frontiers | Research Topics

Advances in metacognition and reflection

Edited by
Igor Bascandziev, Claudia M. Roebers, Stephanie M. Carlson and Loren Marulis

Published in
Frontiers in Developmental Psychology

ISSN 1664-8714
ISBN 978-2-8325-6159-1
DOI 10.3389/978-2-8325-6159-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in metacognition and reflection

Topic editors
Igor Bascandziev — Harvard University, United States
Claudia M. Roebers — University of Bern, Switzerland
Stephanie M. Carlson — University of Minnesota Twin Cities, United States
Loren Marulis — Connecticut College, United States

Citation
Bascandziev, I., Roebers, C. M., Carlson, S. M., Marulis, L., eds. (2025). *Advances in metacognition and reflection*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6159-1

Table of contents

frontiers | Frontiers in Developmental Psychology

TYPE Editorial
PUBLISHED 11 March 2025
DOI 10.3389/fdpys.2025.1579553

OPEN ACCESS

EDITED AND REVIEWED BY
Angeline S. Lillard,
University of Virginia, United States

*CORRESPONDENCE
Loren M. Marulis
loren.marulis@gmail.com

RECEIVED 19 February 2025
ACCEPTED 24 February 2025
PUBLISHED 11 March 2025

CITATION
Marulis LM (2025) Editorial: Advances in metacognition and reflection.
Front. Dev. Psychol. 3:1579553.
doi: 10.3389/fdpys.2025.1579553

Editorial: Advances in metacognition and reflection

Loren M. Marulis*

Department of Human Development, Connecticut College, New London, CT, United States

KEYWORDS

metacognition, reflection, cognitive development, learning, developmental psychology, educational psychology

Editorial on the Research Topic
Advances in metacognition and reflection

Metacognition, first coined in the literature in the 1970s by Flavell (1976) has been the focus of diverse disciplines (e.g., developmental, cognitive, and educational psychology, psychiatry, and criminal justice) because of its substantial, positive impact on development and learning in these fields. We know it is critical for greater and deeper learning and positive life outcomes (e.g., prisoner rehabilitation: Gois and Kane, 2025; academics: He et al., 2024; trauma-related treatment: Wiesepape et al., 2025) but we also know that it is rarely explicitly taught or fostered in formal or informal learning contexts and its development rarely occurs naturally. In fact, in Flavell's unveiling of this term, which has kept us all busy for many decades since, he focused not on its abundance but on how it is most conspicuous (and negatively impactful) in its absence:

> "Resnick and Glaser's research provides us with some striking examples of children failing to solve problems for which they possess the necessary solution procedures. They ought to solve these problems, we think, and yet they do not. Why not? My own guess on the matter originates in the expected place, namely, the area in which I have done most of my recent research and thinking. This area is the development of metacognition." (Flavell, 1976, p. 232).

Historical perspective

Although the official unveiling of the term "metacognition" is relatively new (Flavell, 1976, 1979), there is a long history of references to similar concepts such as reflection or introspection, traces of which can be seen as far back as the musings of Plato, Aristotle, and Simonides. John Locke, in 1690, introduced greater specification by distinguishing "reflection" as a more important and privileged form of thinking than other forms or "sensations" that do not tend to produce "long-lasting ideas" or a deep, reflective type of cognitive processing. Furthermore, early educators such as John Dewey had similar ideas. In his Pedagogic Creed (Dewey, 1897), Dewey stated his belief that the learning process would be disorganized and unsystematic (and thus not "educative") if left unexamined and that looking within one's psychological processes would lead to educative leverage. It is likely that the influx of behaviorism into the field of psychology and education in the early 20th century is related to the hiatus in the focus on research and theorizing about metacognition and reflection and, similarly, to the resurgence of this focus shortly after the shift from behaviorism to cognition with the "cognitive revolution" of the 1950s. This shift resulted in the consequential work of developmental psychologist Flavell and his contemporaries. The zeitgeist was substantial and led to a greater convergence and alliance between the fields of psychology and education, making this body of literature more interdisciplinary and, ultimately, leading to greater contextualization, developmental appropriateness, and ecological validity in the study of metacognition. The rapidly growing body of extant meta-reviews (e.g., Eberhart et al., 2025; He et al., 2024; Norman et al., 2019; Ohtani and Hisasaka, 2018) and primary research (e.g., Coughlin et al., 2022; Desoete and De Craene, 2019; Fu and Qi, 2025; Özçakmak et al., 2021) on metacognition and reflection provide robust evidence of their strong and unique predictive power for important outcomes. Although metacognitive processes have been studied for at least five decades, it is only in recent years that this investigation has included infancy and early childhood, with initially promising and, in 2025, robustly positive and strong results (e.g., Chen et al., 2023; Gourlay et al., 2020; Marulis and Nelson, 2021; van Loon and Roebers, 2024; Whitebread and Neale, 2020). This shift is not only developmentally inclusive but also has critical implications for improving developmental and life trajectories based on the greater cognitive malleability in the early years of development. This Research Topic further elucidates early childhood metacognitive processes contributing to a comprehensive understanding of their developmental trajectory.

Conceptualization and measurement

Observing a set of family portraits, Sir Arthur Conan Doyle's fictional Sherlock Holmes, declaring himself a connoisseur of the arts, remarked on their high artistic quality and, ever the reflective thinker, continued to analyze his assessment of the portraits with the following: "I know what is good when I see it, and I see it now" (Doyle, 1902; p. 93). Perhaps the more known, definitely more modern, and non-fictional instance of this concept occurred in 1964 with U.S. Supreme Court Justice Potter Stewart's explanation of how he determined (i.e., measured) obscene material not protected under the First Amendment, which was essentially, "I know it when I see it." At first glance, these statements elicit something nebulous without a defined set of characteristics, but the idea that "I" (seemingly referring to someone with expertise or authority over a matter) will be able to reliably identify this "something" is also powerful. In the case of this Research Topic, the "something" of focus–advancing our understanding of metacognition and reflection–is particularly important given the consistent, robust, and positive impact these skills have across development and types of learning. The inherent challenge, then, is to reverse engineer this knowledge into operationalized indicators. Since its debut in the literature, there have been calls for achieving a universally agreed-upon conceptualization of what "metacognition" is and is not. The challenges of this endeavor are as great as the rewards. On the one hand, the challenges and difficulties include contradictory findings and limited or no coherence; on the other hand, the benefits include convergent evidence across disparate methods and the emergence of a developmental trajectory for metacognition and reflection. To this end, we have seen decades of rigorous research yet, in some ways, we are no closer to a consensus. I suggest is that we direct our attention to a new charge: Rather than focusing on the struggle to achieve full *unity*, we focus on achieving conditional (contingent and adaptive), calibrated (precise), and *unified* (internally consistent) conceptualizations of metacognitive processes. Collaborative efforts such as this Research Topic reflect this type of pivot and represent metacognition for its complexity and strength.

In the Editorial of a previous Research Topic on metacognition in the *International Electronic Journal of Elementary Education* (Desoete and Özsoy, 2009), capturing metacognition was compared to the murkiness of Scotland's Loch Ness monster. For the sake of argument, I will posit that the authors of this Editorial were referring to the *sightings* (i.e., measurements) of the popular "monster" that are purported to have begun in 565 AD. In this case, it follows that there is something there; something is being seen (previous scientific explanations include boat wakes and other sea creatures such as large eels or water birds, and non-scientific explanations include mythology and intentional hoaxes) and perhaps some would say they would "know it when they see it."

Metacognition was first conceptualized (in the 1970s) as "thinking about thinking," or metacognitive knowledge followed by the addition of regulation of cognition (Brown, 1978, 1987), monitoring and control (Nelson and Narens, 1990) and more recently, motivational, and affective processes (Efklides, 2011). In practice, these conceptualizations translated into a 3-part skill set (plan, monitor, evaluate) (Fogarty, 1994). Their culmination is a broad conceptual agreement of metacognition as the knowledge, regulation, and monitoring of cognitive processes.

An apt analogy for the measurement of metacognition can also be found in black holes within the domain of the physical sciences. The history of the study of black holes has moved from mathematics to physics and from theory (general relativity) to simulations and experiments to telescopic evidence (Oldham and Auger, 2016). Similarly, the study of metacognition has evolved from an abstract conceptualization of the existence of "something" that was hard to pin down but had clear effects to the emergence of (sometimes contradictory) theories and models to the use of more precise and comprehensive measurement tools such as systematic observational coding protocols, computer hardware and software, eye-tracking, and electroencephalogram (EEG). Through these advances, like black holes that have powerful interactions with things around them but can only be seen with special equipment, we have not only been able to fine-tune and calibrate the conceptualization and measurement of metacognitive processes but have also gained a much deeper understanding of their importance for to successful learning and other life outcomes. In both cases, as measurement tools and methods have advanced, so have our understandings and applications.

Specific analogical comparisons between metacognition and black holes or the folklore of the Loch Ness monster may be a bridge too far; nonetheless, these converging ideas across disparate spheres underlie the concept of the existence of an important and impactful "something" (e.g., quality of art; obscenity; Nessie; black holes; metacognition). The important point here is the abstraction of an increasingly measurable "something" at the core of its domain.

TABLE 1 Advancing and calibrating our understanding of metacognition and reflection: important characteristics and findings of nine Research Topic articles.

References	Type of paper	Age/development period	Metacognitive component	Summary of results
Allen et al.	Conceptual	N/A	Knowledge (epistemic reflection)	A model of epistemic reflection based on interactivism (knowing is doing and, subsequently, predicts successful interactions with one's environment rather than information processing) was proposed to better explain new representations and conceptual changes emerging from reflection.
Bascandziev et al.	Empirical	4.75–9.5-year-olds	Skills (cognitive reflection and monitoring)	Children's monitoring skills were associated with an understanding of physical science concepts controlling for age, EF, and cognitive reflection underscoring the importance of metacognitive skills (specifically consistency monitoring) for young children's scientific learning.
Buehler and Oeri	Empirical	5–6-year-olds	Skills (monitoring and control/regulation)	Older children ($M = 5.85$ years old) displayed greater metacognitive control than younger children ($M = 5.05$ years old) on a newly developed, ecologically valid, unsolvable problem-solving task, although no age differences were found for metacognitive monitoring. Children showed more metacognitive monitoring and less control in the solvable than in the unsolvable part of the wooden puzzle.
Dutemple et al.	Empirical	5–6-year-olds	Skills (broad explicit and implicit)	Implicit and explicit metacognition (not EF) significantly predicted school readiness beyond age and sex. Correlations were found between explicit metacognition and EF.
Jacobs et al.	Empirical	7–9-year-olds	Skills (control/regulation)	Significant positive correlations were found longitudinally between metacognitive control and arithmetic accuracy in 7-8 year olds. However, post-error adjustments in arithmetic and the working memory tasks were not correlated.
Kim and Carlson	Mini review	Infancy-Adulthood	Skills (monitoring, cognitive reflection, and control/regulation)	To better understand the development of metacognition and reflection from infancy through adulthood, interactions with the environment were systematically examined. Specifically, children's exploration (experimenting with multiple, familiar and unfamiliar, options) and exploitation (sticking with familiar options or those perceived to be most advantageous for maximum reward) behaviors were investigated focusing on the benefits for adaptive learning and decision-making in children.
Kolloff and Roebers	Empirical	6-year-olds	Skills (monitoring)	Memory and nonverbal intellectual ability were found to be related to metacognitive monitoring, although the impacts of nonverbal intelligence were modest, indicating that young children's nonverbal intellectual ability and metacognitive monitoring skills are relatively independent constructs.
Tomasello	Conceptual	Infants and preschool-aged	Skills (control/regulation)	A developmental (What is regulated? How is it regulated? Where is it regulated?) model integrating executive and metacognitive processes was proposed in which executive processes monitor and control action and attention; in turn, metacognitive processes monitor and control these executive processes. Executive processes emerge between 9-12 months of age; metacognitive processes emerge around 3-4 years of age.
Young and Shtulman	Empirical	5–12-year-olds	Skills (cognitive reflection)	Cognitive reflection strongly predicted children's strategic behaviors and interpretation skills and uniquely predicted children's performance beyond age and EF.

As important as art or black holes are to segments of society, so are metacognitive processes. The core aim of this *Advances in Metacognition and Reflection* Research Topic of *Frontiers in Developmental Psychology* was to build on this foundation and endeavor to fill existing gaps in the past four decades of research on metacognitive processes with a chief focus on reflective processes. As is representative of the literature on metacognition and reflection, the articles in this Research Topic employ diverse theoretical frameworks, methods, and developmental periods yet converge in one critical way: positive and moderate to strong associations and predictions of metacognition and reflection across developmental outcomes, contexts, and perspectives (Table 1). The key contributions thus lie in the elucidation and parsing of specific metacognitive components; the *what*, *why*, *how*, *when*, and for *whom* of detecting effects. In this way, we take a metacognitive approach to the study of metacognition. As we clarify and precisely investigate the conceptualization, operationalization, and measurement of metacognition and its subcomponents, its shape and form will become less amorphous, and we will not only vaguely "know it when we see it" but we will also be able to precisely identify and explicate its elements, associations, and impacts (see Terneusen et al., 2024). Achieving such conditional, calibrated, unified metacognition has important implications at both the basic (creating new knowledge) and applied (teaching, interventions, policies) levels across development, contexts, and individuals, resulting in more efficient and adaptive learning and successful developmental and life outcomes.

Author contributions

LM: Writing – original draft, Writing – review & editing.

Conflict of interest

The author declares 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

References

Brown, A. (1987). "Metacognition, executive control, self-regulation, and other mysterious mechanisms," in *Metacognition, Motivation, and Understanding,* eds. F. E. Weinert and R. H. Kluwe (Hillsdale, NJ: Lawrence Erlbaum Associates), 65–116.

Brown, A. L. (1978). "Knowing when, where, and how to remember: a problem of metacognition," in *Advances in Instructional Psychology,* ed. R. Glaser (Hillsdale, NJ: Lawrence Erlbaum Associates), 77–165.

Chen, S., Guo, M., and Dousay, T. A. (2023). Grow to learn: a metacognitive approach to early childhood teachers' science professional development. *Res. Sci. Technol. Educ.* 2023, 1–21. doi: 10.1080/02635143.2023.2279076

Coughlin, C., Prabhakar, J., D'Esposito, Z., Thigpen, B., and Ghetti, S. (2022). Promoting future-oriented thought in an academic context. *Cogn. Dev.* 62:101183. doi: 10.1016/j.cogdev.2022.101183

Desoete, A., and De Craene, B. (2019). Metacognition and mathematics education: an overview. *ZDM Mathem. Educ.* 51, 565–575. doi: 10.1007/s11858-019-01060-w

Desoete, A., and Özsoy, G. (2009). Introduction: metacognition, more than the Loch Ness monster? *Int. Electr. J. Element. Educ.* 2, 1–6.

Dewey, J. (1897). My pedagogic creed. *School J.* 54, 77–80.

Doyle, A. C. (1902). *The Hound of the Baskervilles.* London: George Newnes Ltd.

Eberhart, J., Ingendahl, F., and Bryce, D. (2025). Are metacognition interventions in young children effective? Evidence from a series of meta-analyses. *Metacogn. Learn.* 20, 1–45. doi: 10.1007/s11409-024-09405-x

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* 46, 6–25. doi: 10.1080/00461520.2011.538645

Flavell, J. H. (1976). "Metacognitive aspects of problem solving," in *The Nature of Intelligence*, ed. L. B. Resnick (Hillsdale, NJ: Erlbaum), 231–236. doi: 10.4324/9781032646527-16

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive developmental inquiry. *Am. Psychol.* 34, 906–911. doi: 10.1037/0003-066X.34.10.906

Fogarty, R. (1994). *The Mindful School: How to Teach for Metacognitive Reflection*. Palatine: IRI/Skylight Publishing, Inc.

Fu, Y., and Qi, C. (2025). The relationship between metacognitive skills and mathematics achievement of Chinese eighth-grade students. *Curr. Psychol.* 2024, 1–12. doi: 10.1007/s12144-024-07216-6

Gois, I., and Kane, E. (2025). Metacognition, philosophy in prisons and the demands of rehabilitation. *Howard J. Crime Just.* 12:592. doi: 10.1111/hojo.12592

Gourlay, C., Mushin, I., and Gardner, R. (2020). Young children's responses to teachers' metacognitive questions. *Int. J. Early Years Educ.* 2020, 1–20. doi: 10.1080/09669760.2020.1742671

He, G., Lin, H., and Su, A. (2024). Longitudinal and reciprocal links between metacognition, mathematical modeling competencies, and mathematics achievement in grades 7–8: a cross-lagged panel analysis. *Metacogn. Learn.* 19, 967–995. doi: 10.1007/s11409-024-09397-8

Marulis, L. M., and Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5-year-olds. *Metacogn. Learn.* 16, 207–231. doi: 10.1007/s11409-020-09244-6

Nelson, T. O., and Narens, L. (1990). "Metamemory: A theoretical framework and new findings," in *The psychology of learning and motivation: Advances in research and theory,* ed. G. H. Bower (London: Academic Press), 125–173. doi: 10.1016/S0079-7421(08)60053-5

Norman, E., Pfuhl, G., Saele, R., Svartdal, F., Låg, T., and Dahl, T. (2019). Metacognition in psychology. *Rev. General Psychol.* 23, 403–424. doi: 10.1177/1089268019883821

Ohtani, K., and Hisasaka, T. (2018). Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacogn. Lear.* 13, 179–212. doi: 10.1007/s11409-018-9183-8

Oldham, L. J., and Auger, M. W. (2016). Galaxy structure from multiple tracers–II. M87 from parsec to megaparsec scales. *Mon. Not. R. Astron. Soc.* 457, 421–439. doi: 10.1093/mnras/stv2982

Özçakmak, H., Köroglu, M., Korkmaz, C., and Bolat, Y. (2021). The effect of metacognitive awareness on academic success. *African Educ. Res. J.* 9, 434–448. doi: 10.30918/AERJ.92.21.020

Terneusen, A., Quaedflieg, C., van Heugten, C., Ponds, R., and Winkens, I. (2024). The many facets of metacognition: comparing multiple measures of metacognition in healthy individuals. *Metacogn. Learn.* 19, 53–63. doi: 10.1007/s11409-023-09350-1

van Loon, M., and Roebers, C. M. (2024). Development of metacognitive monitoring and control skills in elementary school: a latent profile approach. *Metacogn. Learn.* 19, 1065–1089. doi: 10.1007/s11409-024-09400-2

Whitebread, D., and Neale, D. (2020). Metacognition in early child development. *Transl. Issues Psychol. Sci.* 6:8. doi: 10.1037/tps0000223

Wiesepape, C. N., Smith, E. A., Muth, A. J., and Faith, L. A. (2025). Personal narratives in trauma-related disorders: contributions from a metacognitive approach and treatment considerations. *Behav. Sci.* 15:150. doi: 10.3390/bs150 20150

frontiers | Frontiers in Developmental Psychology

TYPE Conceptual Analysis
PUBLISHED 08 March 2024
DOI 10.3389/fdpys.2024.1367381

OPEN ACCESS

EDITED BY
Stephanie M. Carlson,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Sabine Doebel,
George Mason University, United States
Laura Franchin,
University of Trento, Italy

*CORRESPONDENCE
Michael Tomasello
✉ Michael.tomasello@duke.edu

RECEIVED 08 January 2024
ACCEPTED 22 February 2024
PUBLISHED 08 March 2024

CITATION
Tomasello M (2024) An agency-based model of executive and metacognitive regulation. *Front. Dev. Psychol.* 2:1367381.
doi: 10.3389/fdpys.2024.1367381

An agency-based model of executive and metacognitive regulation

Michael Tomasello[1,2]*

[1]Duke University, Durham, NC, United States, [2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

In the context of agentive decision making and action, both executive and metacognitive processes serve self-regulatory functions—just on different hierarchical tiers. In the agency-based model proposed here executive processes monitor and control action and attention from an executive tier of operation, and metacognitive processes monitor and control those executive processes from a second-order metacognitive tier of operation-both with the function of facilitating effective and efficient behavioral decisions. Each is best conceptualized as comprising three key components: (i) what is regulated, (ii) how, via what processes, is it regulated, and (iii) where, in what cognitive workspace, is it regulated—either in individual or in shared agencies. Developmentally, evidence is presented that executive processes for regulating both individual and joint agencies emerge only after 9–12 months of age, and metacognitive processes for regulating both individual and collective agencies emerge only after 3–4 years of age. Cognitive flexibility, as an important outcome, derives from the child's attempts to metacognitively regulate differing social perspectives within shared agencies.

KEYWORDS

executive function, self-regulation, metacognition, agency, decision making

Metacognition is often defined as "thinking about thinking." But why do children (or adults) think about thinking? What psychological function does it serve? Most fundamentally, metacognitive processes serve self-regulative functions monitoring and controlling ongoing cognitive processes as children attempt to solve problems, learn new skills, or achieve challenging goals. Indeed, a term often used in the education literature is "metacognitive regulation."

This self-regulation view of metacognition suggests that it is related to executive function. But there have been few systematic attempts to spell out this relation. Perhaps the most explicit attempt is by Roebers (2017), who claims that executive function and metacognition play quite similar roles in children's behavior and cognition: "Both are higher-order cognitive processes enabling an individual to operate flexibly and adapt efficiently to new and challenging tasks … [Both] similarly encompass dynamic and regulatory functions, which are utilized to optimize information processing of more elementary, first-order tasks" (p. 33). She argues that in the way they are studied in the current literature the two functions comprise different "sub-processes." Paraphrasing slightly to emphasize aspects relevant to the current account, for executive function she identifies such things as attention shifting, behavioral updating, and behavioral inhibition, and for metacognition she identifies cognitive monitoring and cognitive control.

My paraphrases (i.e., adding in explicit reference to "attention," "behavior," and "cognition") are meant to emphasize the proposal I will defend here, namely, that executive function comprises cognitive processes that regulate attention and action, whereas metacognition regulates these executive-level cognitive processes themselves. Both are regulatory processes but operating at different psychological levels.

In this essay, I outline a theoretical approach to executive and metacognitive processes within a theory of human agency and its self-regulation, including processes of shared agency involving cooperative/normative self-regulation. After explicating the evolutionary foundations of the model, I spell out some of its implications for how best to conceptualize executive and metacognitive processes in human ontogeny.

1 Types of human agency

Tomasello (2022, in press) proposes a theory of human agency, decision making, and action that incorporates executive and metacognitive processes as two types of self-regulation. Beginning with a control systems account of agentive action, executive processes monitor and control action and attention in goal pursuit on an executive tier of operation, and metacognitive processes monitor and control those executive processes on a second-order metacognitive tier of operation—both with the function of facilitating effective and efficient behavioral decision making.

Figure 1 graphically illustrates the basic model. Although I know of no existing models of executive function and metacognition that take precisely this two-tiered form, there are existing hierarchical models of executive processes in both the adult (e.g., Koechlin and Summerfield, 2007) and developmental (e.g., Zelazo, 2004, 2015) literatures that focus on different phenomena than the current model. In particular: (i) the main focus of Zelazo's model is on consciousness (whereas I do not mention it); (ii) the structure of his model is detailed information processing (which I do not discuss); and (iii) his focus is on the complexity of rules that children can formulate and follow in adult-structured tasks (whereas I do not focus on rules at all). Also in developmental psychology, Carlson (2023) has recently begun investigating "reflection" (presumably a metacognitive process) in the context of executive function and the effect that children's sense of agency has on their cognition and motivation, which also is not in my model.

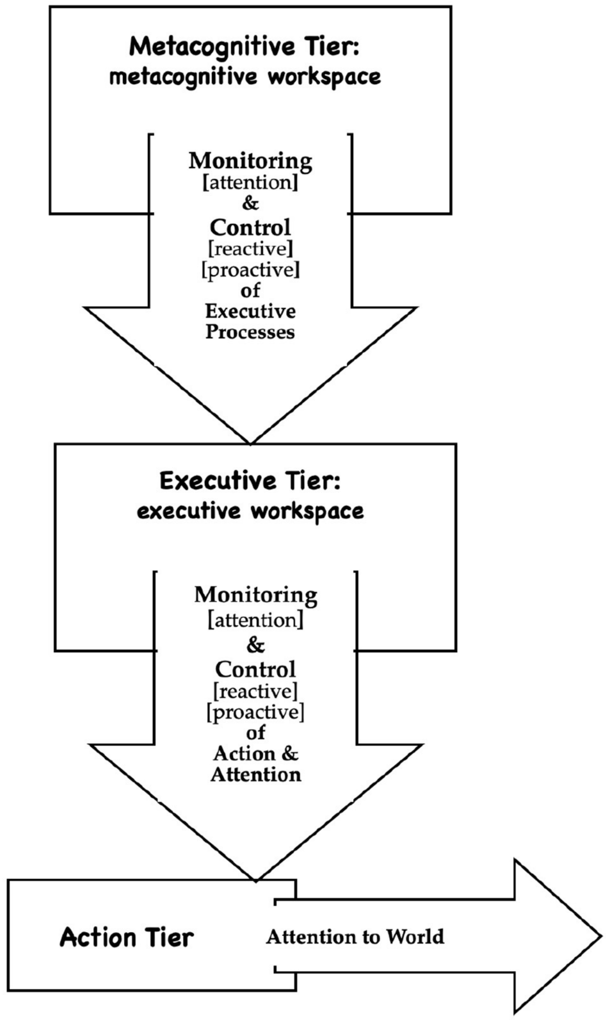


FIGURE 1
Graphic depiction of the hierarchical model of executive and metacognitive regulation.

1.1 Phylogeny

Tomasello (2022) proposes an account of how this human psychological architecture built up over evolutionary time. The model begins with the basic premise that cognitive processes evolved to facilitate agentive decision making and action. Not all organisms operate with cognitive processes, but rather their behavior is reflexive or stimulus driven because natural selection can anticipate the predictable arrival of particular stimuli and needed responses (examples in humans are breathing and swallowing). But in situations of unpredictability and uncertainty, what has evolved is an architecture of agentive decision making in which the individual perceptually and cognitively assesses the situation and makes a decision about what it can do to best pursue its goals. The computational model for agency is cybernetic control systems such as thermostats and self-driving cars that pursue and maintain reference values in dynamically changing circumstances.

If we focus on the species forming an evolutionarily line to humans, there have been three basic forms of individual agentive organization.

- **Goal-directed agency** evolved in the first vertebrates. This architecture is a simple control system sufficient for the organism to make a go/no-go decision (action tier only in Figure 1). These creatures were restricted to this mode of decision making because they had no executive tier of proactive executive control (although they were capable of a simple process of global reactive inhibition or "freeze" response).
- **Intentional agency** evolved in the first mammals. This architecture is a control system supervised by an executive

tier of functioning (executive and action tiers in Figure 1) with skills of thinking and planning sufficient for making an either/or decision between cognitively represented possibilities. These new types of decisions required proactive types of inhibitory control (e.g., suppression of unchosen behavioral options before acting) and executive coordination of attention.

- **Metacognitive (or rational) agency** evolved in the first great apes. This architecture is a control system supervised by an executive tier of functioning supervised by a metacognitive tier of functioning (metacognitive, executive, and action tiers in Figure 1) sufficient for reflecting on decisions already made and assessing their appropriateness given new information. These new types of decisions required metacognitive monitoring and control of executive decision making and metacognitive coordination of thinking and planning.

This hypothesized evolutionary trajectory reflects a natural buildup in complexity over evolutionary time, a common occurrence in biological systems of all types in which subsequent forms build on already existing forms (Bonner, 1988).

In addition, early humans also evolved some species-unique forms of shared agency based on cooperative goal pursuit and cooperative self-regulation, as they evolved more cooperative and cultural lifeways.

- Shared agency evolved in the early humans, who collaborated with others to make shared decisions in pursuit of shared goals. To do this they needed to coordinate with an individual partner (in a joint agency) or with the cultural group at large in terms of its conventions and norms (in a collective agency) and to collaboratively self-regulate these agencies normatively.

Evidence for this overall account comes from a wealth of behavioral experiments with contemporary model species: lizards as exemplars of the first land vertebrates acting as goal-directed agents; squirrels as exemplars of the first mammals acting as intentional agents; and chimpanzees as exemplars of the first great apes acting as metacognitive agents. The two forms of shared agency are connected to two early hominin species: *Homo heidelbergensis* as exemplars of the first joint agents, and *Homo sapiens sapiens* as exemplars of the first collective agents. The hypothesis is that there was a gradual transition from one form of agency to another across species in a perfectly normal process of evolution by means of natural selection.

1.2 Ontogeny

Tomasello (in press) argues that these same basic architectures structure children's cognitive development today, that they emerge at predictable ages, and that they both empower and constrain children's learning at particular ages. They emerge normally along the following general timeline.

- **Goal-directed agency** emerges in early infancy and operates throughout the first 9 months of life. Infants make only go/no-go decisions and operate with no executive processes other than a kind of global inhibition enabling them to freeze whatever they are doing and move on to another go/no-go decision.
- **Both intentional agency and joint agency** emerge at 9 to 12 months of age and predominate in toddlerhood until about 3 to 4 years of age. Toddlers make either/or decisions made possible by the emergence of an executive tier on which the toddler cognitively simulates possible actions and their likely results, regulating her attention and action via proactive thinking and planning. Toddlers also participate in joint agencies coordinating attentional perspectives and actions with others.
- **Both metacognitive agency and collective agency** emerge at 3 to 4 years of age and predominate in early childhood until about 6 years of age. Preschool youngsters make reflective decisions made possible by the emergence of a metacognitive tier on which the child regulates her executive-tier thinking and planning metacognitively. Preschool youngsters also coordinate their thinking, decision making, and perspectives metacognitively with peers in both joint and collective agencies.

The hypothesis is thus that there are qualitative shifts at 9–12 months and at 3–4 years of age in processes of psychological self-regulation. Specifically, from 9 months to 3 years of age children begin to executively regulate their actions and attentional perspectives proactively via thinking and planning—as well as those of partners in joint agencies. From 3 years of age onward children begin to metacognitively regulate their executive-tier thinking and planning via the coordination of conceptual perspectives—as well as normatively regulating others' and their own thinking and conceptual perspectives in both joint and collective agencies.

1.3 Novel features of the model

It is challenging to relate this agency-based model to the developmental literature on executive function and metacognition. The problem is that developmental psychologists have studied a variety of specific processes under these names, but these are typically defined in fairly narrow research contexts, leading to a proliferation of theoretical constructs. Thus, as the broader term, executive function includes such things as behavioral inhibition, cognitive inhibition, inhibitory control, self-control, effortful control, proactive executive function, continuous monitoring, working memory, self-regulation, emotion regulation, attentional control, attention shifting, attention regulation, cognitive flexibility, set shifting, task switching, and others. Although there is no consensus in the field, a widely used typology is that of Diamond (2013), who differentiates: (i) Inhibition (e.g., inhibitory control, self-control, behavioral inhibition, emotion regulation, etc.); (ii) Working Memory (i.e., holding information in mind and mentally working with it in various ways); and (iii) Cognitive Flexibility (e.g., attention shifting, set shifting, mental flexibility, etc.). This typology has proven useful in identifying individual differences in developmental outcomes such as school achievement and emotional adjustment. But many researchers have bemoaned the plethora of terminological jargon in the field, and some have doubted the psychological reality of this menagerie of constructs (e.g., Doebel, 2020).

The main issue is that the types in Diamond's typology are very diverse: "inhibition" is a basic psychological process, "working memory" is a cognitive workspace within which processes operate, and "cognitive flexibility" is a trait that people or processes possess. In contrast, in the current model executive and metacognitive processes are not just a collection of independent mechanisms; they each play a distinct role in a regulatory system evolved to monitor and control agentive decision making and action. We may thus rework Diamond's tripartite typology in the context of the current model in the following way. First, "working memory" is an attentional workspace, and there are two types: one is an executive workspace (on the executive tier in Figure 1) that monitors and controls action and attention, and the other is a metacognitive workspace (the metacognitive tier in Figure 1) that monitors and controls these executive processes.[1] Second, inhibition is one of the main regulatory processes that takes place in these workspaces. But there are others, in particular processes that are more proactive such as planning and the coordination of thoughts and perspectives. Indeed, there can even be reactive and proactive processes of inhibitory control. Therefore, such things as "inhibition" and "cognitive coordination" in Diamond's typology may be recast as the actual regulatory processes by means of which agents monitor and control their decision making, processes such as thinking, planning, inhibitory control, coordination of thoughts, etc. Third, in this context, I would like to make a novel proposal—to be fleshed out in the next section that what Diamond and others call "cognitive flexibility" is about the coordination of perspectives, and this arises mostly in shared agencies in which individuals monitor and control one another's actions, attention, and perspectives. This interactive process is then internalized such that the individual can coordinate perspectives on things flexibly on her own.

To assess this model, in the coming section I empirically evaluate two hypotheses: (1) the hypothesis that there are systematic age-related changes in the organization of agency and decision making that structure the regulatory processes involved: first the executive regulation of attention and action beginning at around 9 to 12 months of age and then the metacognitive regulation of thinking and decision making beginning at 3 to 4 years of age; (2) the hypothesis that important aspects—indeed most of the species-unique aspects of children's cognitive flexibility arise initially from their participation in shared agencies in which they must coordinate their own actions, attention, perspectives, and decision making with those of a partner or a group with whom they are acting interdependently.

1 One could potentially posit emotion as something else to be regulated on the basic tier of action and attention. But what one is monitoring and controlling in such cases is less the involuntary emotions themselves and more their behavioral expressions and/or their effects on one 's actions.

2 The ontogeny of human agency, decision making, and self-regulation

Most research on children's executive function and metacognition uses standardized tasks—often asking children to follow adult-specified rules and focuses on individual differences in children's performance. My focus here, in contrast, is on the kinds of spontaneous self-regulation that characterize all of children's agentive decision making and action throughout their daily lives.

The proposal is that how children make decisions and regulate them depends on the cognitive architecture within which they are working, which includes one or another type of cognitive representation and self-regulative workspace. Further, self-regulation can be more reactive (e.g., inhibiting ongoing action or cognition) or more proactive (e.g., planning and coordinating action and cognition before acting). Finally, shared agency requires flexible interpersonal coordination—sometimes even shared decision making and collaborative and/or normative self-regulation. My focus in this section is on how these things all work together in the agentive decision making and self-regulation of, in turn, young infants (0 to 9 months), toddlers (9 months to 3 years), and preschoolers (3 to 6 years).

2.1 Young infants as goal-directed agents

The capacity for goal-directed action requires young infants (below 9 months) to make decisions about whether or not to execute an action in a particular situation, that is, go/no-go decisions. Despite appearances, they are not making either/or decisions about which action to perform. Thus, at first blush, it would seem that infants do make either/or choices between alternatives. For example, Hamlin et al. (2007) presented 6-month-old infants with two stuffed animals, one of which had behaved more nicely than the other. Infants tended to touch or grab the nice animal, which could be taken as evidence of an either/or decision between the two options. But it is also possible that in their initial observation of the animals' behavior infants developed an attraction to the nicer animal, and as soon as they saw it, they went for it without comparing the relative values of the two different options. Under this interpretation, they are making a go/no-go decision for an attractor, not an either/or choice among alternatives. It is only after 9 months of age that young toddlers make either/or choices among alternatives.

Evidence for this interpretation comes from studies in which infants and toddlers have a prepotent tendency to go for a "wrong" option. The point is that if they succeed in overcoming this prepotent tendency, it suggests that they have attended to both alternatives and made an either/or decision. A good example is action-based object permanence tasks. If a desired object is hidden under a single cloth, 8-month-old infants quickly remove the cloth and retrieve the object. But at this same age they often make the famous A-not-B error. This error occurs in a version of the task in which the infant is confronted with an object hidden under one of two cloths. After she finds it under cloth A, it is placed in plain sight under cloth B. In this two-cloth situation, infants often search for the hidden object under the cloth where they last found it (A), rather than where they last saw it disappear (B). They make this error through the end of early infancy, first searching reliably for the object in its new location (inhibiting any prepotent attraction to the first location) only as toddlers at around 11 months of age (Diamond, 1985; Marcovitch and Zelazo, 2006). The important point is that the single-cloth task only requires the infant to make a go/no-go decision (to remove the cloth or not), whereas in the A-not-B task she is seemingly confronted with an either/or decision between the two cloths, each of which is a salient alternative for good reason. Young infants' behavior in detour tasks is similar. If a desired object is placed behind a transparent glass barrier, infants up to 11 months of age tend to just reach directly for the toy and bang into the glass (Diamond and Gilbert, 1989; Diamond, 1990). They cannot overcome this prepotent tendency and so choose the reach-around alternative, even after seeing this prepotent tendency fail several times, which implies, again, that they are not choosing between the two alternatives but simply seeing an opportunity to grasp an object and going for it. And again toddlers after 11 months of age succeed in choosing the less salient alternative action in this task.

The hypothesis is thus that young infants' actions are generated by a process of decision making that simply determines whether to perform a particular action in the situation at hand: is this an opportunity for a particular goal-directed action? One might propose that the issue for infants is not decision making but inhibitory control, and this would not be totally incorrect. But either/or decision making and inhibitory control go hand-in-hand in the sense that choosing among options means inhibiting the unchosen option before acting. I would thus characterize the issue more broadly. The issue, in the current hypothesis, is that infants before 9 months of age do not have an executive tier of functioning that can simulate alternative action possibilities and their likely outcomes before acting, and so they do not yet have the possibility of either/or decision making with proactive inhibition of unchosen behavioral alternatives. It is interesting that attempts to measure individual differences in inhibition in infants before 9 months of age mostly involve so-called delayed response tasks (e.g., Diamond, 1990), which only measure something like global inhibition of a single action and not selective (proactive) inhibitory control of one alternative in comparison to another before acting.

2.2 Toddlers as intentional and joint agents

In contrast to young infants (before 9 months), toddlers make either/or behavioral decisions in which they imagine behavioral options with their likely outcomes and then choose one before acting. This is what Berkman et al. (2017) call "value-based choice," in which the preferred option is increased in value, and/or the less preferred option is decreased in value, relative to the other(s). Value-based choices require imaginative representations, that is, representations of actions and states of affairs that are not currently the case but could become actually the case.

One can see the origins of 9-month-old toddlers' either/or decision making already in their behavior in the two-cloth object permanence task. Soon after 9 months of age toddlers stop making the A-not-B error: they choose which of the two cloths is likely concealing the desired object and choose that one. This value-based choice involves a more flexible form of inhibitory control than the simple global inhibition characteristic of infants. As they are comparing behavioral options, choosing one involves suppressing the other, often prepotent, tendency such as removing the A cloth where the toy was previously found. In support of this interpretation, much research shows that toddlers' ability to make choices in this manner correlates strongly with other tasks measuring inhibitory control (Marcovitch and Zelazo, 2006). Moreover, either/or comparisons of this kind should take time to execute, and Kim et al. (2020) found that when 12- and 24-month-old toddlers are faced with more uncertainty in their potential choices, they take more time to decide. In general, toddlers seem to be making either/or decisions involving processes of proactive inhibitory control before acting.

Perhaps even clearer evidence for this kind of decision making comes during this same age range as toddlers make decisions in so-called opt-out tasks requiring them to compare options before choosing. A number of mammalian species—including dolphins, rats, and many non-human primates—have been confronted with a choice between an easy-to-obtain low reward and a more difficult to obtain high reward. When chances of obtaining the high reward are high, individuals will go for that; but when chances of obtaining the high reward are low, individuals often opt out and go for the easy-to-obtain low reward. Goupil et al. (2016) tested 20-month-olds in a situation with this logic (the opt-out response in this case was to request adult help) and found that toddlers made efficient choices. Further, Call and Carpenter (2001) found that when 30-month-olds felt uncertain about a decision, they actively sought more information to try to make a better decision, again showing the ability to comparatively evaluate alternative possible actions. The toddlers are monitoring their confidence or uncertainty in a value-based choice, and then responding appropriately.

But perhaps the strongest evidence comes from another experimental paradigm aimed at children's decision-making. The situation is slightly different from uncertainty monitoring in that the costs and risks of both possible choices are clear at the outset (often with one having a kind of prepotent attraction). Thus, Herrmann et al. (2015) confronted 36-month-olds with a spatial discounting task in which the child first spied a nearby small reward and then a farther-off large reward, and they were shown that going for one meant forsaking the other. They had to compare the two situations and make their choice before acting, which prevented a sequential guessing strategy involving only a sequence of go/no-go decisions. In a similar task toddlers had to choose one of two behavioral strategies given that the situation had noticeably changed, which meant inhibiting a previously successful action in favor of a new one demanded by a changed situation (again they had to choose before acting so that a sequential guessing strategy was not possible). In both of these tasks, toddlers were generally successful, equally as good as chimpanzees (but not as good as 6-year-olds).

Toddlers' behavior in all these tasks thus suggests either/or, value-based choices between two simultaneously available courses of action as they imagine them (in imaginative representations). Such value-based decision making among simultaneously available options cannot take place in creatures who operate as a simple goal-directed control system comprising only goals, actions, and attention. Rather, it requires control systems organization with an additional executive tier of monitoring and control to regulate the process of behavioral decision making.

In addition, from around their first birthdays, toddlers are able to form joint agencies with adults to do such things as build a block tower together, get the child dressed together, or walk the dog together. To create such joint agencies the two parties need to coordinate their actions and attention. One- and two-year-old toddlers are notoriously poor at coordinating with same-age peers (Brownell and Carriger, 1990), and they do not seem to participate in joint attention with same-age peers in anything like the way they do with adults either (see Tomasello, 2020a; for a review). The implication is that toddlers cannot really make joint decisions with others, but they can participate in joint agencies when an adult scaffolds the decision-making process. They coordinate actions and attention (but not decisions) with an adult (and not a peer) partner.

Modern conceptions of executive function view it as individual self-regulation, but joint agencies need to be self-regulated as well and this is a social process. In the beginning, toddlers do not participate much if at all in the coordination and self-regulation of the joint agency, as the adult scaffolds the process. But over time they come to coordinate their actions and attention with the adult more actively, sometimes by communicative acts aimed at the partner's actions and attention. My proposal is that it is these attempts at social and mental coordination with adults in joint agencies that create the uniquely human kinds of perspectival flexibility that are measured by the most basic tasks of attentional flexibility such as attention shifting, set shifting, and task switching (i.e., other species show these abilities, but not as flexibly humans). Of special importance are toddlers' newly emerging abilities of joint attention and cooperative communication that help them to establish and maintain joint agencies with others.

The process of establishing joint attention with a partner on some referential situation is not a one-shot, ballistically produced intentional action, but rather a process of cooperative coordination. Thus, indicating and identifying the referent of a pointing gesture (as done already by 12-month-olds) involves the coordination of attentional perspectives. In the prototypical case, one partner initiates things by pointing for the other to a referent that she (the communicator) is already attending to; her referential intention is the aligning of their attention in joint attention. The recipient, if he is being cooperative, goes from his own individual attention elsewhere to jointly attending with his partner. The interpersonal coordination thus involves each partner's sequential shifting from individual to joint attention, as either communicator or recipient, with adjustments as needed (Liszkowski et al., 2007). Unlike simply imagining what another person sees or knows, as occurs in many studies of infant social cognition, negotiating joint attention brings into focus the *relation* between self and other perspectives: to know that perspectives are or are not aligned there must be some imagining of the content of those perspectives and their relationship. Such negotiations require both imaginative representations and an executive workspace in which the two attentional perspectives may be imaginatively compared and coordinated.

From 9 months of age, then, toddlers are operating in a very different way from young infants. Young infants are perceiving and representing the actual world (even if it is behind an occluder at the moment). In contrast, toddlers are imagining possible courses of action and outcomes in the environment and basing their decisions on these imagined possibilities, a process which requires them to employ a kind of proactive inhibitory control in suppressing the imagined alternatives that they do not in the end choose. In addition, toddlers must coordinate attentional perspectives with adults in joint agencies, which requires them to employ a kind of attentional flexibility that is not needed by young infants (and non-human animals) who do not engage in joint agencies. Toddlers are able to do all this, in the current hypothesis, because of the maturation of a new cognitive architecture involving a single tier of executive monitoring and control, operating within a new executive workspace (executive working memory).

