however, generating predictions was not more beneficial for children's learning than generating post hoc judgments. Both predictions and post hoc judgments prompt children to activate relevant prior knowledge, which is known to have a strong positive effect on children's learning because, in contrast to adults, they do not do it spontaneously (Bjorklund et al., 1990; Brod, Lindenberger, et al., 2017; Hasselhorn, 1990). Further evidence for successful prior knowledge activation in the prediction condition comes from the pupillometry data in the geography task, which revealed that children exhibited a pupillary surprise response to events that they had predicted incorrectly. Since having an expectation about an outcome is a prerequisite for being surprised, children can be assumed to have activated at least some relevant prior knowledge in order to generate the prediction. Asking children to generate a prediction can, thus, be seen as a successful method to lead children to activate relevant prior knowledge and to induce cognitive conflict.

The second key finding is that there is an association between EF skills and learning through induced cognitive conflict. This is, to the best of our knowledge, the first demonstration of a relation between children's EFs and their ability to remember unexpected outcomes. We compared the relative influence of inhibition and switching skills while controlling for other important factors like sustained attention, prior knowledge, and gender. These analyses revealed that children's inhibitory control skills were specifically predictive of their ability to remember unexpected outcomes. Thus, it appears that the ability to override a prior belief is more closely related to the ability to override a previously well-learned rule than to switching flexibly back and forth between two equally strong rules. In conclusion, while most children struggled with learning through cognitive conflict, those children with better – that is, more adult-like – inhibitory control skills did better than those with lower inhibitory control skills.

4.1 | Inhibition and learning through cognitive conflict

Learning through cognitive conflict is challenging (see Chinn & Brewer, 1993). What is it that makes it particularly difficult for children? Our findings suggest that it has to do with the rather slow and variable developmental trajectory of EF skills (e.g., Davidson, Amso, Anderson, & Diamond, 2006; Luna et al., 2004). Why might inhibition be particularly important for learning through cognitive conflict? First, this type of learning requires that conflicting information not be discounted right away. That is, sustained attention has to be devoted to the conflicting information, and prior beliefs must be suppressed. Second, upon encountering similar information again, the old belief has to be inhibited – at least initially, until the new one is well established.

Complementary evidence that children struggle to override existing knowledge has been provided in research on the development of episodic memory as well as on learning through feedback. A neuroimaging study on age-related differences in memory retrieval of expectancy-consistent and expectancy-inconsistent events revealed that children (aged 9–11) exhibited less connectivity than adults within a brain network involved in updating behavior during the retrieval of expectancy-inconsistent events, likely reflecting their reduced ability to override their prior beliefs (Brod, Lindenberger, et al., 2017). Corroborating evidence comes from a wealth of research that has explored children's performance on card sorting tasks, in which several rules exist and change without warning. This research has shown that children in mid-to-late childhood use corrective feedback less efficiently than adults, which leads to greater perseveration (e.g., Crone, Jennings, & Van der Molen, 2004; Luciana & Nelson, 2002). Children's slower rule updating is suggested to be due not to an inability to monitor the feedback, but to a failure to inhibit the outdated rule (Crone, Richard Jennings, & Molen, 2004; Welsh, Pennington, & Groisser, 1991; Zelazo, Frye, & Rapus, 1996).

In summary, our findings are in line with prior work involving feedback-based learning tasks by showing that children correctly detected conflicting outcomes, given their pupillary response to expectancy-violating events, but that only those with sufficiently good inhibitory control were able to harness this cognitive conflict to override their prior knowledge.