2.3 Preschool youngsters as metacognitive and normative agents

How animals and children use metacognition to make decisions is often studied using tasks of uncertainty monitoring. For example, when presented with a difficult discrimination or memory problem, many animal species and preschool children opt out and go for a safer alternative: in one interpretation, they know that they do not know. But there is controversy over whether opting out in such cases actually requires metacognition in the strict sense of the term (e.g., see papers in Beran et al., 2012). The key issue in the current context is whether children younger than 3 years of age are able to metacognitively reflect on the decision making process.

In a few studies researchers have claimed metacognitive decision making in 2-year-old toddlers. Specifically, in two studies already described above, when 2-year-olds were uncertain about their ability to solve a behavioral problem, they recruited a parent for help (Goupil et al., 2016), and when 2-year-olds did not see where an adult hid a toy—so they were uncertain where it was—they actively looked behind a barrier to gain needed information (Call and Carpenter, 2001). These two studies are sometimes characterized as involving metacognition under the interpretation that the toddlers "know that they do not know." However, a different interpretation is that the toddlers in these studies are not metacognitively monitoring what they do and do not know, but rather they are executively monitoring what they can and cannot do: whether proceeding with a planned action is or is not likely to be successful in reaching the goal. In the view of Goupil and Proust (2023), monitoring behavioral uncertainty in this manner is not monitoring a thought but rather monitoring a *feeling*. That is, the toddlers are executively monitoring a feeling of uncertainty as they go about choosing an action on the behavioral tier of operation, not metacognitively monitoring the executive-tier cognitive processes they are using to make that decision. Goupil and Proust (2023) actually refer to this type of uncertainty monitoring as a procedural form of metacognition, that is to say, a form that focuses not on cognition proper but on ground-level processes of action and attention. I would thus characterize these two studies as concerned with the executive supervision and control of action and attention.

Then, beginning sometime after 3 years of age, with the development of the metacognitive tier of agentive architecture, young children become able to metacognitively monitor and control not just the feeling of behavioral uncertainty but the cognitive processes involved in executive decision making itself. That is, they become able to metacognitively monitor their executive-tier processes of thinking, planning, and decision making to decide among different possible either/or decisions, including revising already made decisions and beliefs in the light of new evidence or reasons. This takes place in two different forms. One takes place within the agent's mind, as it were, as young children plan and evaluate their own executive-tier decisions before making a final decision, or perhaps reassess things after a decision has been made if new information becomes available. The other takes place between agents' minds, as it were, as young children coordinate decisions with others in joint or collective agencies, jointly attending to the beliefs and reasons involved.

First, within minds, O'Madagain et al. (2022) gave both great apes and human children (3 and 5 years of age) the opportunity to visually locate the best food at location X. The subjects did this, indicating their belief/decision by choosing that location (though not receiving the food as a result). Then, they were exposed to new information that called their initial belief into question, information suggesting that the best food might be in location Y. Subjects then had the possibility to seek further information (or not) that could either confirm or disconfirm their initial belief. Many apes then actively sought more information to resolve the discrepancy between their original belief and the new information, by looking again into location X (and perhaps Y) to check their initial judgment so as to make the best decision. The apes were in this case metacognitively assessing their executive decision *after they had already made it* (which distinguishes the demands of this task from those of the two toddler studies described above); they were reflecting on the belief guiding their decision in the light of newly obtained information and discerning the need to possibly revise that belief and so decision. If this is indeed what they were doing, it is important because attempting to causally diagnose problematic decisions before they are behaviorally executed fulfills a standard criterion for reflective agency, and it clearly is metacognitive.

Like the apes, the human children in this task questioned their own belief and actively attempted to double-check it—but only at 5 years of age. The children at 3 years of age just went with one or the other choice without double-checking. However, in a second study, O'Madagain et al. (2022) provided apes and children with discrepant information in a different manner: the subject made an initial choice, again without receiving anything as a result, and then a conspecific entered and indicated a different choice. In this case, the apes did not double-check their initial choice, presumably because they did not compare the perspectives of themselves and the peer. In contrast, the human children actively double-checked their initial choice, and they did so even at 3 years of age! This suggests that, in contrast with apes, young children find different perspectives emanating from social partners to be more salient indicators of the need for belief revision than new information emanating from the physical world. In their individual decision making, young children are especially attuned to discrepant social perspectives, which prompt them (i.e., more strongly than physical evidence) to metacognitively reflect on and revise their beliefs and so decisions.

Second, between minds, in shared agencies preschool children for the first time begin to mentally coordinate with peers to make truly joint decisions in joint agencies. Whereas 2-year-old toddlers can to some degree coordinate their ongoing actions and attention with adults, preschool children can plan and coordinate their actual *decisions* with others, including peers. The process of coordinating not just actions but decisions is studied formally in game theory in what are called coordination games. A well-known coordination game is the stag hunt. In the classic parable, I am hunting alone for hares when I spy a stag, which is more and better food but which I cannot capture alone. You are in the same situation, and so it is in both our interests to drop our pursuit of hares and collaborate to capture the stag. The problem is that neither of us can be certain that the other will choose to go for the stag (maybe our partner did not see or hear the stag). Chimpanzees do not perceive the stag hunt as a dilemma: they just go for the stag and hope the other will follow. But 4-year-old children perceive the dilemma and so before leaving the hare they communicate to make a joint decision (Duguid et al., 2014). They are monitoring their partner and the possibilities for fruitful collaboration and making their behavioral decisions accordingly. Four- and 5-year-old preschoolers can even coordinate their decisions in situations in which the possibility of communication is eliminated, that is, in games of so-called "pure coordination" (which great apes cannot do; Duguid et al., 2020). That is, they are able to coordinate their decisions if there is some salient feature of one of the choices—e.g., one is red while all the others are white—which they can metacognitively predict will be a salient decision for their partner, whom they know is attempting to metacognitively predict their decision as well (Grüneisen et al., 2015). Moreover, children in this same age range are even able to plan a coordinated decision in a joint problem-solving situation by each partner determining which tool each of them must choose in her role and then coordinating their respective choices accordingly (Warneken et al., 2014).

Once a joint agency with a peer is formed, preschool youngsters attempt to self-regulate it through various forms of action and communication. For example, if the peer does not play her role adequately in their collaboration, the child protests normatively using words such as *should* or *must* or *ought*—to bring the wayward partner back into line. Often preschool peers initiate a collaborative activity with a joint commitment ("Let's do X," agreed to with "OK"), and so the normative protest is then referring the partner back to "our" agreement to collaborate. That is to say, the child is self-regulating the collaborative activity, in an important sense collaboratively, by referencing the original formation of the shared agency and their individual responsibilities in it. This kind of normative self-regulation can be characterized as we > me normative self-regulation (Tomasello, 2020b).

In more discourse-based studies of decision making with peers, pairs of 3- and 5-year-olds are able to coordinate a joint decision by metacognitively comparing their different beliefs and even reasons for their beliefs—through perspective-taking discourse and joint reasoning. For example, in one study peer partners had different information from different sources about what some novel creatures typically eat. To resolve the issue, they metacognitively discussed the validity of the evidential sources from which they each had obtained their information (hearsay vs. direct observation) and came to a reasoned joint decision as a result (Köymen and Tomasello, 2018). The point is that in these joint problem-solving situations peers coordinate not just their actions but their decisions, which requires each of them to metacognitively monitor both their own and the partner's beliefs, as well as their respective reasons for their beliefs (see Köymen and Tomasello, 2020, for a review of these and similar studies).

Preschool youngsters are able to coordinate and regulate their decisions with peers because they are now operating with a new metacognitive tier of functioning that enables them to conceptualize and socially coordinate executive-tier cognitive processes such as beliefs and reasons, with which they, from 3 to 4 years of age, are operating. The Vygotskian hypothesis is that it is precisely this kind of social coordination of beliefs, reasons, and decisions with others that is the original source of preschool children's individual cognitive flexibility and conceptual perspective taking, as they internalize the social process into an internal dialogue which they use to deliberate on their own. The O'Madagain et al. (2022) study described above (in which children metacognitively examined their own beliefs more readily in the face of a discrepant social perspective than discrepant physical information) is generally consistent with this view. Also supportive is the study of Köymen et al. (2020) in which adults trained 3-year-olds in a kind of "meta-talk" about reasons, evidence, and their validity, and this led the children later to engage in more skillful joint decision making with peers.

Relatedly, there are also significant developments in individual cognitive flexibility at around 3 to 4 years of age as well. In preschoolers, cognitive flexibility is classically measured by tasks such as the Dimensional Change Card Sort task (DCCS; Zelazo, 2006). In the DCCS task children are required first to sort cards on one dimension (e.g., color) and then immediately sort them by another (e.g., shape). Early research tended to show that 3-year-old children had trouble classifying objects in a second way (see Doebel and Zelazo, 2015; for a review). But subsequent research employing more child friendly versions of the task has found that performance is quite good at 3 to 3.5 years of age whereas it is very poor at 2.5 years of age (e.g., Blakey et al., 2016). So age 3 would seem to be the key age of transition for successive multiple classification. Simultaneous multiple classification is most often assessed using a matrix completion task. In this task, children must find the missing object in a matrix created by crossing two dimensions, for example, placing a red triangle in the missing space defined by the convergence of a red vertical dimension and a triangle horizontal dimension. Again, early studies showed that 3- and 4-year-old children struggle with this task, but Podjarny et al. (2017) designed a more child friendly version and found that both 3- and 4-year-olds were quite competent. Interestingly and importantly, Podjarny et al. (2022) administered child friendly versions of both a successive and simultaneous task of multiple classification and found that young children were consistently better at the successive version.

What explains this relatively sudden competence at 3 to 4 years? Based on a series of nine studies using the DCCS (as well as a review of relevant literature), Zelazo et al. (2003) concluded that children's performance was not best explained by developments either in memory or in inhibitory control. The best explanation was what they called a "redescription account" (championed most prominently by Perner and Lang, 2002), which attributes growing success to young children's developing ability to appreciate multiple conceptual perspectives on the same object(s) at the same time. But why does this new ability emerge only at around 3 or 4 years of age? In the current account, the obvious reason is that 3 years is the age at which the new metacognitive tier of regulation emerges, and this enables children to re-represent all of the simple categorization activities in which they have been participating for several years already. So perhaps on one occasion the child labeled an object a "bird" and then on another occasion noted that it was a "cardinal," or on one occasion she singled out the ovals from a group of blocks and on another occasion singled out the blue ones. These acts create discrepancies in that the same object is conceptualized as different things on different occasions. Reflective thinking and re-representation on the metacognitive tier use these kinds of discrepant experiences as the raw material to coordinate and perhaps synthesize different conceptual perspectives on the same entities to enable multiple classification of the same object in different ways for all kinds of creative purposes, first successively and then, in certain contexts, simultaneously. This happens most frequently and most saliently in collaborative social interactions with others, including both adults and peers. Thus, if a child came to maturity on a desert island with no social interaction, she would not learn to take different perspectives on things and integrate them.

Finally, a more explicitly social kind of cognitive flexibility comes out in a variety of tasks that Perner et al. (2003) call "perspective problems," that is, problems that bring different conceptual perspectives into conflict (though this is often only apparent and can be resolved), requiring the child to do some kind of coordination of perspectives (perhaps especially with peers) to make sense of things. For example, from her viewing angle a child may see a drawing of a horse as right side up, but a partner on the other side of the table claims that it is upside down. How to resolve the situation? Further, the child may initially believe that something is a rock, but another person uses it as a sponge. How can something appear to be one thing but be used as another? Or the child may know that an object is a tree but hear someone else call it a "bush," or know that an object is a dog and hear someone else call it an "animal." How can an object be two things at once? Or, most famously, the child may know that an object has been hidden in one place but an agent who did not witness the hiding process believes it is in a different place. How to coordinate the child's own perspective, the agent's perspective, and alight on the objective perspective of where the object really is? To construct the necessary concepts to resolve these conflicts, the child needs to flexibly coordinate conceptual perspectives on the world (Tomasello, 2018). Children are typically not successful in any of these tasks (visual perspective taking, appearance reality, dual naming, or false belief) until 3 or 4 years of age. In the current hypothesis that is because they are not able to metacognitive only reflect on their own and others conceptual perspectives from a metacognitive tier of operation and

	Toddlerhood EXECUTIVE/JOINT [Executive Workspace]		Early Childhood METACOGNITIVE/COLLECTIVE [Metacognitive Workspace]	
	Reactive	**Proactive**	**Reactive**	**Proactive**
Monitor *Individual* Agency	Ongoing own action & attention	Anticipated own action & attention	Ongoing own thinking & D-M	Anticipated own thinking & D-M
Control *Individual* Agency	Bolster/inhibit own action & attention	Plan/coordinate own action & attention	Bolster/inhibit own thinking & D-M	Plan/coordinate own thinking & D-M
Monitor *Shared* Agency	Ongoing partner action & attention	Anticipated partner action & attention	Ongoing partner thinking & D-M	Anticipated partner thinking & D-M
Control *Shared* Agency	Praise/protest partner actions & attention	Direct partner action & attention	Praise/protest group thinking & D-M	Direct group thinking & D-M

FIGURE 2
Summary of what is regulated, how it is regulated, and where (in which workspace) it is regulated—for both individual and shared agencies in toddlerhood and early childhood ("D-M" refers to decision making).

coordinate them effectively which, again, occurs most readily and most saliently in their social interactions with others, especially peers, which could in principle be empirically evaluated in some kind of training study.

3 The regulation of agency

In order to bring all of the different aspects of these various self-regulatory process together, I propose focusing on three key components of agentive self-regulation: (i) what is regulated, (ii) how, via what processes, it is regulated, and (iii) where, in what cognitive workspace, it is regulated. First, the proposal is that in individual agencies during toddlerhood what is regulated is basic-level things like action and attention, whereas during early childhood it is also more cognitive things like thinking and decision making. Second, these are all regulated by processes of monitoring (a higher tier attends to one below it) and control in terms of (a) reactive processes (like reactive inhibition), and (b) proactive processes (like anticipatory coordination). Third, this is all done in either an executive workspace on an executive tier during toddlerhood or a metacognitive workspace on a metacognitive tier during early childhood—and either in individual or joint agencies. Figure 2 thus provides a typology of agentive self-regulation in terms of four main dichotomies:

- executive vs. metacognitive workspace
- monitoring vs. controlling as distinct phases of self-regulation
- reactive vs. proactive regulatory processes
- self-regulation in individual vs. shared agencies

Then, in addition, the target of regulation—what is regulated—is shown in the cells of Figure 2 in terms of what behavioral and/or cognitive processes are being regulated (and, in the case of control, a bit about how they are controlled).

The current model thus has three distinctive features relative to other treatments of executive function and metacognition in the literature. First, the model is articulated within the context of an overall theory of human agency and decision making, which provides coherence and functional continuity. Second, the model is hierarchically structured such that executive and metacognitive processes are two analogous control systems operating on different material from different tiers (workspaces) of agentive architecture—emerging at different ages. And third, the account of shared agency provides a principled account of uniquely human processes in the coordination of perspectives—both attentional and conceptual—that integrates what has traditionally been called cognitive flexibility with other processes of agentive self-regulation, as well as specifying unique processes of collaborative or normative self-regulation.

The specific mechanisms of executive and metacognitive function currently posited in the literature (often defined by cognitive tasks) simply reflect a focus on one or another sub-process in this overall regulatory architecture, or else a specific application of these. Particular tasks measure one or more of these processes made more specific in the context of that task. For example, various go/no-go tasks (e.g., delayed response) would be inhibition of action (either reactive or proactive); effortful control and emotion regulation would mostly be inhibition of emotional expression (either reactive or proactive); attention shifting and task switching would be the coordination of attention or action depending on the task; the DCCS would be coordinating conceptual perspectives successively; matrix completion would be coordinating conceptual perspectives simultaneously; and so forth. The model thus provides a theoretical vocabulary for relating specific processes to one another. It is possible that it could be extended to older children to account for some of the phenomena of self-directed cognitive control studied by Frick and Chevalier (2023).

The current proposal is not intended to replace any of the important work that has been done in the study of either executive function or metacognition. Studies of inhibitory control, effortful control, continuous monitoring, working memory, emotion regulation, attention shifting, set shifting, task switching, etc., need to be described at a more detailed level—in terms of the specific task context and demands—than the very coarsely cut categories in the current model in Figures 1, 2. The current model is simply an attempt to provide a larger psychological framework within which current research may be categorized and interrelated. The hope is that keeping the various phenomena in their appropriate theoretical places may provide a unifying framework within which the menagerie of theoretical constructs in the field may be meaningfully interrelated and so spur further research progress.

4 Conclusion

In closing, what I am offering here is a way of unifying executive and metacognitive processes within an overall psychological architecture of agentive decision making and its self-regulation, one that unfolds in a distinctive, two-step developmental pathway. I also believe that integrating social agencies into this account provides additional dimensions of the process of agentive self-regulation especially the proactive coordination of perspectives and normative self-regulation that can broaden the scope of research into executive and metacognitive processes as they emerge and shape young children's cognitive and social development.

Author contributions

MT: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares 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

References

Beran, M., Perner, J., and Proust, J. (2012). *Foundations of Metacognition*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199646739.001.0001

Berkman, E., Hutcherson, C., Livingston, J., Kahn, L., and Inzlicht, M. (2017). Self-control as value-based choice. *Curr. Direct. Psychol. Sci.* 26, 422–428. doi: 10.1177/0963721417704394

Blakey, E., Visser, I., and Carroll, D. (2016). Different executive functions support different kinds of cognitive flexibility: evidence from 2-, 3-, and 4-year-olds. *Child Dev.* 87, 513–526. doi: 10.1111/cdev.12468

Bonner, J. (1988). *The Evolution of Complexity By Means of Natural Selection*. Princeton, NJ: Princeton University Press.

Brownell, C., and Carriger, M. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Dev.* 61, 1164–1174. doi: 10.2307/1130884

Call, J., and Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cogn.* 3, 207–220. doi: 10.1007/s100710100078

Carlson, S. M. (2023). Let me choose: the role of choice in the development of executive function skills. *Curr. Direct. Psychol. Sci.* 32, 220–227. doi: 10.1177/09637214231159052

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Dev.* 56, 868–883. doi: 10.2307/1130099

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching a. *Ann. NY Acad. Sci.* 608, 637–676. doi: 10.1111/j.1749-6632.1990.tb48913.x

Diamond, A. (2013). Executive functions. *Ann. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Diamond, A., and Gilbert, J. (1989). Development as progressive inhibitory control of action: retrieval of a contiguous object. *Cogn. Dev.* 4, 223–249. doi: 10.1016/0885-2014(89)90007-5

Doebel, S. (2020). Rethinking executive function and its development. *Perspect. Psychol. Sci.* 15, 942–956. doi: 10.1177/1745691620904771

Doebel, S., and Zelazo, P. (2015). A meta-analysis of the dimensional change card sort: implications for developmental theories and the measurement of executive function in children. *Dev. Rev.* 38, 241–268. doi: 10.1016/j.dr.2015.09.001

Duguid, S., Wyman, E., Bullinger, A., Herfurth, K., and Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. *Proc. R. Soc. B.* 281:20141973. doi: 10.1098/rspb.2014.1973

Duguid, S., Wyman, E., Grüneisen, S., and Tomasello, M. (2020). The strategies used by chimpanzees (Pan troglodytes) and children (Homo sapiens) to solve a simple coordination problem. *J. Compar. Psychol.* 134:401. doi: 10.1037/com0000220

Frick, A., and Chevalier, N. (2023). A first theoretical model of self-directed cognitive control development. *J. Cogn. Dev.* 24, 191–204. doi: 10.1080/15248372.2022.2160720

Goupil, L., and Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition* 231:105325. doi: 10.1016/j.cognition.2022.105325

Goupil, L., Romand-Monnier, M., and Kouider, S. (2016). Infants ask for help when they know they don't know. *Proc. Natl. Acad. Sci.* 113, 3492–3496. doi: 10.1073/pnas.1515129113

Grüneisen, S., Wyman, E., and Tomasello, M. (2015). Children use salience to solve coordination problems. *Dev. Sci.* 18, 495–501. doi: 10.1111/desc.12224

Hamlin, K., Wynn, K., and Bloom, P. (2007). Social evaluation by preverbal infants. *Nature* 450, 557–559. doi: 10.1038/nature06288

Herrmann, E., Misch, A., and Tomasello, M. (2015). Uniquely human self-control begins at school age. *Dev. Sci.* 18, 979–993. doi: 10.1111/desc.12272

Kim, S., Sodian, B., and Proust, J. (2020). 12-and 24-month-old infants' search behavior under informational uncertainty. *Front. Psychol.* 11:566. doi: 10.3389/fpsyg.2020.00566

Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235. doi: 10.1016/j.tics.2007.04.005

Köymen, B., O'Madagain, C., Domberg, A., and Tomasello, M. (2020). Young children's ability to produce valid and relevant counter-arguments. *Child Dev.* 91, 685–693. doi: 10.1111/cdev.13338

Köymen, B., and Tomasello, M. (2018). Children's meta-talk in their collaborative decision-making with peers. *J. Exper. Child Psychol.* 166, 549–566. doi: 10.1016/j.jecp.2017.09.018

Köymen, B., and Tomasello, M. (2020). The early ontogeny of reason giving. *Child Dev. Perspect.* 14, 215–220. doi: 10.1111/cdep.12384

Liszkowski, U., Carpenter, M., and Tomasello, M. (2007). Pointing out new news, old news, and absent referents at 12 months of age. *Dev. Sci.* 10, F1–F7. doi: 10.1111/j.1467-7687.2006.00552.x

Marcovitch, S., and Zelazo, P. (2006). The influence of number of A trials on 2-year-olds' behavior in two A-not-B-type search tasks: a test of the hierarchical competing systems model. *J. Cogn. Dev.* 7, 477–501. doi: 10.1207/s15327647jcd0704_3

O'Madagain, C., Helming, K., Schmidt, M., Shupe, E., Call, J., and Tomasello, M. (2022). Great apes and human children rationally monitor their decisions. *Proc. R. Soc. B* 289:20212686. doi: 10.1098/rspb.2021.2686

Perner, J., Brandl, J., and Garnham, A. (2003). What is a perspective problem? Developmental issues in understanding belief and dual identity. *Facta Philosophica* 5, 355–378. doi: 10.5840/factaphil20035220

Perner, J., and Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant Child Dev.* 11, 93–105. doi: 10.1002/icd.299

Podjarny, G., Kamawar, D., and Andrews, K. (2017). The multidimensional card selection task: a new way to measure concurrent cognitive flexibility in preschoolers. *J. Exper. Child Psychol.* 159, 199–218. doi: 10.1016/j.jecp.2017.02.006

Podjarny, G., Kamawar, D., and Andrews, K. (2022). Two birds in the hand: concurrent and switching cognitive flexibility in preschoolers. *J. Exper. Child Psychol.* 220:105418. doi: 10.1016/j.jecp.2022.105418

Roebers, C. (2017). Executive function and metacognition: towards a unifying framework of cognitive self-regulation. *Dev. Rev.* 45, 31–51. doi: 10.1016/j.dr.2017.04.001

Tomasello, M. (2018). How children come to understand false beliefs: a shared intentionality account. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8491–8498. doi: 10.1073/pnas.1804761115

Tomasello, M. (2020a). The adaptive origins of uniquely human sociality. *Philos. Trans. R. Soc.* 375:20190493. doi: 10.1098/rstb.2019.0493

Tomasello, M. (2020b). The moral psychology of obligation. *Behav. Brain Sci.* 43, e56–58. doi: 10.1017/S0140525X19002620

Tomasello, M. (2022). *The Evolution of Agency: From Lizards to Humans*. New York: MIT Press. doi: 10.7551/mitpress/14238.001.0001

Tomasello, M. (in press). *Agency and Cognitive Development*. Oxford: Oxford University Press.

Warneken, F., Steinwender, J., Hamann, K., and Tomasello, M. (2014). Young children's planning in a collaborative problem-solving task. *Cogn. Dev.* 31, 48–58. doi: 10.1016/j.cogdev.2014.02.003

Zelazo, P. (2004). The development of conscious control in childhood. *Trends Cogn. Sci.* 8, 12–17. doi: 10.1016/j.tics.2003.11.001

Zelazo, P. (2006). The dimensional change card sort (DCCS): a method of assessing executive function in children. *Nat. Prot.* 1, 297–301. doi: 10.1038/nprot.2006.46

Zelazo, P. (2015). Executive function: reflection, iterative reprocessing, complexity, and the developing brain. *Dev. Rev.* 38, 55–68. doi: 10.1016/j.dr.2015.07.001

Zelazo, P., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., et al. (2003). The development of executive function in early childhood. *Monogr. Soc. Res. Child Dev.* 68:151. doi: 10.1111/j.0037-976X.2003.00260.x

frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 09 May 2024
DOI 10.3389/fdpys.2024.1332358

OPEN ACCESS

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

REVIEWED BY
Loren Marulis,
Connecticut College, United States
Gisella Decarli,
University of Trento, Italy

*CORRESPONDENCE
Elizabeth Dutemple
✉ Elizabeth.dutemple@mail.concordia.ca

RECEIVED 02 November 2023
ACCEPTED 15 April 2024
PUBLISHED 09 May 2024

CITATION
Dutemple E, Brokl C and Poulin-Dubois D (2024) I think therefore I learn: metacognition is a better predictor of school readiness than executive functions.
Front. Dev. Psychol. 2:1332358.
doi: 10.3389/fdpys.2024.1332358

I think therefore I learn: metacognition is a better predictor of school readiness than executive functions

Elizabeth Dutemple *, Carlye Brokl and Diane Poulin-Dubois

Cognitive and Language Development Lab, Department of Psychology, Concordia University, Montreal, QC, Canada

Previous research suggests that metacognition (the knowledge and skills related to knowledge acquisition) and executive functions (skills needed to plan and execute goals) are possible predictors of academic performance, including math and reading abilities. This study sought to clarify the relationship between school readiness and these abilities. A visual identification task was used to measure preschool children's metacognitive skills, specifically their ability to monitor their confidence on their answers (explicit) and ability to ask for a clue only when necessary (implicit). Response time to answering was also measured to obtain a non-verbal implicit measure of metacognition. Executive functions were measured using the Flanker and Dimensional Change Card Sorting (DCCS) tasks from the NIH toolbox. It was hypothesized that both metacognition and executive functions would predict school readiness and that implicit metacognitive skills would be more highly related to school readiness than explicit skills. A hierarchical linear regression was run with age and sex as control variables, and with executive function and metacognition (implicit and explicit) as predictors. Results indicated that both implicit and explicit metacognition remained significant predictors of school readiness scores beyond age and sex. In addition, we found correlations between explicit metacognition and executive functions and a relationship between response time and explicit metacognitive skill. Results highlight the importance of early metacognitive abilities beyond other cognitive skills and the importance of being able to effectively use metacognitive strategies from a young age. The implications relating to academic abilities are discussed.

KEYWORDS

school readiness, metacognition, executive functions, cognition, development

1 Introduction

School readiness or preparedness can be defined as the social, cognitive, and emotional skills required to succeed in an academic environment (sometimes grouped under the concept of self-regulation; Blair, 2002; Mashburn and Pianta, 2006; Blair and Raver, 2015) and whether their family and community environment is ready to support them (Williams et al., 2019). Because school readiness is such a broad concept, it is important to define which aspect is being measured in any assessment. The most commonly measured aspect of school readiness is whether children have accumulated the basic knowledge needed to understand what is taught in school (e.g., language, numbers). These types of assessments are routinely used to determine whether younger children are ready to enter either

Kindergarten or Grade 1. It has been argued that such assessments put certain economic or racial minorities at a disadvantage (Bierman et al., 2008; Evans and Schamberg, 2009; Blair et al., 2011; Ursache et al., 2013; Fitzpatrick et al., 2014; Blair and Raver, 2015; Micalizzi et al., 2019). To remediate such discrepancies and close this gap, it may be beneficial to study the skills that knowledge accumulation requires to encourage their development. Indeed, understanding what skills contribute to these individual differences in school readiness scores is essential, as they have been shown to predict academic, social, and behavioral outcomes later in life (La Paro and Pianta, 2000; Lonigan et al., 2000; Lonigan, 2006; McClelland et al., 2006; Duncan et al., 2007; Bernier et al., 2017; Mariano et al., 2019; Williams et al., 2019).

Although the environment has a significant role to play in precocious academic competence, children's cognitive skills can have a significant effect on academic achievement, making them an interesting focus of study (McClelland et al., 2006; Bernier et al., 2017; Mariano et al., 2019). The present study focuses on the main cognitive abilities that have been linked to early school performance: metacognition (e.g., Veenman and Spaans, 2005; Schraw et al., 2006; Schneider, 2015; Freeman et al., 2017) and executive function (e.g., Bierman et al., 2008; Fitzpatrick and Pagani, 2012; Bernier et al., 2020) by comparing their respective contribution to school readiness in a cross-sectional design. This differs from most other studies on this topic as it measures both abilities in a number of ways, which will be detailed below.

Metacognition is the ability to reflect upon personal thoughts, or "thinking about thinking" (Flavell, 1979; Schneider, 2008; Heyes et al., 2020). People display metacognition in their daily lives when they assert confidence in their knowledge or beliefs, or reflect upon their own emotions, for example. Broadly, metacognition can be separated into two distinct skills: metacognitive knowledge and metacognitive skills (Nelson, 1990; Nelson and Narens, 1994; Efklides, 2008; Schneider, 2008). Metacognitive knowledge, or declarative metacognition, encapsulates the knowledge one has about learning strategies and factors that can affect learning. Metacognitive skill, also known as procedural metacognition, includes the ability to monitor ones confidence and apply learning strategies when appropriate (e.g., Flavell, 1979; Nelson, 1990; Schneider, 2008). Metacognitive skills can be measured explicitly ("Are you sure?") or implicitly, by measuring behaviors such as eye-gaze, response time to a decision, or persistence behavior (Roderer and Roebers, 2014; Goupil and Kouider, 2016; Leckey et al., 2020; Resendes et al., 2021). Indeed, to make decisions based on their knowledge, individuals can rely on various cues such as their explicit monitoring skills (e.g., "Am I sure? Am I ready?") or implicit monitoring skills, which amounts to relying on performance cues such as response fluency (Simmons and Nelson, 2006; Thompson, 2010; Geurten and Willems, 2016), or how long it takes for one to answer, and perceptual fluency (Alter et al., 2007), or how easily a problem can be perceived and interpreted. Metacognitive skills can also be measured by assessing how efficiently one puts learning strategies to appropriate use. The latter can be measured by asking individuals if they want to use a particular metacognitive strategy in the context of a learning activity, for example if they want help on a task or not (e.g., Geurten and Bastin, 2019). Some consider this an implicit measure of metacognition, as it does not necessarily rely on individuals' explicit ability to articulate confidence but can instead rely on an implicit 'feeling-of-knowing' (Koriat, 1993, 2007) or even on response fluency cues (Geurten and Willems, 2016) which even toddlers and infants may have access to (e.g., Balcomb and Gerken, 2008; Hembacher and Ghetti, 2013; Lyons and Ghetti, 2013; Goupil and Kouider, 2016).

Together, the metacognitive abilities described above allow individuals to self-regulate by identifying mistakes or errors and selecting and applying strategies to improve their performance (Hembacher and Ghetti, 2013; Destan et al., 2014; Geurten and Bastin, 2019). In summary, much like many cognitive mechanisms, metacognition involves an implicit process in addition to a more deliberate and conscious one (Thompson, 2010; Henrich and Broesch, 2011). Individuals would then be able to weigh both implicit and deliberate metacognitive cues to make decisions (Stanovich, 2009). Studies have reported significant concurrent relationships between metacognitive skills where monitoring one's confidence correctly may allow for more efficient strategy use (Roderer and Roebers, 2014; Roebers and Spiess, 2017; Marulis and Nelson, 2021; Dutemple et al., 2023) but this relationship is inconsistent in young children (e.g., Roebers and Spiess, 2017). Though they might be measuring related constructs, they may therefore be distinct from one another. This strongly suggests that they should be investigated separately in relation to other cognitive skills and outcomes, especially when measuring younger children.

Research on metacognition has a long history (Flavell, 1979; Efklides, 2008) including research about metacognition's role in learning and school achievement (e.g., Sternberg, 1998; Efklides and Misailidi, 2010; Efklides, 2011). Metacognition has been related to better theory of mind (Feurer et al., 2015), better executive function and motivation (Marulis and Nelson, 2021), and better academic success as measured by language or mathematical skills (Duckworth and Seligman, 2005; Veenman and Spaans, 2005; Schraw et al., 2006; Dunlosky and Metcalfe, 2008; Dunlosky and Rawson, 2012; Schneider, 2015; Freeman et al., 2017), among other things. These studies mostly measured metacognitive control, so implicit metacognition, as they noted how well children engaged in strategies that fostered learning (e.g., selecting relevant information to review, planning studying, etc.; Veenman and Spaans, 2005). As far as we know, no one has yet tested the impact of explicit versus implicit metacognitive skills on most of these outcomes in children who have not yet entered grade school. Evidence suggests that it might be a better predictor of learning performance than general measures of intelligence (Veenman and Spaans, 2005), making metacognition an ideal candidate for cognitive skills to work on and improve from a young age. Metacognitive development has also been known to be intertwined with the development of executive function because it relates to regulation and planning (Marulis et al., 2020).

In the educational literature, metacognition has sometimes been referred to as calibration, which refers to a person's sensitivity to their knowledge, as expressed, for example, in better confidence on correct than on incorrect answers (e.g., Hattie, 2013; Roebers and Spiess, 2017). Calibration has been found to be linked to overall educational performance (Duncan et al., 2007; Hadwin and Webster, 2013), reading comprehension (Dabarera et al., 2014), as well as learning disabilities (Klassen, 2002). In infants and toddlers, metacognition impacts from whom children decide to learn, where better metacognition leads to choosing knowledgeable sources over ignorant ones (Kuzyk et al., 2020; Resendes et al., 2021). As Heyes (2020) suggested, the ability to think about our own thoughts may therefore give us insight into other people's competence and expertise, which in turn informs our decision to learn from them. This suggests that early development and nourishment of these abilities may be pivotal to improving quality of life and learning.

Executive functions are typically split into three abilities, namely inhibition (i.e., controlling impulses), shifting (i.e., switching quickly and efficiently between tasks), and working memory (i.e., holding and manipulating information in your mind; also known as updating) (Diamond, 2013; Weintraub et al., 2013). Together, they allow the planning and execution of actions flexibly (Miyake et al., 2000; Zelazo and Carlson, 2012). In children, however, some have found that executive functions cannot be reliably parsed into three separate abilities (Hughes et al., 2009; Wiebe et al., 2011; Diamond, 2013; Willoughby et al., 2016), but can rather be grouped into a single unitary construct which differentiate when the children reach older childhood (Lerner and Lonigan, 2014). Despite this, researchers typically evaluate two or three out of these skills when assessing executive functions in children and average them (e.g., Jacob and Parkinson, 2015). Executive functions are also sometimes referred to as "self-regulation" (not to be confused with the broader model of self-regulation; Efklides, 2008) or "effortful control" (Rothbart and Bates, 2006).

Studies have shown that executive functions are related to several positive outcomes including school achievement (e.g., Blair and Diamond, 2008; Razza and Blair, 2009; Monette et al., 2011; Fitzpatrick et al., 2014; Spiegel et al., 2021), mathematics ability (e.g., Espy et al., 2004; Clark et al., 2010; Bull and Lee, 2014; Fuhs et al., 2014), early reading skills (Kieffer et al., 2013), school readiness (specifically updating and set shifting; Bierman et al., 2008; Vitiello et al., 2011; Fitzpatrick and Pagani, 2012; Bernier et al., 2020), effortful control (Blair and Razza, 2007), theory of mind (Sabbagh et al., 2006; Hughes and Ensor, 2007; Carlson et al., 2015; Brock et al., 2019), and better metacognitive abilities (Bryce et al., 2015; Roebers, 2017). Some have also suggested that school readiness may in turn promote better executive functioning (Bierman et al., 2008).

The relationship between executive functions and metacognition has been a topic of recent discussion because of the similarities noted between the two abilities (Roebers, 2017; Filippi et al., 2020; Marulis et al., 2020). Indeed, both contribute to a child's ability to self-regulate (Efklides, 2008; Lyons and Zelazo, 2011) and behave in a goal-directed manner, however one is thought of as slower and more deliberate (metacognition) and other more automatic (executive functions). Self-regulation as a broader concept, which has also been known to relate to other factors such as a child's temperament (Chae, 2022), has also been linked to higher school achievement (Blair and Razza, 2007; Pianta et al., 2017; Weimer et al., 2021). In her framework, Roebers (2017) argued that executive functions lay the groundwork for metacognitive abilities; indeed, inhibition may explicitly contribute to metacognitive monitoring, as it allows an individual to pause and reflect on their answer (Bryce et al., 2015), and shifting and updating may be needed to keep in mind the goal of the task and decide, based on what was monitored, whether any control strategies need to be implemented to improve performance. Executive function and metacognition would therefore be highly correlated until children begin attending school, during which time the more deliberative type 2 metacognitive skills become more dependent on feedback from one's environment to improve those skills. In sum, the two cognitive abilities are closely related but grow apart as children begin attending school, which is why assessing their separate contributions to school readiness and performance is essential.

Given the state of the literature, the main goal of the present study is to elucidate with greater specificity the relationship between different subcomponents of metacognition and executive function, and how they differentially contribute to school readiness as measured by children's early arithmetic and language skills using a cross-sectional design. Specifically, it examined whether executive function and metacognition (monitoring, control, and implicitly measured) are longitudinally related to school readiness. School readiness was measured with the Lollipop Test (Chew and Morris, 1984, 1987). The metacognition task consisted of a perceptual discrimination task in which children had to recognize blurry pictures (Geurten and Bastin, 2019) thanks to which we could measure metacognitive monitoring and control in addition to an indirect measure of metacognition. Finally, executive functions (inhibition and shifting) were measured using child-appropriate versions of the Flanker task and of the Dimensional Change Card Sorting Task from the NIH toolbox (Weintraub et al., 2013). Three hypotheses were tested: (1) executive functions and metacognition will be related, (2) executive functions and metacognition will predict school readiness, and (3) metacognition will predict school readiness beyond executive functions. Given previous studies, implicit forms of metacognition, especially control, were expected to be more related to school readiness than implicit metacognition. Other than this, the link between implicit metacognition and school readiness has not been explored, therefore there is no specific hypothesis to be outlined. Together, these results aimed to shed light on the mechanisms that help children become self-sufficient, confident, and successful learners.

2 Methods

2.1 Participants

Participants lived in a large Canadian city and were recruited from a laboratory's database of past participants and through recruitment on social media. Informed consent was obtained before testing. An a priori statistical analysis for a linear regression using G*Power 3.1.9.7 suggested a sample size of 129 (six predictors, four tested predictors, power = 0.95, $\alpha = 0.05$). We tested 136 participants but had to exclude 6 due to undisclosed developmental delays ($N = 5$) or excessive distractiveness ($N = 1$) leaving a final sample of 130 ($Mage = 68.6$, $SD = 4.12$; 63 males). Eighty-three participants were tested in English and 47 in French. Multilingual children were allowed to answer in whichever language they felt most comfortable in. No significant difference was found on any of the reported variables based on language of testing or language status (monolingual vs. multilingual) so these variables were ignored in the main analyses. The median income of the families was between 100,000 and 150,000 CAD per year, making our sample upper middle class. Families primarily identified their children as Caucasian (57%), however 21% identified as Asian, 6% as Latin/Central/South American, 8% as African, 5% as Caribbean, and 4% Middle Eastern. Participants were allowed to choose up to 3 ethnicities with which they identified.

2.2 Measures

2.2.1 School readiness

School readiness was measured using the Lollipop test (Chew and Morris, 1984, 1987). It assesses four aspects of school readiness, specifically children's knowledge of colors and shapes, letters, numbers, and spatial recognition. Each of these four subscales is separately scored, and the total test is scored on 69. This test has good convergent validity with other school readiness tools (Chew and Morris, 1984) and been shown to predict academic achievement in both English and French (Chew and Morris, 1989; Chew and Lang, 1990; Venet et al., 2003; Boivin et al., 2014), lending credence to its validity as a tool to measure school ability. The French translation was also found to have good internal consistency ($\alpha = 0.89$), test-retest reliability (Venet et al., 2003), and was found to be the variable most related to various academic achievement measurements (Hammes et al., 2016).

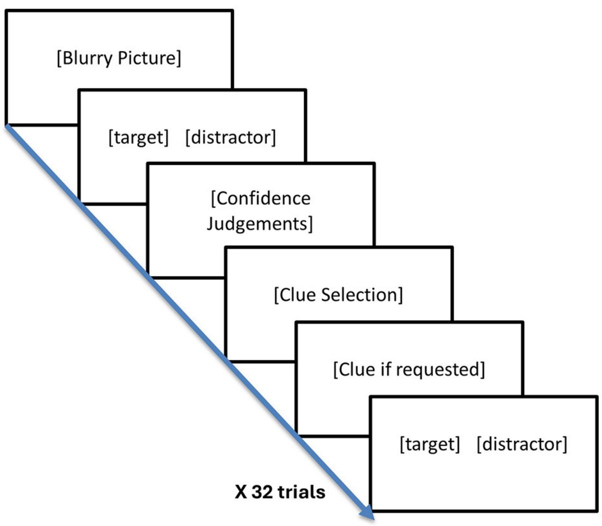


FIGURE 1
Schematic of the metacognition task.

2.2.2 Metacognition

We adapted the visual discrimination task from Geurten and Bastin (2019) and translated the procedure into French (Dutemple et al., 2023). A visual representation of the task can be found in Figure 1. The task was run on PsychoPy (version 2022.2.2). Children first practiced the task on three trials, after which they were given test trials and standardized metacognitive feedback (see Table 1). This also allowed the experimenter to explain to the child what a clue was if they did not know (i.e., "a clue is something that can help you decide which picture looked the most like the blurry picture.") If the child still did not understand they were given additional support (e.g., "the clue was a candle! Which picture is like the candle? Yes, because the light bulb makes light!") We also used the synonym "hint" if they preferred and understood that word. They were not given any feedback during the 32 following test trials.

Children were seated at a table roughly 60 cm (arm's length) from the computer screen. They were then shown blurry pictures which appeared on the screen for 1 s. Two clear but similar pictures appeared on the screen. The experimenter asked the children "which of the two pictures looks the most like the blurry picture you just saw?" Their time to answer was automatically recorded by Python to get an indication of their answer fluency. After their selection came the metacognitive monitoring trial. Two drawings of a boy appeared on the screen (e.g., Lyons and Ghetti, 2013; Geurten and Bastin, 2019). They were asked whether they were "very sure" or "not sure at all" about their answer, like the boy on the right or the boy on the left. Next, we tested their metacognitive control by assessing if they could appropriately apply the strategy of asking for help. The image of a gift with a question mark appeared on the screen with the words "yes" and "no". They were given the option to ask for a clue (or a "hint" if they preferred that word) if they believed they made a mistake. The hints were semantically or visually related to the blurry pictures (e.g., grass as a hint for a flower) thus pointing the child toward the correct answer. Finally, they were offered the opportunity to change their initial answer. In total, therefore, three measures of

TABLE 1 Metacognition task feedback according to children's answers.

	Correct	Incorrect
Confidence	You selected the correct picture and that's great because you told me you were really sure of your answer.	You selected the wrong picture, but you told me you were really sure of your answer. Here, maybe you were not so sure of your answer.
Not confident	You selected the correct picture, but you told me that you were not sure of your answer. Here, maybe you were really sure of your answer.	You selected the wrong picture but that's okay because you told me you were not sure of your answer.
Selected a clue	You selected the correct picture, but you asked for a clue. Maybe here, you did not need a clue.	You selected the wrong picture, but that's okay because you asked me for a clue to help you.
Did not select a clue	You selected the correct picture and what's great is that you didn't ask me for a clue because you didn't need help.	You selected the wrong picture, but you didn't ask me for a clue to help choose. Here, you could have asked me for help.

metacognition were recorded: (1) implicit metacognition (response fluency), (2) confidence monitoring, (3) metacognitive control (clue selection/help seeking).

The difficulty of the trials varied according to the child's ability, where more correct responses incurred more difficult trials, and more incorrect trials incurred easier trials. Specifically, there were three levels of difficulty. All children started with "medium" difficulty pictures as defined in the original study. If the child answered two out of three visual discrimination trials correctly, they were given more difficult pictures ("Hard" pictures). If they did not answer at least two out of three correctly, they were given easier pictures ("Easy" pictures). This was repeated throughout the whole task, where children moved between categories as they made mistakes. This served two purposes. For one, it was to ensure that every child would provide both correct and incorrect trials to compare their confidence. In addition, it allowed to more confidently assert that children's individual visual discrimination skills were considered, and that each participant was exercising a similar amount of effort. The only significant changes made between this iteration and the original task were to (1) omit one version of the task where children were exposed to the clue selection before the confidence selection and (2) to reduce the number of practice trials from 6 to 3. This was done to first ensure children would see the clue as soon as they requested it rather than doing their confidence judgement before the clue was revealed to them, and second to cut the time this task would take. Indeed, pilot testing suggested 6 trials were not necessary to understand the task (the original researchers used this task on a younger population and therefore were justified in a longer practice) and shortening testing time increased engagement throughout. In addition, Geurten and Bastin (2019) did not report any order effects, suggesting this deviation would not have major effects on performance.

2.2.3 Executive function

2.2.3.1 Shifting

The NIH toolbox (Weintraub et al., 2013) Dimensional Change Card Sorting Task (DCCS) for 4–7-year-olds was administered. The children were seated approximately 8 inches away from an iPad, which was set up at an angle on the table. During the task, participants were shown pictures of balls or trucks that were yellow or blue. They had to classify them according to shape or color on randomly alternating trials. Children were required to pass 3 out of 4 practice trials before beginning the test trials. The first test block required children to pass 4 out of 5 trials for both the color and shape trials, which were kept apart. The second block was mixed (30 trials), meaning children had to actively alternate between sorting pictures according to color or shape. The children were given raw scores, computed scores, and standard scores, as calculated by the NIH toolbox. Raw scores indicate how many trials were correctly answered by the participant. Computed scores calculate two vectors, one for accuracy and one for speed, both scored on five and then combined, resulting in scores out of 10. Finally, standard scores compare the children's computed scores to a normed sample of scores from other children (see Zelazo et al., 2013 for the mathematical equations used to determine the computed scores).

2.2.3.2 Inhibition

The NIH toolbox (Weintraub et al., 2013) Flanker task for 4–7-year-olds was administered. The set up was identical to that of the DCCS described above. The children first experienced sixteen trials with a row of fish, where the middle fish sometimes pointed in the same or different direction than the fish around it. The children were told that to feed the fish, they had to select the button with the arrow that pointed in the same direction as the middle fish was swimming. The children were given 4 practice trials, three of which had to be correct before moving on to the first block of 20 test trials. To proceed to the second block, children had to commit fewer than two mistakes. The second block of 20 trials replaced the fish with arrows. Once again, children received three scores as described above (raw, computed, and standard).

2.3 Procedure

Parents were seated in the testing room and filled out the demographic sheet while their children participated. The tasks were administered as part of a larger battery of tasks. There were eight possible orders in which the children could complete the tasks. Between 14 and 18 participants completed each counterbalance order. Each task was set up on a different table to allow the child small breaks between the games. The two executive function tasks on the iPad were always split up to ensure children did not need to sit at the same table for more than 15 min. Observation of the children during piloting confirmed that the iPad tasks were the least engaging and therefore not ideal to encourage participation and engagement early in the session, therefore none of the counterbalanced orders began with them. The original design also included the NIH working memory task, however piloting revealed that this task was too difficult for the children to grasp and therefore was dropped. This also had the benefit of keeping the length of the sessions to an hour, which was deemed a reasonable length of testing for children that age. To ensure that no notable order effects were present, we ran t-tests between the groups that were administered each task either early or late (i.e., we assessed whether those who completed the Lollipop early or late showed a difference in their scores, etc.) Between those who were administered the metacognition task early or late in the assessment, there was no difference between their accuracy on the task [$t_{(125)} = -1.54$, p =0.126, d = – 0.27], their confidence judgements according to their accuracy [$t_{(126)} = 1.15$, p =0.126, d = 0.20], their clue selection according to their accuracy [$t_{(125)} = 1.24$, p =0.218, d = 0.22], or their response latency according to their accuracy [$t_{(119)} = 0.55$, p =0.583, d =0.10]. Finally, between those who were administered one executive function task before the other, there was no difference between the two groups [$t_{(128)} = 0.83$, p =0.41, d = 0.15]. None were therefore significant.

3 Results

Univariate statistical outliers (± 3 standard deviations from the mean) were determined per measure and excluded. Two participants did not complete the metacognition task and one participant's data was partially lost due to a computer malfunction

($N = 127$). Two outliers were identified for the school readiness task (> 44.46, $N = 128$), no outliers were identified for the executive function tasks, one outlier was identified on the clue selection measure of the metacognition task (> 39.06, $N = 126$), and seven were identified for the response time measure (>1.7, $N = 120$).

3.1 Chance analyses

First, it was important to determine whether the metacognitive and school readiness tasks were too difficult for the sample. Chance analyses were therefore run. Overall, participants performed above chance on the visual discrimination task {M =0.63, SD =0.11, $t_{(126)} = 13.36$, p <0.001, 95%CI [0.11-0.15], $d = 1.19$}. For the metacognitive component of the task, chance analyses were performed according to the proportion of correct and incorrect trials (see Table 2). Difference scores were calculated to obtain a measure of calibration. Confidence on incorrect trials was subtracted from confidence on correct trials; clue selection on correct trials was subtracted from clue selection on incorrect trials; and response time on correct trials was subtracted from response time on incorrect trials. Positive scores are indicative of better metacognition. Participants performed above chance on all three difference score measures (see Table 2), suggesting they could differentiate between correct and incorrect trials based on both explicit and implicit metacognitive measures. Children also performed above chance on the school readiness task [$M = 61.12$, $SD = 4.62$, $N = 128$, $t_{(127)} = 65.2$, $p < 0.001$, $d = 5.76$].

3.2 Task-specific results

To determine the extent to which there were significant differences between children's answers on accurate and inaccurate trials, a series of within-subjects repeated measures one-way ANOVAs were run (accuracy: correct and incorrect) on confidence judgements (metacognitive monitoring), clue selection (metacognitive control), and response time (implicit measure of metacognitive monitoring). Children were more confident on correct than on incorrect trials [$F_{(1,127)} = 24.8$, p <0.001, η^2 =0.01]. Children chose a clue more often on incorrect trials than on correct trials [$F_{(1,127)} = 26.62$, p <0.001, η^2 =0.01] and children were slower on incorrect trials than on correct trials [$F_{(1,122)} = 47.70$, p <0.001, η^2 =0.07].

Performance on the executive function tasks is reported in Table 3. The NIH toolbox provides three scores, and reported in the table below are the computed scores and the standard scores (mean of 100 and standard deviation of 15) comparing the children's performance to that of their same-aged peers. The latter are derived from the computed scores, which take into consideration both the child's performance (scored on 5) and the speed at which they answered (scored on 5) when they performed above a certain threshold (see Zelazo et al., 2013 for more details on the scoring procedure).

3.3 Intertask correlations

Performance on all tasks of interest is included in the following correlation matrix to determine whether they were related in this sample (see Table 4, Figure 2). Bivariate Pearson correlations were run with missing data removed pairwise. School readiness was positively correlated with both explicit ($r_{(124)} = 0.248$, $p = 0.005$) and implicit ($r_{(123)} = 0.222$, $p = 0.013$) metacognition as measured by confidence judgements and clue selection, respectively. Explicit and implicit metacognition were also significantly related to one another ($r_{(125)} = 0.426$, $p < 0.001$). Executive function was significantly correlated with explicit metacognition ($r_{(126)} = 0.239$, $p = 0.006$). This remained true if the average computed scores or the averaged standard scores were used (the standard scores are reported here). Finally, implicit metacognition as measured by response time was related to explicit metacognition ($r_{(119)}$ =0.215, p =0.018). All correlations remained significant after correction for pairwise comparisons with False Discovery Rate analyses (Benjamini and Hochberg, 1995).

3.4 Regression analyses

Multivariate outliers were identified and removed. They were defined as those who had Cook's distances above 4/n (4 divided by our sample size, or 0.032). The final sample for the regression was 118, accounting for 125 participants without any missing data (Mage = 68.69) and 7 multivariate outliers (Mage = 66.86). The two groups are not different in age {$t_{(123)} = 1.14$, p =0.129, d =0.44, 95%CI [−1.35; 5.01]}. All assumptions were then verified and met. Specifically, visual inspection of the residuals confirmed multivariate linearity and homoscedasticity; no collinearity was detected as measured by tolerance values (0.882–0.995 >0.1) and VIF statistic (1.01–1.26 < 5); and autocorrelation was not present as measured by a Durbin-Watson test (1.5 < 2.05 < 2.5).

Following the removal of the multivariate outliers, a hierarchical regression with age and sex as baseline demographic factors was performed. The models were run based on the initial prediction that both executive function and metacognition would contribute to school readiness scores. Model 1 with demographic factors only accounted for 9% of the variance in school readiness scores. Model 2 with executive functions predicted an identical 9% of the variance. Model 3 with all predictors included explained 19% of the variance in school readiness scores (see Table 5). Age remained significantly predictive of school readiness ($B = 0.24$, $\beta = 0.25$, $p = 0.006$). Implicit metacognition remained significant beyond age ($B = 6.01$, $\beta = 0.20$, $p = 0.037$; see Table 6 for model details) as did explicit metacognition ($B = 6.31$, $\beta = 0.19$, $p = 0.047$) though that latter was only marginally significant.

4 Discussion

This study sought to investigate the cognitive predictors of school readiness. Findings suggest significant relationships between implicit and explicit metacognition and school readiness but no such link between school readiness and executive functions. This relationship mostly holds in a hierarchical regression which included age and sex as demographic factors, suggesting metacognition is an important contributor to school readiness. Metacognitive monitoring and control were significantly correlated, and aspects of executive functioning (inhibition and shifting) were related to metacognitive monitoring. Metacognitive monitoring was also related to our implicit metacognitive measure,

TABLE 2 Chance analyses for the metacognition task.

	Confidence (monitoring)		Clue selection (control)		Confidence difference score	Clue difference score	Response time difference score (implicit)
	Incorrect	Correct	Incorrect	Correct			
M	0.74	0.80	0.45	0.39	0.06	0.06	0.63
SD	0.29	0.24	0.37	0.35	0.13	0.13	0.99
N	128	128	128	128	128	127	121
t	9.58	13.83	−1.41	−3.62	4.98	5.16	7.03
df	127	127	127	127	127	126	120
p	<0.001	<0.001	0.08	<0.001	<0.001	<0.001	<0.001
Cohen's d	0.85	1.22	−0.13	−0.32	0.44	0.46	0.64
95% CI	0.19–0.29	0.26–0.34	−0.11–0.02	−0.17–−0.05	0.03–0.08	0.04–0.08	0.45–0.81

Chance was operationalized as 0.5 for the Correct and Incorrect trials and at 0 for the difference scores.

TABLE 3 Descriptive statistics for the executive function tasks.

	Flanker computed	DCCS computed	Executive function mean computed	Flanker standard	DCCS standard	Executive function mean standard
N	130	128	130	130	128	130
M	5.76	4.29	5.03	74.38	69.92	72.14
SD	1.47	2.36	1.52	14.88	20.69	14.10

TABLE 4 Intertask bivariate correlations.

		1	2	3	4	5
1. School readiness	*r*	1	-	-	-	-
	N	128				
2. Executive function	*r*	0.06	1	-	-	-
	N	128	130			
3. Confidence judgements (explicit)	*r*	**0.25***	**0.24***	1	-	-
	N	126	128	128		
4. Clue selection (implicit)	*r*	**0.22***	0.14	**0.43***	1	-
	N	125	127	127	127	
5. Response time (implicit)	*r*	0.10	0.14	**0.22***	0.13	1
	N	119	121	121	120	121

Results that remain significant following the false discovery rate are in bold. *$p < 0.05$.

suggesting 5-year-old children may be able to use processing fluency as a metacognitive cue.

Previous research had shown that metacognition plays an important role in academic achievement and learning, but no study had yet studied how early this link can be observed with a school readiness measure. Another innovative feature of the current study was to investigate this link with a preschool population with explicit and implicit measures of metacognition. Indeed, our children demonstrated that despite the fact that they were somewhat overconfident (as expected in this population; see Lipowski et al., 2013), the more they were able to discriminate between their correct and incorrect answer by asking for help only, when necessary, the better they performed on the school readiness task. This strongly suggests that being able to act upon personal thoughts and knowledge scaffolds early academic development and interest in gaining knowledge. Indeed, some have suggested that metacognition fuels curiosity, or the desire to learn, which implies being able to discern what one knows or does not know (Goupil and Proust, 2023). Curiosity can also be thought of as some sort of implicit metacognitive process as it does not rely on people directly reflecting about their knowledge but asking for information. This may also be related to children's positive approaches to learning (i.e., children's motivation, persistence, and initiative toward learning; Kagan et al., 1995) which have been

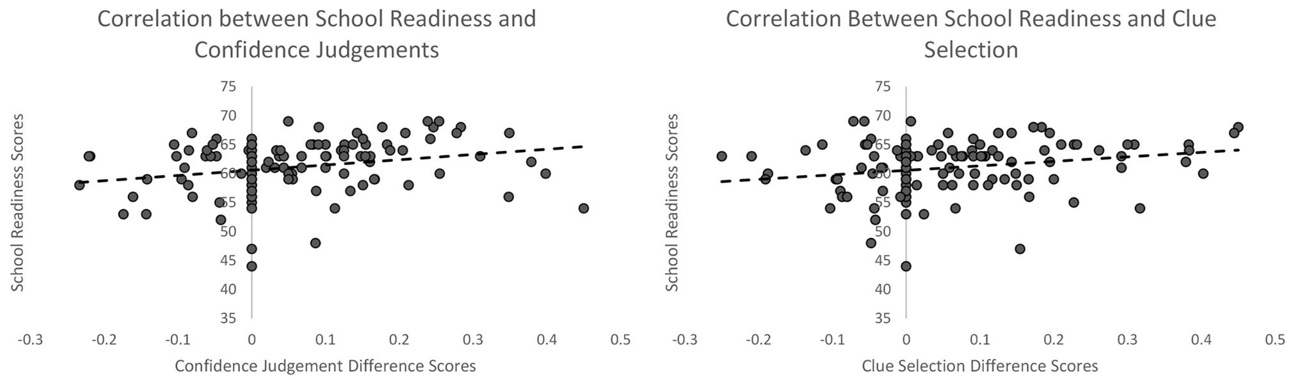


FIGURE 2
Correlations between school readiness and metacognition variables.

TABLE 5 Hierarchical model summary with school readiness as dependent variable.

Model	R	R^2	Adjusted R^2	Std. error of the estimate
1	0.30[a]	0.09	0.08	3.76
2	0.30[b]	0.09	0.07	3.77
3	0.44[c]	0.19	0.16	3.59

[a]Model with sex and age; [b]model with sex, age, and executive functions c. model with sex, age, executive functions, explicit metacognition, implicit metacognition.

TABLE 6 Hierarchical model summary by predictor.

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	β		
1	(Constant)	41.82	5.88		7.11	<0.001
	Age (months)	**0.29**	**0.09**	**0.30**	**3.37**	**0.001**
	Sex	−0.36	0.69	−0.05	−0.52	0.602
2	(Constant)	41.86	5.91		7.08	<0.001
	Age (months)	**0.29**	**0.09**	**0.31**	**3.29**	**0.001**
	Sex	−0.38	0.70	−0.05	−0.54	0.591
	EF Average	−0.00	0.03	−0.02	−0.19	0.847
3	(Constant)	46.00	5.74		8.02	<0.001
	Age (months)	**0.24**	**0.09**	**0.25**	**2.80**	**0.006**
	Sex	−0.62	0.67	−0.08	−0.92	0.358
	EF average	−0.02	0.03	−0.08	−0.88	0.379
	Confidence judgements	**6.31**	**3.15**	**0.19**	**2.00**	**0.047**
	Clue selection	**6.01**	**2.85**	**0.20**	**2.11**	**0.037**

Significant predictors are bolded.

found to positively predict school readiness (McClelland et al., 2000; Fantuzzo et al., 2004; McWayne et al., 2004). An awareness of what one knows and does not know would therefore encourage aspiring students to seek out and remember new information. It also implies a motivational component or an assumption that the children desire to seek out unknown information, which was not controlled for in this study. An interesting future direction may be to follow a longitudinal sample further into their academic progress and investigate whether this relationship strengthens or disappears with time (e.g., Roebers et al., 2012; Tibken et al., 2022) and how motivation may play a role.

Because the Lollipop did not provide a large variability in scores, it limited our ability to detect a relationship between the skills. Another option as a measure of school readiness would have been the Bracken School Readiness Assessment (Bracken, 1998, 2002; Panter and Bracken, 2009), as it includes 88 items, some of which may have been a little more difficult, thus providing even more variability in scores. It is important to note that these school readiness tasks were designed to identify individuals with difficulties rather than those who have none, meaning high scores on these tasks with a typically developing population is to be expected.

Moreover, we were able to look at the relationship between the different potential predictors of school readiness. We found that metacognitive monitoring and control measured verbally were related and that response fluency (also an index of implicit metacognition), which replicates previous research by demonstrating an even earlier link between verbal monitoring and control skills than previously recorded (e.g., Roebers and Spiess, 2017). This is consistent with results showing that children seek out more information when not confident (Call and Carpenter, 2001) or opt-out of answering when they believe they cannot answer (Balcomb and Gerken, 2008; Bernard et al., 2015), meaning children after age 5 may be able to use their lack of knowledge to guide whether they seek help to obtain information (Hembacher and Ghetti, 2013; Destan et al., 2014; Coughlin et al., 2015; Destan and Roebers, 2015; Goupil and Kouider, 2019; Lapidow et al., 2022). Although other studies have shown children's limited ability to use certain cues such as retrieval fluency to guide their metacognitive monitoring (e.g., Koriat and Ackerman, 2010), the results from this study extend the current research in this area by showing that children may be able to use answering time implicitly to guide monitoring, but maybe not control. In addition, the monitoring metacognition component was correlated with the executive function measure, replicating previous work drawing a link between these abilities (Garner, 2009; Bryce et al., 2015; Spiess et al., 2016; Roebers, 2017). Surprisingly however, it was metacognitive monitoring that was most strongly related to executive function, in conflict with prior research suggesting that metacognitive control was most dependent on skills such as inhibition and shifting (e.g., Bryce et al., 2015; Spiess et al., 2016; Roebers, 2017). It is possible that monitoring in the context of the visual discrimination task required children to inhibit their initial impulse to answer to reflect more on their accuracy. Furthermore, shifting skills may have allowed them to evaluate each image on its own merit rather than automatically declare themselves confident. This relationship is worth investigating in more detail and is a rich area for future work.

This study has many strengths. Adequate statistical power was obtained by testing a large sample size; therefore, results can be confidently interpreted. Next, the impact of metacognition was thoroughly investigated by including explicit and implicit measures of metacognition. To our knowledge, this is the first study to parse metacognition's impact on school readiness this way. In addition, the measurement of executive functions alongside metacognition provides a comprehensive picture of the possible cognitive influences on school readiness. This study is amongst the first to measure these relationships in preschoolers and to directly compare the relative importance of these skills for school readiness. Indeed, these results will hopefully lead developmental psychologists to encourage parents and educators to engage in interactive activities developing metacognition, as the latter has been found to be a trainable skill (De Jager et al., 2005; Michalsky et al., 2009) to set their preschoolers on a path of success in the academic realm. Interestingly however, recent studies have shown that feedback from teachers may not be enough to train metacognitive skills in school-aged children (Buehler et al., 2023) which may further suggest that environmental and motivational factors may be at play in the development of metacognition. Metacognition may operate in a similar way to executive function; Zelazo (2015) indeed makes a distinction between "cold" and "hot" executive functions, the latter being more influenced by an individual's emotional or motivational state during the task (e.g., the marshmallow task; Munakata and Michaelson, 2021). Instead of training metacognition alone, perhaps self-regulation skills (Efklides, 2008) need to be trained in tandem to result in long-term benefits. Future studies may want to further explore the longitudinal relationship between these skills in the context of a training study and extend the findings to other outcomes; for instance, does the training indeed have an impact on metacognitive skills that translates onto school readiness? In addition, because children in lower SES are more likely to have lower school readiness scores, speculating about how this training may be implemented at the community or preschool levels would be essential (e.g., Roberts, 2011; Weiland and Yoshikawa, 2013; Blair and Raver, 2015; Bierman et al., 2020; Joo et al., 2020; Shaw et al., 2021).

Despite our best efforts, this study also has limitations. For one, the scores on the Lollipop test were generally high and somewhat limited in variability. This was surprising, as Venet et al. (2003) used this task in a similar age group (Mean age of 67 months) and had obtained a lower average score of 47.4/69. This task was chosen because it is available in both French and English and has been validated in both languages. The marked difference in scores may be because we were measuring school readiness in a higher SES sample (e.g., Geoffroy et al., 2010). As discussed above, parents from higher SES tend to foster more learning in their young children, perhaps setting them up for better success on these types of school readiness tasks. For the reasons outlined above.

Next, a complete battery of executive function measures generally includes updating, or working memory, as an important component; however, it was not included here as pilot testing suggested the working memory task available was too difficult for children to complete. Future research may wish to consider simpler tasks to measure working memory in preschoolers even if executive functions at that age are not entirely differentiated (Miyake et al., 2000; Jurado and Rosselli, 2007; Garon et al., 2008; Miller et al., 2012; Brydges et al., 2014; Willoughby et al., 2016). In addition, though some have found associations between executive functions and academic performance, the causality of this relationship is less certain, suggesting a link between the two skills need not always be apparent (Jacob and Parkinson, 2015). Finally, the academic performance task was a school readiness task rather than the reading- or mathematic-only tasks reported in other studies, suggesting EF may be important for specific subcomponents of academics rather than overall readiness to learn and participate in school.

Finally, as discussed previously, school readiness is a broad concept that includes more than the basic knowledge children have acquired. Future studies may want to include other factors that some have included in their definitions of school readiness, such as measures of social adjustment, or even measures related to the school or parenting environments, to extend our understanding of which factors are found to be reliably linked to school readiness and later school performance (e.g., McCallan, 2010; Denham et al., 2012; Flook et al., 2015; Darling-Hammond et al., 2020; Joo et al., 2020). Indeed, this would allow researchers to answer interesting questions related to the cognitive correlates of school readiness; does better metacognition compensate for certain environmental deficiencies? Do executive functions play a bigger role in gaining social skills and integrating inside the classroom? Further research on the longitudinal correlates of metacognition (e.g., does it lead to better academic and social success? Is it associated with better health?) may cement metacognition itself as a component of school readiness as a "readiness to learn" and something to be more explicitly encouraged in classrooms from a young age. Furthermore, future studies may want to explore the growth of metacognitive knowledge in parallel with metacognitive skills in the context of school readiness.

In conclusion, school readiness is related to metacognitive control beyond the effects of age. Explicit metacognition was correlated with executive functions and implicit metacognition as measured by asking for clues and reaction time. Finally, executive functions were not related to school readiness in this population. This study aimed to clarify the link between these cognitive skills and school readiness with the hope to better understand which skills are best to nourish early in preschool.

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 below: https://osf.io/7v5gs/.

Ethics statement

The studies involving humans were approved by Institutional Review Board at Concordia University. 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

ED: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. CB: Formal analysis, Investigation, Writing – original draft. DP-D: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, 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 funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada via a grant awarded to DP-D.

Acknowledgments

The authors would like to thank their funding agency and all the wonderful families for participating. They would also like to thank Anna Baumann, Mihaela Zlatanovska, and Gal Zohar for their help at various stages of the 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

Alter, A. L., Oppenheimer, D. M., Epley, N., and Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *J. Exp. Psychol.: General* 136, 569. doi: 10.1037/0096-3445.136.4.569

Balcomb, F. K., and Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Dev. Sci.* 11, 750–760. doi: 10.1111/j.1467-7687.2008.00725.x

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statist. Soc.: Series B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bernard, S., Proust, J., and Clément, F. (2015). Procedural metacognition and false belief understanding in 3-to 5-year-old children. *PLoS ONE* 10, e0141321. doi: 10.1371/journal.pone.0141321

Bernier, A., Beauchamp, M. H., and Cimon-Paquet, C. (2020). From early relationships to preacademic knowledge: a sociocognitive developmental cascade to school readiness. *Child Dev.* 91, e134–e145. doi: 10.1111/cdev.13160

Bernier, A., St-Laurent, D., Matte-Gagné, C., Milot, T., Hammond, S. I., and Carpendale, J. I. M. (2017). "Parenting and young children's executive function development," in *Executive Functions in Children's Everyday Lives: A Handbook for Professionals in Applied Psychology*, eds. M. J. Hoskyn, G. Iarocci, and A. R. Young (Oxford University Press), 70–87. doi: 10.1093/acprof:oso/9780199980864.003.0006

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., and Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Dev. Psychopathol.* 20, 821–843. doi: 10.1017/S09545794080 00394

Bierman, K. L., Sanders, M., and Ho, L. C. (2020). "Addressing socioeconomic disparities in school readiness with preschool programming and professional development support," in *Healthy Development in Young Children: Evidence-Based Interventions for Early Education*, eds. V. C. Alfonso and G. J. DuPaul (American Psychological Association), 67–84. doi: 10.1037/0000197-004

Blair, C. (2002). School readiness: integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *Am. Psychol.* 57, 111. doi: 10.1037/0003-066X.57.2.111

Blair, C., and Diamond, A. (2008). Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. *Dev. Psychopathol.* 20, 899–911. doi: 10.1017/S0954579408000436

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., et al. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Dev.* 82, 1970–1984. doi: 10.1111/j.1467-8624.2011.01643.x

Blair, C., and Raver, C. C. (2015). School readiness and self-regulation: a developmental psychobiological approach. *Annu. Rev. Psychol.* 66, 711–731. doi: 10.1146/annurev-psych-010814-015221

Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x

Boivin, M., Desrosiers, H., Lemelin, J.-P., and Forget-Dubois, N. (2014). "Assessing the predictive validity and early determinants of school readiness using a population-based approach," in *Promoting School Readiness and Early Learning: Implications of Developmental Research for Practice*, eds. M. Boivin and K. L. Bierman (The Guilford Press), 46–72.

Bracken, B. A. (1998). *Bracken Basic Concept Scale-Revised*. San Antonio, TX: The Psychological Corporation.

Bracken, B. A. (2002). *Bracken School Readiness Assessment*. San Antonio, TX: The Psychological Corporation.

Brock, L. L., Kim, H., Kelly, C. L., Mashburn, A. J., and Grissmer, D. W. (2019). Theory of mind, directly and indirectly, facilitates kindergarten adjustment via verbal ability, executive function, and emotion knowledge. *Psychol. Sch.* 56, 176–193. doi: 10.1002/pits.22216

Bryce, D., Whitebread, D., and Szucs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognit. Learn.* 10, 181–198. doi: 10.1007/s11409-014-9120-4

Brydges, C. R., Fox, A. M., Reid, C. L., and Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: a longitudinal latent-variable analysis. *Intelligence* 47, 34–43. doi: 10.1016/j.intell.2014.08.010

Buehler, F. J., Ghetti, S., and Roebers, C. M. (2023). *Training Primary School Children's Uncertainty Monitoring*. Charlottesville, VA: OSF Preprints.

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

Call, J., and Carpenter, M. (2001). Do apes and children know what they have seen?. *Anim. Cogn.* 3, 207–220. doi: 10.1007/s100710100078

Carlson, S. M., Claxton, L. J., and Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *J. Cognit. Dev.* 16, 186–197. doi: 10.1080/15248372.2013.824883

Chae, S. E. (2022). Executive function and effortful control-Similar and different evidence from big data analysis. *Front. Psychol.* 13, 1004403. doi: 10.3389/fpsyg.2022.1004403

Chew, A. L., and Lang, W. S. (1990). Predicting academic achievement in kindergarten and first grade from prekindergarten scores on the Lollipop Test and DIAL. *Educ. Psychol. Meas.* 50, 431–437. doi: 10.1177/0013164490502022

Chew, A. L., and Morris, J. D. (1984). Validation of the lollipop test: a diagnostic screening test of school readiness. *Educ. Psychol. Meas.* 44, 987–991. doi: 10.1177/0013164484444022

Chew, A. L., and Morris, J. D. (1987). Investigation of the Lollipop Test as a pre-kindergarten screening instrument. *Educ. Psychol. Meas.* 47, 467–471. doi: 10.1177/0013164487472019

Chew, A. L., and Morris, J. D. (1989). Predicting later academic achievement from kindergarten scores on the Metropolitan Readiness Tests and the Lollipop Test. *Educ. Psychol. Meas.* 49, 461–465. doi: 10.1177/0013164489492019

Clark, C. A., Pritchard, V. E., and Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Dev. Psychol.* 46, 1176. doi: 10.1037/a0019672

Coughlin, C., Hembacher, E., Lyons, K. E., and Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev. Sci.* 18, 957–971. doi: 10.1111/desc.12271

Dabarera, C., Renandya, W. A., and Zhang, L. J. (2014). The impact of metacognitive scaffolding and monitoring on reading comprehension. *System* 42, 462–473. doi: 10.1016/j.system.2013.12.020

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., and Osher, D. (2020). Implications for educational practice of the science of learning and development. *Appl. Dev. Sci.* 24, 97–140. doi: 10.1080/10888691.2018.1537791

De Jager, B., Jansen, M., and Reezigt, G. (2005). The development of metacognition in primary school learning environments. *Sch. Eff. Sch. Improv.* 16, 179–196. doi: 10.1080/09243450500114181

Denham, S. A., Bassett, H. H., and Zinsser, K. (2012). Computerizing social-emotional assessment for school readiness: first steps toward an assessment battery for early childhood settings. *J. Appl. Res. Child.: Inform. Policy Child. Risk* 3:1091. doi: 10.58464/2155-5834.1091

Destan, N., Hembacher, E., Ghetti, S., and Roebers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5-to 7-year-old children. *J. Exp. Child Psychol.* 126, 213–228. doi: 10.1016/j.jecp.2014.04.001

Destan, N., and Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence?. *Metacognit. Learn.* 10, 347–374. doi: 10.1007/s11409-014-9133-z

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Duckworth, A. L., and Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol. Sci.* 16, 939–944. doi: 10.1111/j.1467-9280.2005.01641.x

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Dunlosky, J., and Metcalfe, J. (2008). *Metacognition*. Thousand Oaks: Sage Publications.

Dunlosky, J., and Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learn. Instruct.* 22, 271–280. doi: 10.1016/j.learninstruc.2011.08.003

Dutemple, E., Hakimi, H., and Poulin-Dubois, D. (2023). Do I know what they know? Linking metacognition, theory of mind, and selective social learning. *J. Exp. Child Psychol.* 227, 105572. doi: 10.1016/j.jecp.2022.105572

Efklides, A. (2008). Metacognition: defining its facets and levels of functioning in relation to self-regulation and co-regulation. *Eur. Psychol.* 13, 277–287. doi: 10.1027/1016-9040.13.4.277

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* 46, 6–25. doi: 10.1080/00461520.2011.538645

Efklides, A., and Misailidi, P. (2010). "Introduction: the present and the future in metacognition," in *Trends and Prospects in Metacognition Research*, eds. A. Efklides, and P. Misailidi (Boston: Springer).

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., and Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Dev. Neuropsychol.* 26, 465–486. doi: 10.1207/s15326942dn2601_6

Evans, G. W., and Schamberg, M. A. (2009). Childhood poverty, chronic stress, and adult working memory. *Proc. Nat. Acad. Sci.* 106, 6545–6549. doi: 10.1073/pnas.0811910106

Fantuzzo, J., Perry, M. A., and McDermott, P. (2004). Preschool approaches to learning and their relationship to other relevant classroom competencies for low-income children. *School Psychol. Quart.* 19, 212. doi: 10.1521/scpq.19.3.212.40276

Feurer, E., Sassu, R., Cimeli, P., and Roebers, C. (2015). Development of meta-representations: Procedural metacognition and the relationship to theory of mind. *J. Educ. Dev. Psychol.* 5, 6–18. doi: 10.5539/jedp.v5n1p6

Filippi, R., Ceccolini, A., Periche-Tomas, E., and Bright, P. (2020). Developmental trajectories of metacognitive processing and executive function from childhood to older age. *Quart. J. Exp. Psychol.* 73, 1757–1773. doi: 10.1177/1747021820931096

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., and Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children?. *Learn. Instruct.* 30, 25–31. doi: 10.1016/j.learninstruc.2013.11.003

Fitzpatrick, C., and Pagani, L. S. (2012). Toddler working memory skills predict kindergarten school readiness. *Intelligence* 40, 205–212. doi: 10.1016/j.intell.2011.11.007

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am. Psychol.*34, 906–911. doi: 10.1037/0003-066X.34.10.906

Flook, L., Goldberg, S. B., Pinger, L., and Davidson, R. J. (2015). Promoting prosocial behavior and self-regulatory skills in preschool children through a mindfulness-based Kindness Curriculum. *Dev. Psychol.* 51, 44. doi: 10.1037/a0038256

Freeman, E. E., Karayanidis, F., and Chalmers, K. A. (2017). Metacognitive monitoring of working memory performance and its relationship to academic achievement in Grade 4 children. *Learn. Individ. Differ.* 57, 58–64. doi: 10.1016/j.lindif.2017.06.003

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., and Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Dev. Psychol.* 50, 1698. doi: 10.1037/a0036633

Garner, J. K. (2009). Conceptualizing the relations between executive functions and self-regulated learning. *J. Psychol.* 143, 405–426. doi: 10.3200/JRLP.143.4.405-426

Garon, N., Bryson, S. E., and Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychol. Bull.* 134, 31–60. doi: 10.1037/0033-2909.134.1.31

Geoffroy, M. C., Côté, S. M., Giguère, C. É., Dionne, G., Zelazo, P. D., Tremblay, R. E., et al. (2010). Closing the gap in academic readiness and achievement: the role of early childcare. *Journal of Child Psychology and Psychiatry* 51, 1359–1367. doi: 10.1111/j.1469-7610.2010.02316.x

Geurten, M., and Bastin, C. (2019). Behaviors speak louder than explicit reports: implicit metacognition in 2.5-year-old children. *Dev. Sci.* 22, e12742. doi: 10.1111/desc.12742

Geurten, M., and Willems, S. (2016). Metacognition in early childhood: fertile ground to understand memory development?. *Child Dev. Perspect.* 10, 263–268. doi: 10.1111/cdep.12201

Goupil, L., and Kouider, S. (2016). Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Curr. Biol.* 26, 3038–3045. doi: 10.1016/j.cub.2016.09.004

Goupil, L., and Kouider, S. (2019). Developing a reflective mind: from core metacognition to explicit self-reflection. *Curr. Dir. Psychol. Sci.* 28, 403–408. doi: 10.1177/0963721419848672

Goupil, L., and Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition* 231, 105325. doi: 10.1016/j.cognition.2022.105325

Hadwin, A. F., and Webster, E. A. (2013). Calibration in goal setting: examining the nature of judgments of confidence. *Learn. Instruct.* 24, 37–47. doi: 10.1016/j.learninstruc.2012.10.001

Hammes, P. S., Bigras, M., and Crepaldi, M. A. (2016). Validity and bias of academic achievement measures in the first year of elementary school. *Int. J. Res. Method Educ.* 39, 3–18. doi: 10.1080/1743727X.2014.933473

Hattie, J. (2013). Calibration and confidence: where to next? *Learn. Instruct.* 24, 62–66. doi: 10.1016/j.learninstruc.2012.05.009

Hembacher, E., and Ghetti, S. (2013). How to bet on a memory: developmental linkages between subjective recollection and decision making. *J. Exp. Child Psychol.* 115, 436–452. doi: 10.1016/j.jecp.2013.03.010

Henrich, J., and Broesch, J. (2011). On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philosoph. Trans. Royal Soc. B: Biol. Sci.* 366, 1139–1148. doi: 10.1098/rstb.2010.0323

Heyes, C. (2020). Psychological mechanisms forged by cultural evolution. *Curr. Dir. Psychol. Sci.* 29, 399–404. doi: 10.1177/0963721420917736

Heyes, C., Bang, D., Shea, N., Frith, C. D., and Fleming, S. M. (2020). Knowing ourselves together: the cultural origins of metacognition. *Trends Cogn. Sci.* 24, 349–362. doi: 10.1016/j.tics.2020.02.007

Hughes, C., and Ensor, R. (2007). Executive function and theory of mind: predictive relations from ages 2 to 4. *Dev. Psychol.* 43, 1447–1459. doi: 10.1037/0012-1649.43.6.1447

Hughes, C., Ensor, R., Wilson, A., and Graham, A. (2009). Tracking executive function across the transition to school: a latent variable approach. *Dev. Neuropsychol.* 35, 20–36. doi: 10.1080/87565640903325691

Jacob, R., and Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: a review. *Rev. Educ. Res.* 85, 512–552. doi: 10.3102/0034654314561338

Joo, Y. S., Magnuson, K., Duncan, G. J., Schindler, H. S., Yoshikawa, H., and Ziol-Guest, K. M. (2020). What works in early childhood education programs? A meta–analysis of preschool enhancement programs. *Early Educ. Dev.* 31, 1–26. doi: 10.1080/10409289.2019.1624146

Jurado, M. B., and Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychol. Rev.* 17, 213–233. doi: 10.1007/s11065-007-9040-z

Kagan, S. L., Moore, E., and Bredekamp, S. (1995). *Reconsidering Children's Early Development and Learning: Toward Common Views and Vocabulary*. National Education Goals Panel.

Kieffer, M. J., Vukovic, R. K., and Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Read. Res. Q.* 48, 333–348. doi: 10.1002/rrq.54

Klassen, R. (2002). A question of calibration: a review of the self-efficacy beliefs of students with learning disabilities. *Learn. Disab. Quart.* 25, 88–102. doi: 10.2307/1511276

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychol. Rev.* 100, 609–639. doi: 10.1037/0033-295X.100.4.609

Koriat, A. (2007). "Metacognition and consciousness," in *The Cambridge Handbook of Consciousness*, eds. P. D. Zelazo, M. Moscovitch and E. Thompson (Cambridge, UK: Cambridge University Press), 289–325.

Koriat, A., and Ackerman, R. (2010). Metacognition and mindreading: judgments of learning for self and other during self-paced study. *Conscious. Cogn.* 19, 251–264. doi: 10.1016/j.concog.2009.12.010

Kuzyk, O., Grossman, S., and Poulin-Dubois, D. (2020). Knowing who knows: metacognitive and causal learning abilities guide infants' selective social learning. *Dev. Sci.* 23, e12904. doi: 10.1111/desc.12904

La Paro, K. M., and Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Rev. Educ. Res.* 70, 443–484. doi: 10.3102/00346543070004443

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, e13178. doi: 10.1111/desc.13178

Leckey, S., Selmeczy, D., Kazemi, A., Johnson, E. G., Hembacher, E., and Ghetti, S. (2020). Response latencies and eye gaze provide insight on how toddlers gather evidence under uncertainty. *Nat. Human Behav.* 4, 928–936.v doi: 10.1038/s41562-020-0913-y

Lerner, M. D., and Lonigan, C. J. (2014). Executive function among preschool children: unitary versus distinct abilities. *J. Psychopathol. Behav. Assess.* 36, 626–639. doi: 10.1007/s10862-014-9424-3

Lipowski, S. L., Merriman, W. E., and Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Dev. Psychol.* 49, 1505–1516. doi: 10.1037/a0030614

Lonigan, C. J. (2006). Development, assessment, and promotion of preliteracy skills. *Early Educ. Dev.* 17, 91–114. doi: 10.1207/s15566935eed1701_5

Lonigan, C. J., Burgess, S. R., and Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Dev. Psychol.* 36, 596–613. doi: 10.1037/0012-1649.36.5.596

Lyons, K. E., and Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Dev.* 84, 726–736. doi: 10.1111/cdev.12004

Lyons, K. E., and Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: elucidating the role of self-reflection in the development of self-regulation. *Adv. Child Dev. Behav.* 40, 379–412. doi: 10.1016/B978-0-12-386491-8.00010-4

Mariano, M., Santos-Junior, A., Lima, J. L., Perisinotto, J., Brandão, C., Surkan, P. J., et al. (2019). Ready for school? A systematic review of school readiness and later achievement. *Global J. Human-Soc. Sci. Res.* 19, 56–64. doi: 10.34257/GJHSSGVOL19IS10PG57

Marulis, L. M., Baker, S. T., and Whitebread, D. (2020). Integrating metacognition and executive function to enhance young children's perception of and agency in their learning. *Early Child. Res. Q.* 50, 46–54. doi: 10.1016/j.ecresq.2018.12.017

Marulis, L. M., and Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacogn. Learn.* 16, 207–231. doi: 10.1007/s11409-020-09244-6

Mashburn, A. J., and Pianta, R. C. (2006). Social relationships and school readiness. *Early Educ. Dev.* 17, 151–176. doi: 10.1207/s15566935eed1701_7

McCallan, S. A. (2010). *Influences of school, classroom, and teacher characteristics on children's school readiness* (Doctoral dissertation, University of Maryland). Digital Repository at the University of Maryland.

McClelland, M. M., Acock, A. C., and Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Child. Res. Q.* 21, 471–490. doi: 10.1016/j.ecresq.2006.09.003

McClelland, M. M., Morrison, F. J., and Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Child. Res. Q.* 15, 307–329. doi: 10.1016/S0885-2006(00)00069-7

McWayne, C. M., Fantuzzo, J. W., and McDermott, P. A. (2004). Preschool competency in context: an investigation of the unique contribution of child competencies to early academic success. *Dev. Psychol.* 40, 633. doi: 10.1037/0012-1649.40.4.633

Micalizzi, L., Brick, L. A., Flom, M., Ganiban, J. M., and Saudino, K. J. (2019). Effects of socioeconomic status and executive function on school readiness across levels of household chaos. *Early Child. Res. Q.* 47, 331–340. doi: 10.1016/j.ecresq.2019.01.007

Michalsky, T., Mevarech, Z. R., and Haibi, L. (2009). Elementary school children reading scientific texts: effects of metacognitive instruction. *J. Educ. Res.* 102, 363–376. doi: 10.3200/JOER.102.5.363-376

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., and Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *J. Cognit. Dev.*13, 395–423. doi: 10.1080/15248372.2011.585478

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

Monette, S., Bigras, M., and Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *J. Exp. Child Psychol.* 109, 158–173. doi: 10.1016/j.jecp.2011.01.008

Munakata, Y., and Michaelson, L. E. (2021). Executive functions in social context: implications for conceptualizing, measuring, and supporting developmental trajectories. *Annu. Rev. Dev. Psychol.* 3, 139–163. doi: 10.1146/annurev-devpsych-121318-085005

Nelson, T. O. (1990). "Metamemory: A theoretical framework and new findings," in *Psychology of Learning and Motivation* (Cambridge, MA: Academic Press), 125–173.