4.2 | Future directions

In this study, we used tasks that enabled us to compare generating predictions to closely matched control conditions in a within-subject design. These task designs allowed us to generate a large number of unique trials, which is necessary for analyses on a trial level and for the acquisition of reliable eye-tracking data. As a consequence, these tasks enabled us to examine simple belief revision, not complex conceptual change; we think that it is important to extend these findings to more complex scenarios, such as those found in science education. Indeed, asking students to generate a prediction is a popular instructional technique, most frequently as part of the 'Scientific Method' (i.e. hypothesis testing using experiments) in science education (e.g. Champagne, Klopfer, & Gunstone, 1982; Hardy, Jonen, Möller, & Stern, 2006; Liew & Treagust, 1995). It is suggested to fulfill various functions, including activating and exposing the learners' prior knowledge, stimulating curiosity for the correct answer (Brod & Breitwieser, 2019), and – if the prediction is wrong – triggering belief revision because the learners are surprised and realize that there is a flaw in their concept (Brod et al., 2018). Thus, we believe that our research using well-controlled and simplistic scenarios in the laboratory could lay a foundation for further research that looks at changes in more complex concepts and their relation with EF skills.

Do our results suggest that eliciting predictions to induce belief revision is not worth the effort with children? No. But they do suggest that the younger the learners are, the more help they will need to overcome existing misconceptions. We view prediction generation as a good first step to help learners overcome their misconceptions, because activating prior knowledge before presenting new information is key (Bransford & Johnson, 1972), and our results suggest that prediction generation serves this function well. Indeed, generating a prediction is usually just the first step in a sequence, followed by a group discussion of the results and, if necessary, an explanation by the teacher (e.g., Liew & Treagust, 1995).

The data from our study cannot speak to what would be good next steps after prediction generation. However, recent developmental studies have convincingly demonstrated that already by kindergarten age, children benefit from generating explanations, and that it helps them to overcome misconceptions (e.g. Crowley & Siegler, 1999; Legare & Lombrozo, 2014; Wellman & Liu, 2010, for overviews, see Legare, 2014; Wellman, 2011). These findings suggest that prompting children to generate explanations for why the outcome was different than predicted could be a good next step. Future research should assess the efficacy of eliciting both predictions and explanations. Additionally, future research should examine the link between EFs and learning through cognitive conflict in a wider age range, and test whether this link can be generalized to a wider range of tasks.

A further important avenue for future research concerns the relation between children's ability to leverage surprise for belief revision and their capacity for metacognitive monitoring and control – that is, the ability to accurately represent and control one's current cognitive activities. Metacognitive skills follow a similarly late-maturing developmental trajectory as EFs, and depend on – but are not fully determined by – EFs (Roebers, 2017). It is therefore plausible that metacognitive skills are similarly related to children's belief revision to what was found for inhibitory control in the current study. This finding would open exciting new avenues for interventions, such as strategy training for children that tackle their deficient metacognitive monitoring and control abilities (see Schneider, 2008).

4.3 | Limitations of this study

First and foremost, an important caveat is that our findings provide correlational rather than causal evidence of the importance of EFs for learning from incorrect predictions. While it seems implausible that the direction of causality is reversed (i.e. that learning through wrong predictions improved EFs), the observed correlation may result from a third factor that is common to both learning through conflict and EFs. To establish a causal role of EFs, studies are needed which manipulate EF demands in addition to prediction generation. We have tried to control for a number of potential cognitive mediators – in particular, overall processing speed and memory performance. However, we have not explicitly tested working memory or metacognitive abilities such as the use of memory strategies (Bjorklund, 1987; Roebers, Cimeli, Röthlisberger, & Neuenschwander, 2012; Schneider, 2015). There are several other factors such as learners' motivation, epistemological beliefs, and metacognitive knowledge that are likely to be important as well (for an overview, see Limón, 2001). Thus, we posit that EFs are necessary, albeit not sufficient, for learning through cognitive conflict.

In our soccer task, memory performance was clearly better for expectancy-consistent than expectancy-inconsistent results, in both the prediction and postdiction conditions. This result is in line with a rich literature on the memory congruency effect. This effect is commonly observed in memory tasks in which participants can successfully guess based on their prior knowledge (see Brod et al., 2013). Guessing, then, benefits memory for expectancy-consistent events in the absence of true recollection. It is thus unclear how much of the enhanced memory for expectancy-consistent events is due to guessing, and how much due to better actual memory. We have not addressed this issue with the current study design because our main goal was to allow a fair comparison between the two conditions as well as with the baseline condition, in which participants did not give expectancy ratings.