Nelson, T. O., and Narens, L. (1994). Why investigate metacognition. *Metacognition* 13, 1–25. doi: 10.7551/mitpress/4561.003.0003

Panter, J. E., and Bracken, B. A. (2009). Validity of the bracken school readiness assessment for predicting first grade readiness. *Psychol. Sch.* 46, 397–409. doi: 10.1002/pits.20385

Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., Locasale-Crouch, J., et al. (2017). Early childhood professional development: coaching and coursework effects on indicators of children's school readiness. *Early Educ. Dev.* 28, 956–975. doi: 10.1080/10409289.2017.1319783

Razza, R. A., and Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: a longitudinal analysis. *J. Appl. Dev. Psychol.* 30, 332–343. doi: 10.1016/j.appdev.2008.12.020

Resendes, T., Benchimol-Elkaim, B., Delisle, C., René, J. L., and Poulin-Dubois, D. (2021). What I know and what you know: the role of metacognitive strategies in preschoolers' selective social learning. *Cogn. Dev.* 60, 101117. doi: 10.1016/j.cogdev.2021.101117

Roberts, Y. H. (2011). *School Readiness in Children Attending Public Preschool: Implications for Public Policy, School Programming and Clinical Practice*. Cincinnati, OH: University of Cincinnati.

Roderer, T., and Roebers, C. M. (2014). Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. *Metacognit. Learn.* 9, 1–23. doi: 10.1007/s11409-013-9109-4

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

Roebers, C. M., Cimeli, P., Röthlisberger, M., and 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. *Metacognit. Learn.* 7, 151–173. doi: 10.1007/s11409-012-9089-9

Roebers, C. M., and Spiess, M. (2017). The development of metacognitive monitoring and control in second graders: A short-term longitudinal study. *J. Cognit. Dev.* 18, 110–128. doi: 10.1080/15248372.2016.1157079

Rothbart, M. K., and Bates, J. E. (2006). "Temperament," in *Handbook of Child Psychology: Vol. 3. Social, Emotional, Personality Development, 6th,* eds. W. Damon, and N. Eisenberg (York: Wiley), 99–166.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind: a comparison of Chinese and US preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x

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

Schneider, W. (2015). "The development of metamemory," in *Memory Development from Early Childhood Through Emerging Adulthood*. Cham: Springer.

Schraw, G., Kauffman, D. F., and Lehman, S. (2006). Self-regulated learning. *Encyclop. Cognit. Sci.* 1063–1073. doi: 10.1002/0470018860.s00671

Shaw, D. S., Mendelsohn, A. L., and Morris, P. A. (2021). Reducing poverty-related disparities in child development and school readiness: the smart beginnings tiered prevention strategy that combines pediatric primary care with home visiting. *Clin. Child Fam. Psychol. Rev.* 24, 669–683. doi: 10.1007/s10567-021-00366-0

Simmons, J. P., and Nelson, L. D. (2006). Intuitive confidence: choosing between intuitive and nonintuitive alternatives. *J. Exp. Psychol.: General* 135, 409. doi: 10.1037/0096-3445.135.3.409

Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., and Lonigan, C. J. (2021). Relations between executive functions and academic outcomes in elementary school children: a meta-analysis. *Psychol. Bull.* 147, 329–351. doi: 10.1037/bul0000322

Spiess, M. A., Meier, B., and Roebers, C. M. (2016). Development and longitudinal relationships between children's executive functions, prospective memory, and metacognition. *Cogn. Dev.* 38, 99–113. doi: 10.1016/j.cogdev.2016.02.003

Stanovich, K. E. (2009). "Distinguishing the reflective, algorithmic, and autonomous minds: it is time for a tri-process theory?" in *In Two Minds: Dual Processes and Beyond*, eds. J. St. B. T. Evans and K. Frankish (New York, NY: Oxford University Press), 55–88.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: what makes an expert student? *Instruct. Sci.* 1998, 127–140.

Thompson, V. A. (2010). "Towards a metacognitive dual process theory of conditional reasoning," in *In Cognition and Conditionals: Probability and Logic in Human Thinking*, eds. M. Oaksfor, and N. Chater (Oxford: Oxford University Press), 335–54.

Tibken, C., Richter, T., von Der Linden, N., Schmiedeler, S., and Schneider, W. (2022). The role of metacognitive competences in the development of school achievement among gifted adolescents. *Child Dev.* 93, 117–133. doi: 10.1111/cdev.13640

Ursache, A., Blair, C., Stifter, C., and Voegtline, K. (2013). Emotional reactivity and regulation in infancy interact to predict executive functioning in early childhood. *Dev. Psychol.* 49, 127. doi: 10.1037/a0027728

Veenman, M. V. J., and Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learn. Individ. Differ.* 15, 159–176. doi: 10.1016/j.lindif.2004.12.001

Venet, M., Normandeau, S., Letarte, M. J., and Bigras, M. (2003). Les propriétés psychométriques du Lollipop. *Revue de Psychoéducation* 32, 165–176.

Vitiello, V. E., Greenfield, D. B., Munis, P., and George, J. L. (2011). Cognitive flexibility, approaches to learning, and academic school readiness in Head Start preschool children. *Early Educ. Dev.* 22, 388–410. doi: 10.1080/10409289.2011.538366

Weiland, C., and Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Dev.* 84, 2112–2130. doi: 10.1111/cdev.12099

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., and Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: an integrated model. *Dev. Review* 59:100945. doi: 10.1016/j.dr.2020.100945

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., et al. (2013). Cognition assessment using the NIH Toolbox. *Neurology* 80(, S54–S64. doi: 10.1212/WNL.0b013e3182872ded

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., and Espy, K. A. (2011). The structure of executive function in 3-year-olds. *J. Exp. Child Psychol.* 108, 436–452. doi: 10.1016/j.jecp.2010.08.008

Williams, P. G., Lerner, M. A., Council on Early Childhood and Council on School Health (2019). School readiness. *Pediatrics* 144:e20191766. doi: 10.1542/peds.2019-1766

Willoughby, M. T., Blair, C. B., and The Family Life Project Investigators. (2016). Measuring executive function in early childhood: a case for formative measurement. *Psychol. Assess.* 28, 319–330. doi: 10.1037/pas0000152

Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Dev. Rev.* 38, 55–68. doi: 10.1016/j.dr.2015.07.001

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., and Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs Soc. Res. Child Dev.* 78, 16–33. doi: 10.1111/mono.12032

Zelazo, P. D., and Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: development and plasticity. *Child Dev. Perspect.* 6, 354–360. doi: 10.1111/j.1750-8606.2012.00246.x

TYPE Original Research
PUBLISHED 02 July 2024
DOI 10.3389/fdpys.2024.1417197

OPEN ACCESS

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

REVIEWED BY
Elena Escolano-Pérez,
University of Zaragoza, Spain
Ana Clara Ventura,
National Scientific and Technical Research Council (CONICET), Argentina

*CORRESPONDENCE
Claudia M. Roebers
✉ claudia.roebers@unibe.ch

RECEIVED 14 April 2024
ACCEPTED 14 June 2024
PUBLISHED 02 July 2024

CITATION
Kolloff K and Roebers CM (2024) The relationship between metacognitive monitoring, non-verbal intellectual ability, and memory performance in kindergarten children. *Front. Dev. Psychol.* 2:1417197.
doi: 10.3389/fdpys.2024.1417197

The relationship between metacognitive monitoring, non-verbal intellectual ability, and memory performance in kindergarten children

Kristin Kolloff and Claudia M. Roebers*

Institute of Psychology, University of Bern, Bern, Switzerland

When assessing their certainty, children are often poor at accurately monitoring their level of learning. The study examined the relationships between memory performance, intellectual ability, and metacognitive monitoring accuracy in kindergarten children. We also explored whether specific thresholds in memory performance and non-verbal intellectual ability influence metacognitive monitoring accuracy to identify group-specific patterns that might be masked by an overall linear analysis. We assessed the monitoring discrimination of 290 kindergarteners (M_{age} 6 years) using a paired-associates learning task. Results showed small correlations between task performance, intellectual ability, and metacognitive monitoring. Non-verbal intellectual ability provided explanatory value for monitoring accuracy beyond memory performance. We observed group-specific results consistent with the unskilled-and-unaware effect; children with the highest memory skills were more effective at discriminating between correct and incorrect answers than their peers with the lowest memory skills. However, kindergarteners with the highest non-verbal intellectual abilities did not demonstrate greater cognitive adaptability in novel tasks, as their monitoring accuracy was comparable to that of peers with average or lower intellectual abilities. Findings indicate that both task performance and non-verbal intellectual ability are relevant for monitoring accuracy, but the impact of non-verbal intellectual ability was less significant than anticipated. The modest correlation suggests that kindergarteners' non-verbal intellectual ability and metacognitive monitoring abilities operate relatively independently.

KEYWORDS

monitoring, memory, non-verbal intellectual ability, metacognition, discrimination, unskilled-and-unaware, intelligence

1 Introduction

Imagine a kindergarten child playing a memory game with peers. When trying to remember the correct location of the matching card, she may demonstrate metacognitive monitoring skills by evaluating different candidate positions and selecting the card for which she is most certain. However, her performance in the game may also benefit from overall and generally good memory skills or superior intellectual abilities compared to her peers. While in the adult literature, memory skills and intellectual abilities have been found to influence metacognitive monitoring, very little is known about the impact of one or both factors on children's emerging metacognitive monitoring skills. In the present approach, therefore, we will shed light on these intertwined and interacting cognitive processes in kindergarten children. Metacognitive monitoring is a fundamental part of metacognition (Flavell, 1979), referring to the ability to accurately monitor the certainty or uncertainty of one's ongoing cognitive activity and is critical for detecting errors and making informed decisions about the strategic regulation of behavior (Nelson and Narens, 1990; Lyons and Ghetti, 2011; Ghetti et al., 2013; Coughlin et al., 2015). Metacognitive monitoring is important in many everyday life situations, such as remembering shopping lists, memory cards' locations, peers' names, trains' departure times, and the like. In early childhood, children's metacognitive monitoring skills are shaped by their everyday experiences. These experiences, such as interactions with caregivers (Fukkink et al., 2024), play activities (Moore et al., 1986), and early childhood educational programs (Eckhardt and Egert, 2020), form the basis for cognitive and metacognitive skills. Individual differences in developmental trajectories can be attributed, at least in part, to contextual factors such as the interactions children have at home or at school (Ornstein et al., 2008). For example, with increasing experience, children become more skillful at using conscious strategies to remember sets of words, objects, and pictures (Schneider and Pressley, 1997). Metacognitive monitoring skills are also considered a prerequisite for self-regulated learning as it enables an individual to identify knowledge gaps, recognize and correct errors, and control and orchestrate the different cognitive processes involved (Flavell et al., 1997; Efklides, 2008; Schneider and Löffler, 2016).

There is no doubt that metacognitive monitoring processes depend on the underlying memory being monitored and the person's general intellectual abilities. When considered separately, both intellectual ability and metacognitive monitoring have consistently been shown to be strong predictors of performance across different studies involving children and adults (Neisser et al., 1996; Roth et al., 2015; Roebers, 2017; Ohtani and Hisasaka, 2018). Given that these cognitive processes undergo rapid development in children aged 5 to 7 and are not yet functioning optimally (Blair and Raver, 2015; Roebers, 2017), it prompts the question: to what extent are these interrelated processes associated during early development? Might general and abstract intellectual abilities potentially serve as a driving force in this dynamic interplay? Or are individual differences in the task at hand, the memory task, more important for children's emerging metacognitive monitoring abilities?

Indeed, research has shown that students with higher intellectual abilities not only excel in task performance but also possess a superior ability to monitor performance (Sternberg, 1985, 1999; Alexander et al., 1995; Efklides, 2019). At the same time, metacognitive monitoring appears to be distinctly influenced by the individual's level of task performance, indicating that better task performance is often associated with more accurate monitoring (Roderer and Roebers, 2013; Destan and Roebers, 2015; Händel and Dresel, 2018). However, the influence of intellectual differences on task performance, and consequently on metacognitive monitoring, has often been overlooked in previous research (Roebers, 2017). To understand the early development of metacognitive monitoring, researchers might want to consider a child's cognitive resources alongside her memory skills. This is especially the case as both metacognitive monitoring and intelligence are considered higher-order cognitive processes. It has been suggested that metacognitive monitoring ability in adolescents and adults may not entirely depend on their intellectual abilities (Veenman and Elshout, 1999; Veenman and Beishuizen, 2004).

In the present study, we aimed to investigate the relationship between metacognitive monitoring and non-verbal intellectual ability during the early stages of cognitive development, as these different cognitive processes unfold. In other words, this study aimed to understand factors contributing to differences in metacognitive monitoring among kindergarten children, especially because children in this age range typically show pronounced difficulties in assessing their memory abilities. To do so, we examined the relationships between non-verbal intellectual ability and task performance in a memory task, on the one hand, with monitoring accuracy, on the other hand, in a large sample of kindergarten children. In the following, we consistently refer to non-verbal intellectual ability as *intellectual ability*. We sought to explore to what extent memory performance, intellectual ability, and metacognitive monitoring are interrelated and how individual differences in intellectual ability influence metacognitive monitoring over and above task performance. Understanding this interplay may – in the long run – be informative for the development of kindergarten programs that strengthen not only children's cognitive but also metacognitive skills.

One way to study the impact of performance on metacognitive monitoring abilities is to compare students with lower and higher achievement levels. Research examining a variety of outcome variables suggests that high performers within a sample, defined as those with scores above the median, tend to provide more accurate metacognitive monitoring judgments than students performing below the median in that sample (e.g., Hacker et al., 2000; Händel and Fritzsche, 2016; Serra and DeMarree, 2016; Händel and Bukowski, 2019), but slightly underestimate their actual performance (Kruger and Dunning, 1999). Conversely, low performers often show less accurate monitoring, along with a tendency to overestimate their performance (Hacker et al., 2008; Roderer and Roebers, 2013). Kruger and Dunning (1999) referred to this cognitive bias as the *"unskilled-and-unaware-effect,"* whereby poor performers not only lack sufficient task-relevant knowledge but also experience a deficit in metacognitive monitoring ability. This effect has been documented in studies with adults in a variety of contexts, such as logical reasoning tasks, grammar (Kruger and Dunning, 1999), card games (Simons, 2013), medicine (Hodges et al., 2001), or mathematics (Händel and Dresel, 2018).

The "unskilled-and-unaware effect" is also evident in children. Lucangeli et al. (1997) demonstrated that students who achieved higher assessment scores had a more comprehensive understanding of the sequential steps and were more familiar with the rules and criteria required to complete tasks effectively compared to lower-achieving age-mates. This deeper understanding not only helped higher achievers to complete tasks efficiently, but also improved their ability to judge the accuracy and quality of their own work, which in turn contributed to their more accurate monitoring judgments.

A body of literature focusing on children with and without learning disabilities has highlighted the importance of possessing task-relevant knowledge for accurate monitoring (Klassen, 2007;

Job and Klassen, 2012; Crane et al., 2017). These studies revealed that children facing learning difficulties who lack such task-specific knowledge tend to exhibit lower accuracy levels and are more prone to overconfidence in their judgments. For example, Desoete et al. (2006) conducted a study with third graders. They found that children with learning disability in math, regardless of their intelligence level, not only had difficulty understanding the knowledge required for the tasks but also made significantly more errors and generally had lower monitoring accuracy than children without learning disabilities. By providing children with learning difficulties with both task-relevant and metacognitive knowledge (i.e., awareness of their own strengths and weaknesses as learners), Lucangeli et al. (2019) aimed to train the children's ability to accurately monitor their performance and identify errors. Post-training results showed that students with mathematical difficulties outperformed the control group in both performance and monitoring accuracy. After the training, these children were more active and independent in applying metacognitive monitoring, better at recognizing difficult tasks, and more engaged in identifying errors and finding solutions compared to the untrained control group.

To date, research on the unskilled-and-unaware effect in typically developing samples of young children is limited. Roderer and Roebers (2013) addressed this gap by conducting research involving fifth graders in a real-world school setting over a year. Their study focused on students' performance estimations and the deviation from their actual mathematics and science performance. Based on their test scores, these students were categorized as low, average, or high achievers. The results revealed significant differences in monitoring accuracy between lower and higher achievers, with high-achieving children consistently demonstrating higher monitoring accuracy across most of the tests. Importantly, only the two extreme groups differed from one another; that is, children with average performance did not significantly differ from either group with respect to their monitoring accuracy.

In addition to task performance, *intellectual ability* is known to have a direct impact on metacognitive monitoring skills. Several studies have shown that intellectual ability is a significant factor in memory and learning situations (Shore and Dover, 1987; Alexander et al., 1995; Hannah and Shore, 1995; Sternberg, 1999). A review of the literature reveals a variety of theoretical assumptions regarding the relationship between metacognition and its facets (metacognitive knowledge, metacognitive monitoring, and control) and intelligence (see for a review Alexander et al., 1995). A prominent example is Sternberg's Theory of Adaptive Intelligence, which views metacognition as a manifestation of intelligent thinking. According to Sternberg's definition of intelligence, a key ability is to adapt effectively to the environment (Sternberg, 1980, 1988, 2019). Intellectual differences determine how effectively individuals can approach new, challenging tasks or adapt to new situations. In this context, Sternberg's (1999) concept of metacognition (referred to as metacomponent) is of particular importance, as it considers metacognition as a fundamental skill essential for navigating and excelling in complex tasks. This concept emphasizes a profound link between metacognitive monitoring and intellectual differences, ultimately suggesting that these interrelated aspects converge into a single, overarching skill essential for effective task performance. Empirical support for Sternberg's theoretical considerations originates from giftedness research showing that gifted children tend to outperform their peers in metacognitive knowledge (i.e., knowledge about cognition see Flavell, 1979; Schneider et al., 1987; Alexander et al., 2006) and metacognitive control strategies (Carr et al., 1994). Interestingly, gifted children typically display advanced metacognitive monitoring abilities when presented with new and challenging tasks, confirming the assumed adaptive nature of metacognitive processes (Carr et al., 1996).

Studies examining intelligence and metacognition across non-gifted samples of participants, however, consistently report only small amounts of shared variance, suggesting two rather than one overarching ability (Veenman and Spaans, 2005; van der Stel and Veenman, 2008). For example, a meta-analysis examined the influence of metacognition on intelligence across various age groups, including adults, adolescents, as well as primary school and kindergarten children (Ohtani and Hisasaka, 2018). The study revealed a moderate overall correlation between metacognition and fluid intelligence ($r = 0.27$). The relationship between metacognition and general intelligence, which also encompasses the component of crystallized intelligence, has been shown to vary depending on age. That is, the effect sizes for primary school children ($r = 0.25$) and kindergarteners ($r = 0.29$) were smaller than those for adolescents ($r = 0.38$) and adults ($r = 0.34$). When focusing solely on fluid intelligence and offline measures of metacognition (i.e., typically measured after a cognitive task or learning, e.g., questionnaires), consistently weaker correlations were found. A correlation of $r = 0.23$ was found for children in primary school, $r = 0.28$ for adolescents, and $r = 0.22$ for adults. However, these findings are based on only four studies, and no study was identified examining kindergarten children. Nevertheless, the overall findings of this meta-analysis suggest that correlations between fluid intelligence and monitoring accuracy in kindergarteners are most likely low.

The limited research underscores a critical need to examine how metacognitive monitoring and intellectual ability and their mutual influence within specific learning contexts, such as memory learning, among kindergarten children. Including young children offers the advantage of exploring the interdependence of these central information-processing skills while each of these components is emerging and developing. Is their development during early ages characterized by greater independence, influenced primarily by specific everyday life experiences, while later on, they develop a mutual influence? This question is of theoretical and practical relevance as it helps to better understand developmental progression in each of the components and may inform educators how to best support children who face difficulties in either of these processes.

As to the existing methods of assessing metacognitive monitoring, there is large heterogeneity across studies. Prospective monitoring judgments, such as judgments-of-learning or feeling-of-knowing judgments, are typically assessed beforehand. In contrast, our study used retrospective monitoring judgments (confidence judgments) provided after completing a task. We chose confidence judgments because prior studies have shown that retrospective judgments are more accurate than prospective judgments (Hacker et al., 2000; van Loon and Roebers, 2021). Repeatedly, retrospective judgments have been documented to become (a) more precise with age (for reviews, see Schneider and Löffler, 2016; Roebers, 2017) and (b) task experience (Hacker et al., 2000; Bol and Hacker, 2012). That is, children become increasingly better at discriminating in their monitoring judgments between correct and incorrect task performance by giving higher judgments after correct and lower judgments after incorrect responses. Monitoring accuracy, namely, the ability to metacognitively discriminate between correct and incorrect performance, was used as the primary monitoring measure in the present study (Schraw, 2009).

1.1 Current study

Since both task performance and intellectual abilities have been shown to have a unique influence on metacognitive monitoring (e.g., Neisser et al., 1996; Veenman and Spaans, 2005; van der Stel and Veenman, 2008; Destan and Roebers, 2015; Schneider and Löffler, 2016), their interplay is of particular interest in this study. In our study, we could only build on the previously inconsistent and scarce findings on the relationship between intellectual abilities, memory performance, and monitoring performance in young children. In the context of a paired-associates learning task, a typical sample of 6-year-old kindergarteners studied Japanese symbols and their meanings and were then tested on their ability to recognize these paired-associates and provide confidence judgments about their accuracy.

Firstly, we investigated whether and how intellectual ability and memory performance are related to the ability to monitor one's own thought processes. Based on theoretical considerations and the aforementioned empirical studies showing interrelations between metacognitive monitoring and memory performance (Klassen, 2007; Crane et al., 2017), on the one side, and metacognitive monitoring and intellectual ability (van der Stel and Veenman, 2008), on the other side, we expected metacognitive monitoring, memory performance, and intellectual ability to be weakly but positively related to each other.

Secondly, we paid particular attention to whether the nature of these relationships changes when a specific threshold of either memory performance or intellectual ability is exceeded or not reached.

In the context of memory performance, our study aimed to determine whether we could observe a pattern similar to the "unskilled-and-unaware effect" proposed by Kruger and Dunning (1999) within our sample of typically developing kindergarten children. In other words, we wanted to find out whether children in the highest quartile of memory performance are better at detecting errors and distinguishing between correct and incorrect answers than children in the lowest threshold range of memory performance. Based on existing studies with adults and older children that have examined the difference between estimated and actual performance (Hacker et al., 2000; Job and Klassen, 2012; Simons, 2013), we anticipated that discrimination ability would follow a similar pattern. Specifically, we expected children whose memory performance ranked in the lowest quartile to display a substantially lower level of monitoring accuracy than those age mates whose memory performance ranked in the highest quartile.

Regarding intellectual ability, we aimed to find out, for example, whether children within the highest quartile of the distribution of intellectual abilities in our sample, similar to gifted children, have a superior ability to recognize errors and thus an advantage in monitoring their cognitive processes. This follows the findings of some intelligence research studies suggesting such an advantage (Swanson, 1992; Alexander et al., 1995), as well as Sternberg's (1985, 1988) assumption that exceptional intelligence is relevant for dealing effectively with novel situations in different cognitive domains. Thus, we expected that children with high levels of intellectual ability would exhibit high monitoring accuracy, as their cognitive abilities enable them to evaluate their performance more effectively and vice versa.

Thirdly, we wanted to go beyond existing research by investigating the extent to which intellectual abilities can predict the ability to metacognitively discriminate between correct and incorrect answers *beyond* task performance. That is, we aimed to explore whether intellectual abilities and task performance interact in predicting monitoring accuracy, as higher intellectual abilities may support task performance, which – in turn – positively impacts metacognitive monitoring. To date, there has been a lack of research on this specific issue in the context of kindergarten children. Previous research encompassing all three variables—intellectual ability, metacognitive monitoring, and task performance—primarily investigated whether intellectual ability predicts academic achievement over and above what is predicted by monitoring (van der Stel and Veenman, 2008). In contrast, we focused on the predictive value of memory performance and intellectual ability for monitoring accuracy. Despite the sparse empirical evidence available, we expected that intellectual ability would significantly contribute to monitoring accuracy, over and above task performance, uncovering interactional effects of intellectual ability and task performance on monitoring discrimination.

2 Methods

2.1 Participants

We recruited 294 children from local kindergartens in the central German-speaking part of Switzerland, each completing their mandatory second year in a kindergarten. Teachers confirmed that all but one child were sufficiently proficient in the German language to follow the verbal instructions. Two children declined participation, and one was excluded due to insufficient German language proficiency. One child was excluded from the analysis due to a technical error in a task. The final sample consisted of 290 children ($M_{age} = 6$ years, 4 months, $SD = 0.3$ years; age range between 5.5 to 7.5 years, 50% girls), with 63% being native German speakers and 30% non-native German speakers. Demographic data for 7% of the children were unavailable. Parents provided written consent, and children gave oral assent. The study was approved by the Faculty's Ethics Committee of the University of Bern and was conducted in accordance with the Declaration of Helsinki.

2.2 Materials and procedures

We utilized pretest data stemming from a broader intervention study. The pretest spanned two non-consecutive days, separated by a minimum of one day and a maximum of two days interval, with each session lasting between 45 and 60 min. We tested children in small groups ranging from three to eight children within a quiet environment at their kindergarten. Children individually completed a memory learning task, as well as a subscale of a non-verbal intelligence assessment on Samsung Galaxy Tablets A7 (10.4″) and S4 (10.5″). All the children completed the tasks on the tablets independently and easily and reported to have prior experiences with such devices. The design of the tasks, which only required simple touch gestures, was intuitive and enabled easy interaction with the technology. In addition, the tasks had no time limits, which made it even easier for the children to use the technology.

Data were anonymized and securely transferred to a server. Two trained assistants gave children general instructions about test material and provided technical support. Children received task-specific instructions through headphones.

2.2.1 Paired-associates learning task

We used a paired-associates learning task (Kanji) with a recognition test and confidence judgments to assess children's metacognitive monitoring. This task was previously used in other studies and has been proven effective as learning materials for younger children (Bayard et al., 2021; Buehler et al., 2021). We selected 32 paired associates of varying difficulty from a pool of 66 items. We randomly divided these selected items into two sets, each containing 16 Japanese characters (Kanjis). Children were then randomly assigned to either set A or B for the pretest and received the opposite set for the posttest (not included in this study). Each item was presented for 5 s in a random order. We provided items varied in difficulty for each measurement point to ensure sufficient variability in the monitoring judgments. The different difficulty levels were established based on prior studies and based on the perceptual demands required for memory encoding, as illustrated in the Supplementary Figure S1 with an example of a complex and a simple Kanji character. The examples demonstrate the variance in difficulty and the associated cognitive load required to encode them into memory. *Post-hoc* item difficulty for the Kanji task was determined by the number of correct answers divided by the total number of participants. The mean item difficulty was $M = 0.33$, $SD = 0.08$, and ranged from 0.22 (difficult) to 0.48 (easy). The tasks comprised 19% easy, 59% average, and 22% difficult items. Both sets of items demonstrated high internal consistency $\alpha > 0.92$.

Prior to the test, children were introduced to four short stories illustrating varying levels of certainty and uncertainty in everyday situations (see Supplementary Figure S2). These stories aimed to familiarize them with a 7-point Likert scale presented as a thermometer (adapted from Koriat and Shitzer-Reichert, 2002, see Supplementary Figure S3). They were instructed to use this scale to rate their feelings of certainty or uncertainty, such as feeling unsure when the content of a jar was unknown or invisible *versus* feeling sure when the content was known or visible. The thermometer scale consisted of seven colored fields, each corresponding to a different level of certainty, ranging from blue (very unsure = 1) to red (very sure = 7). To ensure that the children could accurately apply the scale and become familiar with the task steps, they conducted a practice trial on the tablet, which they completed successfully. Children understood the rationale of the scale with ease. On average, participants completed the task within 15 min.

The paired-associates learning task consisted of four phases: a study phase, a filler task, a recognition phase, and a monitoring phase. In the study phase, children were required to learn 16 Japanese characters (Kanji) and their corresponding pictorial interpretations (see Figure 1 for the schematic task procedure). Following the study phase, a 1-min filler task was introduced to discourage rehearsal strategies. Children used their fingers to interact with a moving cat on the screen. Immediately after that, the recognition phase started. A Kanji appeared on the left side of the screen, accompanied by four pictures on the right side. Among the four randomly presented alternatives, one picture matched the meaning of the Kanji, while the other three corresponded to different learned picture pairs. Each test trial concluded with the monitoring phase, in which children provided confidence judgments in their given answer being correct by using the thermometer scale. We calculated the percentage of correctly recognized items out of 16 items to measure memory performance.

To assess monitoring, we calculated a monitoring discrimination score that quantifies the ability to discriminate between confidence judgments for correct and incorrect answers (Schraw, 2009; Fleming and Lau, 2014). Monitoring discrimination measures the accuracy with which participants metacognitively discriminate between correct and incorrect answers in their monitoring judgments. Thus, we computed the difference between the mean confidence judgment after correct answers and the mean confidence judgments after incorrect answers. A positive monitoring discrimination score suggests that an individual is able to discriminate the accuracy of given answers by giving higher judgments after correct than after incorrect responses.

2.2.2 Intellectual ability

To assess children's intellectual ability, we administered the computerized Odd-Item Out RIAS subscale (Reynolds and Kamphaus, 2009, German adaptation: Hagmann-von Arx and Grob, 2014). This subtest not only assesses non-verbal reasoning skills but also necessitates the application of spatial ability, visual imagery, and a range of other non-verbal skills across various items. It essentially serves as a reverse form of non-verbal analogy. We chose this subscale due to its capability to assess intellectual ability from the earliest stages of development effectively, its excellent psychometric properties (Andrews, 2007), and its adaptability for computerized administration, ensuring the test's independence from reading skills.

In this test, the children were required to identify one stimulus out of five to six possible alternatives by choosing an unrelated item (see Supplementary Figure S3 for an example item). Patterns progressively increased in difficulty. The test consisted of four practice trials and 51 test items. If the given answer was correct and provided within 30 s, the participants received two points. If

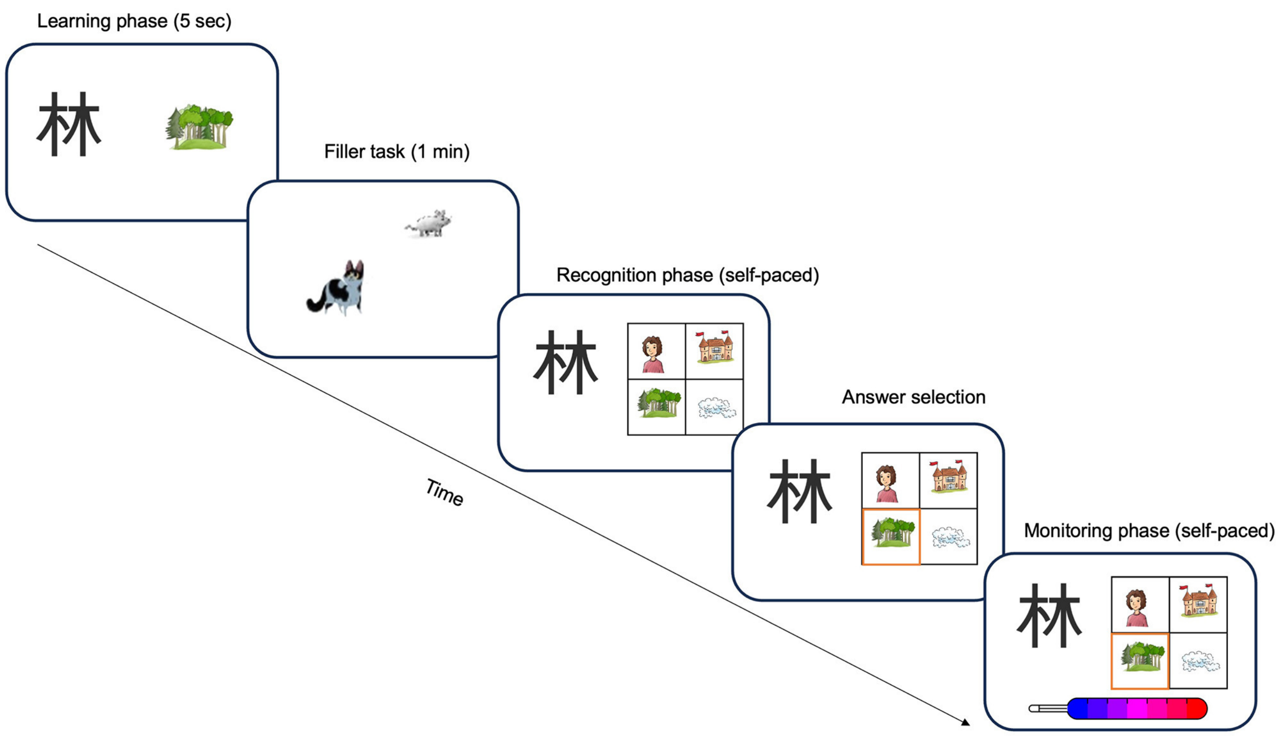


FIGURE 1
Task procedure at pretest and posttest. Schematic representation of Kanji task procedure. Children learned 16 items.

the answer was correct and provided within 50 s, the participants received one point. Otherwise, the participant received zero points, and the next array was shown. After three consecutive incorrect answers, the test ended. Test scores could range from 0 to 102 points. The score for intellectual ability was calculated as the sum of received points.

2.3 Statistical analyses

We conducted all analyses in R version 4.2.3 (R Core Team, 2022). We ran Pearson correlations in the *correlation* package version 0.8.4 (Makowski et al., 2020) and hierarchical linear regression models in the *stats* package version 4.2.3. We used an alpha level of 5% for significance tests. For the effect sizes, we reported eta squared (η^2). Values are defined as small = 0.01, medium =0.06, and large =0.14 effects, respectively. For the effect sizes Cohen's d, values are defined as small = 0.2, medium = 0.5, and large = 0.8, respectively (Field et al., 2012).

3 Results

The descriptive statistics for children's memory performance, non-verbal intelligence, confidence judgments following correct and incorrect memory performance, and monitoring discrimination are presented in Table 1.

Although the memory task turned out to be rather difficult, kindergarteners demonstrated a significant ability to distinguish

TABLE 1 Descriptive statistics for study variables ($N = 290$).

Measures	*M*	*SD*	Range
CJ correct recognition	5.41	1.67	1–7
CJ incorrect recognition	4.93	1.76	1–7
Monitoring discrimination	0.49	0.93	−2.14–3.79
Non-verbal intellectual ability	32.87	10.72	4–66
Memory performance (%)	0.33	0.13	0.06–0.94

M, mean; *SD*, standard deviation; CJ, confidence judgments.

between correct and incorrect answers [$t(289) = -7.33$, $p < 0.001$, $d = 0.27$]. They consistently assigned higher confidence judgments to correct responses compared to incorrect ones, suggesting an emerging ability to metacognitively differentiate between what they could recognize and what not. Despite these early signs of monitoring skills, inspection of Table 1 showed that the mean confidence judgments of incorrect answers were still quite high, leaving ample room for more accurate monitoring of uncertainty.

3.1 Interrelations between metacognitive monitoring, memory performance, and intellectual ability

Pearson correlations revealed significant relationships among monitoring discrimination, intellectual ability, and memory

TABLE 2 Average memory performance and monitoring discrimination across performance quartiles.

Quartile	*n*	Memory performance *M* (*SD*)	Monitoring discrimination *M* (*SD*)
Q1	73	0.17 (0.05)	0.17 (1.13)
Q2	73	0.28 (0.03)	0.41 (1.07)
Q3	72	0.39 (0.04)	0.48 (0.94)
Q4	72	0.55 (0.11)	0.86 (1.16)

N, number of participants per group; *M*, mean; *SD*, standard deviation.

performance. A significant correlation was found between monitoring and intellectual ability ($r = 0.16$, $p = 0.01$), as well as between monitoring and memory performance ($r = 0.22$, $p < 0.001$). Furthermore, a significant correlation was also observed between memory performance and intellectual ability ($r = 0.12$, $p = 0.04$). That is, children who either had higher intellectual ability or higher memory accuracy were better able to discriminate between correct and incorrect answers. In sum, our expectations that relations between metacognitive and cognitive study variables would be positively intercorrelated were confirmed.

3.2 Impact of memory performance levels on monitoring accuracy

To investigate the influence of variations in memory performance on monitoring, that is, to explore whether there is a certain threshold of memory performance under or above which differences in monitoring will occur (unskilled-and-unaware-effect) or a non-linear relationship, we analyzed the variation in children's monitoring across memory performance quartiles. Descriptive statistics for monitoring accuracy as a function of memory performance quartile are presented in Table 2. We conducted a one-way ANOVA with monitoring discrimination as the dependent variable and memory performance (four performance quartiles) as the independent variable. The results revealed a significant difference in memory performance on monitoring discrimination, $F_{(3,286)} = 5.04$, $p = 0.01$, $\eta^2 = 0.05$, showing that monitoring discrimination seemed to be more accurate in high compared to low performers (see Figure 2). Subsequent analyses using Tukey HSD *post-hoc* tests revealed a significant contrast in monitoring accuracy between groups only at the extremes of memory performance (see Supplementary Table S1). This finding supports our hypothesis that children with higher memory performance are more capable of discriminating between correct and incorrect answers than children with the lowest memory performance.

3.3 Impact of intellectual ability levels on monitoring

To explore whether there is a non-linear relationship or whether there is a certain threshold of intellectual ability under or above which differences in monitoring will occur, we analyzed children's levels of intellectual ability by segmenting intellectual ability into quartiles. Descriptive statistics are presented in Table 3. We conducted a one-way ANOVA with monitoring discrimination as the dependent variable and intellectual ability (four quartiles) as the independent variable. The results showed no specific threshold of intellectual ability influencing monitoring discrimination [$F_{(3,286)} = 2.25$, $p = 0.08$, $\eta^2 = 0.02$] (see Figure 3). Unexpectedly, children with high intellectual ability did not outperform those with the lowest memory performance in distinguishing between correct and incorrect answers.

3.4 Role of memory performance and intellectual abilities in predicting monitoring discrimination

To investigate the extent to which intellectual abilities can predict the monitoring of one's own thought processes *beyond* task performance, we conducted a three-stage hierarchical linear regression with monitoring discrimination as the dependent variable. Memory performance was introduced at stage one, followed by intellectual ability at stage two, and their interaction at stage three. Prior to conducting hierarchical regression analysis, several assumptions need to be met to ensure the validity of the results. Therefore, we reviewed the linearity and homogeneity of our final hierarchical model by reviewing Q-Q plots, which plot residuals against predicted values. The plot has no visible patterns, indicating that our model's assumption was not met. The Breusch-Pagan test was conducted to assess the presence of heteroscedasticity. No evidence was found $\chi^2(2) = 0.03$, $p > 0.05$. For the predictor variables memory performance and intellectual ability, the VIF values were found to be VIF = 1.014, indicating no issues of multicollinearity. As anticipated, Model 1 revealed that memory performance significantly predicted monitoring discrimination independently ($b = 1.60$, $p < 0.001$, $R^2 = 0.05$), aligning with our expectations. Similarly, including intellectual ability in Model 2 explained an additional 2% of the variance ($b = 0.02$, $p = 0.02$), which is also consistent with our hypothesis. However, contrary to our expectations, the interaction between memory performance and non-verbal intelligence in Model 3 was not significant and did not yield a significant increase in explained variance ($b = 0.01$, $p = 0.24$). Both memory performance and intellectual ability were predictive of monitoring discrimination, suggesting that higher levels of memory performance and intellectual ability are associated with increased monitoring accuracy. However, the non-significant interaction indicated that the relationship between performance and monitoring discrimination does not significantly differ across levels of memory performance and intellectual ability.

4 Discussion

Our investigation explored the relationship between metacognitive monitoring, memory performance, and non-verbal intellectual ability in kindergarten children, using a

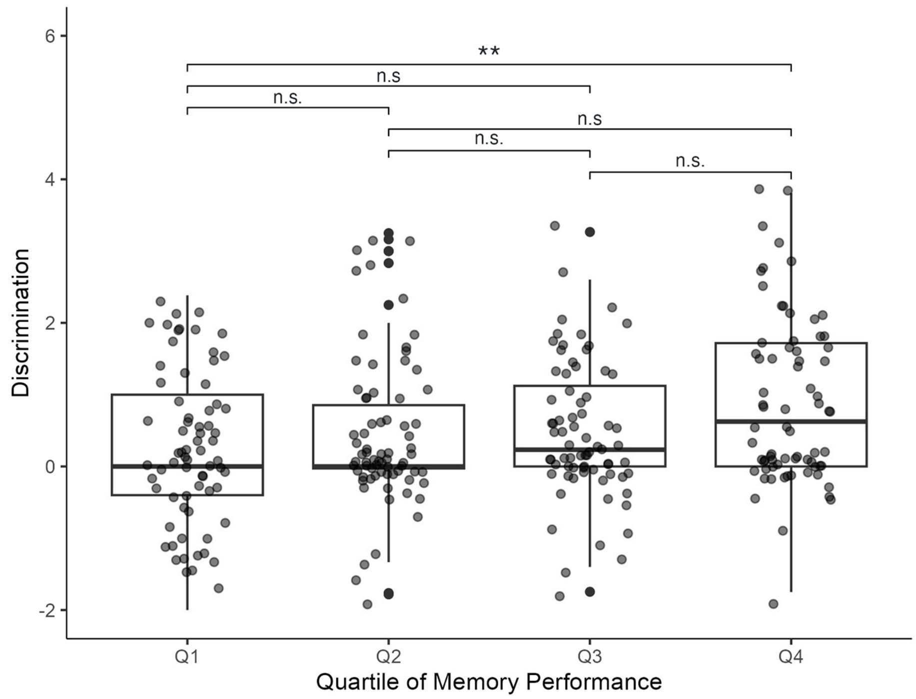


FIGURE 2
Distribution of monitoring discrimination as a function of memory performance quartiles. n.s, not significant, $^{**}p = 0.01$.

TABLE 3 Mean intellectual ability and monitoring discrimination as a function of intellectual ability quartiles.

Quartile	*n*	Intellectual ability *M* (*SD*)	Monitoring discrimination *M* (*SD*)
Q1	73	20.01 (3.76)	0.31 (1.07)
Q2	73	28.52 (2.31)	0.30 (1.04)
Q3	72	36.03 (2.09)	0.66 (1.15)
Q4	72	47.17 (6.04)	0.63 (1.12)

N, number of participants per group; *M*, mean; *SD*, standard deviation.

paired-associate paradigm, allowing them to metacognitively monitor the accuracy of their responses in a recognition test. We investigated how task performance and intellectual ability relate to monitoring accuracy by focusing on children's ability to metacognitively discriminate between correctly and incorrectly identified pairs of items. Additionally, we sought to assess the extent to which intellectual abilities can predict the metacognitive discrimination of correct and incorrect responses beyond task performance. Finally, we explored the possibility of non-linear relationships and therefore investigated whether there are specific thresholds for memory performance and intellectual ability that, when exceeded or not exceeded, have different effects on monitoring accuracy.

In line with our expectations, we found memory children's memory performance to be positively related to metacognitive discrimination, indicating that children who recognized more item pairs demonstrated an improved ability to distinguish between their correct and incorrect responses than children with lower memory performance. This is in line with previous work showing that a higher level of test performance is linked to higher monitoring accuracy (Roebers et al., 2009; Roderer and Roebers, 2013, 2014). This underscores the significance of task performance for metacognitive judgments, even among kindergarteners (Ohtani and Hisasaka, 2018).

Additionally, our findings revealed a small correlation between children's intellectual ability and monitoring accuracy, suggesting that those with higher intellectual ability exhibited greater precision in their metacognitive monitoring. This finding supports the results of recent studies that have also found small to moderate effects between metacognitive processes and intelligence in primary school children, suggesting that these two cognitive processes are only partially dependent (Sarac et al., 2014; Ohtani and Hisasaka, 2018). Yet, the finding contrasts with Sternberg's (1999) proposition of a singular overarching higher-order cognitive construct. Furthermore, we explored whether intellectual ability could independently account for variance in metacognitive monitoring ability beyond memory performance. Consistent with our expectations and previous research, our results revealed that intellectual ability has additional explanatory value for the monitoring accuracy in kindergarten children beyond memory performance (Veenman and Beishuizen, 2004; van der Stel and Veenman, 2008, 2010). Nevertheless, the unique contribution of intellectual ability to the prediction of discrimination performance was noticeably lower compared to memory performance, possibly implying that task performance holds greater significance than intellectual ability. The absence of a significant interaction between memory performance and intellectual ability in predicting monitoring accuracy was somewhat

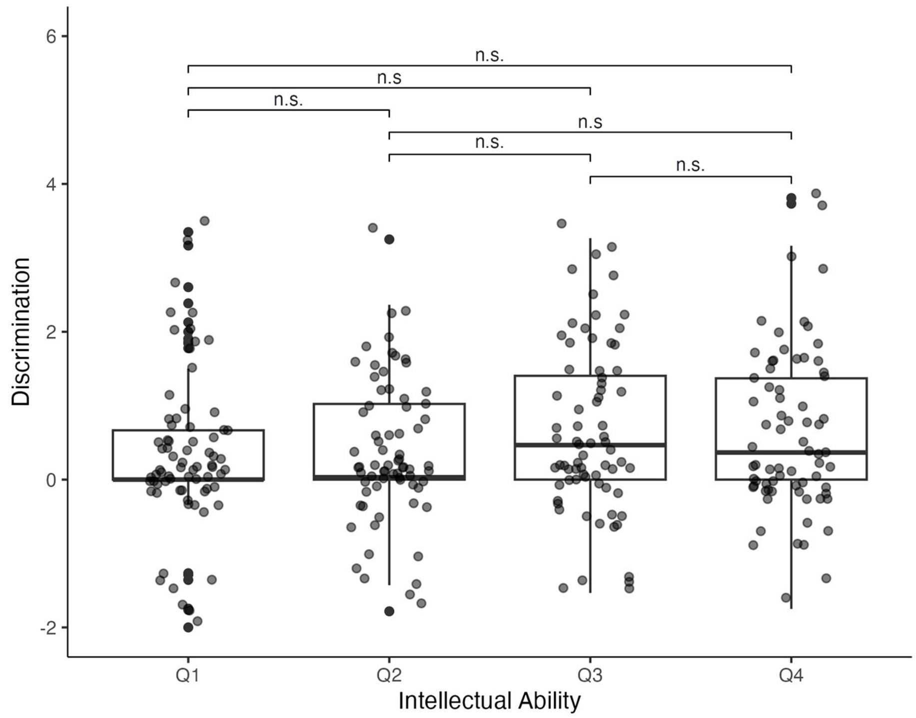


FIGURE 3
Distribution of monitoring discrimination as a function of intellectual ability quartiles. n.s, not significant.

unexpected, especially considering that both variables individually contributed to monitoring accuracy (Hannah and Shore, 1995; Neisser et al., 1996; Roth et al., 2015; Roebers, 2017).

Our results suggest that although there is a relationship between memory performance, intellectual ability, and the ability for metacognitive monitoring in young children, this is only a weak association. Thereby, our findings support previous research that cognitive and metacognitive abilities function largely independently during childhood (Veenman and Elshout, 1999; Veenman and Beishuizen, 2004). In other words, children's ability to recognize information from memory, their intellectual ability, and their ability to monitor the accuracy of their own answers appear to operate more in parallel than in close dependence on each other.

Another aim of our study was to explore the existence of specific thresholds in memory performance and intellectual ability above or below which metacognitive monitoring accuracy might be significantly affected. Our approach aimed, firstly, to identify group-specific patterns within the distributions of memory and intellectual ability that might remain masked in overall linear analysis and, secondly, to quantify the magnitude of group differences necessary to have a meaningful impact on monitoring accuracy. Thus, the sample was divided into quartiles based on their memory performance and intellectual ability.

The results obtained for memory performance levels were consistent with the pattern of the often-replicated unskilled-and-unaware effect (Kruger and Dunning, 1999). Even among kindergarten children, those who ranked in the top quartile in memory performance not only made fewer errors but were also more accurate in their metacognitive discrimination compared to their peers with the lowest memory performance quartile. This means that children with substantially better memory skills were better able to metacognitively distinguish between their correct and incorrect answers than children who struggled with the memory task. However, in this young age group, the unskilled-and-unaware-effect was only detectable between the two extreme groups of memory performance. Our results thus support Kruger and Dunning's assumption that a high level of task performance, such as knowing how to master a particular task, may be necessary to accurately judge one's own performance, which may lead to more accurate monitoring between correct and incorrect responses (Kruger and Dunning, 1999; Dunning et al., 2003; Ehrlinger et al., 2008). Another possible explanation for the present results could be that the children's metacognitive judgments are influenced by different cues at the level of the individual tasks (e.g., Koriat, 1997; see also Koriat and Levy-Sadot, 1999). It is conceivable that high-achieving vs. low-achieving children use different cues to assess their confidence (Koriat et al., 2009a,b). For example, high-achieving children may have more task-specific experience that enables them to make judgments about their own learning or work processes based on concrete information and data.

In contrast, low-performing children may rely more on intuitive judgments because they lack the necessary knowledge or experience to analyze their assessments more systematically (Schneider and Löffler, 2016). Applied to our sample, it might be that high-performing children rely on the content of information that they may know from previous experiences (Schneider, 1993; Bjorklund and Schneider, 1996). It is known that in this age group, features of the home environment, such as playing memory games with family members and discussing these experiences with parents and other relatives, play a crucial role in children's cognitive development (Gottfried, 1984; Harris and Almutairi, 2016). For example, Carr et al. (1989) showed that the amount and type of parental interaction, understanding of game rules, and task-related knowledge are positively related in second graders. In particular, children who learned strategies in situations relevant to everyday life and had access to games promoting strategic thinking showed more pronounced metacognition than their peers from families with less metacognitive, strategy-based interactions. The findings from previous studies, along with our results, suggest that early experiences with task-based knowledge and strategic thinking are likely to positively influence memory and recall.

Furthermore, young children seem to develop a kind of "expertise" through prior experiences, potentially helping them discover and apply memory strategies later on (Ornstein et al., 2006, 2008). For example, Schneider (2015) showed that children show more strategic behavior in familiar environments and in familiar tasks, such as memory tasks, compared to new tasks. This means that a wide range of experiences can help to quickly apply previous experience to new situations. Through these interactions, children learn to reflect on their own thought processes, identify effective strategies, and adapt their learning approaches accordingly. This hands-on engagement facilitated by parents is instrumental in nurturing children's metacognitive skills and enhancing their overall cognitive development (Moore et al., 1986; Laakso, 1995).

Low-performing children, on the other hand, may lack the task mastery skills to utilize information-based processing. In other words, children with lower performance levels may base their confidence on subjective feelings, such as the ease with which information comes to mind or their motivation and wishful thinking to give their best. However, this approach may not be as reliable as confidence judgments based on the processing of memory information and could, therefore, be more prone to error and result in inaccurate monitoring judgments (Efklides, 2008). However, as we did not directly measure children's previous experiences with memory games at home in our study — for example, by asking parents how often they play them at home — this explanation remains speculative. Nevertheless, future research could investigate the home environment in more detail to find out how it influences error detection and the acquisition of metacognitive monitoring skills (Schneider, 2015; Roebers, 2022). In summary, the findings from our sample suggest that during the preschool years, understanding how to effectively tackle a task is important for success in memory tasks. This implies that having at least a fundamental understanding of task-specific knowledge, likely acquired through prior experiences with memory tasks, is necessary for children to learn how to recognize errors and accurately assess resulting uncertainties (Schneider et al., 1989; Schneider, 1993; Brod, 2021).

The assumption that metacognitive discrimination in task-specific confidence judgments might also have varied as a function of intellectual ability was based on theoretical assumptions on the adaptive nature of intellectual thinking (Sternberg, 1985, 1999, 2019). This perspective suggests that children with higher intellectual ability are able to adapt their thinking more quickly to new and challenging tasks, possibly using both their advanced mental activities and metacognitive skills to respond flexibly to new and challenging memory tasks. Our results did not align with this assumption. That is, our findings indicated that in our sample kindergarteners with the highest intellectual abilities did not demonstrate advanced metacognitive adaptability in the novel task by showing higher monitoring accuracy compared to children with average or below-average intellectual abilities. This finding is noteworthy considering previous research exploring differences in metacognition between gifted and non-gifted children (Alexander et al., 1995; Alexander and Schwanenflugel, 1996; Snyder et al., 2011; Efklides, 2019; Straka et al., 2021; Tibken et al., 2022). Several explanations could be considered for our unexpected finding. Firstly, our results do not necessarily contradict previous research on giftedness, as it could be that our sample did not include children who fall into this specific range of giftedness (Alexander et al., 1995; Efklides, 2019), and it was also not our primary interest to include specifically gifted children. Secondly, despite the apparent homogeneity of the overall sample, variations or disparities within the quartiles could have influenced the relationship between the variables. Such intra-quartile variability may preclude the identification of consistent linear patterns in the relationships between these variables, which we did find in the regression analysis (van der Stel and Veenman, 2014). Thirdly, the methodology for measuring intellectual ability might not have adequately captured the specific relationships between the quartiles. Relying solely on non-verbal abilities as a measure of intelligence is a limitation in our approach (Ohtani and Hisasaka, 2018). This restricted approach may fail to capture other, more crucial aspects of intelligence, such as verbal abilities, problem-solving skills, and cognitive flexibility. Future research could explore alternative methodologies for measuring intellectual ability that encompass a broader range of cognitive functions, including both verbal and non-verbal abilities, for achieving a more comprehensive measurement of intelligence.

It is also worth noting that participants in previous studies were older than those in our study, which may have implications for the interpretation of our findings. Specifically, young children's confidence judgments may be more strongly influenced by motivational factors compared to other participants (Efklides, 2019). For example, young children are known to base their monitoring – at least in part – on their desire for good performance (wishful thinking; Schneider, 1998) or their effort in completing the task well (effort heuristic; Koriat et al., 2009a). The often observed monitoring accuracy increases in older children can thus also be attributed to decreasing self-protective biases, especially as they typically and more frequently acknowledge the possibility of errors (Efklides and Tsiora, 2002; van Loon et al., 2017; van Loon and Roebers, 2020).

4.1 Limitations

Our study, while providing important and new insights, is not without limitations. Firstly, our study's correlational design necessitates a cautious interpretation of the findings, particularly given the assumed interplay between cognitive and metacognitive factors (Flavell, 1979; Zimmerman, 1995; Boekaerts, 1999; Dignath and Büttner, 2008). While we observed a relationship between monitoring accuracy and memory performance, we did not investigate a possible bidirectional nature of cognitive and metacognitive processes, as discussed in self-regulation models (e.g., Efklides, 2008, 2011). It is equally plausible that high monitoring accuracy fosters improved memory performance, potentially through enhanced metacognitive regulation and more effective use of learning strategies (Lockl and Schneider, 2007; Roebers et al., 2014; Godfrey et al., 2023). This possibility aligns with the notion that metacognitive experiences and internal feedback loops generated through repeated exposure to learning materials—as was the case with the 16 items in our study—play a critical role in shaping learning behavior and performance (Efklides, 2006; Efklides and Metallidou, 2020). Therefore, our correlational findings could reflect a dynamic and reciprocal relationship between monitoring accuracy and memory performance rather than a unidirectional influence. To disentangle these complex interactions and clarify the directionality of these relationships, future research should employ longitudinal or experimental designs, which would allow for a more detailed exploration of the causality and bidirectionality between these key factors in learning.

Secondly, given the practical feasibility constraints of conducting a large-scale intervention study (as a reminder, this study utilized pretest data), we were compelled to strike a balance between conducting a comprehensive assessment and addressing the practical challenges associated with testing young children. As a result, we chose to assess intellectual ability exclusively through the Odd-Item Out subscale of the Reynolds scale (i.e., RIAS; Hagmann-von Arx and Grob, 2014) to provide an economical and largely stress-free test experience for our young participants. We acknowledge this choice as a limitation and recommend that future research should encompass a broader range of intellectual ability measures. Finally, although the paired-associates test was designed to be challenging, the actual performance was somewhat low (M = 33%, SD = 13%). Although there were no pronounced floor effects, it is possible that the somewhat poor memory performance, on average, has impacted the results.

5 Conclusion

In summary, this study analyzed the relationships between metacognitive monitoring, memory performance, and intellectual abilities in a sample of kindergarten children based on their metacognitive ability to discriminate between correct and incorrect responses in a memory task. While the findings revealed that children's monitoring accuracy did relate to memory performance and intellectual ability, the extent of this relationship, especially for intellectual ability, was less significant than expected. The small correlation suggests that cognitive and metacognitive abilities function rather independently in kindergarteners. Furthermore, our analysis of group differences revealed that only the extreme ranges of task performance—specifically, very low and very high levels—affected kindergarteners' ability to accurately distinguish between correct and incorrect answers. In contrast, variations in intellectual ability, whether at lower or higher levels, did not impact children's monitoring accuracy. Thus, our findings emphasize the importance of task performance yielding direct and positive effects on metacognitive monitoring, while intellectual abilities appear to play a more subordinate role. Against this background, the present study underscores the necessity to include different cognitive factors operating at the task at hand to increase our understanding regarding children's early stages of metacognitive development.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee Department of Psychology, University of Bern. 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

KK: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft. CR: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The research leading to these results received funding from the Swiss National Science Foundation under Grant agreement no. 197336. The grant was awarded to CR.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdpys.2024.1417197/full#supplementary-material

References

Alexander, J. M., Carr, M., and Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: directions for future research. *Dev. Rev.* 15, 1–37. doi: 10.1006/drev.1995.1001

Alexander, J. M., Johnson, K. E., Albano, J., Freygang, T., and Scott, B. (2006). Relations between intelligence and the development of metaconceptual knowledge. *Metacognit. Learn.* 1, 51–67. doi: 10.1007/s11409-006-6586-8

Alexander, J. M., and Schwanenflugel, P. J. (1996). Development of metacognitive concepts about thinking in gifted and non-gifted children: recent research. *Learn. Ind. Diff.* 8, 305–325. doi: 10.1016/S1041-6080(96)90021-7

Andrews, J. J. W. (2007). Test Reviews: Reynolds, C. R., and Kamphaus, R. W. (2003). RIAS: Reynolds Intellectual Assessment Scales. Lutz, FL: Psychological Assessment Resources, Inc. *J. Psychoeduc. Assessm.* 25, 402–408. doi: 10.1177/0734282907300381

Bayard, N. S., van Loon, M. H., Steiner, M., and Roebers, C. M. (2021). Developmental improvements and persisting difficulties in children's metacognitive monitoring and control skills: cross-sectional and longitudinal perspectives. *Child Dev.* 92, 1118–1136. doi: 10.1111/cdev.13486

Bjorklund, D. F., and Schneider, W. (1996). The interaction of knowledge, aptitude, and strategies in children's memory performance. *Adv. Child Dev. Behav.* 26, 59–89. doi: 10.1016/S0065-2407(08)60506-6

Blair, C., and Raver, C. C. (2015). School readiness and self-regulation: a developmental psychobiological approach. *Ann. Rev. Psychol.* 66, 711–731. doi: 10.1146/annurev-psych-010814-015221

Boekaerts, M. (1999). Self-regulated learning: where we are today. *Int. J. Educ. Res.* 31, 445–457. doi: 10.1016/S0883-0355(99)00014-2

Bol, L., and Hacker, D. (2012). Calibration research: Where do we go from here? *Front. Psychol.* 3:229. doi: 10.3389/fpsyg.2012.00229

Brod, G. (2021). Generative learning: Which strategies for what age? *Educ. Psychol. Rev.* 33, 1295–1318.

Buehler, F. J., van Loon, M. H., Bayard, N. S., Steiner, M., and Roebers, C. M. (2021). Comparing metacognitive monitoring between native and non-native speaking primary school students. *Metacogn. Learn.* 16, 749–768. doi: 10.1007/s11409-021-09261-z

Carr, M., Alexander, J., and Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Appl. Cognit. Psychol.* 8, 583–595. doi: 10.1002/acp.2350080605

Carr, M., Alexander, J., and Schwanenflugel, P. (1996). Where gifted children do and do not excel on metacognitive tasks. *Roeper Rev.* 18, 212–217. doi: 10.1080/02783199609553740

Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., and Borkowski, J. G. (1989). Strategy acquisition and transfer among American and German children: environmental influences on metacognitive development. *Dev. Psychol.* 25, 765–771. doi: 10.1037/0012-1649.25.5.765

Coughlin, C., Hembacher, E., Lyons, K. E., and Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev. Sci.* 18, 957–971. doi: 10.1111/desc.12271

Crane, N., Zusho, A., Ding, Y., and Cancelli, A. (2017). Domain-specific metacognitive calibration in children with learning disabilities. *Contemp. Educ. Psychol.* 50, 72–79. doi: 10.1016/j.cedpsych.2016.09.006

Desoete, A., Roeyers, H., and Huylebroeck, A. (2006). Metacognitive skills in Belgian third grade children (age 8 to 9) with and without mathematical learning disabilities. *Metacognit. Learn.* 1, 119–135. doi: 10.1007/s11409-006-8152-9

Destan, N., and Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognit. Learn.* 10, 347–374. doi: 10.1007/s11409-014-9133-z

Dignath, C., and Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognit. Learn.* 3, 231–264. doi: 10.1007/s11409-008-9029-x

Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003). Why people fail to recognize their own incompetence. *Curr. Direct. Psychol. Sci.* 12, 83–87. doi: 10.1111/1467-8721.01235

Eckhardt, A. G., and Egert, F. (2020). Predictors for the quality of family child care: a meta-analysis. *Childr. Youth Serv. Rev.* 116:105205. doi: 10.1016/j.childyouth.2020.105205

Efklides, A. (2006). Metacognitive experiences: the missing link in the self-regulated learning process. *Educ. Psychol. Rev.* 18, 287–291. doi: 10.1007/s10648-006-9021-4

Efklides, A. (2008). Metacognition: defining its facets and levels of functioning in relation to self-regulation and co-regulation. *Eur. Psychol.* 13, 277–287. doi: 10.1027/1016-9040.13.4.277

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* 46, 6–25. doi: 10.1080/00461520.2011.538645

Efklides, A. (2019). Gifted students and self-regulated learning: the MASRL model and its implications for SRL. *High Ability Stu.* 30, 79–102. doi: 10.1080/13598139.2018.1556069

Efklides, A., and Metallidou, P. M. (2020). *Applying Metacognition and Self-Regulated Learning in the Classroom. The Oxford Encyclopedia of Educational Psychology*. Oxford: Oxford University Press. doi: 10.1093/acrefore/9780190264093.013.961

Efklides, A., and Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychol. Int. J. Psychol. Orient* 45, 222–236. doi: 10.2117/psysoc.2002.222

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. (2008). Why the unskilled are unaware: further explorations of (absent) self-insight among the incompetent. *Org. Behav. Hum. Dec. Proc.* 105, 98–121. doi: 10.1016/j.obhdp.2007.05.002

Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. London: Sage Publications.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* 34, 906–911. doi: 10.1037/0003-066X.34.10.906

Flavell, J. H., Green, F. L., Flavell, E. R., and Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Dev.* 68, 39–47. doi: 10.2307/1131923

Fleming, S. M., and Lau, H. C. (2014). How to measure metacognition. *Front. Hum. Neurosci.* 8:443. doi: 10.3389/fnhum.2014.00443

Fukkink, R. G., Sluiter, R. M. V., and Fekkes, M. (2024). Transition from childcare to school: surgency, center-based care and caregiver-child relationship predict self-regulation, social competence and well-being. *Learn. Ind. Diff.* 110:102409. doi: 10.1016/j.lindif.2024.102409

Ghetti, S., Hembacher, E., and Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: a metacognitive approach. *Child Dev. Persp.* 7, 160–165. doi: 10.1111/cdep.12035

Godfrey, M., Casnar, C., Stolz, E., Ailion, A., Moore, T., Gioia, G., et al. (2023). A review of procedural and declarative metamemory development across childhood. *Child Neuropsychol.* 29, 183–212. doi: 10.1080/09297049.2022.2055751

Gottfried, A. W. W. (1984). "Home environment and early cognitive development: Integration, meta-analyses, and conclusions," in *Home Environment and Early Cognitive Development: Longitudinal Research*, ed. A. W. Gottfried (London: Academic Press), 329–342.

Hacker, D. J., Bol, L., and Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognit. Learn.* 3, 101–121. doi: 10.1007/s11409-008-9021-5

Hacker, D. J., Bol, L., Horgan, D. D., and Rakow, E. A. (2000). Test prediction and performance in a classroom context. *J. Educ. Psychol.* 92, 160–170. doi: 10.1037/0022-0663.92.1.160

Hagmann-von Arx, P., and Grob, A. (2014). *RIAS - Reynolds Intellectual Assessment Scales and Screening: Deutschsprachige Adaptation der Reynolds Intellectual Assessment Scales (RIAS).* London: Hans Huber.

Händel, M., and Bukowski, A. K. (2019). The gap between desired and expected performance as predictor for judgment confidence. *J. Appl. Res. Mem. Cognit.* 8, 347–354. doi: 10.1016/j.jarmac.2019.05.005

Händel, M., and Dresel, M. (2018). Confidence in performance judgment accuracy: the unskilled and unaware effect revisited. *Metacognit. Learn.* 13, 265–285. doi: 10.1007/s11409-018-9185-6

Händel, M., and Fritzsche, E. S. (2016). Unskilled but subjectively aware: metacognitive monitoring ability and respective awareness in low-performing students. *Mem. Cognit.* 44, 229–241. doi: 10.3758/s13421-015-0552-0

Hannah, C. L., and Shore, B. M. (1995). Metacognition and high intellectual ability: Insights from the study of learning-disabled gifted students. *Gifted Child Q.* 39, 95–109. doi: 10.1177/001698629503900206

Harris, Y. R., and Almutairi, S. (2016). A commentary on parent–child cognitive learning interaction research: What have we learned from two decades of research? *Front. Psychol.* 7:1210. doi: 10.3389/fpsyg.2016.01210

Hodges, B., Regehr, G., and Martin, D. (2001). Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad. Med.* 76:S87. doi: 10.1097/00001888-200110001-00029

Job, J. M., and Klassen, R. M. (2012). Predicting performance on academic and non-academic tasks: a comparison of adolescents with and without learning disabilities. *Contemp. Educ. Psychol.* 37, 162–169. doi: 10.1016/j.cedpsych.2011.05.001

Klassen, R. M. (2007). Using predictions to learn about the self-efficacy of early adolescents with and without learning disabilities. *Contemp. Educ. Psychol.* 32, 173–187. doi: 10.1016/j.cedpsych.2006.10.001

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *J. Exp. Psychol. Gen.* 126, 349–370. doi: 10.1037/0096-3445.126.4.349

Koriat, A., Ackerman, R., Lockl, K., and Schneider, W. (2009a). The easily learned, easily remembered heuristic in children. *Cognit. Dev.* 24, 169–182. doi: 10.1016/j.cogdev.2009.01.001

Koriat, A., Ackerman, R., Lockl, K., and Schneider, W. (2009b). The memorizing effort heuristic in judgments of learning: a developmental perspective. *J. Exp. Child Psychol.* 102, 265–279. doi: 10.1016/j.jecp.2008.10.005

Koriat, A., and Levy-Sadot, R. (1999). "Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge," in *Dual-Process Theories in Social Psychology*, eds. S. Chaiken and Y. Trope (London: Guilford), 483–502.

Koriat, A., and Shitzer-Reichert, R. (2002). "Metacognitive judgments and their accuracy," in *Metacognition: Process, Function and Use*, eds. P. Chambres, M. Izaute, and P.-J. Marescaux (New York, NY: Springer US), 1–17.

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121

Laakso, M. L. (1995). Mothers' and fathers' communication clarity and teaching strategies with their school-aged children. *J. Appl. Dev. Psychol.* 16, 445–461. doi: 10.1016/0193-3973(95)90029-2

Lockl, K., and Schneider, W. (2007). Knowledge about the mind: links between theory of mind and later metamemory. *Child Dev.* 78, 148–167. doi: 10.1111/j.1467-8624.2007.00990.x

Lucangeli, D., Coi, G., and Bosco, P. (1997). Metacognitive awareness in good and poor math problem solvers. *Learn. Disab. Res. Prac.* 12, 209–212.

Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., et al. (2019). Metacognition and errors: the impact of self-regulatory trainings in children with specific learning disabilities. *ZDM* 51, 577–585. doi: 10.1007/s11858-019-01044-w

Lyons, K. E., and Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Dev.* 82, 1778–1787. doi: 10.1111/j.1467-8624.2011.01649.x

Makowski, D., Ben-Shachar, M. S., Patil, I., and Lüdecke, D. (2020). Methods and algorithms for correlation analysis in R. *J. Open Source Softw.* 5:2306. doi: 10.21105/joss.02306

Moore, J. J., Mullis, R. L., and Mullis, A. K. (1986). Examining metamemory within the context of parent-child interactions. *Psychol. Rep.* 59, 39–47. doi: 10.2466/pr0.1986.59.1.39

Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., et al. (1996). Intelligence: knowns and unknowns. *Am. Psychol.* 51:77. doi: 10.1037/0003-066X.51.2.77

Nelson, T. O., and Narens, L. (1990). "Metamemory: a theoretical framework and new findings," in *Psychology of Learning and Motivation: Advances in Research and Theory*, *Vol. 26,* ed. G. H. Bower (London: Academic Press), 125–173.

Ohtani, K., and Hisasaka, T. (2018). Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacogni. Learn.* 13, 179–212. doi: 10.1007/s11409-018-9183-8

Ornstein, P. A., Haden, C. A., and Elischberger, H. B. (2006). *Children's Memory Development: Remembering the Past and Preparing for the Future. Lifespan Cognition: Mechanisms of Change* (Oxford: Oxford University Press), 143–161.

Ornstein, P. A., Haden, C. A., and San Souci, P. (2008). "The development of skilled remembering in children," in *Earning and Memory: A Comprehensive Reference, Vol. 2*, eds. H. L. Roediger III and J. H. Byrne (Amsterdam: Elsevier), 715–744.

R Core Team (2022). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (Version 4.2.1) [Computer software]./

Reynolds, C. R., and Kamphaus, R. W. (2009). "Development and application of the Reynolds intellectual assessment scales (RIAS)," in *Practitioner's Guide to Assessing Intelligence and Achievement*, eds. J. A. Naglieri and S. Goldstein (New York, NY: John Wiley and Sons Inc), 95–126.

Roderer, T., and Roebers, C. (2013). Children's performance estimation in mathematics and science tests over a school year: a pilot study. *Electr. J. Res. Educ. Psychol.* 11, 5–24.

Roderer, T., and Roebers, C. M. (2014). Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. *Metacognit. Learn.* 9, 1–23. doi: 10.1007/s11409-013-9109-4

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

Roebers, C. M. (2022). "The development of semantic memory: the role of memory strategies and metacognition," in *The Development of Memory in Infancy and Childhood*, eds. M. L. Courage and N. Cowan (London: Routledge).

Roebers, C. M., Krebs, S. S., and Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: their interrelations and their role for test performance. *Learn. Ind. Diff.* 29, 141–149. doi: 10.1016/j.lindif.2012.12.003

Roebers, C. M., Schmid, C., and Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *Br. J. Educ. Psychol.* 79, 749–767. doi: 10.1348/978185409X429842

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., Spinath, F. M., et al. (2015). Intelligence and school grades: a meta-analysis. *Intelligence* 53, 118–137. doi: 10.1016/j.intell.2015.09.002

Sarac, S., Önder, A., and Karakelle, S. (2014). The relations among general intelligence, metacognition and text learning performance. *Egitim ve Bilim* 39:173.

Schneider, W. (1993). Domain-specific knowledge and memory performance in children. *Educ. Psychol. Rev.* 5, 257–273. doi: 10.1007/BF01323047

Schneider, W. (1998). Performance prediction in young children: effects of skill, metacognition and wishful thinking. *Dev. Sci.* 1, 291–297. doi: 10.1111/1467-7687.00044

Schneider, W. (2015). *Memory Development From Early Childhood Through Emerging Adulthood, 1st Edn*. Cham: Springer International Publishing.

Schneider, W., Körkel, J., and Weinert, F. (1987). The effects of intelligence, self-concept, and attributional style on metamemory and memory behaviour. *Int. J. Behav. Dev.* 10, 281–299. doi: 10.1177/016502548701000302

Schneider, W., Körkel, J., and Weinert, F. E. (1989). Domain-specific knowledge and memory performance: a comparison of high- and low-aptitude children. *J. Educ. Psychol.* 81, 306–312. doi: 10.1037/0022-0663.81.3.306

Schneider, W., and Löffler, E. (2016). "The development of metacognitive knowledge in children and adolescents," in *The Oxford Handbook of Metamemory*, eds. J. Dunlosky and Sarah (Uma) K. Tauber (Oxford: Oxford University Press), 491–518.

Schneider, W., and Pressley, M. (1997). *Memory Development Between Two and Twenty, 2nd Edn*. Psychology Press. doi: 10.4324/9780203774496

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognit. Learn.* 4, 33–45. doi: 10.1007/s11409-008-9031-3

Serra, M. J., and DeMarree, K. G. (2016). Unskilled and unaware in the classroom: college students' desired grades predict their biased grade predictions. *Memory Cognit.* 44, 1127–1137. doi: 10.3758/s13421-016-0624-9

Shore, B. M., and Dover, A. C. (1987). Metacognition, intelligence and giftedness. *Gifted Child Q.* 31, 37–39. doi: 10.1177/001698628703100108

Simons, D. J. (2013). Unskilled and optimistic: overconfident predictions despite calibrated knowledge of relative skill. *Psychon. Bullet. Rev.* 20, 601–607. doi: 10.3758/s13423-013-0379-2

Snyder, K. E., Nietfeld, J. L., and Linnenbrink-Garcia, L. (2011). Giftedness and metacognition: A short-term longitudinal investigation of metacognitive monitoring in the classroom. *Gifted Child Q.* 55, 181–193. doi: 10.1177/0016986211412769

Sternberg, R. J. (1980). Sketch of a componential subtheory of human intelligence. *Behav. Brain Sci.* 3, 573–584. doi: 10.1017/S0140525X00006932

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge: Cambridge University Press.

Sternberg, R. J. (1988). *The Triarchic Mind: A New Theory of Human Intelligence*. London: Viking Penguin.

Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemp. Educ. Psychol.* 24, 359–375. doi: 10.1006/ceps.1998.0998

Sternberg, R. J. (2019). A theory of adaptive intelligence and its relation to general intelligence. *J. Int.* 7:23. doi: 10.3390/jintelligence7040023

Straka, O., Portešová, Š., Halámková, D., and Jaburek, M. (2021). Metacognitive monitoring and metacognitive strategies of gifted and average children on dealing with deductive reasoning task. *J. Eye Movem. Res.* 14:4. doi: 10.16910/jemr.14.4.1

Swanson, H. L. (1992). The relationship between metacognition and problem solving in gifted children. *Roeper Rev.* 15:43. doi: 10.1080/02783199209553457

Tibken, C., Richter, T., von der Linden, N., Schmiedeler, S., and Schneider, W. (2022). The role of metacognitive competences in the development of school achievement among gifted adolescents. *Child Dev.* 93, 117–133. doi: 10.1111/cdev.13640

van der Stel, M., and Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learn. Ind. Diff.* 18, 128–134. doi: 10.1016/j.lindif.2007.08.003

van der Stel, M., and Veenman, M. V. J. (2010). Development of metacognitive skillfulness: a longitudinal study. *Learn. Ind. Diff.* 20, 220–224. doi: 10.1016/j.lindif.2009.11.005

van der Stel, M., and Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: a longitudinal study from a developmental perspective. *Eur. J. Psychol. Educ.* 29, 117–137. doi: 10.1007/s10212-013-0190-5

van Loon, M., Destan, N., Spiess, M. A., de Bruin, D, and Roebers, A. (2017). Developmental progression in performance evaluations: Effects of children's cue-utilization and self-protection. *Learn. Instr.* 51, 47–60. doi: 10.1016/j.learninstruc.2016.11.011

van Loon, M. H., and Roebers, C. M. (2020). Using feedback to improve monitoring judgment accuracy in kindergarten children. *Early Childhood Res. Q.* 53, 301–313. doi: 10.1016/j.ecresq.2020.05.007

van Loon, M. H., and Roebers, C. M. (2021). "Using feedback to support children when monitoring and controlling their learning," in *Trends and Prospects in Metacognition Research Across the Life Span*, eds. D. Moraitou and P. Metallidou (Cham: Springer International Publishing), 161–184.

Veenman, M., and Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. *Eur. J. Psychol. Educ.* 14, 509–523. doi: 10.1007/BF03172976

Veenman, M. V. J., and Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. *Learn. Instr.* 14, 621–640. doi: 10.1016/j.learninstruc.2004.09.004

Veenman, M. V. J., and Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learn. Ind. Diff.* 15, 159–176. doi: 10.1016/j.lindif.2004.12.001

Zimmerman, B. J. (1995). Self-regulation involves more than metacognition: a social cognitive perspective. *Educ. Psychol.* 30, 217–221. doi: 10.1207/s15326985ep3004_8

frontiers | Frontiers in Developmental Psychology

TYPE Conceptual Analysis
PUBLISHED 15 August 2024
DOI 10.3389/fdpys.2024.1449705

OPEN ACCESS

EDITED BY
Stephanie M. Carlson,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Michael Tomasello,
Duke University, United States
Philip D. Zelazo,
University of Minnesota Twin Cities,
United States

*CORRESPONDENCE
Jedediah W. P. Allen
jallen@bilkent.edu.tr

RECEIVED 15 June 2024
ACCEPTED 29 July 2024
PUBLISHED 15 August 2024

CITATION
Allen JWP, Mirski R and Bickhard MH (2024) Beyond the mirror: an action-based model of knowing through reflection.
Front. Dev. Psychol. 2:1449705.
doi: 10.3389/fdpys.2024.1449705

Beyond the mirror: an action-based model of knowing through reflection

Jedediah W. P. Allen[1]*, Robert Mirski[2] and Mark H. Bickhard[3]

[1]Bil-Ge Lab, Department of Psychology, Bilkent University, Ankara, Türkiye, [2]Department of Cognitive Science, Nicolaus Copernicus University in Toruń, Toruń, Poland, [3]Department of Philosophy and Psychology, Lehigh University, Bethlehem, PA, United States

Epistemic reflection involves the creation of qualitatively new knowledge. Different models have been proposed to account for new knowing through reflection that have typically been grounded in an information-processing framework. However, there are in-principle arguments that information-processing approaches preclude the emergence of new representation altogether. Accordingly, any information-processing account of knowing through reflection is plagued by emergence issues. After discussing some of these emergence issues for four prominent models in the cognitive science literature, an alternative action-based model of representation and reflection is presented called interactivism. Interactivism's model of representation, as grounded in action anticipations, serves as the foundational emergence needed to account for subsequent knowing through reflection. After introducing the interactivist models of representation and reflection through knowing levels, some of the implications for consciousness, enculturation, language, and developmental methodology are discussed.

KEYWORDS

epistemic reflection, action-based approach, interactivism, emergent representation, interactive knowing, enculturation, language as interaction

1 Introduction

Reflection is often characterized as serving one of two functions: the creation of qualitatively new knowledge, or qualitatively new capabilities involving self-/emotion-regulation through some sort of distancing process. While most researchers incorporate some role for language in the reflection process, a basic division can also be drawn between approaches that emphasize the developmental origins of reflection as a cognitive activity vs. those who argue that language is the original locus through which reflection takes place. In the current paper, we will explore efforts to explain the development of reflection as a cognitive activity for emergent knowing, but we will also indicate the subsequent role that language plays in this process. The paper will proceed by briefly discussing several different models that are all united in trying to explain how reflection enables the creation of qualitatively new knowing: these include Mandler's (2004) perceptual analysis, Karmiloff-Smith's (1992) representational redescription, Perner's (1991) meta-representation, and finally, Zelazo's (2004) levels of consciousness model. This discussion will ultimately reject the adequacy of these models due to their information-processing assumptions and inability to account for representational emergence. The alternative interactivist model is an action-based framework that contrasts with an information-processing ontology (Bickhard and Terveen, 1995; Bickhard, 2024)[1]. This model will be introduced and discussed in the context of interactive vs. reflective knowing, primary consciousness vs. reflective consciousness, and internalization vs. enculturation as the process of socialization. Finally, the role of language for reflection will be addressed in terms of its differential relevance for both pre-reflective and reflective development with some implications for developmental methodology.

2 Qualitatively new knowing: existing models (Mandler, Karmiloff-Smith, Perner, and Zelazo)

At the core of the developmental sciences are issues of origins in general and the issue of representational/knowing origins in particular. Nativist positions generally side-step the issue of origins by assuming that essential knowledge structures are provided to the species through some unspecified evolutionary process. The theoretical motivation for nativism comes from learnability arguments that innateness is *necessary* for learning to get started (Chomsky, 1959; Fodor, 1975). Contemporary empiricist positions actually agree with the need for some innateness but disagree about the amount and type (e.g., feature representations or full concepts, a few concepts or many concepts, for a full discussion see Allen and Bickhard, 2013). However, empiricist positions are more developmental and pursue more powerful possibilities for learning such that they assume that qualitatively new representations/knowing are produced during ontogeny. That said, both nativist and empiricist positions tend to assume a background information-processing framework with implications for the nature of representation that make qualitatively new representations (i.e., emergence) impossible. For information-processing approaches, the nature of representation is in terms of some sort of encoding relationship with the world (Bickhard and Terveen, 1995; Bickhard, 2009a).

Encodingism is the assumption that foundational representations are encodings. Encodings are constituted by a correspondence relationship with what they represent, and these correspondences are often assumed to be causal, nomological, or informational. Regardless of the specific nature of the relationship, encodings are representational stand-ins such that they must derive their content from some other source of representation. For example, the rings in a tree encode its age in years. This is a factual/informational[2] relationship that is only representational if there is an epistemic agent who already knows about rings in trees, annual growth, and the relationship between the two. Without an interpreting agent, there is no content for the encoding and its relationship to what it represents. Thus, as an account of foundational representations, encoding approaches are incoherent. What's needed is an account of emergent representation in which representation is emergent within a foundation that is not already representational and only action-based approaches have offered to provide such account (Allen and Bickhard, 2013). While Piaget is the best known action-based approach, it is the interactivist model that will be presented in Section 2. Before that discussion, we present four empiricist models that all assume that learning and development involve qualitative changes in the nature of the representations that can be constructed through reflection; however, all four models are also committed to an information-processing framework that precludes the possibility of emergent representation.

2.1 Perceptual analysis (Mandler)

Jean Mandler's model has the laudable goal of trying to account for the foundations of meaning itself (Mandler, 2000, 2004). We refer to this issue as the *foundational emergence* problem. One avenue for resolving this issue is to take a nativist stance (i.e., foundational meanings/representations are innate). However, such an approach does not solve the issue so much as it ignores it. This leaves empiricism as the alternative—and in the current state of the field, some form of information-processing empiricism. While modern empiricist approaches to development also start with some amount of innate conceptual/representational base (Gopnik, 2003), Mandler's model attempts to address the developmental emergence of such a base. This model commits to a sharp distinction between perception and (meaningful) conception, in which the latter is supposed to derive from the former through an abstraction process—*perceptual analysis*. Perceptual analysis is a volitional process involving attention to certain aspects of the perceptual data stream in order to abstract a simplified rendering of the input. This process also involves "recoding" the format of the abstracted content into "explicit" form which enables "...one to describe, recall, or think about something new, not just recognize it" (Mandler, 2004, p. 18). Mandler grounds much of her theorizing in a set of empirical findings in which very young infants seem to have abstract (i.e., conceptual) categorization abilities that include the functions of objects (in addition to their perceptual features).

While these empirical findings should be a constraint on any adequate theory of foundational meaning, Mandler's account has a number of theoretical problems that appear to be unsolvable (see Müller and Overton, 1998 for a full treatment of the model and its limitations). Two of the most relevant of these problems concern the abstraction process. Abstraction is supposed to produce meaningful/conceptual categories. But how can the correct features be abstracted without already knowing what the

1 Tomasello (2024) has recently proposed an agency-based model of reflection that is also more pragmatically oriented in its background assumptions. Interestingly, it also shares the basic distinction between what we call interactive vs. reflective knowing in terms of executive vs. metacognitive regulation.

2 We contrast the meaning of "information" in information theory with the semantic meaning of it—as mental representation with truth value, intentionality, and content. Informational relationships in the former sense are integrated into the interactivist model in terms of epistemic contact (i.e., differentiation/detection) not epistemic content (i.e., representation, Bickhard, 2009a). Detected correlations need to be accompanied by anticipations, otherwise they mean nothing to the organism. In this paper, when we refer to "information," it is in the information theoretic sense, unless stated otherwise.

category is supposed to be? Mandler's empirical works suggests that perceptual features like object salience or similarity are not sufficient—correct abstraction requires already knowing which features should be abstracted. For the second issue, abstracting relevant features means taking a subset of the perceptual data, but a subset of the input stream does not give new data. A subset may give new access to volitional processes, but there is no new data *per se*. Further, both problems assume that feature representations are available as distinct "pieces" of an overall representation (i.e., a feature nativism). However, perceptual analysis was presented as an alternative that could avoid the pitfalls of nativism. This means that in addition to the problems with abstraction as an account of new content, the need for a feature innateness/foundationalism means that such a model is open to the same emergence limitation as standard nativist accounts—that representations *must* already be assumed in order to explain the origins of new representations, whether in development or in evolution (Allen and Bickhard, 2013). Lastly, a third issue is that changing formats does not change the content of the data, nor does it make the data more/less explicit (more on this in the discussion of Karmiloff-Smith). The interactivist account of representation in the current perspective will provide a model for foundational emergence that does not have the above problems.

2.2 Representational redescription (Karmiloff-Smith)

Karmiloff-Smith's model builds on the work by Mandler in terms of foundational meanings to explain a process for the subsequent development of new representations (Karmiloff-Smith, 1992). Karmiloff-Smith accepted much of the cutting-edge empirical conclusions coming out of the nativist research program in the 80 and 90's while also trying to transcend the contrast between domain-*specific* and domain-*general* learning processes (Karmiloff-Smith, 2018). Her domain-*relevant* approach attempted to explain how innate biases could result in a cascade of emergent developmental outcomes. Accordingly, Karmiloff-Smith provides an account for the emergence of new forms of representation that go beyond the foundational emergence of conceptual from perceptual. We refer to this issue as the *subsequent emergence* problem. Similar to Mandler, Karmiloff-Smith also appeals to changes in format to account for new content. Different from Mandler, Karmiloff-Smith adopts a more robust constructivism in that there are *internal sources* of change such that cognitive processes derive new content from the overall organization of old content (e.g., information embedded in problem-solving procedures). This is a process of *Representational Redescription* (RR) in which the implicit content of prior knowing is made explicit and constitutes new representational content.

The main function of RR is to facilitate flexibility, and thus, control of behavior relative to new purposes. This is a consequence of the increased operations that can *access* the more explicit representations that eventuate in conscious access, linguistic access, and theory construction processes (cross-domain integration). The RR process suggests four types of representations: one implicit (I) and three explicit (E1—unconscious, E2—conscious access, and E3—conscious and linguistic access). Implicit representations are procedures (or sensorimotor encodings) that have a sequential organization that is encapsulated and inflexible. These representations are used in response to external stimuli (i.e., they are not internally driven). The RR process involves *reformatting* the sensorimotor encodings through abstraction so that more operations can access their contents. It is an abstraction in the sense of extracting the sensorimotor information while losing the perceptual details. Once the newly formed E1 representations are available, they can be used in more flexible ways (e.g., understanding the analogy between a Zebra and a crosswalk sign). This means that the creation of E1 representations precedes any sort of reflection about potential relations embedded in the sensorimotor procedures. At E2, representations are in a format accessible to consciousness but not verbal report. Finally, E3 representations are needed to use language because they involve a "cross-system code." This makes language a tool that can be used after two or three iterations of the RR process have abstracted them into the correct format.