5 | CONCLUSIONS

Revising one's existing knowledge is hard. Our results suggest that this is particularly true for children, and that this is related to the fact that – on average – they exhibit lower inhibitory control than adults. It is therefore no surprise that cognitive conflict as an instructional strategy often fails in classroom intervention studies (Limón, 2001). Our study indicates that younger students in particular will need considerable support to successfully leverage the cognitive conflict induced by wrong predictions for revising their knowledge.

ACKNOWLEDGEMENTS

We thank Freya Moßig, Solveig Gresselmeyer, and Manuel Lieb for their help in collecting the data and Ilonca Hardy for valuable discussions. G.B. was supported by a Jacobs Foundation Early Career Research Fellowship. S.A.B. was supported by a Jacobs Foundation Research Fellowship.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

ORCID

Garvin Brod https://orcid.org/0000-0002-7976-5609
Jasmin Breitwieser https://orcid.org/0000-0002-0337-7159
Marcus Hasselhorn https://orcid.org/0000-0002-6617-2556
Silvia A. Bunge https://orcid.org/0000-0001-8431-7627

REFERENCES

Alexander, P. A. (1996). The past, present, and future of knowledge research: A reexamination of the role of knowledge in learning and instruction. *Educational Psychologist*, *31*(2), 89–92. https://doi.org/10.1080/00461520.1996.10524941

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

Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review*, *7*(2), 93–130. https://doi.org/10.1016/0273-2297(87)90007-4

Bjorklund, D. F., Muir-Broaddus, J. E., & Schneider, W. (1990). The role of knowledge in the development of strategies. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 93–128). Hillsdale, NJ: Erlbaum.

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: Generating a prediction stimulates curiosity. *Npj Science of Learning*, *4*, 17. https://doi.org/10.1038/s41539-019-0056-y

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science*, *28*(7), 967–978. https://doi.org/10.1177/0956797617699838

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Brod, G., Lindenberger, U., & Shing, Y. L. (2017). Neural activation patterns during retrieval of schema-related memories: Differences and commonalities between children and adults. *Developmental Science*, *20*(6), 1–16. https://doi.org/10.1111/desc.12475

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*, 139. https://doi.org/10.3389/fnbeh.2013.00139

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press. https://doi.org/10.1016/S0016-6995(85)80176-5

Carter, C. S., & Van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective and Behavioral Neuroscience*, *7*(4), 367–379. https://doi.org/10.3758/CABN.7.4.367

Champagne, A. B., Klopfer, L. E., & Gunstone, R. F. (1982). Cognitive research and the design of science instruction. *Educational Psychologist*, *17*(1), 31–53. https://doi.org/10.1080/00461528209529242

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, *106*(14), 5529–5533. https://doi.org/10.1073/pnas.0810002106

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49. https://doi.org/10.3102/00346543063001001

Crone, E. A., Richard Jennings, J., & Van Der Molen, M. W. (2004). Developmental change in feedback processing as reflected by phasic heart rate changes. *Developmental Psychology*, *40*(6), 1228–1238. https://doi.org/10.1037/0012-1649.40.6.1228

Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, P. M., & Rombouts, S. A. R. B. (2008). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective and Behavioral Neuroscience*, *8*(2), 165–177. https://doi.org/10.3758/CABN.8.2.165

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*(2), 304–316. https://doi.org/10.1111/1467-8624.00023

Davidson, M.C., Amso, D., Anderson, L.C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388. https://doi.org/10.1126/science.1151148

Fiske, A., & Holmboe, K. (2019). Neural substrates of early executive function development. *Developmental Review*, *52*, 42–62. https://doi.org/10.1016/j.dr.2019.100866

Gropen, J., Clark-Chiarelli, N., Hoisington, C., & Ehrlich, S. B. (2011). The importance of executive function in early science education. *Child Development Perspectives*, *5*(4), 298–304. https://doi.org/10.1111/j.1750-8606.2011.00201.x

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of “floating and sinking”. *Journal of Educational Psychology*, *98*(2), 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Hasselhorn, M. (1990). The emergence of strategic knowledge activation in categorical clustering during retrieval. *Journal of Experimental Child Psychology*, *50*(1), 59–80. https://doi.org/10.1016/0022-0965(90)90032-4