Karmiloff-Smith's theorizing involves developmental elaboration beyond the model of foundational meaning provided by Mandler's account. This elaboration is both at the broader level of development and at the specific level of RR. Similar to the perceptual analysis account, the issues for abstraction as a source of new content apply here as well. However, the focus of RR is on how changes in format affect explicitness, which enables new forms of knowing. There are two issues here: (1) does format affect the explicitness of the representational content? (2) does format enable new forms of knowing? For (1), as Fodor (1998) indicated, all encodings are explicit about something and so the idea of implicit representation cannot be with respect to the content of the representation itself. For example, changing the format of the letter "S" to "..." does not alter the explicitness of either representation[3]. What changes from "S" to "..." are the sorts of things one can do with the new encoding (Bickhard and Terveen, 1995). The dots can be sent over telegraph lines while the letter cannot. Accordingly, for the RR model, the implicitness is in terms of how the overall systems can (or cannot) make use of the (explicit) content of the "implicit" representations. This means that the changes in format from sensorimotor encodings to E1 do not involve new content for the E1 representation (or E1 to E2 or E3). However, if the changes in format through the RR process do not involve the emergence of new content, then the increasing access does not involve new forms of knowing. That is, issue (2) is also answered in the negative.

At the broader level of development, Karmiloff-Smith has captured several important features. Her theorizing suggests that the internal dynamics of cognition are a source for change with recurrent phases of development that oscillate between behavioral mastery and cognitive reorganization. This makes it important to consider how the same behavioral performance at two different ages may in fact be a consequence of different cognitive processes. This means that U-shaped development is not noise to be averaged away but an important constraint on developmental

3 Encoding content is borrowed from or defined in terms of other already available content—e.g., "..." from "S." It cannot create new content. That is the central problem with "information" processing models.

explanations (see also Gershkoff-Stowe and Thelen, 2004). The theory also makes multiple distinctions about different forms of knowing. Representational multiplicity is important because there is a strong tendency in development psychology to ignore the possibility that children at different ages have qualitatively different ways of knowing (adultocentricism) and to thus not control for such possibilities in "empirical" research (Allen and Bickhard, 2013). Finally, Karmiloff-Smith's theory attempts to reconcile the emergent constructivism of Piaget's theory with the representational innateness of nativist research programs. In this respect, it shares an overall goal and structure with Carey's (2009) more recent model of how to reconcile innateness with qualitative development. However, in both cases, the requirement of an innate representational foundation for learning and development involves a notion of representation that precludes the possibility of genuinely new representational content (i.e., encodingism). Further, an adequate account of new content through learning obviates the necessity for an innate foundation. Thus, either qualitative emergence in development is impossible, or, there is no necessity for (homuncular) innateness (Allen and Bickhard, 2011).

2.3 Meta-representation (Perner)

Perner (1991) has developed a model of meta-representational development to account for changes in false-belief understanding and a number of other qualitative changes around age 4. This model suggests that meta-representational development involves new knowledge in that children become able to represent representational relationships, and this has cascading developmental consequences. In particular, children with meta-representational abilities are able to understand misrepresentation (of people with false-beliefs or objects like signs and photographs), the representational nature of language (i.e., that words are not properties of what they represent), and the distinction between sense and reference as manifest in understanding that Clark Kent and Superman are the same person (Perner et al., 2002; Iao et al., 2011). Although this model has some basic convergence with the interactivist model to be presented below, it has been discussed in detail from the interactivist perspective previously (Bickhard, 1992). The most relevant conclusions from that discussion are that no account of foundational emergence will be possible given the (encoding) assumptions about representation and that reflection seems to already be needed for even the basic representations of infants (not just the meta-representing of preschoolers).

2.4 Levels of consciousness (Zelazo)

A more recent model for how the development of reflection enables new forms of knowing, representing, and acting comes from Zelazo (1999, 2004, 2015). This model is similar to Karmiloff-Smith's in that it is: focused on levels of subsequent emergence, developmentally rich, conceptually coherent, and grounded in both behavioral and neural data. It is also unique in terms of the focus on consciousness as being relevant for modeling changes in knowing. Nonetheless, as with Mandler, Karmiloff-Smith, and Perner, the underlying information-processing empiricism creates limitations for how well the model can account for epistemic reflection (i.e., the emergence of new knowing through reflection).

Much of the recent empirical motivation for the "emergence" process in this model comes from brain studies in which there seems to be "iterative reprocessing" of information within and between areas of the brain (Zelazo, 2015). However, if the technical sense of information relevant for brain studies cannot account for the semantic sense of information relevant for cognition, then the implications of these data are unclear. Further, the myriad reciprocal projections of the brain can also be characterized as supporting oscillatory processes, rather than semantic "re-entrant" processes, and oscillatory processes have been argued as a neural foundation for the anticipatory processes that constitute the core of an action-based "semantics" (Bickhard, 2015, 2024).

Regardless of the status of re-entrant processing, the original reflection model is mostly explicated in terms of theoretical considerations, and that will be the focus of our analysis. The Levels of Consciousness (LoC) model is an account of changes in the reflective capabilities of children (Zelazo, 1999, 2004). New reflective capabilities enable more complex representing through the creation of new representations (i.e., of relations between lower-level representations and of hierarchical control structures). Zelazo highlights intentionality as the key feature of any form of consciousness. This is intentionality in Brentano's sense of being *about* something and for motivating action [1973 as cited by Zelazo (2004)], but there is no account of the emergence of intentionality itself. Instead, intentional representations/descriptions of objects in semantic Long-Term Memory (LTM) are triggered by actual objects from the environment. These representations then trigger the most salient action pattern that has been learned through association (e.g., a rattle might trigger the action pattern of sucking at one age and shaking at another). This form of representing is supposed to constitute basic consciousness (i.e., minimal Consciousness or minC).

Although the mechanism for ascent in the LoC model is the same (i.e., recursion), the most qualitative change in representing takes place in the transition from level 1 to level 2 at the end of the 1st year of development. This change involves a constitutive role for language in terms of labeling. Labels are supposed to provide an enduring trace to segments of the perceptual input stream that constitutes basic conscious experience (i.e., minC). These traces are representations proper in that they can be "decoupled" from the ephemeral flow of experience and manipulated in working memory as part of top-down control (e.g., representation of an occluded object that can serve as a goal). However, for labeling to serve this decoupling function requires level 2 consciousness to create identity relations between two moments in the input stream from first-level consciousness. Thus, the construction of these identity relations require reflection through *recursion*. Recursion is understood in the sense of a computer program that calls on itself as a parameter [e.g., Factorial (n) = n * Factorial (n-1)]. More recent discussion about reflection is in terms of *iterative reprocessing* where information output is fed back into the system to be combined and integrated with existing representations to create a new interpretation of the situation (Zelazo, 2015).

Our concern with this model can be divided into two issues: (1) how do semantic representations/descriptions work such that labels liberate the infant from the flow of first-level consciousness? (2) how does recursion enable new levels of consciousness? We suggest that the answers given by the model presuppose a rich innate representational base as well as the reflection capability that was meant to be explained. First, labels (from semantic LTM) are supposed to be attached to identity relations that connect the contents (also from semantic LTM) from two moments in the input stream. However, this process seems to be creating a linguistic encoding of the content of the identity relation with the label—instead of "... " there is "dog" whose content is dog, and the content of dog came from semantic LTM. This means that all of the content for the encoding relationship is coming from semantic LTM with no account of its origins or how the semantic descriptions are being interpreted in the first place. Further, if reflection is needed to make the new linguistic encodings (in addition to it being needed to create the identity relations and perhaps for interpreting the descriptions in the first place), then this leaves recursion to account for all of the functionality of reflection[4].

Second, if reflection is required to both interpret semantic descriptions and attach them to labels (recC) or to objects (minC), then reflection is present from the very beginning, and this would make it homuncular (Bickhard and Terveen, 1995). If reflection were already present, then perhaps recursive/re-entrant processing could construct something "new." That is, if semantic information contents are re-entered into a consciousness that is already reflective, then a homunculus can survey all those contents (with all of the consequences at each level that the model posits). However, this would not create new content, instead, different levels of detail are being selected with different levels of reflection. This makes the development of "new" representation a matter of *selection* amongst existing content rather than the *emergence* of new content[5]. If our analysis is correct, the LoC model is not able to fulfill its epistemic function. This is because recursion does not yield a higher level of consciousness *per se*, but yields a hierarchy of levels of "content" within reflection. This may be the best option available within an information-processing framework but that is not the only alternative for modeling development.

As an account of emergent forms of knowing through reflection, the LoC model appears problematic; however, the descriptions, properties, and functions attributed to the different levels of consciousness may still capture something important about development. That is, the LoC model may be adequate for certain aspects of the developmental changes in consciousness even if it is not adequate as a model of epistemic reflection. Further, a core feature of all of the models reviewed above is the idea that lower-level representations serve as the foundation for new representations at higher levels through reflection. The current interactivist model of reflection shares this basic idea but the crucial difference concerns its action-based foundation (Allen and Bickhard, 2013). In contrast, all of the above models are developed within an information-processing empiricist framework. This framework is incapable of accounting for emergent representation and precludes the possibility of an emergent constructivism (Bickhard and Terveen, 1995; Allen and Bickhard, 2022). Without an emergent constructivism, learning and development cannot result in new knowing, and any model of reflection will ultimately fail as an explanation for such an outcome.

3 Interactive knowing and reflection

Interactivism is an action-based model of cognition and persons in which knowing is doing, and competent knowing means successful interaction (Bickhard, 2009b, 2024). Perhaps the best known action-based approach in developmental psychology was Piaget's sensorimotor theory (Piaget, 1954). However, misinterpretations and misguided methodology side-lined Piagetian theory in general and its action-orientation along with it (Smith, 1993; Allen and Bickhard, 2013). Rejections of computationalism for some strands of cognitive science have seen a move toward embodiment and most recently an explicitly pragmatist turn (Engel et al., 2016). However, interactivism differs from these embodied/pragmatist approaches, including Piaget's, in terms of the underlying models of representation (i.e., interactive knowing) and reflection (Bickhard, 1978; Campbell and Bickhard, 1986; Bickhard and Terveen, 1995).

For interactivism, representation is constituted in terms of anticipating potential interactions with the world. The anticipations are discovered to be true or false once enacted (i.e., they have *truth-value*) and they involve presupposition that the world will cooperate (i.e., they are *about* the world). For example, to anticipate that a coffee cup can be picked up presupposes that the cup is not broken. Being unbroken is usually presupposed by our interactions with cups, but it is not indicated within the anticipation and therefore it is not represented explicitly. However, if that presupposition is relevant (i.e., the cup is in fact broken), then the interaction will fail (or at least break down) and thus, presuppositions can be functionally important for the interaction without being explicitly represented. In this model, presupposition provides the implicit content that is about the world (note that presupposition is an aboutness that is not homuncular) while the explicit content is constituted in the internal anticipations or indications of potential interactions *per se* [e.g., a "pointer" indication of a subsystem that could engage in the anticipated-as-possible interaction(s)][6].

Let us stress the point that interactivist mental content is constituted by what is *implicitly presupposed* by the anticipatory dynamics, which contrasts with the criticized model of

4 There are also potential empirical reasons for caution about the role of labeling in this model as it is not clear that infants use labels to succeed on tasks like A-not-B at the end of the 1st year, or what kind of labels those would be Müller and Kerns (2015); also, non-human animals seem to have rather sophisticated forms of top-down control although they do not use language (Penn et al., 2008).

5 Further, how could reflection explain the origins for how we represent non-observables like mental-states. No amount of reprocessing at any level of resolution is going to enable the extraction of something that is not already present in the input stream of conscious experience.

6 The possibility of pointers show that indications pose no particular problem, although that is not how the CNS actually does it. See Bickhard (2015, 2024) for how the indicating/anticipation function is served in the CNS.

encodingism. As we have discussed earlier, encodingism views mental content as constituted by information in information theoretic sense, i.e., by correlation between the agent's internal states and some feature of the world (see text footnote 2). In the interactivist critique of encodingism, the issue is not whether or not information plays a role in cognition. Information understood as correlation is a property of the world and it naturally matters to agents. Rather, the problem is the ontological assumption that information *constitutes mental content*. One of the critical points we made earlier is especially relevant here—correlation needs to be known in order to be representationally utilized by the agent and so it cannot be what constitutes that knowledge itself. In contrast, content as implicit presupposition makes no such problematic assumption; as a natural consequence of learning to effectively interact with the world, the organism's anticipatory knowledge comes to "agree" with how the world is, to implicitly presuppose how it is.

For a developmental example, consider object representation. Object representation for the 2-year-old is constituted in the web of anticipated possibilities for interaction remaining constant with respect to other sorts of changes (e.g., occlusion or displacement). While the basic properties of representing are present in the anticipations (i.e., truth-value and aboutness), the permanence is a property of the overall organization of the web of anticipated possibilities. Such permanence is functional for the 2-year-old in that they can act in accordance with the presupposition that the object has a continued existence, but the permanence *per se*, the presupposition, is not itself represented by the toddler. This is because the toddler cannot directly interact with the permanence of the object and therefore cannot have anticipations directly about it. Instead, reflection will be the process that enables the implicit content/presupposition to become explicit (i.e., reflection is required to know about permanence *per se*).

Interactive knowing is constituted in the organism/system interacting with the environment (i.e., first level knowing). Reflection requires a second interactive system that can interact with the first system/organism (i.e., second level knowing). In humans, this means that the development of reflection involves an architectural change to the CNS—maturational development of the brain—to enable interaction between regions (i.e., second level knowing) in a fashion similar to how the CNS of the toddler interacts with the world (first level knowing, Bickhard, 2015, 2024)[7]. With reflection comes the possibility of knowing about the system (its internal functional organization) that interactively knows the world. That is, the properties and relations implicit in first level knowing (i.e., the presuppositions of interactive knowing) become knowable through reflection (i.e., second level knowing). While there are no a priori constraints on the age of development for reflection, there are ample empirical reasons to think that it is around age 3.5–4 (Allen and Bickhard, 2018). This is the age at which there seems to be developmental transitions in abilities within and across domains. There is also evidence that uniquely supports the interactivist model of reflection over other domain-general explanations for such changes at this age (see discussion of Allen et al., 2021 below).

To further illustrate the contrast between interactive and reflective knowing, let us consider the development of an empirical test specific to the interactivist model of reflection. Any such test is difficult for three general reasons: first, given that any task can, in principle, be interactively learned through non-reflective knowing, it is important that the task have sufficient novelty. Second, if all the different interactions of a toddler[8] are already consistent with the implicitly presupposed properties like permanence, then what difference does it make to have explicit knowledge of those presupposed properties? Third, as adults, our reflectively conscious experience of objects can always be explicit, and so it can seem as if infant interactions that are consistent with our explicit representations are also explicit for the infant (i.e., adultocentrism).

To address these issues, a test of reflection was developed that turned on being able to explicitly represent the relationship between two objects—a mutual support relationship (Allen and Bickhard, 2018). Similar to the permanence of objects, relations amongst objects cannot be directly interacted with and therefore cannot be explicitly represented by toddlers. Without representing relations *per se*, children should not be able to anticipate their consequences in a sufficiently novel situation. Accordingly, the Leaning Blocks (LB) task involves asking children what will happen to a block being held at a 45° angle when released (i.e., "fall" or "stay up"). After asking the same question for a second block, the test question involves holding the two blocks such that they are *leaning* against each other. Children are again asked what will happen upon release. Three-year-old's fail the question while 4- and 5-year-olds are basically at ceiling. These findings suggest that the older children can explicitly represent the mutual support relationship between the two blocks, and in so doing, correctly determine the consequence given the relative novelty of the situation. A follow-up study, that included a second reflection task (i.e., Candy Monster, CM) and three EF measures, suggests that the results from LB are not due to changes in executive functions. Specifically, inhibition, working memory, and cognitive flexibility interpretations were tested against the reflection interpretation and the results favored the later (Allen et al., 2021). Importantly though, reflection is an enabling constraint which means that learning relevant to any specific task must still take place before the "reflective ability" can be measured. The design intention of the LB and CM tasks are as relatively "pure" measures of reflection because they do not seem to involve many additional abilities beyond explicit representing *per se*.

7 For example, a maturation of a neural loop from pre-frontal to basal ganglia to thalamus and back to cortex (Bickhard, 2015, 2024), thus enabling interaction with other regions of the CNS.

8 It is not until toddlerhood that children show a coherent set of interactions consistent with the permanence of object. At earlier ages, infants show only a limited set of interactions consistent with permanence (Baillargeon, 2008). For example, small changes in whether a looking measure involves occluding an object vs. covering it, and later, containing it, affects performance such that the same aged infants fail one version while passing the other(s).

3.1 Consciousness and reflection

"Consciousness" is often used in a crucially equivocal manner: (1) as an "awareness" of the potentialities that constitute the world, and (2) as a kind of reflection on those first level processes and organizations. Failing to distinguish these yields aporetic problems in understanding consciousness (Bickhard, 2005). For example, as Dewey pointed out about Russell's "sense data" (Dewey, 1915, 1941; Tiles, 1990), sense data (today's descendent is "qualia") are supposed to *constitute* "consciousness" of the world, but in fact sense data (qualia) are products of *analysis* of (reflection on) primary awareness—they are generated in analysis, not constitutive of what is being analyzed,

In the interactivist model, there is a clear distinction between first level interactions with the environment and anticipations of possible such interactions, and second level interactive reflections on those first level processes and properties (and relations). The model of primary awareness has already been outlined: anticipations of (organizations of) possible interactions and their intrinsically related presuppositions. The model of reflection is that of a second level of such interactive "knowing" that interacts with the first level. The first iteration of such reflection is not possible in all species—it requires the macro-evolution of a special functional organization in the brain, and a developmental maturation of that functional organization in the individual. Further levels can be constructed in a strictly functional manner through language and culture (Bickhard, 2024), which will be discussed briefly in what follows.

4 Internalization vs. enculturation

While psychology today generally accepts that human minds are largely shaped by culture, the actual models of how that happens remain problematic (Turner, 1994, 2018; Christopher and Bickhard, 2007). Culture tends to be framed in terms of a set of beliefs and practices that the child "internalizes" as she undergoes the process of enculturation. The concept of internalization can be traced back to both Piaget (1952), Piaget and Inhelder (2000), and Vygotsky (1978), but its current uses usually draw on the latter. Vygotsky was especially interested in internalization of culture. His idea was that culture is dialectically externalized and internalized by any individual interacting socially. Children, being newcomers to social reality, were said to internalize into their minds the ways of thinking instantiated in social interactions, which made for the central mechanism of enculturation in his theory.

The details of the presumed internalization process remain vague; most fundamentally, the question arises as to what it actually means—how something that is out there in the world can get into the child's mind? And once it gets there, what kind of phenomenon is it? Potential answers to these questions depend on one's wider ontology of the mind. In encodingist models, which still dominate the field, the internalization process has been argued to be a conceptually incoherent proposal (Christopher and Bickhard, 2007). This incoherence is a consequence of the wider problems with encodingism discussed earlier: In order to internalize anything that is outside of the agent—such as a norm or custom—an encodingist agent would have to already know the thing in order to be able to internalize it, which means that internalization cannot be the basic mechanism for how cultural knowledge is formed (cf. the similar critique by Piaget, 1971). The interactivist model of enculturation, in contrast, follows naturally from the principles on which the interactivist ontology is based, and has no need for the concept of internalization.

Enculturation in interactivism follows the same basic principles as development of interactive knowledge of the physical reality—what differs is the object of interaction and resulting anticipatory organization: While knowledge of the physical world is constructed by engaging with and learning, for instance, the interactive stabilities of physical objects, cultural knowledge originates in the child's interaction with cultural or conventional objects of social ontology, such as norms governing dinner or nighttime routines (for the interactivist model of social ontology as convention see Bickhard, 1980; Mirski and Bickhard, 2021). Consequently, the pre-reflective knowledge of a child developing within a culture involves implicit presuppositions about cultural phenomena—it is organized in a way that "honors" cultural phenomena such as values or customs, but the child does not represent them explicitly as such; culture is implicit within the child's anticipatory organization, it is part of how the person views the world and interacts with it. Rather than internalization, the process is that of construction constrained and guided by the socio-cultural milieu.

Implicit presuppositions concerning the socio-cultural world, similarly to those concerning physical reality, can be represented explicitly once reflection is available to the child. For example, at knowing level 1, the child can interactively differentiate him or herself from other agents and the rest of the world, but she will not be able to represent that differentiation explicitly. In other words, the child will have a self, but will not know it. This implicit self will be greatly constrained and guided by culture as it will involve all types of presuppositions about the social world and its norms, such as, for instance, a preference to play with toy cars rather than dolls. Reflection, or level-2 knowing, allows the child to examine the self-embodied in level-1 organization and develop, for instance, meta-strategies for navigating the social world, such as heuristics for successfully creating play situations with toy cars rather than dolls. These reflective representations and strategies will constitute the child's self-representation, or its *identity*—a set of ways of being in the world. However, this self-representation will not be known explicitly, the child will not be able to represent the way it represents him or herself—for that, a third level reflection is needed. The self-representation will have their own set of implicit presuppositions, which again can be only explicitly known by a higher level of knowing; once that is available, the child will be able to, for instance, compare her own identity with alternatives or examine it in terms of values and perhaps reconstruct it to agree with them (Campbell and Bickhard, 1986, p. 118–127). The epistemic climb up the knowing levels need not stop at level 3—every epistemic level involves its own implicit presuppositions, which can be potentially known by a level higher than that. A level 4 examination of one's identity may involve a discovery that one has a tendency to frequently switch between identities, which can then be duly addressed by the agent. Importantly, even though every level leads to the emergence of qualitatively new knowledge, it too involves implicit presuppositions that remain unknown before a higher-level reflection makes them explicitly. While there is not an in-principle limit to how high in the reflective levels the agent can climb, there naturally are various factors that influence the process[9]. Among them, language seems to be a major one, to which we turn below.

5 Does language serve a reflective function?

Interactivism models language as a system for interacting with social situations, or situation conventions, which constitute social reality (Bickhard, 1980). The basic idea is that language is a meta-convention—a convention for interacting with conventions—that allows the agent to coordinate action with its conspecifics. For instance, consider the child's early developing use of the utterance "no!" and how he or she uses it to negotiate or modify social situations—even though at first the child uses it simply to protest the current situation, it is understood by both the child and the caregiver in a similar way and thus succeeds in communicating the desired change to the situation (i.e., that it should stop or change). Importantly, such early uses of language are fully implicit and do not amount to a symbolic understanding of utterances—they are part of knowing level 1[10].

However, pre-reflective mastery of language is limited: language is not in this early form understood symbolically, i.e., as representing some part or aspect of reality, but only as yet another way of interacting with the world. As such, it does have presuppositions about it, and—just like any other knowing in interactivism—those presuppositions are not explicitly represented. Once reflection is available, it becomes possible for the child to start constructing explicit representations of what utterances actually mean and how they fit into the social world—i.e., what the presuppositions are of and how they modify situation conventions. This process takes time and effort, but by age 4, when reflection seems to emerge, the child has already constructed considerable knowledge of the linguistic realm of interaction, whose implicit presuppositions can be examined and represented. That is, content is there, but it is not as-of-yet represented explicitly.

More mature linguistic interaction, such as having a conversation about things that are not there, requires its participants to exercise reflection and to understand the meaning of utterances symbolically. That is, a toddler can have a conversation of that kind—e.g., about clouds and pets—but will be incapable of representing and considering in the conversation the abstract properties of those objects, such as the "hidden" causes of their behavior. In other words, language (i.e., situation conventions involving linguistic interaction) constitutes a realm of interaction that can be fully successfully navigated only with proper reflective understanding. As such, it imposes a selective pressure on the child's budding reflection—language-based interaction tests out the child's attempted constructions of reflective understanding and selects only those that afford successful anticipation of the interactive flow. Naturally, the child is aided in this developmental task by caregivers who engage in all kinds of functional scaffolding to lower the selective pressures inherent in language (Bickhard, 1992): Repeating things, narrating while demonstrating and so on. Language, then, is a realm of interaction that serves both as a motivator for reflective construction and as a testing ground for it. Without an opportunity to interact linguistically, reflective understanding is critically impaired, as the tragic cases of language deprived children attest (Fromkin et al., 1974).

Further, as success in linguistic interaction drives the child's reflective construction (once enabled by CNS developments), by the same token it imparts some organization onto the child's resulting reflective knowledge. Not only due to its formal properties such as syntax or morphology, but also in terms of associations, symbolic tropes, or generally speaking—ways of thinking—that abide in a given language or culture more broadly. Indeed, it is hard to imagine how an organism would show culturally-constituted reflection without a language scaffolding the process, and thus it can be difficult to disentangle properties of our reflective thinking that stem from its linguistic formatting and those that characterize reflection as such. Perhaps due to this entanglement, many scholars in history have declared thought to have a language-like structure (e.g., Fodor, 1975), which from the interactivist perspective amounts to misattributing properties of language to the nature of reflective thought as such.

It needs to be stressed that cultures and languages differ, and that they do so to some extent in terms of what kind of reflective abstraction is needed to enter them; this can be both in terms of types of content—like mental state concepts vs. behavioral concepts—or ways of thinking about some content—like theory vs. narrative. These differences in interactive realms likely lead to children from those cultures to exercise their reflection in accordance with them and thus do better on tests that presuppose competence in those terms. For instance, the explicit change of location False Belief Task (FBT) is passed at different ages depending on culture—in the West it is around age 4, but in Japan at 6+ (Naito, 2014). Whereas, multiple factors can be evoked to explain this difference, the specificity of the folk conceptualizations about the social world that dominates in the two cultures might be a significant one. As Naito argues, rather than a theory of mind, Japanese folk theory is that of relations between people. To be sure, both of these conceptualizations are true in the sense that they abstract real aspects of the social world—individuality and epistemic separateness in the former case, and the interconnectedness in the latter—but the difference in emphasis seems to lead to differing developmental trajectories in what is reflectively represented, which seems to be reflected in children's performance on socio-cognitive tests. The FBT arguably requires the child to have a clear reflective understanding of how perceptual contact of an individual mind relates to their knowledge of the world—the kind of reflective understanding that American children steeped in Western folk psychology would develop early and Japanese children would find rather foreign. However, things are different with other socio-cognitive tests, such as ones that involve aspects more aligned with the Japanese theory of relations. For instance, in one such task the object about which the protagonist of the FBT forms a false belief is changed from a

9 Empirically speaking, there does not seem to be evidence for development beyond level 4. However, the issue has not been directly investigated.

10 The term "symbolic" is usually understood in an encodingist way; here, instead, we mean it in the interactivist sense, as explicit representation of implicit presuppositions about what words refer to.

physical object (e.g., a toy) to a person who has promised to stay in one place rather than the other, but moves unbeknownst to the protagonist (Symons et al., 1997; Naito, 2014, p. 390). Japanese children seem to do better than their Western counterparts on that test, and when they are asked to motivate their answers, they tend to cite social obligations such as "he promised he'd be there" rather than individual epistemic states of the protagonist such as their mistaken belief.

Finally, once understood symbolically, language greatly facilitates reflective abstraction; that is, symbolic and systematic language provides a format that externalizes thought, which facilitates the climb up the knowing levels. The fundamental principle of interactivist knowledge formation is that only that which can be interacted with can be represented. For levels 1 and 2, the epistemic access is direct—level 1 interacts with the structure of reality, both physical and social, via the senses; and level 2 interacts with the organization of level 1 knowledge, via the physiological links in the CNS. This leaves the question of how reflection can climb beyond these two levels of representing—how to represent the implicit presuppositions of level 2 knowledge and higher?

Action involving level 2 reflection will leave a mark on the organization of level 1, both indirectly by influencing how the agent acts in the world and directly via internal thought. Consequently, the reorganized level 1 knowledge will come to involve some of the presuppositions of the reflective processes, which will make it possible for those presuppositions to be represented, leading to the emergence of level 3 knowledge—an explicit characterization of level 2 presuppositions.

While in principle, this loop of externalization and reflective abstraction could proceed indefinitely, having a symbolic system that provides an external systematic format for mental content greatly aids the process. Knowing processes that are put in language can be examined in terms of their presuppositions regardless of the level of reflection. As discussed by Campbell and Bickhard (1986), Aristotle's development of syllogistic logic forms an illustrative example here: He started to use abbreviations for names in syllogistic sentences, which later became variables in the general form—reflective abstraction of the logical properties of level 2 reasoning into an explicit representation of those properties. Once that happened, it became possible to examine the presuppositions of that abstracted framework and construct a representation of them as Aristotle's syllogistic calculus—level 4.

5.1 Language and developmental methodology

Thinking about how language operates for pre-reflective thought has implications for methodological design and interpretation of empirical results. In general, language does not operate for 3-year-olds as it does for 4- and 5-year-olds. This means that the same instructions or manipulations have different consequences for the two groups. For example, consider social learning research focused on testimony (Harris et al., 2012). The canonical version of the *trust* paradigm involves someone (mis)labeling familiar objects to induce (un)reliability in one of two informants. From the interactivist perspective, the nature of this manipulation is different for pre-reflective 3-year-olds than it is for 4- and 5-year-olds. For 3-year-olds, the mislabeling cannot be a *reliability* manipulation *per se*. Reliability is a reflective attribute that can only potentially be represented by around age 4. The manipulation clearly has consequences for 3-year-olds in terms of their informant preferences, but we would suggest that the proper interpretation of those preferences is in terms of 3-year-olds avoiding the "unreliable" informant rather than selecting the "reliable" informant. This would mean that trust research is more appropriately characterized as being about "mistrust" for children under age four. Further, a scientific explanation of the reasons for their (mis)trust can be modeled in ways that go beyond dispositional explanations about credulity and skepticism.

For example, consider testimony paradigms with a single informant who makes a claim that differs from the child's own experience (Ma and Ganea, 2010). In this situation, an object is placed in an occluded location. An informant then claims it is actually at a second location. Three-year-olds, but not older children, chose to rely on the informant's information over their own experience. The explanation for this is that 3-year-olds are overly credulous. However, other evidence suggests that 3-year-olds are overly skeptical (Woolley and Ghossainy, 2013). This raises two issues: (1) which characterization is accurate; (2) being credulous or skeptical does not explain behavior so much as it describes a tendency to behave a certain way. From the interactivist perspective, 3-year-olds "credulity/skepticism" are both a consequence of language as transforming social realities. In the case of credulity, the informant's claim transforms the 3-year-olds interactive characterization into one in which the object is indeed at the second location. This happens because they cannot yet evaluate the utterance separate from its transformative consequences. In the case of skepticism, the testimony applies for claims about contents for which the child does not have interactive experience (e.g., fantastical/historical characters). Accordingly, the utterance in such situations has too little interactive characterization to transform. This is like trying to manipulate physical objects that do not exist. Accordingly, reflection will be required to represent fantastical objects in the first place such that an utterance can then serve its transformative function.

6 Conclusion

The current proposal sought to critically evaluate extant models of the emergence of representation during development (both for *foundational* emergence as well as for *subsequent* emergence). It was concluded that the limitations of these models ultimately derive from their own development within an information-processing framework. Interactivism was introduced as an action-based alternative to information-processing and its specific models of representation (foundational emergence) and reflection (subsequent emergence) were presented. Implications for the model of reflection were discussed in terms of some empirics, thinking about consciousness, enculturation as a construction process on the part of the child, and the role of language in that process with some examples involving the sociality of theory of mind. A final discussion opened the door to considerations about how language may affect developmental methodology and interpretation for preschooler with reflective vs. pre-reflective capabilities.

Author contributions

JA: Conceptualization, Writing – original draft, Writing – review & editing. RM: Conceptualization, Writing – original draft, Writing – review & editing. MB: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank Stephanie Carlson for helping us work through the open access process and two reviewers for their comments. We would also like to thank Ezgi Ozgan, Deniz Hasan, and Yağmur Esendemir for their comments on a prior draft.

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

References

Allen, J. W., and Bickhard, M. H. (2011). You can't get there from here: foundationalism and development. *Behav. Brain Sci.* 34:124. doi: 10.1017/S0140525X10002311

Allen, J. W., and Bickhard, M. H. (2013). Stepping off the pendulum: why only an action-based approach can transcend the nativist-empiricist debate. *Cogn. Dev.* 28, 96–133. doi: 10.1016/j.cogdev.2013.01.002

Allen, J. W., and Bickhard, M. H. (2018). Stage fright: internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cogn. Dev.* 45, 77–91. doi: 10.1016/j.cogdev.2017.12.005

Allen, J. W., and Bickhard, M. H. (2022). Emergent constructivism: theoretical and methodological considerations. *Hum. Dev.* 66, 276–294. doi: 10.1159/000526220

Allen, J. W., Çelik, B., and Bickhard, M. H. (2021). Age 4 transitions: reflection as a domain-general development for explicit reasoning. *Cogn. Dev.* 59:101091. doi: 10.1016/j.cogdev.2021.101091

Baillargeon, R. (2008). Innate ideas revisited: for a principle of persistence in infants' physical reasoning. *Psychol. Sci.* 3, 2–13. doi: 10.1111/j.1745-6916.2008.00056.x

Bickhard, M. H. (1978). The nature of developmental stages. *Hum. Dev.* 21, 217–233. doi: 10.1159/000271586

Bickhard, M. H. (1980). *Cognition, Convention, and Communication*. Westport, CT: Praeger.

Bickhard, M. H. (1992). Commentary on the age 4 transition. *Hum. Dev.* 35 182–192. doi: 10.1159/000277151

Bickhard, M. H. (2005). Consciousness and reflective consciousness. *Philos. Psychol.* 18, 205–218. doi: 10.1080/09515080500169306

Bickhard, M. H. (2009a). The interactivist model. *Synthese* 166, 547–591. doi: 10.1007/s11229-008-9375-x

Bickhard, M. H. (2009b). Interactivism: a manifesto. *New Ideas Psychol.* 27, 85–95. doi: 10.1016/j.newideapsych.2008.05.001

Bickhard, M. H. (2015). Toward a model of functional brain processes II: central nervous system functional macro-architecture. *Axiomathes* 25, 377–407. doi: 10.1007/s10516-015-9276-9

Bickhard, M. H. (2024). *The Whole Person: Toward a Naturalism of Persons*. Amsterdam: Elsevier.

Bickhard, M. H., and Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution, Vol. 109.* New York, NY: Elsevier.

Campbell, R. L., and Bickhard, M. H. (1986). *Knowing Levels and Developmental Stages. Contributions to Human Development, Vol. 16.* Berlin: S. Karger.

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

Chomsky, N. (1959). A review of Skinner's verbal behavior. *Language* 35, 26–58. doi: 10.2307/411334

Christopher, J. C., and Bickhard, M. H. (2007). Culture, self and identity: interactivist contributions to a metatheory for cultural psychology. *Cult. Psychol.* 13, 259–295. doi: 10.1177/1354067X07079881

Dewey, J. (1915). "The existence of the world as a logical problem," in *The Collected Works of John Dewey, 1882–1953. 37 Vols. The Middle Works. Vol. 8, 83–97* (Carbondale, IL: Southern Illinois University Press), 1967–1987.

Dewey, J. (1941). "Propositions, warranted assertibility, and truth," in *The Collected Works of John Dewey, 1882–1953. 37 Vols. The Later Works of John Dewey, 1925–1953 (168–188)* (Carbondale, IL: Southern Illinois University Press), 1967–1987.

Engel, A. K., Friston, K. J., and Kragic, D. (2016). *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science, Vol. 18.* Cambridge, MA: MIT Press.

Fodor, J. (1998). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind.* Cambridge, MA: The MIT Press.

Fodor, J. A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., and Rigler, M. (1974). The development of language in genie: a case of language acquisition beyond the "critical period". *Brain Lang.* 1, 81–107. doi: 10.1016/0093-934X(74)90027-3

Gershkoff-Stowe, L., and Thelen, E. (2004). U-shaped changes in behavior: a dynamic systems perspective. *J. Cogn. Dev.* 5, 11–36. doi: 10.1207/s15327647jcd0501_2

Gopnik, A. (2003). "The theory theory as an alternative to the innateness hypothesis," in *Chomsky and His Critics*, eds. L. Antony and N. Hornstein (New York, NY: Basil Blackwell), 238–254.

Harris, P. L., Corriveau, K. H., Pasquini, E. S., Koenig, M., Fusaro, M., and Clément, F. (2012). Credulity and the development of selective trust in early childhood. *Foundat. Metacogn.* 193–210. doi: 10.1093/acprof:oso/9780199646739.003.0013

Iao, L. S., Leekam, S., Perner, J., and McConachie, H. (2011). Further evidence for nonspecificity of theory of mind in preschoolers: training and transferability in the understanding of false beliefs and false signs. *J. Cogn. Dev.* 12, 56–79. doi: 10.1080/15248372.2011.539523

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (2018). "An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis," in *Thinking Developmentally From Constructivism to Neuroconstructivism* (London: Routledge), 289–304.

Ma, L., and Ganea, P. A. (2010). Dealing with conflicting information: young children's reliance on what they see vs. what they are told. *Dev. Sci.* 13, 151–160. doi: 10.1111/j.1467-7687.2009.00878.x

Mandler, J. (2004). *The Foundations of Mind: Origins of Conceptual Thought*. New York, NY: Oxford University Press.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *J. Cogn. Dev*. 1, 3–36. doi: 10.1207/S15327647JCD0101N_2

Mirski, R., and Bickhard, M. H. (2021). Conventional minds: an interactivist perspective on social cognition and its enculturation. *New Ideas Psychol*. 62:100856. doi: 10.1016/j.newideapsych.2021.100856

Müller, U., and Kerns, K. (2015). The development of executive function. Handbook of child psychology and developmental science. *Cogn. Process*. 2, 571–623. doi: 10.1002/9781118963418.childpsy214

Müller, U., and Overton, W. F. (1998). How to grow a baby: a reevaluation of image-schema and Piagetian action approaches to representation. *Hum. Dev*. 41, 71–111. doi: 10.1159/000022570

Naito, M. (2014). "From theory of mind to theory of relation: sociocultural perspectives on Japanese Children's Social Understanding," in *Contemporary Perspectives in Early Childhood Education. Contemporary Perspectives on Research in Theory of Mind in Early Childhood Education*, ed. O. N. Saracho (Cambridge, MA: Information Age Publishing), 381–408.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130. doi: 10.1017/S0140525X08003543

Perner, J. (1991). *Understanding the Representational Mind.* Cambridge, MA: MIT Press.

Perner, J., Stummer, S., Sprung, M., and Doherty, M. (2002). Theory of mind finds its Piagetian perspective: why alternative naming comes with understanding belief. *Cogn. Dev.* 17, 1451–1472. doi: 10.1016/S0885-2014(02)00127-2

Piaget, J. (1952). Play, dreams and imitation in childhood. *Int. Libr. Psychol.* 1999:87.

Piaget, J. (1954). *The Construction of Reality in the Child.* New York, NY: Norton.

Piaget, J. (1971). Genetic epistemology. *Norton* 1971:piag91272. doi: 10.7312/piag91272

Piaget, J., and Inhelder, B. (2000). *The Psychology of the Child*. New York, NY: Basic Books.

Smith, L. (1993). *Necessary Knowledge: Piagetian Perspectives on Constructivism (Essays in Developmental Psychology)* (Hove: Lawrence Elrbaum Associates Ltd, Publishers), 155.

Symons, D., McLaughlin, E., Moore, C., and Morine, S. (1997). Integrating relationship constructs and emotional experience into false belief tasks in preschool children. *J. Exp. Child Psychol*. 67, 423–447. doi: 10.1006/jecp.1997.2416

Tiles, J. E. (1990). *Dewey*. London: Routledge.

Tomasello, M. (2024). An agency-based model of executive and metacognitive regulation. *Front. Dev. Psychol.* 2:1367381. doi: 10.3389/fdpys.2024.1367381

Turner, S. P. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge and Presuppositions/Stephen Turner*. Cambridge: Polity Press.

Turner, S. P. (2018). *Cognitive Science and the Social: A Primer/Stephen P. Turner (1st)*. London: Routledge.

Vygotsky, L. S. (1978). *Mind in Society: the Development of Higher Psychological Processes/L. S. Vygotsky; Edited By Michael Cole ... [et al.]; [Translated From the Russian]*. Cambridge, MA: Harvard University Press.

Woolley, J. D. E., and Ghossainy, M. (2013). Revisiting the fantasy-reality distinction: children as naïve skeptics. *Child Dev*. 84, 1496–1510. doi: 10.1111/cdev.12081

Zelazo, P. D. (1999). "Language, levels of consciousness, and the development of intentional action," in *Developing Theories of Intention* (London: Psychology Press), 95–118.

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends Cogn. Sci*. 11:1. doi: 10.1016/j.tics.2003.11.001

Zelazo, P. D. (2015). Executive function: reflection, iterative reprocessing, complexity, and the developing brain. *Dev. Rev*. 38, 55–68. doi: 10.1016/j.dr.2015.07.001

frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 23 August 2024
DOI 10.3389/fdpys.2024.1424754

OPEN ACCESS

EDITED BY
Claudia M. Roebers,
University of Bern, Switzerland

REVIEWED BY
Mari Van Loon,
University of Zurich, Switzerland
Valerio Santangelo,
University of Perugia, Italy

*CORRESPONDENCE
Eveline Jacobs
✉ eveline.jacobs@kuleuven.be

RECEIVED 28 April 2024
ACCEPTED 08 August 2024
PUBLISHED 23 August 2024

CITATION
Jacobs E, Bellon E and De Smedt B (2024) Adjusting to errors in arithmetic: a longitudinal investigation of metacognitive control in 7–9-year-olds.
Front. Dev. Psychol. 2:1424754.
doi: 10.3389/fdpys.2024.1424754

Adjusting to errors in arithmetic: a longitudinal investigation of metacognitive control in 7–9-year-olds

Eveline Jacobs*, Elien Bellon and Bert De Smedt

Faculty of Psychology and Educational Sciences, Parenting and Special Education Research Group, University of Leuven (KU Leuven), Leuven, Belgium

Introduction: Monitoring and controlling one's performance are essential skills for children's cognitive development and academic success. Metacognitive control, operationalized as post-error adjustments, is, however, often measured in conflict tasks, but the findings of such studies may not be readily generalizable to academic domains, such as arithmetic. Yet, investigating how children control their performance in arithmetic is crucial in understanding the large individual differences within this specific academic domain. This longitudinal study investigated how children control their performance through post-error slowing and accuracy improvement in arithmetic. We additionally examined this development of metacognitive control in a working memory task, to further unravel its domain-generality or the lack thereof.

Methods: A cohort of 127 typically developing children, followed up longitudinally from 7–8 years old (2nd grade of primary school) to 8–9 years old (3rd grade of primary school), completed an arithmetic and working memory task at two time points.

Results and discussion: Meticulous comparison of response times and accuracy rates following errors with those following correct answers revealed the presence of metacognitive control at each time point. We observed significant positive correlations between children's metacognitive control and their arithmetic accuracy at 7–8 years old, underscoring a possible adaptive role of metacognitive control in the learning phase of arithmetic. No correlations were found between the post-error adjustments in the arithmetic task and those in the working memory task, challenging previous evidence for domain-generality of post-error adjustments.

KEYWORDS

metacognitive control, post-error slowing, post-error improvement in accuracy, metacognitive monitoring, mathematical cognition, mental arithmetic, working memory

1 Introduction

Imagine a student taking a test. He feels confident answering the questions due to his thorough preparation. Yet, as the test progresses, he encounters a more challenging part, making him uncertain and less confident about his answers. The student, therefore, decides to slow down his thought process to answer the questions with increased focus. The awareness and regulation of one's own cognitive processes, as the student presented in this example,

10.3389/fdpys.2024.1424754

are also known as metacognitive monitoring and metacognitive control, respectively, both subskills of metacognitive regulation or procedural metacognition (Flavell, 1979; Nelson and Narens, 1990). Metacognitive regulation is thought to be of great importance to academic learning (e.g., Efklides and Misailidi, 2010). Yet, most of the existing body of work has examined metacognitive control in simple tasks, such as perceptual or conflict tasks, leaving it unresolved as to how it operates in academic tasks. The present study aimed to address this gap by exploring metacognitive control in the context of arithmetic.

Metacognitive control encompasses the actions individuals take to enable cognitive adaptations to increase task performance (Roebers et al., 2014). There are many possible manifestations of metacognitive control, such as allocation of study time and information-seeking, which occur during different stages of cognitive performances (Nelson and Narens, 1990). These manifestations of metacognitive control are often measured explicitly by giving participants the option to control their performance (e.g., asking participants whether they need help, Coughlin et al., 2015). It is, however, also possible to assess metacognitive control in an implicit way through post-error adjustments, which are thought to reflect the cognitive adaptations individuals make following errors (Danielmeier and Ullsperger, 2011). Two prominent manifestations of post-error adjustments are post-error slowing (PES) and post-error improvement in accuracy (PEIA). PES refers to the phenomenon that people tend to slow down their response speed after committing an error (Notebaert et al., 2009). PEIA is the phenomenon that individuals show increased accuracy in performance immediately after committing an error (Danielmeier and Ullsperger, 2011). It is important to note that PES and PEIA are often considered to be measures of cognitive control rather than metacognitive control due to their immediate nature, with the main difference between these two types of control being the consciousness involved in these processes (Roebers, 2017). However, the extent to which individuals consciously engage in these adjustments might vary, both across tasks and among people, making it possible that post-error adjustments are situated on a continuum from cognitive to metacognitive control. As it is beyond the scope of this study to disentangle these two conceptualizations, findings from both the cognitive and metacognitive control literature are integrated in the current study.

Although it is generally agreed that PES and PEIA reflect cognitive control, alternative interpretations especially regarding PES, have been suggested. Specifically, the orienting account posits PES as a reaction to a surprising event (i.e., an error) that prompts the individual to slow down (Notebaert et al., 2009). However, studies giving evidence for PES as an orienting response are restricted to conflict tasks (e.g., Fiehler et al., 2005; Hajcak and Simons, 2008; King et al., 2010; Notebaert et al., 2009; Notebaert and Verguts, 2011), in which opportunities for behavioral post-error adjustments are limited. This interpretation may, therefore, not apply to academic tasks, which have a more complex nature and, therefore, allow for multiple possible ways to adapt behavior following errors.

One such an academic domain is arithmetic. Studies investigating PES and PEIA in arithmetic are surprisingly scarce. The few studies that have examined this have found both PES and PEIA to be present in children (de Mooij et al., 2022) and adults (Desmet et al., 2012; Van der Borght et al., 2016). Both Desmet et al. (2012) and Van der Borght et al. (2016) studied PES and PEIA in a verification multiplication task in university students, and concluded that post-error adjustments can be observed in arithmetic in adults. Van der Borght et al. (2016) additionally highlighted the role of changing strategies after errors as a way to improve task performance in arithmetic. This finding suggests that, in contrast to conflict tasks, arithmetic does allow for multiple ways to adjust behavior after committing an error. Regarding post-error adjustments in children, to our knowledge, only one study has been performed in the domain of arithmetic. de Mooij et al. (2022) investigated PES in children from 5 to 13 years old in an adaptive learning environment including both mathematical and language activities. They found PES to be present in almost all learning activities, and found it to be positively associated with PEIA and the children's ability level. This latter finding suggests that PES could play an important role in explaining the large individual differences in mathematical ability, and more specifically in arithmetic skills.

While there is knowledge—albeit limited—on post-error adjustments in arithmetic, knowledge about its development in this particular domain is close to non-existent. Studies investigating the development of metacognitive control in other domains, such as spelling and memory, agree that metacognitive control undergoes substantial development during primary school (Krebs and Roebers, 2010; Roebers et al., 2014; Roebers and Spiess, 2017; Selmeczy et al., 2021), and continues to develop until late adolescence (Crone and Steinbeis, 2017). However, these studies investigated explicit operationalizations of metacognitive control, such as withdrawal of wrong answers and information-seeking. Findings regarding the development of post-error adjustments specifically remain to be mixed, as some studies found the magnitude of PES to decrease between 7 and 19 years old (Dubravac et al., 2022; Schachar et al., 2004), while Smulders et al. (2016) found it to remain stable across development until adulthood and Gupta et al. (2009) reported a non-linear developmental trend between the ages of 6 and 11. In the domain of arithmetic, there is, to our knowledge, only one study that examined the development of post-error adjustments. de Mooij et al. (2022) investigated PES cross-sectionally from 5 until 13 years old in mathematical activities. They found a non-linear developmental trend with an increase in PES from 6 to 9 followed by a decrease from 9 until 13 years old, suggesting that children from 6 to 9 years old are in an important developmental phase regarding metacognitive control in mathematical activities. The authors interpreted the decrease in PES from 9 to 13 years old as a shift from reactive control, which accounts for greater PES, to more proactive control, as previous research has provided evidence for such a shift around the age of 8 years old (Niebaum et al., 2021). Nevertheless, to our knowledge, no longitudinal investigations on metacognitive control in arithmetic have been performed as of now.

Another question that arises, especially in a developmental perspective, is whether post-error adjustments are domain-specific or domain-general. Research has only recently begun to address this question. Ger and Roebers (2023) and Dubravac et al. (2022)

argue for a domain-general nature, as they found similarities in PES between different tasks at various ages ranging from 4 years old until adulthood. However, both of these studies compared conflict tasks. Studies examining this issue in academic tasks are scarce. van Loon et al. (2024) investigated the withdrawal of wrong answers as a measure of metacognitive control in three different language-based learning tasks in 8- to 12-year-olds. While they mainly observed evidence for a domain-general nature, they also found evidence for a task-specific factor. However, what is similar in all the above-described studies is that they all compared tasks that are different versions of the same task, and, therefore, are, quite similar in task-requirements and cognitive demands, for which reason correlations between tasks, which is taken as evidence for domain-generality, are likely to occur. To the best of our knowledge, studies investigating the domain-generality of metacognitive control in tasks that involve more distinct cognitive domains are non-existent. Such studies might provide a more appropriate test of the idea of domain-generality of metacognitive control, and more specifically of post-error adjustments. We will address this issue in the current study.

Extending the discussion of control as a separate skill, according to the theoretical framework of Nelson and Narens (1990), this skill is closely intertwined with monitoring, which involves the self-awareness and judgement of one's own task performance and has been shown to be a fundamental skill in diverse domains, such as memory, reading, spelling, and arithmetic (e.g., Bellon et al., 2019, 2020; Efklides and Misailidi, 2010; Rinne and Mazzocco, 2014; Schneider and Artelt, 2010; Touron et al., 2010). This theoretical assumption is supported by empirical evidence in adults. For example, adults allocate their study time based on how well they think they know the subject (Souchay et al., 2003), and seem to slow down more when they are uncertain about their answer (Dali et al., 2022). From a developmental perspective, however, the evidence for this hypothesis is less conclusive. While most studies have found an association between monitoring and control in primary school children between the ages of 8 and 12 (e.g., Destan et al., 2014; Hoffmann-Biencourt et al., 2010; Krebs and Roebers, 2010; Roebers and Spiess, 2017; Steiner et al., 2020; van Loon et al., 2024), a few studies have observed monitoring and control to operate independently from each other in that same age range (O'Leary and Sloutsky, 2017, 2019). In younger age groups, from 5 until 7 years old, most studies have failed to find an association between the two skills (e.g., Destan et al., 2014; Roebers and Spiess, 2017), while other studies observed an association between the two skills in even younger children in preschool (e.g., Coughlin et al., 2015; Gardier and Geurten, 2024). These findings suggest that, while it appears that monitoring and control become increasingly intertwined across development, their association is complex and results might depend on measurement methods, as the described studies used diverse measures of metacognitive control (e.g., Destan et al., 2014; Hoffmann-Biencourt et al., 2010; Krebs and Roebers, 2010; Roebers and Spiess, 2017). Furthermore, to our knowledge, no studies have investigated implicit measures, such as post-error adjustments, in relation to explicit metacognitive monitoring in academic tasks, which will be addressed in the current study. As a measure of metacognitive monitoring, the current study focuses on task-specific retrospective monitoring, identified as an important, unique predictor of children's concurrent (Bellon et al., 2019) and future arithmetic skills (Bellon et al., 2021; Rinne and Mazzocco, 2014). Assessing confidence judgements retrospectively is particularly interesting in relation to post-error adjustments, as this allows us to examine how these judgments are related to immediate subsequent behavior.

In the present study, we longitudinally examined metacognitive control, operationalized as PES and PEIA, in arithmetic in children of 7–9 years old, as this age range is considered an important developmental period for metacognitive regulation (de Mooij et al., 2022; Geurten et al., 2018). To do so, we had four aims. Firstly, we wanted to examine the presence of PES and PEIA during an arithmetic task in 7–8- and 8–9-year-olds, that is 2nd and 3rd grade of primary school (Research question 1). We expected to observe both PES and PEIA at both ages and predicted to observe greater PES and PEIA in 3rd grade than in 2nd grade. As an exploratory analysis, we additionally investigated the association between PES and PEIA to further unravel the underlying mechanisms of metacognitive control. Secondly, we wanted to investigate whether an association was present between PES and PEIA on the one hand and overall task performance on the other hand (Research question 2). We expected PES and PEIA to be positively correlated with overall task performance. Thirdly, we aimed to examine the association between metacognitive control (operationalized via PES and PEIA) and metacognitive monitoring (Research question 3). Given the age range of the children under study, we did not expect control to be correlated with monitoring. Fourth, the present study aimed to examine domain-generality of metacognitive control (Research question 4). To do so, we examined PES and PEIA in a working memory task to verify whether results differed with the ones found in the arithmetic task. Comparing two tasks that reflect distinct domains could yield new insights beyond those obtained from studies comparing tasks within similar domains (e.g., Dubravac et al., 2022; Ger and Roebers, 2023). While we expected to observe PES and PEIA in the working memory task as well as correlations with overall task performance, we did not expect correlations with PES and PEIA in the arithmetic task, challenging previous found evidence for domain-generality of post-error adjustments in children at this young age.

2 Methods

This study involves a secondary data analysis of the studies by Bellon et al. (2019, 2021). These studies focused on the cross-sectional associations of numerical magnitude processing, executive functions and metacognitive monitoring during arithmetic (Bellon et al., 2019) and the longitudinal associations between metacognitive monitoring, math anxiety and arithmetic (Bellon et al., 2021). None of these studies reported data on metacognitive control. As a result, measures of PES and PEIA, indices of metacognitive control, have never been analyzed and reported before, which makes the current study unique.

10.3389/fdpys.2024.1424754

2.1 Participants

No a-priori sample size calculation was performed. At the outset of the longitudinal study, a total of 127 typically developing Flemish children from 2nd grade of primary school participated (64 girls; M_{age} = 7 years 11 months, *SD* = 4 months, range = 7 years 4 months to 8 years 5 months). Of these participants, 121 were followed up 1 year later in 3rd grade (63 girls; M_{age} = 8 years 8 months, *SD* = 3 months, range = 8 years 2 months to 9 years 2 months). None of them had a diagnosis of a developmental disorder, nor did any of them repeat a grade. All the participants had a predominantly middle- to high-socioeconomic background. Written informed parental consent was obtained for every participant. This study was approved by the social and societal ethics committee of KU Leuven.

2.2 Materials and measures

Materials consisted of custom computerized tasks designed with E-Prime 2.0 (Schneider et al., 2002) and a standardized paper-and-pencil test.

2.2.1 Arithmetic task

2.2.1.1 General performance

A single-digit computerized production task with addition and multiplication problems was administered. Because the accuracy rate of the addition problems was too high for the scope of this study, the current study focused only on the second part of this task, namely the multiplication problems. The multiplication problems consisted of all possible combinations of the numbers 2 through 9 as operands. Problems with 0 or 1 as one of the operands were excluded, yielding a total of 64 multiplication problems. To ensure the children were familiar with the task, six practice trials were performed at the start. After fixation, each item was presented in white on a black background for 2,000 ms. The children were instructed to respond verbally as quickly and accurately as possible as soon as the item was presented. Once the 2,000 ms passed, a black screen appeared, during which the children were still allowed to response. RTs and answers were registered by the experimenter through a key press on the computer. The task was pseudo-randomly divided into two blocks (i.e., no commutative pairs in the same block). During the first block the children were presented with the multiplication items as described above. During the second block, each arithmetic item was followed by a metacognitive monitoring measure (see below). The performance metric used was the average response time on correct multiplication trials and the total number of correct multiplication answers across the two blocks.

2.2.1.2 Metacognitive control

Metacognitive control was measured trough PES and PEIA in the computerized arithmetic task. Response times on a trial-by-trial basis were used to quantify PES. Trial-by-trial accuracy rates were used to quantify PEIA.

2.2.1.2.1 Post-error slowing

There are two prominent ways of measuring PES in the current body of literature, namely the traditional method and a robust method of Dutilh et al. (2012). The traditional method quantifies PES as the difference between the mean RT of correct trials following errors and the mean RT of correct trials following correct trials. However, according to Dutilh et al. (2012) this method can be biased because of fluctuations in attention and motivation during the task. Therefore, these authors proposed a more robust method, in which PES is quantified as the average difference between the RT of correct trials following errors and the RT of trials preceding an error. In the current study, however, the stimuli used in the computerized arithmetic task were multiplication problems, which are known to differ in the level of difficulty due to the problem size and interference effects (De Visscher et al., 2018; Imbo and Vandierendonck, 2008). This results in longer RTs and lower accuracy rates on harder problems compared to easier ones. Additionally, as Derrfuss et al. (2022) pointed out with congruent and incongruent trials in interference tasks, these differences in difficulty could account for imbalances in the percentage of post-correct, pre-error, and post-error trials. These two considerations call for the need of a quantification of PES that is corrected for these imbalances.

The current study, therefore, pioneers in using two quantifications of PES based on Derrfuss et al. (2022) in arithmetic: the corrected traditional method and the corrected robust method. This implied that we divided the multiplication problems in categories of equal difficulty based on the problem size effect (i.e., large problems are harder than small problems, Imbo and Vandierendonck, 2008) and the interference effect (i.e., problems that have more overlap in digits with previously learned problems are harder to retrieve than low interfering problems, De Visscher et al., 2018), resulting in three categories: (1) the easiest category, which consisted of problems with a problem size below or equal to 25 and an interference effect below 8, (2) the middle category, which included problems with a problem size below or equal to 25 and an interference effect above or equal to 8, or vice versa, and (3) the hardest category, which consisted of problems with a problem size above 25 and an interference effect above or equal to 8. For the corrected traditional method, PES was calculated by computing the difference between the mean RT of post-error correct trials and the mean RT of post-correct correct trials for each level of difficulty separately to control for imbalances in it, and then taking the average across these three measures for each participant. Similarly, the corrected robust method was calculated by computing the difference between the mean RT of post-error correct trials and the mean RT of pre-error correct trials for each level of difficulty separately, before averaging across these three measures for each participant. In order to not completely rely on these two quantifications of PES and because RT data are typically skewed with large variability, we additionally repeated the same quantifications making use of the median instead of the mean. This resulted in four different quantifications of PES for each participant in the computerized arithmetic task. However, during analyses we encountered some challenges related to the robust method, which are discussed in more detail in Section 3.

$$PES_{trad,\ corr} = \frac{\sum(\overline{RT}_{PostError_{correct},diff_i} - \overline{RT}_{PostCorrect_{correct},diff_i})}{n_{diff}}$$

$$PES_{robust,\ corr} = \frac{\sum(\overline{RT}_{PostError_{correct},diff_i} - \overline{RT}_{PreError_{correct},diff_i})}{n_{diff}}$$

2.2.1.2.2 Post-error improvement in accuracy

As the above-described difficulty differences might also influence accuracy rates, PEIA in the computerized arithmetic task was quantified in a similar way to PES. We calculated the difference between the proportion of post-error trials that were answered correctly (out of the total number of post-error trials) and the proportion of post-correct trials that were answered correctly (out of the total number of post-correct trials). This was done for each difficulty level separately before averaging these differences to obtain a single PEIA measure for each participant controlled for the influence of difficulty differences.

$$PEIA_{corr} = \frac{\sum(\frac{n_{PostError_{correct},diff_i}}{n_{PostError_{total},diff_i}} - \frac{n_{PostCorrect_{correct},diff_i}}{n_{PostCorrect_{total},diff_i}})}{n_{diff}}$$

2.2.1.3 Metacognitive monitoring

Metacognitive monitoring was measured similarly to Rinne and Mazzocco (2014). In the second block of the computerized arithmetic task, a question was added after each item to measure task-specific retrospective metacognitive judgements. After each multiplication item, participants were asked to indicate their confidence in the accuracy of their answer. They did so by verbally choosing between "correct", "not sure", or "incorrect". These three answer options were presented simultaneously on the screen accompanied by a happy, neutral, and sad smiley, respectively. The participants were presented with six practice trials to familiarize themselves with the task. Calibration of confidence scores, which represent the alignment between the participant's confidence and the actual accuracy of their arithmetic answer, were calculated on a trial-by-trial basis, as in Bellon et al. (2019, 2020, 2021). Participants got a score of 2 when they made a correct judgement (i.e., said they were correct when they were correct, or reversed), a score of 1 when they answered, "not sure", and a score of 0 when they made an incorrect judgment (i.e., said they were correct when they were incorrect, or reversed). These scores were then averaged for each participant, yielding one calibration score per child.

2.2.2 Working memory task

2.2.2.1 General performance

Working memory was assessed using a standard 2-back task (adapted from Pelegrina et al., 2015). Participants were presented with a sequence of colored images on a computer screen. For each item, they needed to indicate whether the presented stimulus matched the one that occurred two trials back. To do so, participants pressed a green or red key, corresponding to "yes" or "no", respectively. After fixation, the items were presented in the center of a white screen for 3,000 ms. This was followed by a black screen for 1,000 ms. The participants were allowed to answer both during the white screen and the black screen. They were instructed to answer as quickly and accurately as possible. In total, 40 items, divided into two blocks, were presented. An additional practice block of 20 trials was added to the beginning of the task to familiarize the children with the requirements. Each block started with three non-target trials (correct answer = no) and 30% of the trials in each block were target trials (correct answer = yes). The total number of correct answers served as a performance measure reflecting the participants' working memory skills.

2.2.2.2 Metacognitive control

Metacognitive control, measured trough PES and PEIA, was also assessed in the domain of working memory. Response times on a trial-by-trial basis of the 2-back task were used to quantify PES. Trial-by-trial accuracy rates were used to quantify PEIA.

2.2.2.2.1 Post-error slowing

PES in the 2-back task was quantified in the same ways as earlier described in the arithmetic task. However, as all the trials in the 2-back task are expected to be of a similar difficulty level, no correction for difficulty differences was applied. This resulted in four uncorrected measures of PES: the traditional method using the mean, the robust method using the mean, the traditional method using the median, and the robust method using the median.

2.2.2.2.2 Post-error improvement in accuracy

For the 2-back task, PEIA was quantified in the same way as in the computerized arithmetic task, except that we did not control for differences in difficulty. However, the quantification of PEIA in the 2-back task poses a challenge, as the accuracy on a trial directly depends on the participant's performance two trials before. Therefore, participants are not able to actively improve their accuracy on the trial immediately after the error but might be able to do so two or three trials after committing the error. Thus, PEIA in the 2-back task was quantified in two ways: (1) as the difference between the proportion of correct answers two trials after an error and the proportion of correct answers two trials after a correct answer, and (2) as the difference between the proportion of correct responses on trials that were completed three trials after an error and the proportion of correct answers three trials after a correct answer.

2.2.3 Control variables

The Raven's Standard Progressive Matrices (Raven et al., 1992) were used to assess intellectual ability. This standardized test was used as a control measure in our study to make sure that any associations observed between the various variables could not be explained by the intellectual ability of the children, as all of the variables of interest in our study are assumed to be associated with intellectual ability to some extent (e.g., Veenman and Spaans, 2005). Children were instructed to complete 60 patterns. To do so, they had to choose the correct answer out of the provided possibilities. The number of correctly solved patterns within the time limit of 40 min was the performance metric.

2.3 Procedure

The administered tasks were divided into three sessions. The first session was an individual session, in which each participant completed the arithmetic task. In the second session, groups of five children were tested individually on the working memory task. During this session the children were also administered other tasks, from which the data were not used in the current study. These tasks included a motor speed task, a symbolic numerical magnitude processing task, and three other executive functioning tasks. Lastly, there was a group-administered session for the intellectual ability task. This last session also included three other paper and pencil tasks, namely the Tempo Test Arithmetic, a metacognitive questionnaire, and a math anxiety questionnaire, from which the data were not used in the current study. All these sessions took place at the school of the participants during regular school hours. Each child went through the exact same order of tasks. The duration of the sessions was 40, 45, and 60 min, respectively. The participants were tested again 1 year later on the same tasks in the same order.

2.4 Analyses

We employed a combination of frequentist and Bayesian analyses to examine the data. For the Bayesian analyses, a default prior provided by the statistical program JASP (JASP Team, 2024) was used. Prior to conducting the main analyses, ANOVAs were performed to check whether the proposed difficulty categories in the arithmetic task differed in average RT and accuracy rate. The main analyses aimed to investigate the presence of PES and PEIA in primary school children in the domain of arithmetic. To do so, we ran one-sample *t*-tests for the various quantifications of PES and PEIA, and paired *t*-tests to assess developmental changes (Research question 1). Furthermore, we assessed correlations between post-error adjustments, calibration scores, and overall task performance (Research questions 2 and 3). Additionally, the same analyses were performed in the working memory task to assess the presence of metacognitive control and its correlations with overall task performance. These results were then compared with the results obtained in the arithmetic domain to investigate domain-generality of post-error adjustments (Research question 4).

3 Results

As the current study controlled for difficulty differences in the arithmetic task, we encountered some challenges during the analyses regarding the robust quantification of PES. The robust method proposed by Dutilh et al. (2012) assumes trials of similar difficulty levels, as each post-error trials needs to be compared with the pre-error trial of that same error. As there was not always a pre-error trial of the same difficulty level to compare with the post-error trials in the current study, this quantification resulted in few trials to compare within each participant, ultimately leading to a less reliable measure of PES compared to the traditional quantification. Therefore, only results from the traditional quantification are reported and discussed. Results regarding the robust method are available in the Supplementary material.

3.1 Preliminary analyses

Some participants were excluded from the analyses due to the following reasons. On the arithmetic task, 3 participants in 2nd grade and 13 participants in 3rd grade made no errors. Their data were, therefore, removed for the analyses regarding post-error adjustments. An additional 13 children committed only one error on the arithmetic task in 3rd grade, which poses an issue for the interpretation of the post-error adjustments, as PEIA would always be a positive value regardless of the actual presence of metacognitive control. These participants were, thus, also removed from the analyses of post-error adjustments. Due to a lot of participants who made only two errors on the arithmetic task in 3rd grade, possibly accounting for unreliable measures of post-error adjustments, we decided to repeat all the analyses after removing these participants. Results remained unchanged. Thus, all reported results include participants that made two or more errors on the computerized arithmetic task. Finally, data of 3 participants from 2nd grade on the 2-back task were removed due to too many non-responses.

Of the remaining participants in the computerized arithmetic task, the mean accuracy rate was 0.83 ($SD = 0.38$) in 2nd grade and 0.91 ($SD = 0.28$) in 3rd grade. A paired sample *t*-tests revealed significant improvement in overall accuracy from 2nd to 3rd grade on the arithmetic task [$t_{(120)} = -10.62$, $p < 0.001$, Cohen's $d = -0.97$]. The mean RT was 9,243.11 ms ($SD = 13{,}469.68$) in 2nd grade and 5,161 ms ($SD = 6{,}287.48$) in 3rd grade. A paired sample *t*-test revealed a significant decrease in RT from 2nd to 3rd grade [$t_{(120)} = 10.70$, $p < 0.001$, Cohen's $d = 0.97$]. The mean calibration of confidence score of the participants in the arithmetic task was 1.74 ($SD = 0.18$) in 2nd grade and 1.86 ($SD = 0.13$) in 3rd grade. In the 2-back task, participants had a mean accuracy rate of 0.71 ($SD = 0.45$) in 2nd grade and 0.75 ($SD = 0.43$) in 3rd grade, and a mean RT of 1,151.77 ($SD = 490.57$) in 2nd grade and 1,162.07 ($SD = 585.12$) in 3rd grade. A paired sample *t*-test revealed a significant improvement in accuracy from 2nd to 3rd grade on the 2-back task [$t_{(118)} = -4.51$, $p < 0.001$, Cohen's $d = -0.41$].

Using ANOVA, we tested whether the chosen difficulty categories in which we divided the multiplication problems differed in RT and accuracy. The proposed difficulty categories based on the problem size and interference effect did indeed differ significantly in average RT, both in 2nd grade, $F_{(2,7,933)} = 461.67$, $p < 0.001$, $\eta^2 = 0.10$, and in 3rd grade, $F_{(2,6,077)} = 312.94$, $p < 0.001$, $\eta^2 = 0.09$. *Post-hoc* tests using Bonferroni correction indicated significant differences between all categories in both grades, with the slowest RTs in the hardest category and the fastest RTs in the easiest category. The results of these *post-hoc* tests can be found in Appendix A. The proposed difficulty categories also differed significantly in accuracy rate, both in 2nd grade, $F_{(2,7933)} = 284.88$, $p < 0.001$, $\eta^2 = 0.07$ and in 3rd grade, $F_{(2,6077)} = 75.95$, $p <$

TABLE 1 Mean/median RT's (in ms) for post-error and post-correct trials of the arithmetic task.

	Post-error trials		Post-correct trials	
2nd grade ($n = 124$)	$M = 7{,}423.08$	$SD = 5{,}697.17$	$M = 7{,}580.49$	$SD = 4{,}669.87$
	$Mdn = 6{,}903.16$	$SD = 5{,}335.83$	$Mdn = 5{,}660.01$	$SD = 3{,}544.61$
3rd grade ($n = 95$)	$M = 4{,}674.01$	$SD = 3{,}226.72$	$M = 4{,}580.35$	$SD = 2{,}282.21$
	$Mdn = 4{,}407.92$	$SD = 3{,}121.80$	$Mdn = 3{,}557.81$	$SD = 1{,}801.91$

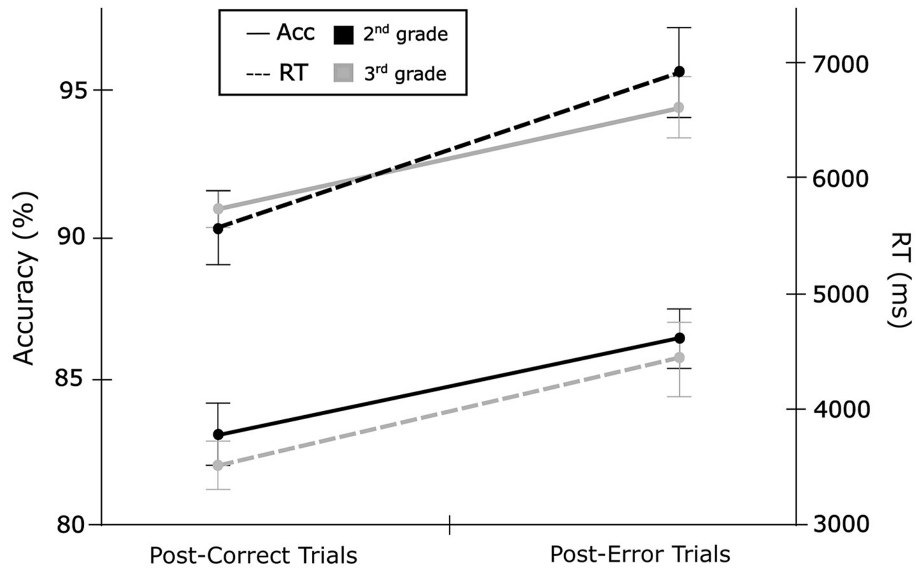


FIGURE 1
Mean accuracy rates and median response times after errors and corrects trials for 2nd and 3rd grade in the arithmetic task.

0.001, $\eta^2 = 0.02$. Similar to the RTs, *post-hoc* tests using Bonferroni correction, of which the results can be found in Appendix A, revealed significant differences between all categories in both grades, with the lowest accuracy rate in the hardest category and the highest accuracy rate in the easiest category. These results indicate both the effectiveness of our categorization and the necessity of accounting for these differences when quantifying PES and PEIA.

3.2 Metacognitive control in arithmetic

3.2.1 Post-error slowing

The mean and median RTs for post-error and post-correct trials at both time points are shown in Table 1 and Figure 1. To test whether there was PES in the arithmetic task (Research question 1), one-sample t-tests were performed for the different PES quantifications. When using the traditional quantification with the mean, there was no significant PES present, not in 2nd grade [$t_{(123)} = -0.38$, $p = 0.70$, Cohen's $d = -0.03$], nor in 3rd grade [$t_{(94)} = 0.38$, $p = 0.70$, Cohen's $d = 0.04$]. Thus, the children did not significantly respond slower on post-error trials compared to post-correct trials. The Bayes factor indicated moderate evidence in favor of the null hypothesis at both time points ($0.10 < BF_{10} < 0.33$). The corrected traditional quantification using the median, however, did reveal significant PES. This was the case in 2nd grade [$t_{(123)} = 3.12$, $p = 0.002$, Cohen's $d = 0.28$], as well as in 3rd grade [$t_{(94)} = 3.49$, $p < 0.001$, Cohen's $d = 0.36$]. Thus, using this metric, the children did significantly respond slower on post-error trials compared to post-correct trials. The Bayes factors indicated moderate to strong evidence for this effect in 2nd grade ($BF_{10} = 9.83$), and very strong evidence in 3rd grade ($BF_{10} = 30.21$).

Using paired t-tests, we investigated whether the magnitude of PES changed from 2nd to 3rd grade. The analyses revealed no significant difference between the two time points, for neither of the traditional quantifications of PES [$t_{(94)} = -0.11$, $p = 0.91$, Cohen's $d = -0.01$ for the corrected traditional method using the mean; $t_{(94)} = 1.22$, $p = 0.22$, Cohen's $d = 0.13$ for the corrected traditional method using the median]. The Bayes factor indicated moderate evidence for the null hypothesis ($0.10 < BF_{10} < 0.33$) for both quantifications.

3.2.2 Post-error improvement in accuracy

Accuracy rates of post-error and post-correct trials at both time points are shown in Table 2 and Figure 1. To investigate the presence of PEIA in the arithmetic task (Research question 1), one-sample t-tests were performed. The analysis revealed significant PEIA in arithmetic, both in 2nd grade, $t_{(123)} = 3.12$, $p = 0.002$, Cohen's $d = 0.28$, and in 3rd grade, $t_{(94)} = 3.32$, $p < 0.001$, Cohen's $d = 0.34$. The children were, thus, significantly more accurate on

TABLE 2 Accuracy rates (in %) for post-error and post-correct trials of the arithmetic task.

	Post-error trials		Post-correct trials	
	M	*SD*	*M*	*SD*
2nd grade ($n = 124$)	86.67	14.78	83.39	12.75
3rd grade ($n = 95$)	94.65	11.34	91.01	6.58

trials following an error compared to trials following a correct response. The Bayes factor indicated moderate to strong evidence in 2nd grade ($BF_{10} = 9.64$) and strong evidence in 3rd grade ($BF_{10} = 18.33$). A paired t-test revealed that the magnitude of PEIA did not significantly change from 2nd to 3rd grade, $t_{(94)} = 0.08$, $p = 0.94$, Cohen's $d = 0.01$. The Bayes factor ($BF_{10} = 0.11$) indicated moderate evidence for the null hypothesis.

3.2.3 Correlations

Pearson correlation coefficients, controlled for performance on the Raven's Standard Progressive Matrices, were run to assess associations between PES, PEIA, calibration of confidence, and overall task performance in the arithmetic task. A full correlation matrix can be found in Appendix B. No significant correlations were found between PES and PEIA, neither in 2nd grade nor in 3rd grade (Research question 1). For each quantification at both time points, the Bayes factor indicated moderate evidence in favor of the null hypothesis ($0.10 < BF_{10} < 0.33$).

Regarding the association between metacognitive control and overall task performance (Research question 2), both the corrected traditional quantifications of PES were found to be positively correlated with the number of accurate answers on the arithmetic task in 2nd grade [$r_{(121)} = 0.22$, $p = 0.02$ for the traditional quantification using the mean; $r_{(121)} = 0.21$, $p = 0.02$ for the traditional quantification using the median]. The Bayes factor, however, indicated only anecdotal evidence for these associations ($1 < BF_{10} < 3$). The traditional quantification of PES using the median was also found to be positively correlated with the RT on correct trials in 3rd grade, $r_{(92)} = 0.21$, $p = 0.04$, but the Bayes factor indicated no evidence either way ($BF_{10} = 1$). No other significant correlations between PES or PEIA and overall task performance were found.

Moving on to the association between metacognitive monitoring and control (Research question 3), it is important to note that calibration of confidence is typically influenced by task performance (i.e., accurate responses are easier to judge than inaccurate responses, Fleming and Lau, 2014). This was also observable in the current study, as the mean calibration of correct trials was 1.91 ($SD = 0.12$) in 2nd grade and 1.91 ($SD = 0.13$) in 3rd grade, while for error trials this was only 0.82 ($SD = 0.54$) in 2nd grade and 0.93 ($SD = 0.74$) in 3rd grade. We, therefore, controlled for overall accuracy on the arithmetic task when running these correlations. No significant correlations were found between monitoring and post-error adjustments, not in 2nd grade nor in 3rd grade.

3.3 Metacognitive control in working memory

3.3.1 Post-error slowing

The mean and median RTs for post-error and post-correct trials at both time points are shown in Table 3 and Figure 2. One-sample t-tests revealed significant PES in the 2-back task for both traditional quantifications. This was the case in 2nd grade [$t_{(123)} = 10.14$, $p < 0.001$, Cohen's $d = 0.91$ for the traditional quantification with the mean; $t_{(123)} = 9.70$, $p < 0.001$, Cohen's $d = 0.87$ for the traditional quantification with the median] and also in 3rd grade [$t_{(120)} = 8.92$, $p < 0.001$, Cohen's $d = 0.81$ for the traditional quantification with the mean; $t_{(120)} = 9.80$, $p < 0.001$, Cohen's $d = 0.89$ for the traditional quantification with the median]. The Bayes factor indicated decisive evidence for both quantifications in both grades ($BF_{10} > 100$).

Paired t-tests revealed that the magnitude of PES did not change significantly from 2nd to 3rd grade for neither of the quantifications [$t_{(118)} = -0.35$, $p = 0.73$, Cohen's $d = -0.03$ for the traditional quantification with the mean; $t_{(118)} = 0.44$, $p = 0.66$, Cohen's $d = 0.04$ for the traditional quantification with the median]. The bayes factor indicated moderate evidence in favor of the null hypothesis for both quantifications of PES ($0.1 < BF_{10} < 0.33$).

3.3.2 Post-error improvement in accuracy

Accuracy rates of post-error and post-correct trials at both time points are shown in Table 4 and Figure 2. One-sample t-tests revealed no significant PEIA in the 2-back task, for neither of the quantifications of PEIA. To the contrary, for both quantifications, there was a significant decrease in accuracy. This was the case in 2nd grade as well as in 3rd grade [$t_{(123)} = -7.03$, $p < 0.001$, Cohen's $d = -0.63$ for PEIA two trials after the error in 2nd grade; $t_{(123)} = -10.30$, $p < 0.001$, Cohen's $d = -0.92$ for PEIA three trials after the error in 2nd grade; $t_{(120)} = -9.06$, $p < 0.001$, Cohen's $d = -0.82$ for PEIA two trials after the error in 3rd grade; $t_{(120)} = -13.49$, $p < 0.001$, Cohen's $d = -1.23$ for PEIA three trials after the error in 3rd grade]. The Bayes factor indicated decisive evidence for all these effects ($BF_{10} > 100$).

Paired t-tests revealed a significant difference in the magnitude of PEIA 2 trials after an error between 2nd and 3rd grade [$t_{(118)} = 2.48$, $p = 0.02$, Cohen's $d = 0.23$]. However, the Bayes factor indicated only anecdotal evidence for this effect ($BF_{10} = 1.91$). In contrast, no significant difference was found in the magnitude of PEIA 3 trials after an error between 2nd and 3rd grade [$t_{(118)} = 1.11$, $p = 0.27$, Cohen's $d = 0.10$]. The Bayes factor indicated moderate evidence for the null hypothesis ($BF_{10} = 0.19$).

3.3.3 Correlations

Pearson correlation coefficients, controlled for performance on the Raven's Standard Progressive Matrices, were run to assess associations between PES, PEIA, and overall task performance, as well as with PES and PEIA in the arithmetic task. A full correlation matrix can be found in Appendix C. The traditional quantification of PES with the mean was found to be negatively correlated with PEIA 2 trials after an error in 2nd grade [$r_{(121)} = -0.19$,

TABLE 3 Mean/median RT's (in ms) for post-error and post-correct trials on the 2-back task.

	Post-error trials		Post-correct trials	
2nd grade ($n = 124$)	$M = 1{,}294.72$	$SD = 302.784$	$M = 1{,}098.43$	$SD = 231.55$
	$Mdn = 1{,}256.05$	$SD = 332.45$	$Mdn = 1{,}032.98$	$SD = 239.71$
3rd grade ($n = 121$)	$M = 1{,}304.24$	$SD = 330.77$	$M = 1{,}098.16$	$SD = 245.88$
	$Mdn = 1{,}216.73$	$SD = 296.66$	$Mdn = 1{,}011.42$	$SD = 225.22$

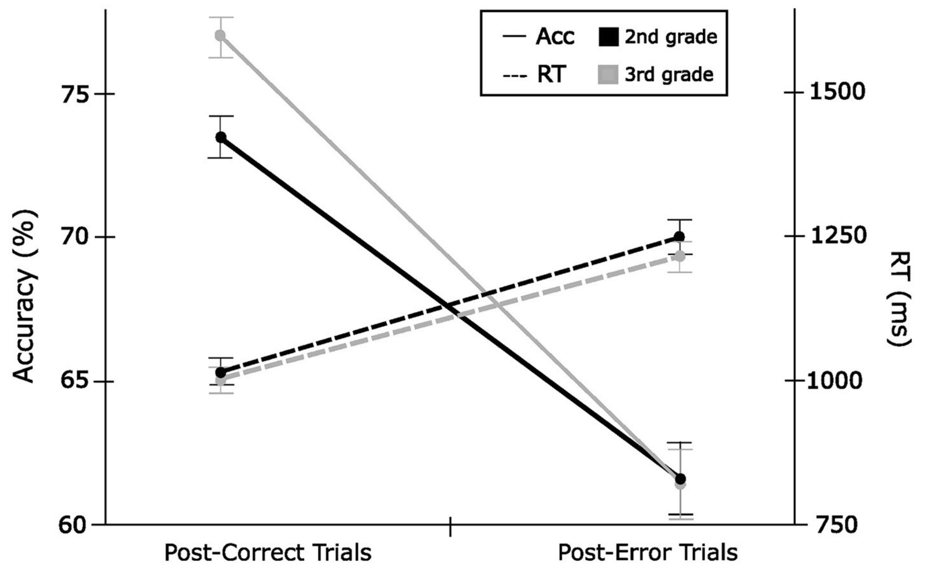


FIGURE 2
Mean accuracy rates and median response times after errors and correct trials for 2nd and 3rd grade in the working memory task.

TABLE 4 Accuracy rates (in %) for post-error and post-correct trials on the 2-back task.

	2 trials post-error		2 trials post-correct	
	M	*SD*	*M*	*SD*
2nd grade ($n = 124$)	61.67	14.48	73.91	8.90
3rd grade ($n = 121$)	61.50	14.44	77.94	8.75
	3 trials post-error		**3 trials post-correct**	
	M	*SD*	*M*	*SD*
2nd grade ($n = 124$)	60.74	11.68	73.62	7.58
3rd grade ($n = 121$)	62.21	9.52	77.13	7.34

$p = 0.03$]. However, the Bayes factor indicated only anecdotal evidence ($BF_{10} = 1.02$). No other significant correlations between the quantifications of PES and the two quantifications of PEIA were found, not in 2nd grade nor in 3rd grade.

Regarding the association between metacognitive control and overall task performance, no significant correlations between PES or PEIA and performance on the 2-back task were found in 2nd grade. Surprisingly, in 3rd grade, the two traditional quantifications of PES and PEIA 2 trials after an error were negatively correlated with performance on the 2-back task [$r_{(118)} = -0.20$, $p = 0.03$ for the traditional quantification of PES using the mean; $r_{(118)} = -0.19$, $p = 0.04$ for the traditional quantification of PES using the median; $r_{(118)} = -0.30$, $p < 0.001$ for PEIA 2 trials after an error]. While the Bayes factor indicated anecdotal evidence for the null hypothesis for the correlations with PES ($0.33 < BF_{10} < 1$), it did reveal strong evidence for the correlation with PEIA ($BF_{10} = 24.37$).

Moving on to the associations between metacognitive control in the working memory domain and metacognitive control in the arithmetic domain (Research question 4), PES in the 2-back task was not significantly correlated with PES in the arithmetic task, not in 2nd grade nor in 3rd grade, except for the traditional quantifications with the median in 2nd grade [$r_{(118)} = 0.20$, $p = 0.03$]. However, the Bayes factor indicated only anecdotal evidence for this association ($BF_{10} = 1.30$). The Bayes factor indicated moderate evidence in favor of the null hypothesis for most of the other correlations, both in 2nd and in 3rd grade ($0.10 < BF_{10} < 0.33$). For the correlation between the traditional quantifications with the mean in 3rd grade, the Bayes factor indicated only anecdotal evidence in favor of the null hypothesis ($0.33 < BF_{10} < 1$). Similarly, neither of the two quantifications of PEIA in the 2-back task were significantly correlated with PEIA in the computerized arithmetic task, not in 2nd grade nor in 3rd grade. The Bayes factor indicated moderate evidence in favor of the null hypothesis for these correlations in both grades ($0.1 < BF_{10} < 0.33$), except for the correlation with PEIA 3 trials after an error in the 2-back task in 3rd grade, for which the Bayes factor indicated only anecdotal evidence in favor of the null hypothesis ($BF_{10} = 0.64$).

4 Discussion

This study longitudinally investigated metacognitive control, operationalized as post-error slowing (PES) and post-error improvement in accuracy (PEIA), in an arithmetic and working memory task in children from 2nd to 3rd grade of primary school. No strong evidence for PES in arithmetic was found, with only the traditional method using the median revealing significant PES effects. The great variability in the RT's between and within the children in the current study might explain why PES was only found when using the median. In contrast, we found strong evidence for PES in the working memory domain. This is a surprising result, as PES has previously been observed in children both in simple conflict tasks (e.g., Dubravac et al., 2022; Smulders et al., 2016), as well as in arithmetic (de Mooij et al., 2022).

There are two points worth mentioning regarding the absence of PES in arithmetic. First, previous studies that found PES in arithmetic, such as de Mooij et al. (2022) in children and Desmet et al. (2012) in adults, included feedback immediately following responses, while this was not the case in our study. One possibility is, therefore, that an external feedback signal after an error is necessary and a driving force for PES, especially in young children, who exhibit a stronger reaction to external feedback than adults (Ferdinand and Kray, 2014). We did, however, observe strong evidence for PES in the working memory task without feedback. Moreover, previous research has also observed PES in tasks without immediate feedback in adults (e.g., Allain et al., 2009; Houtman et al., 2012). Taken together, it is plausible that the necessity of feedback to elicit PES is greater in more complex tasks, such as arithmetic, and is, therefore, task-dependent. Research directly comparing the magnitude of PES between feedback vs. non-feedback conditions in arithmetic is, therefore, needed to gain more insights into the role of feedback in post-error adjustments.