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development*, *17*(3), 359–377. https://doi.org/10.1080/15248372.2015.1135801

Johnson, S. C., & Carey, S. (1998). Knowledge enrichment and conceptual change in folkbiology: Evidence from Williams syndrome. *Cognitive Psychology*, *37*(2), 156–200. https://doi.org/10.1006/cogp.1998.0695

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026. https://doi.org/10.1126/science.1089910

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068–1078. https://doi.org/10.1111/ejn.12859

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, *8*(2), 101–106. https://doi.org/10.1111/cdep.12070

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, *126*, 198–212. https://doi.org/10.1016/j.jecp.2014.03.001

Liew, C. W., & Treagust, D. F. (1995). A predict-observe-explain teaching sequence for learning about students' understanding of heat and expansion of liquids. *Australian Science Teachers' Journal*, *41*(1), 68–71.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*(4–5), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Loewenfeld, I. E., & Lowenstein, O. (1993). *The pupil: Anatomy, physiology, and clinical applications*. Ames, IA: Iowa State University Press.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-year-old children. *Developmental Neuropsychology*, *22*(3), 595–624. https://doi.org/10.1207/S15326942DN2203_3

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

Mandler, G. (1990). A constructivist theory of emotion. In N. L. Stein, B. Leventhal & T. R. Trabasso (Eds.), *Psychological and biological approaches to emotion* (pp. 21–45). Hillsdale, NJ: Lawrence Erlbaum.

Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, *12*(1), 1–25. https://doi.org/10.1111/j.1467-7687.2008.00754.x

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*(3), 251–274. https://doi.org/10.1023/A:1024422330338

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Morton, J. B., & Munakata, Y. (2009). Connectionist approaches to perseveration: Understanding universal and task-specific aspects of children's behavior. In J. P. Spencer, M. S. C. Thomas & J. L. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 141–164). New York, US: Oxford University Press.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago, IL: University of Chicago Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*, 1–12. https://doi.org/10.3389/fnins.2011.00115

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion*, *14*(1), 1–38. https://doi.org/10.1080/026999300378978

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001

Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, *7*(3), 151–173. https://doi.org/10.1007/s11409-012-9089-9

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schneider, W. (2015). *Memory development from early childhood through emerging adulthood*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-09611-7

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, *52*(1), 290–301. https://doi.org/10.1016/j.neuroimage.2010.04.009

Waxer, M., & Morton, J. B. (2011). The development of future-oriented control: An electrophysiological investigation. *NeuroImage*, *56*(3), 1648–1654. https://doi.org/10.1016/j.neuroimage.2011.02.001

Wellman, H. M. (2011). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives*, *5*(1), 33–38. https://doi.org/10.1111/j.1750-8606.2010.00154.x

Wellman, H. M., & Liu, D. (2010). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 261–279). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176803.003.0017

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131–149. https://doi.org/10.1080/87565649109540483

White, R., & Gunstone, R. (1992). *Probing understanding*. London, UK: Routledge.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, *5*, 213. https://doi.org/10.3389/fpsyg.2014.00213

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*(1), 160–175. https://doi.org/10.1111/cdev.12145

Zaitchik, D., & Solomon, G. E. A. (2008). Animist thinking in the elderly and in patients with Alzheimer's disease. *Cognitive Neuropsychology*, *25*(1), 27–37. https://doi.org/10.1080/02643290801904059

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences*, *8*(1), 12–17. https://doi.org/10.1016/j.tics.2003.11.001

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Zielinski, B. A., Gennatas, E. D., Zhou, J., & Seeley, W. W. (2010). Network-level structural covariance in the developing brain. *Proceedings of the National Academy of Sciences of the USA*, *107*(42), 18191–18196. https://doi.org/10.1073/pnas.1003109107

How to cite this article: Brod G, Breitwieser J, Hasselhorn M, Bunge SA. Being proven wrong elicits learning in children – but only in those with higher executive function skills. *Dev Sci*. 2020;23:e12916. https://doi.org/10.1111/desc.12916