Second, the absence of PES in arithmetic in combination with the presence of PEIA suggests that children might use control mechanisms other than slowing down to control their performance. A possibility could be that children decide to switch to a more effective strategy. Van der Borght et al. (2016) revealed that only adults who repeat the same strategy following an error exhibit PES. Adults that do not slow down after errors are the ones that change strategies after an error and are also the ones that are more accurate on post-error trials. Such patterns of performance might also be observed in children, as children have been shown to be able to select and switch strategies in arithmetic within the same task (Ardiale and Lemaire, 2013; Imbo and Vandierendonck, 2007), yet this has not been examined empirically. Studies investigating this in children are needed to obtain empirical evidence for this hypothesis. In the working memory task, individuals are more limited in post-error behavioral adjustments compared to the arithmetic task. Other than slowing down, there are not many other possibilities to control behavior in this task, which could explain why PES was observed in this domain in contrast to what was found in arithmetic.

Although we did not observe substantial evidence for the presence of PES in arithmetic, we did find strong evidence for PEIA in both grades. This suggests that children control their performance after committing errors, resulting in improved accuracy on the trial following the error. This aligns with de Mooij et al. (2022) who also revealed PEIA in children during mathematical activities. In contrast, many studies investigating PEIA in more simple tasks, such as conflict tasks, failed to observe PEIA (e.g., Hajcak and Simons, 2008; King et al., 2010; Notebaert and Verguts, 2011) or even found a decrease in accuracy on the trials following errors (e.g., Fiehler et al., 2005; Hajcak and Simons, 2008; Notebaert et al., 2009). The latter was also observed in the current study for the working memory domain. This discrepancy suggests that arithmetic allows for more ways to control behavior and improve accuracy following errors than more simple tasks. The fact that we observed a decrease in accuracy in the working memory task suggests that, in this specific task where performance on one trial is partly dependent of performance on another trial, children seem to lose control after committing an error or might be thrown off by their error, resulting in a pattern of subsequent errors after the initial error. However, it is important to note that PEIA measures in tasks where accuracy streaks are task-inherent (e.g., trials depending on each other) should be interpreted with caution (Danielmeier and Ullsperger, 2011).

PES and PEIA were not associated, neither in arithmetic nor in working memory, suggesting that slowing down is not effective in improving accuracy following errors. This is not a surprising finding considering the ongoing debate about the functionality of PES (Danielmeier and Ullsperger, 2011; Notebaert et al., 2009). The absence of an association between PES and PEIA might suggest that PES is not a reflection of cognitive control but rather a reaction to a surprising error prompting the individual to slow down, also referred to as the orienting account (Notebaert et al., 2009). Moreover, this interpretation can be backed-up with empirical evidence, as many studies have observed an absence of PEIA in combination with PES (e.g., Hajcak and Simons, 2008; King et al., 2010; Notebaert and Verguts, 2011) or a lack of association between these two post-error adjustments (e.g., King et al., 2010). What is, however, surprising is that the scarce studies in the domain of arithmetic have found PES and PEIA to be associated (Desmet et al., 2012), even in children (de Mooij et al., 2022). Moreover, most studies support the idea that PES only functions as an orienting response in simple tasks, such as conflict tasks, while it is more likely to reflect cognitive control in more complex tasks, such as arithmetic (e.g., de Mooij et al., 2022; Desmet et al., 2012). It is, therefore, plausible that the orienting account can explain the findings of the current study in the working memory domain, but not in the arithmetic domain. Perhaps more likely for the arithmetic domain is that children do slow down with the goal to control and improve their performance, but slowing down may not be the most effective way to do so. As mentioned previously, Van der Borght et al. (2016) found that switching strategies, rather than slowing down, is a more effective control mechanism in arithmetic in adults and does not necessarily occur in combination with PES.

We did not observe any developmental differences in the magnitude of the post-error adjustments between 2nd and 3rd grade, neither in arithmetic nor in working memory. While the 1-year follow-up period of the current study might be too short to capture any significant changes, as metacognitive regulation and post-error adjustments specifically are presumed to have a longer developmental trajectory (e.g., Dubravac et al., 2022), these findings are surprising. This is because previous research—albeit not all of them in the domain of arithmetic or working memory nor operationalized as post-error adjustments—have depicted the period of 7 until 9 years old as vital in the development of metacognitive control (de Mooij et al., 2022; Krebs and Roebers, 2010; Roebers et al., 2014; Roebers and Spiess, 2017; Selmeczy et al., 2021; Steiner et al., 2020). However, apart from Roebers and Spiess (2017) and Steiner et al. (2020), these studies all encompassed cross-sectional investigations rather than longitudinal ones, which could explain the difference in results with the current study. Moreover, while van Loon et al. (2024) did observe age-related differences in metacognitive control between the ages of 8 until 10 in one out of three tasks, these differences were not apparent in the other two. Steiner et al. (2020) also found age-related developmental differences to differ across tasks, suggesting that developmental differences could be task- and domain-specific.

It is, however, important to note that, while we did observe significant improvement in overall performance from 2nd to 3rd grade, the children in the current study got exactly the same tasks in 2nd grade as in 3rd grade. While other studies, although not in the domain of arithmetic, have also administered the exact same task to children from different ages and found age-related differences (e.g., Roebers and Spiess, 2017), the children in the current study were administered multiplication problems, which are known to go through major developmental progression in the age range under study in the Flemish school system. The 2nd graders in our study were still in a learning phase for multiplication, while the 3rd graders already automatized these multiplications, as is evidenced by the notable differences in accuracy rates and RTs between the two time points on the arithmetic task. More specifically, in 3rd grade we observed ceiling effects for many of the children. This could account for smaller PES and PEIA than what we might observe if the task was adjusted to their skill level. In other words, if the children were administered a task that reflected their increasing skill level, increases in the magnitude of PES and PEIA might have been observed. This hypothesis is strengthened by the study of de Mooij et al. (2022) who investigated PES in an adaptive learning environment and found the magnitude of PES to increase from the age of 6 until 9 years old. Moreover, they found PES to be greater in children that chose the highest difficulty level and, therefore, had the highest error rate.

The current study revealed a small association between PES and the overall accuracy on the arithmetic task in 2nd grade, even after controlling for intellectual ability. No such association was found in 3rd grade. The latter finding is not surprising, given that previous research suggests that metacognitive control, although not operationalized as PES or PEIA, at a young age is not always associated with overall task performance (Ger and Roebers, 2023), and that this association only emerges from the age of 10 years old onwards (van Loon and Oeri, 2023). While de Mooij et al. (2022) found PES to be positively associated with ability level in mathematical activities between the ages of 5 and 13 years old, they did not investigate the influence of age on this association, leaving it unresolved whether this association is present throughout this whole age range or, for example, only in older children. The current study found an association between PES and overall task performance in 2nd grade. Even though no strong conclusions can be drawn from this result, as the Bayes factor only indicated anecdotal evidence, there are two plausible explanations for the difference in results between 2nd and 3rd grade. First, the ceiling effects observed in 3rd grade, as mentioned earlier, could explain the lack of association between PES and overall accuracy. Second, as mentioned before, 2nd graders in Flemish schools are still in the learning phase regarding single-digit multiplication. It could, therefore, be that slowing down following errors helps children learn and memorize the material better, resulting in greater ability and better overall task performance. This hypothesis is strengthened by the findings from de Mooij et al. (2022), who, as mentioned before, found an association with ability level; importantly, they found this association in an adaptive environment that is more focused on learning than performance. Research specifically investigating post-error adjustments in an arithmetic learning protocol could provide more insights into how these behavioral adjustments might support the learning process for new arithmetic problems.

Regarding the working memory domain, the current study revealed a negative association between PEIA two trials after an error and performance on the working memory task, indicating that children who are more accurate two trials after an error perform worse on the task. This is a surprising result considering that PEIA is thought to reflect cognitive control with the goal to increase overall task performance (Danielmeier and Ullsperger, 2011). One possibility is that children might notice the error, resulting in an orientation toward and increased recall of the presented item, which ultimately results in better performance two trials later, as that trial is inherently dependent on the trial two steps before. This could, however, ultimately result in worse task performance overall, as an increased orientation to the error might make it harder for the child to regain focus on the other trials (Notebaert et al., 2009). In this situation, PEIA is, thus, not a reflection of cognitive control but rather an orienting response to the error.

Moving on to the association between metacognitive control and metacognitive monitoring, no significant associations were found. These results align with previous research, as studies suggest that an association between monitoring and control is only just emerging at this age (Hoffmann-Biencourt et al., 2010; Roebers and Spiess, 2017). A hypothesis for the lack of association between monitoring and PES is that other skills, such as executive functions, are needed to slow down after committing an error. Given that these types of skills are still developing in this age group (Diamond, 2013), children who pick up on their errors might not be able to translate this in an immediate control response, resulting in a lack of association between these two skills.

The discussion up until this point made clear that findings regarding post-error adjustments in arithmetic and post-error adjustments in working memory differed. Furthermore, we also found measures of PES and PEIA in arithmetic to be uncorrelated with PES and PEIA in working memory. These results challenge previous evidence for domain-generality of post-error adjustments found in studies by Dubravac et al. (2022) and Ger and Roebers (2023). A possible explanation for this discrepancy in results is that these studies compared different types of conflict tasks that reflect a similar domain, while the current study compared two tasks that reflect two distinct domains. Domain-generality of post-error adjustments might, therefore, only hold evidence when assessing and comparing relatively similar domains. It is, however, important to note that the two tasks used in the current study differed on more characteristics than solely the domain. First, as mentioned earlier, trials are dependent on each other in the 2-back task, while they are independent in the arithmetic task. To account for the dependency of trials in the 2-back task, we calculated PEIA in a different way than in the arithmetic task, raising challenges regarding the interpretation of the lack of association between the two tasks. Second, in contrast to the arithmetic task, PEIA was found to go in the opposite direction in the 2-back task, making the interpretation regarding correlations between the two more complicated. Such differences make it difficult to draw strong conclusions on the domain-generality of post-error adjustments and the results should, therefore, be interpreted with caution. Research investigating and comparing tasks that reflect distinct domains but are similar in task-requirements is, thus, needed.

The findings of this study should be interpreted with knowledge about its limitations, which offer opportunities for future research. First, ceiling effects present in the arithmetic task in 3rd grade might have biased or hidden possible effects and correlations. Future studies investigating post-error adjustments should avoid high accuracy rates by using tasks that reflect the participants' skill level. Second, while the current study provides new insights in the development of post-error adjustments due to its longitudinal nature, it only covered a short developmental period, which might have been insufficient to capture developmental changes. Future longitudinal studies should cover a larger age range to capture the long developmental trajectory that metacognitive control is presumed to have. Third, while the current study provides insights into post-error adjustments in an academic task, the tasks were still to some extent controlled. Although the controlled nature of the tasks is needed to isolate and obtain a clear understanding of metacognitive processes, it should be noted that there are still differences with tasks used in real classroom settings. Finally, the provided hypotheses regarding the underlying mechanisms of PES and PEIA could not be empirically evaluated in the current study. In other words, the design of this study did not allow us to investigate why children slow down or how they manage to improve their accuracy on trials. Moreover, these underlying mechanisms could be different depending on the task. Further research investigating other post-error adjustments, such as strategy switches, in combination with PES and PEIA could provide more insight into the underlying mechanisms.

In summary, the current study provides some evidence for the presence of metacognitive control, as indicated by measures of PES and PEIA, among 7–8-year-old children who were longitudinally followed up until 8–9 years old, both in arithmetic and working memory tasks. Nevertheless, notable distinctions emerged between the two domains, challenging previous evidence for domain-generality of post-error adjustments. Modest associations between metacognitive control and overall task performance in arithmetic were found at 7–8 years old, suggesting a potential adaptive role of post-error adjustments in the learning phase of arithmetic. It is, however, yet to be empirically investigated what the precise underlying mechanisms of the observed post-error adjustments are. Further research is necessary to advance our understanding of metacognitive control in arithmetic.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://osf.io/943va/.

Ethics statement

The studies involving humans were approved by Social and Societal Ethics Committee KU Leuven. 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

EJ: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. EB: Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing – review & editing, Project administration. BD: Funding acquisition, Project administration, Supervision, Writing – review & editing, Conceptualization, Resources, Methodology.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by a WEAVE project G.0041.23 of the Research Foundation Flanders (FWO) and Austrian Research Fund (FWF) and by the Research Fund KULeuven (METH/24/003). Elien Bellon was supported by a post-doctoral fellowship of the Research Foundation Flanders (FWO).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdpys.2024.1424754/full#supplementary-material

References

Allain, S., Burle, B., Hasbroucq, T., and Vidal, F. (2009). Sequential adjustments before and after partial errors. *Psychon. Bull. Rev.* 16, 356–362. doi: 10.3758/PBR.16.2.356

Ardiale, E., and Lemaire, P. (2013). Within-item strategy switching in arithmetic: a comparative study in children. *Front. Psychol.* 4:924. doi: 10.3389/fpsyg.2013.00924

Bellon, E., Fias, W., and De Smedt, B. (2019). More than number sense: the additional role of executive functions and metacognition in arithmetic. *J. Exp. Child Psychol.* 182, 38–60. doi: 10.1016/j.jecp.2019.01.012

Bellon, E., Fias, W., and De Smedt, B. (2020). Metacognition across domains: is the association between arithmetic and metacognitive monitoring domain-specific? *PLoS ONE* 15:e0229932. doi: 10.1371/journal.pone.0229932

Bellon, E., Fias, W., and De Smedt, B. (2021). Too anxious to be confident? A panel longitudinal study into the interplay of mathematics anxiety and metacognitive monitoring in arithmetic achievement. *J. Educ. Psychol.* 113, 1550–1564. doi: 10.1037/edu0000704

Coughlin, C., Hembacher, E., Lyons, K. E., and Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev. Sci.* 18, 957–971. doi: 10.1111/desc.12271

Crone, E. A., and Steinbeis, N. (2017). Neural perspectives on cognitive control development during childhood and adolescence. *Trends Cogn. Sci.* 21, 205–215. doi: 10.1016/j.tics.2017.01.003

Dali, G., Orr, C., and Hester, R. (2022). Error awareness and post-error slowing: the effect of manipulating trial intervals. *Conscious. Cogn.* 98:103282. doi: 10.1016/j.concog.2022.103282

Danielmeier, C., and Ullsperger, M. (2011). Post-error adjustments. *Front. Psychol.* 2, 1–10. doi: 10.3389/fpsyg.2011.00233

de Mooij, S. M., Dumontheil, I., Kirkham, N. Z., Raijmakers, M. E., and van der Maas, H. L. (2022). Post-error slowing: large scale study in an online learning environment for practising mathematics and language. *Dev. Sci.* 25:e13174. doi: 10.1111/desc.13174

De Visscher, A., Vogel, S. E., Reishofer, G., Hassler, E., Koschutnig, K., De Smedt, B., et al. (2018). Interference and problem size effect in multiplication fact solving: individual differences in brain activations and arithmetic performance. *Neuroimage* 172, 718–727. doi: 10.1016/j.neuroimage.2018.01.060

Derrfuss, J., Danielmeier, C., Klein, T. A., Fischer, A. G., and Ullsperger, M. (2022). Unbiased post-error slowing in interference tasks: a confound and a simple solution. *Behav. Res. Methods* 54, 1416–1427. doi: 10.3758/s13428-021-01673-8

Desmet, C., Imbo, I., De Brauwer, J., Brass, M., Fias, W., and Notebaert, W. (2012). Error adaptation in mental arithmetic. *Q. J. Exp. Psychol.* 65, 1059–1067. doi: 10.1080/17470218.2011.648943

Destan, N., Hembacher, E., Ghetti, S., and Roebers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5-to 7-year-old children. *J. Exp. Child Psychol.* 126, 213–228. doi: 10.1016/j.jecp.2014.04.001

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Dubravac, M., Roebers, C. M., and Meier, B. (2022). Age-related qualitative differences in post-error cognitive control adjustments. *Br. J. Dev. Psychol.* 40, 287–305. doi: 10.1111/bjdp.12403

Dutilh, G., Van Ravenzwaaij, D., Nieuwenhuis, S., Van der Maas, H. L., Forstmann, B. U., and Wagenmakers, E. J. (2012). How to measure post-error slowing: a confound and a simple solution. *J. Math. Psychol.* 56, 208–216. doi: 10.1016/j.jmp.2012.04.001

Efklides, A., and Misailidi, P. (2010). *Trends and Prospects in Metacognition Research*. New York, NY: Springer Science + Business Media.

Ferdinand, N. K., and Kray, J. (2014). Developmental changes in performance monitoring: how electrophysiological data can enhance our understanding of error and feedback processing in childhood and adolescence. *Behav. Brain Res.* 263, 122–132. doi: 10.1016/j.bbr.2014.01.029

Fiehler, K., Ullsperger, M., and Von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology* 42, 72–82. doi: 10.1111/j.1469-8986.2005.00265.x

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am. Psychol.* 34:906. doi: 10.1037/0003-066X.34.10.906

Fleming, S. M., and Lau, H. C. (2014). How to measure metacognition. *Front. Hum. Neurosci.* 8:443. doi: 10.3389/fnhum.2014.00443

Gardier, M., and Geurten, M. (2024). The developmental path of metacognition from toddlerhood to early childhood and its influence on later memory performance. *Dev. Psychol.* 60, 1244–1254. doi: 10.1037/dev0001752

Ger, E., and Roebers, C. (2023). Hearts, flowers, and fruits: all children need to reveal their post-error slowing. *J. Exp. Child Psychol.* 226:105552. doi: 10.1016/j.jecp.2022.105552

Geurten, M., Meulemans, T., and Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cogn. Dev.* 48, 62–81. doi: 10.1016/j.cogdev.2018.08.002

Gupta, R., Kar, B. R., and Srinivasan, N. (2009). Development of task switching and post-error-slowing in children. *Behav. Brain Funct.* 5, 1–13. doi: 10.1186/1744-9081-5-38

Hajcak, G., and Simons, R. F. (2008). Oops!.. I did it again: an ERP and behavioral study of double-errors. *Brain Cognit.* 68, 15–21. doi: 10.1016/j.bandc.2008.02.118

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., and Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *Br. J. Dev. Psychol.* 28, 767–784. doi: 10.1348/026151009X479042

Houtman, F., Castellar, E. N., and Notebaert, W. (2012). Orienting to errors with and without immediate feedback. *J. Cogn. Psychol.* 24, 278–285. doi: 10.1080/20445911.2011.617301

Imbo, I., and Vandierendonck, A. (2007). The development of strategy use in elementary school children: working memory and individual differences. *J. Exp. Child Psychol.* 96, 284–309. doi: 10.1016/j.jecp.2006.09.001

Imbo, I., and Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychol. Res.* 72, 331–346. doi: 10.1007/s00426-007-0112-8

JASP Team (2024). *JASP (Version 0.19.0)* [Computer software]. Available at: https://jasp-stats.org/

King, J. A., Korb, F. M., von Cramon, D. Y., and Ullsperger, M. (2010). Post-error behavioral adjustments are facilitated by activation and suppression of task-relevant and task-irrelevant information processing. *J. Neurosci.* 30, 12759–12769. doi: 10.1523/JNEUROSCI.3274-10.2010

Krebs, S. S., and Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *Br. J. Educ. Psychol.* 80, 325–340. doi: 10.1348/000709910X485719

Nelson, T. O., and Narens, L. (1990). "Metamemory: a theoretical framework and some new findings," in *The Psychology of Learning and Motivation*, ed. G. H. Bower (New York, NY: Academic Press), 125–173.

Niebaum, J. C., Chevalier, N., Guild, R. M., and Munakata, Y. (2021). Developing adaptive control: age-related differences in task choices and awareness of proactive and reactive control demands. *Cogn. Affect. Behav. Neurosci.* 21, 561–572. doi: 10.3758/s13415-020-00832-2

Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., and Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition* 111, 275–279. doi: 10.1016/j.cognition.2009.02.002

Notebaert, W., and Verguts, T. (2011). Conflict and error adaptation in the Simon task. *Acta Psychol.* 136, 212–216. doi: 10.1016/j.actpsy.2010.05.006

O'Leary, A. P., and Sloutsky, V. M. (2017). Carving metacognition at its joints: Protracted development of component processes. *Child Dev.* 88, 1015–1032. doi: 10.1111/cdev.12644

O'Leary, A. P., and Sloutsky, V. M. (2019). Components of metacognition can function independently across development. *Dev. Psychol.* 55, 315–328. doi: 10.1037/dev0000645

Pelegrina, S., Lechuga, M. T., García-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras, M., et al. (2015). Normative data on the n-back task for children and young adolescents. *Front. Psychol.* 6:1544. doi: 10.3389/fpsyg.2015.01544

Raven, C. J., Court, J. H., and Raven, J. (1992). *Standard Progressive Matrices*. Oxford: Oxford Psychologist Press.

Rinne, L. F., and Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: calibration of confidence predicts the development of accuracy. *PLoS ONE* 9:e98663. doi: 10.1371/journal.pone.0098663

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

Roebers, C. M., Krebs, S. S., and Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: their interrelations and their role for test performance. *Learn. Individ. Differ.* 29, 141–149. doi: 10.1016/j.lindif.2012.12.003

Roebers, C. M., and Spiess, M. (2017). The development of metacognitive monitoring and control in second graders: a short-term longitudinal study. *J. Cognit. Dev.* 18, 110–128. doi: 10.1080/15248372.2016.1157079

Schachar, R. J., Chen, S., Logan, G. D., Ornstein, T. J., Crosbie, J., Ickowicz, A., et al. (2004). Evidence for an error monitoring deficit in attention deficit hyperactivity disorder. *J. Abnorm. Child Psychol.* 32, 285–293. doi: 10.1023/B:JACP.0000026142.11217.f2

Schneider, W., and Artelt, C. (2010). Metacognition and mathematics education. *ZDM Int. J. Math. Educ.* 42, 149–161. doi: 10.1007/s11858-010-0240-2

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime Computer Software and Manual.* Pittsburgh, PA: Psychology Software Tools.

Selmeczy, D., Ghetti, S., Zheng, L. R., Porter, T., and Trzesniewski, K. (2021). Help me understand: adaptive information-seeking predicts academic achievement in school-aged children. *Cogn. Dev.* 59:101062. doi: 10.1016/j.cogdev.2021.101062

Smulders, S. F., Soetens, E., and van der Molen, M. W. (2016). What happens when children encounter an error?. *Brain Cogn.* 104, 34–47. doi: 10.1016/j.bandc.2016.02.004

Souchay, C., Isingrini, M., Pillon, B., and Gil, R. (2003). Metamemory accuracy in Alzheimer's disease and frontotemporal lobe dementia. *Neurocase* 9, 482–492. doi: 10.1076/neur.9.6.482.29376

Steiner, M., van Loon, M. H., Bayard, N. S., and Roebers, C. M. (2020). Development of children's monitoring and control when learning from texts: effects of age and test format. *Metacognit. Learn.* 15, 3–27. doi: 10.1007/s11409-019-09208-5

Touron, D. R., Oransky, N., Meier, M. E., and Hines, J. C. (2010). Metacognitive monitoring and strategic behaviour in working memory performance. *Q. J. Exp. Psychol.* 63, 1533–1551. doi: 10.1080/17470210903418937

Van der Borght, L., Desmet, C., and Notebaert, W. (2016). Strategy changes after errors improve performance. *Front. Psychol.* 6:2051. doi: 10.3389/fpsyg.2015.02051

van Loon, M., Orth, U., and Roebers, C. (2024). The structure of metacognition in middle childhood: evidence for a unitary metacognition-for-memory factor. *J. Exp. Child Psychol.* 241:105857. doi: 10.1016/j.jecp.2023.105857

van Loon, M. H., and Oeri, N. S. (2023). Examining on-task regulation in school children: interrelations between monitoring, regulation, and task performance. *J. Educ. Psychol.* 115, 446–459. doi: 10.1037/edu0000781

Veenman, M. V., and Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: age and task differences. *Learn. Individ. Differ.* 15, 159–176. doi: 10.1016/j.lindif.2004.12.001

frontiers | Frontiers in Developmental Psychology

TYPE Mini Review
PUBLISHED 23 September 2024
DOI 10.3389/fdpys.2024.1467880

OPEN ACCESS

EDITED BY
Catherine Sandhofer,
University of California, Los Angeles,
United States

REVIEWED BY
Elena Escolano-Pérez,
University of Zaragoza, Spain

*CORRESPONDENCE
Seokyung Kim
✉ kim01426@umn.edu

RECEIVED 21 July 2024
ACCEPTED 26 August 2024
PUBLISHED 23 September 2024

CITATION
Kim S and Carlson SM (2024) Understanding explore-exploit dynamics in child development: current insights and future directions. *Front. Dev. Psychol.* 2:1467880.
doi: 10.3389/fdpys.2024.1467880

Understanding explore-exploit dynamics in child development: current insights and future directions

Seokyung Kim* and Stephanie M. Carlson

Institute of Child Development, University of Minnesota, Minneapolis, MN, United States

Examining children's decisions to explore or exploit the environment provides a window into their developing metacognition and reflection capacities. Reinforcement learning, characterized by the balance between exploring new options (exploration) and utilizing known ones (exploitation), is central to this discussion. Children initially exhibit broad and intensive exploration, which gradually shifts toward exploitation as they grow. We review major theories and empirical findings, highlighting two main exploration strategies: random and directed. The former involves stochastic choices without considering information or rewards, while the latter is driven by reducing uncertainty for information gain. Behavioral tasks such as n-armed bandit, horizon, and patch foraging tasks are used to study these strategies. Findings on the n-armed bandit and horizon tasks showed mixed results on whether random exploration decreases over time. Directed exploration consistently decreases with age, but its emergence depends on task difficulty. In patch-foraging tasks, adults tend to overexploit (staying too long in one patch) and children overexplore (leaving too early), whereas adolescents display the most optimal balance. The paper also addresses open questions regarding the mechanisms supporting early exploration and the application of these strategies in real-life contexts like persistence. Future research should further investigate the relation between cognitive control, such as executive function and metacognition, and explore-exploit strategies, and examine their practical implications for adaptive learning and decision-making in children.

KEYWORDS

reinforcement learning, explore-exploit dynamics, executive function, metacognition, child development

When a child is born, the world around them is new and unpredictable. However, they gradually learn about their environment through contingency, forming associations between their behaviors and either positive or negative consequences, and start to use these contingencies to guide their future behaviors. This type of learning is known as reinforcement learning (e.g., Nussenbaum and Hartley, 2019). For example, infants as young as 2 months old quickly increase their kicking behavior in an experiment where a ribbon is attached to their ankle and connected to a mobile hanging overhead (Rovee-Collier, 1997). This behavior occurs because they explore the object attached to their ankle and learn the associations between their leg movements and the mobile's movements. In the beginning, this kind of *exploration* aims at improving and expanding knowledge. However, choosing whether and when to explore is a genuinely complex decision, as more options become available, varying in value. For instance, if the infants also were given an attractive toy to grasp, they could explore the new toy for potential enjoyment or continue playing with the mobile, which already provides them with joy. As children grow, they face decisions ranging from trivial ones, such as what to eat for dinner or where to play, to more significant ones, such as whether to go to college and whom to be friends with. In such situations, they must either search for better options (explore) or utilize their known options (exploit). Developmental psychologists actively research how children balance the competing demands of exploration and exploitation when faced with two or more options, yet much is still unknown.

In this paper, we aim to review the major theories and empirical findings regarding explore-exploit strategies and how they shift across development. Indeed, young children do explore intensively and broadly, often at the cost of exploitation, and the exploration tendency decreases with age (see Gopnik, 2020, and Nussenbaum and Hartley, 2019 for review). Below, we will overview the definition of exploration and exploitation, the explore-exploit tradeoff/dilemma, and one optimal solution. Next, we will summarize exploration across development in the reinforcement learning literature. Finally, we will highlight directions for future explore-exploit developmental research, with a focus on its potential to advance our understanding of executive function, metacognition, and reflection.

While this is not a systematic review, our methodology is consistent with utilizing PsychInfo and Google Scholar as primary sources. The search was conducted using the following keywords: 1. explore-exploit; development, 2. explore-exploit; development; and task-specific terms (e.g., bandit, horizon, patch-foraging), 3. exploration; reinforcement learning, 4. exploration; reinforcement learning; and task-specific terms (e.g., bandit, horizon, patch-foraging). The search was restricted to articles published between 2010 and 2024, with exceptions made for seminal articles that introduce key concepts, focusing on studies involving human participants from infancy through early adulthood (see Table 1).

Key concepts in explore-exploit learning

Exploration involves experimenting with various options and is typically favored under conditions of low knowledge and high uncertainty (Daw et al., 2006). Conversely, *exploitation* involves adhering to the most lucrative option to maximize rewards and is typically favored under conditions of high knowledge and low uncertainty. Exploration and exploitation represent endpoints along a spectrum–ranging from broad to narrow, noisy to efficient, and information-seeking to reward-seeking–rather than a strict dichotomy (Frankenhuis and Gopnik, 2023). An explore-exploit tradeoff naturally occurs in the decision-making process because choosing to seek new information (exploration) means forgoing an opportunity to choose a familiar option and secure a known reward (exploitation). This dilemma is prevalent not only in human lives but also across the animal kingdom and within society in general (Cohen et al., 2007; Hills et al., 2015; Mehlhorn et al., 2015).

One strategy used by organisms, including humans, animals, and machines to tackle the explore-exploit dilemma, is *balancing exploration and exploitation*. This balance refers to initially preferring exploration and gradually transitioning toward exploitation (Cohen et al., 2007; Hills et al., 2015; Mehlhorn et al., 2015). Exploration is prioritized at the onset of the learning process and diminishes over time as the agent accumulates knowledge and reduces uncertainty (Auer, 2002). This pattern is sensible for two reasons, according to Gopnik (2020). Firstly, agents cannot effectively exploit the reward structure of their environment until they have sufficiently explored it. As agents learn more, it becomes more rational to rely on existing knowledge and reduce the drive to acquire new information. Secondly, if there is a limited timeframe to solve a task, as time passes, there are fewer chances to leverage the information acquired through exploration. There is substantial empirical evidence to believe that explore-first and exploit-later strategies may be embodied in our typical developmental trajectories.

Exploration and exploitation across development

There is considerable evidence of children's increased exploration during play in their early years (e.g., Bonawitz et al., 2012; Doan et al., 2020; Golinkoff et al., 2006; Schulz and Bonawitz, 2007). However, this paper focuses specifically on reinforcement learning literature, as it provides the most compelling evidence of developmental transitions in exploration, explicitly showing adaptive decision-making with age (Table 1). Two major exploration strategies are *random exploration* and *directed exploration*. Random exploration follows a stochastic choice policy, without considering information or rewards (Giron et al., 2023; Meder et al., 2021). Directed exploration, on the other hand, is driven by a strong desire to gain information and resolve high uncertainty (Giron et al., 2023; Meder et al., 2021; Schulz et al., 2019). Although they are conceptually distinct (Wilson et al., 2014), with dissociable neural signatures (Zajkowski et al., 2017), random and directed exploration are not mutually exclusive. For example, systematic switching in random exploration appears to approximate directed exploration. Behavioral tasks used to study exploration strategies include n-armed bandit tasks (e.g., Gittins and Jones, 1979; Speekenbrink, 2022), horizon tasks (e.g., Wilson et al., 2014), and patch-foraging tasks (e.g., Charnov, 1976; Lloyd et al., 2023).

An n-armed bandit task is like a slot machine with multiple levers. In a 4-armed bandit task, individuals choose from four options, receive feedback on the reward, and make the next selection (Daw et al., 2006). They must balance between exploiting the highest-value option and exploring others to confirm that the known highest-value option remains the best choice. The reward probability can stay constant or change over time (Speekenbrink, 2022). Studies using n-armed bandit tasks have mixed results on whether the randomness of choices decreases over time. Using a spatially correlated multi-armed bandit task (where rewards of different options are correlated to their spatial proximity, meaning that close-distance options have similar reward probabilities), a study comparing 6- and 8-year-olds found high levels of random exploration only in the 6-year-olds group, suggesting a decline in random exploration by middle childhood (Meder et al., 2021). Similar results were observed in a broad age range of participants from 5 to 55 years old, showing a

TABLE 1 Summary of developmental explore/exploit research findings.

Explore-exploit (EE) tasks	Reference	*N*	Age(s)	Specific EE measure	EE results
N-armed bandit tasks	Daw et al., 2006	14	Adults	Four-armed bandit task with dynamic rewards	• Brain regions for exploratory and exploitative decisions were identified - Exploratory decisions: frontopolar cortex, intraparietal sulcus - Exploitative decisions: striatum, ventromedial prefrontal cortex
	Meder et al., 2021	102	• 4–9 years 54 Younger children: 6 years ($M = 72.6$ months) 48 Older children: 8 years ($M = 93.1$ months)	Spatially correlated multi-armed bandit task	• Random exploration decreases with age • Directed exploration was found at all ages, decreasing slightly with age
	Giron et al., 2023	281	• 5–55 years ($M = 14.46$ years)	Spatially correlated multi-armed bandit task	• Random exploration decreases with age • Directed exploration decreases with age
	Schulz et al., 2019	160	• 55 Younger children: 7–8 years ($M = 7.53$ years) • 55 Older children: 9–11 years ($M = 9.95$ years) • 50 Adults: 18–55 years ($M = 33.76$ years)	Spatially correlated multi-armed bandit task	• No reliable differences in random exploration between age groups • Directed exploration decreases with age
	Blanco and Sloutsky, 2020	218	• 110 Children: 48–67 months ($M = 57$ months) • 108 Adults: 18–29 years ($M = 19$ years)	Simplified 4-armed bandit task with static rewards	• Children switched between options more frequently than adults, which characterizes their systematic exploration, although salience disrupted this pattern • Adults showed consistent exploitation
	Blanco and Sloutsky, 2021	139	• Experiment 1 32 Children: 4 years ($M = 54.8$ months) 34 Adults • Experiment 2 36 Children: 4–5 years ($M = 58.9$ months) 37 Adults	Simplified 4-armed bandit task with static rewards	• Experiment 1: - Children showed high levels of systematic exploration - Adults maximized rewards through exploitation • Experiment 2: - Children's exploration was influenced by uncertainty: some preferred a hidden option with an unknown reward, while others actively avoided it
	Blanco and Sloutsky, 2024	214	• 188 Children: 38 years ($M = 64$ months) • 26 Adults: 18–21 years ($M = 19$ years)	Simplified 4-armed bandit task with static rewards	• Exploration decreases with age • Children predominantly explore, with even 3- to 4-year-olds systematically avoiding repeated choices • Adults predominantly exploit the highest reward option
	Wu et al., 2018	241	• Experiment 1 81 Adults: 22–44 years ($M = 33$ years) • Experiment 2 80 Adults: 23–41 years ($M = 32$ years) • Experiment 3 80 Adults: 25–45 years ($M = 35$ years)	Spatially correlated multi-armed bandit task	• Experiments 1, 2, 3: - Adults balanced exploration and exploitation, achieving higher rewards by sampling locally and using generalization in spatially correlated environments through Gaussian process function learning and an optimistic upper confidence bound sampling strategy

(Continued)

TABLE 1 (Continued)

Explore-exploit (EE) tasks	Reference	*N*	Age(s)	Specific EE measure	EE results
Horizon tasks	Wilson et al., 2014	31	• Adults: 18–24 years (*M* = 19.7 years)	Horizon task with short and long horizons	• Adults showed more random and directed exploration in long horizons than in short horizons
	Somerville et al., 2017	147	• 12–28 years	Horizon task with short and long horizons	• No reliable differences in the strategic use of random exploration across ages • Strategic use of directed exploration emerges in adolescence and stabilizes into adulthood - The age difference is partly because adolescents favor immediate rewards over new information
	Zhuang et al., 2023	132	• 43 Younger children: 4–5 years (*M* = 5.5 years) • Older children: 11–12 years (*M* = 11.5 years) • 49 Adults: 18–31 years (*M* = 19.4 years)	Simplified horizon task with short, long, and ambiguous horizons	• Adaptation to time horizons increased with age • Adult levels of adaptation are evident by ages 11–12, but not at ages 5–6. • Under short and ambiguous horizons, older children and adults exploited, while younger children did not
Patch foraging tasks	Constantino and Daw, 2015	52	• Experiment 1A 11 Adults (19–35 years) • Experiment 1B 11 Adults (19–35 years) • Experiment 2 30 Adults (19–35 years)	Virtual apple patch-foraging task with varying travel times and depletion rates	• Adults adapted their foraging behavior to environmental changes but tended to overharvest as well • Trial-by-trial decisions were better explained by the marginal value theorem than by temporal-difference learning
	Harms et al., 2024	121	• 62 Early adolescents: 10–13 years (*M* = 11.1 years) • 59 Young adults: 18–32 years (*M* = 19.3 years)	Orchard Task, Grid Task, Chain Task, Horizon Task	• Orchard task: More exploration in shorter travel time conditions; No reliable differences between age groups (Both overexplored, leaving a patch earlier than optimal for reward maximization) • Grid task: Early adolescents explored more than adults • Chain task: No reliable differences between age groups • Horizon task: A more random and directed exploration in longer horizon tasks; Early adolescents showed less directed exploration than adults
	Lloyd et al., 2021	137	• 68 Adolescents: 16–17 years (*M* = 16.57 years) • 69 Adults: 21–50 years (*M* = 30.77 years)	Virtual apple patch-foraging task with varying travel times and depletion rates	• Adolescents explored more than adults - Adolescents accumulated more rewards (though not statistically significant) - Adolescents explored more optimally (i.e., leave a patch at the right time for maximizing rewards) than adults who overexploited (i.e., overharvest in a patch longer than optimal for reward maximization)
	de Liaño et al., 2022	279	• 179 Children from junior kindergarten and elementary school • 67 Adolescents from middle and high school • 33 University college students	Hybrid visual foraging task	• Optimal quitting behavior improves with age • 4–5-years-olds quit slightly earlier

decrease in random exploration with age (Giron et al., 2023). These findings support the "cooling off" theory (Gopnik, 2020), drawing an analogy from statistical physics (Kirkpatrick et al., 1983). Random exploration is likened to a "higher-temperature" (noisier) search, and the "cooling off" process is likened to a simulated annealing algorithm. Just as heating and cooling metal strengthens its structure, children—naïve learners—begin with broad, "high temperature" exploration to avoid local optima and gradually shift to narrow, "low temperature" exploitation by reducing randomness. However, other studies reported no significant differences between children and adults in the amount of random exploration (Schulz et al., 2019) or found that children's exploration is even "systematic" from a young age. In a simplified 4-armed bandit task, Blanco and Sloutsky (2020, 2021, 2024) found that 3–4-year-old children frequently switched their responses and specifically prioritized choosing options they had visited the least recently, making their exploration pattern systematic. These findings may suggest that children are engaging in uncertainty-based directed exploration.

Unlike random exploration, there is more consensus that directed exploration decreases across ages. Relative to adults, children have a bias toward directed exploration and sample options with an intrinsic goal of maximizing the information gain. In a simplified 4-armed bandit task, 4-year-old children preferred options with hidden rewards over visually explicit ones, although there was significant variability within the group (Blanco and Sloutsky, 2021). Using a spatially correlated multi-armed bandit task, studies with children ages 4 to 11 showed higher levels of directed exploration than adults (Meder et al., 2021; Schulz et al., 2019; Wu et al., 2018). For individuals implementing directed exploration, obtaining information is inherently rewarding, and the exploration is encouraged by an information bonus (Auer, 2002).

It is important to note, however, that in n-armed bandit tasks and similar explore-exploit tasks, there is a reward-information confound, making it hard to distinguish between random and directed exploration. Participants only receive feedback on their chosen options and often select the rewarding options to maximize their rewards. This results in an abundance of information about rewarding options, obscuring whether participants' choices were random or aimed at reducing uncertainty. To address this concern, novel tasks like the horizon task have been developed (Wilson et al., 2014). A horizon task is a 2-armed bandit task that includes initial forced-choice trials revealing information about one bandit, followed by free-choice trials where participants choose between two bandits. This design clearly parses between random and directed exploration by removing reward-information confounds in forced-choice trials and manipulating the number of free-choice trials with varying time horizons (e.g., one free-choice trial for a short horizon vs. six for a long horizon).

Several studies have used horizon tasks to investigate how individuals strategically use random and directed exploration. In strategic learning, individuals should select the option with lower means of rewards across trials and the uncertain option more often in the long horizon than in the short horizon. This is because, on the long horizon, individuals have more opportunities to utilize the rewards they explored and learned.

Concerning how this strategic use matures with age, the existing literature does not clearly indicate *when* children start to show the adult level of mature adaptation to the time horizon or strategic uses of random and directed exploration based on the utility of the environment. Adults increased both directed exploration (by choosing the uncertain option) and random exploration (by choosing the lower-mean option) in the long horizon relative to the short horizon (Wilson et al., 2014). However, adolescents were less flexible in guiding their exploration based on the horizon length, often choosing less uncertain options in the long horizon and preferring high-mean options instead (Somerville et al., 2017). This behavior suggests adolescents value immediate rewards more than new information that holds potential long-term benefits. No age-related changes in random exploration were observed. While Somerville et al. (2017) reported 12-year-olds did not exhibit mature adaptation like adults, another study using a simplified horizon task found that adult-like adaptation can be acquired by ages 11–12, but not at ages 5–6 years old (Zhuang et al., 2023).

The last explore-exploit behavior task is a patch foraging task (e.g., Orchard task in Constantino and Daw, 2015; Harms et al., 2024; Lloyd et al., 2021), which simulates the animal foraging scenario where an individual must decide how long to exploit a resource patch (e.g., a bush with apples) before exploring a new one (Lloyd et al., 2023). As time spent in a patch increases, the resources (apples) become scarcer. Moving to a new patch incurs time costs, and so during the limited time, the best strategy is to optimize harvest per patch. The marginal value theorem (MVT) suggests that the optimal time to explore new patches is when the expected rewards from the current patch drop below the background reward rate, or the average reward rate of the environment.

In patch foraging tasks, exploration decreases from childhood through adulthood (Lloyd et al., 2023). As children grow, they become adept at adjusting their foraging behavior to the environment's richness, aligning with MVT (Lloyd et al., 2023). Adolescents and adults explore more in richer environments and exploit patches more in poorer ones (Lloyd et al., 2023). In some foraging tasks, mature "leaving" even emerges as early as age 6, indicating the early development of optimal threshold identification (de Liaño et al., 2022). However, in classic patch foraging tasks like the Orchard task, middle adolescence seems to be the peak period for optimal foraging behavior. Early adolescents around 11 years old and young adults aged 19 displayed more exploration by leaving earlier than was optimal for reward maximization (Harms et al., 2024). In contrast, using a similar task, 16–17-year-old middle adolescents explored more than adults aged 30 (Lloyd et al., 2021), whereas adults tended to overexploit patches, showing suboptimal performance (Constantino and Daw, 2015). Middle adolescents' optimal-like foraging, garnering more rewards compared to adults, contrasts with the "cooling off" theory, which posits that adults should be more effective at acquiring rewards. Researchers attribute adults' overexploitation to their risk sensitivity, placing too much value on immediate rewards

(Constantino and Daw, 2015). Adolescents' reduced aversion to ambiguity may explain their greater exploration and faster adaptation to new environments (Conley and Baskin-Sommers, 2023).

Open questions and future directions

We have reviewed key literature on the dynamics of exploration and exploitation from the preschool period through adulthood. In this section, we highlight two significant questions that remain underexplored and suggest directions for future research.

The first question is: *How can explore-exploit strategies be studied in relation to more real-life contexts characterized by uncertainty, complex reward structures, and constraints on time, money, and effort? One* relevant context is persistence. Traditionally, the persistence literature has focused on whether individuals persist by repeating the same action until achieving a goal or quitting (e.g., Leonard et al., 2017, 2020, 2021). Recent studies have begun to view persistence as a dynamic process, incorporating the temporal-behavioral aspects of persistence (Lucca et al., 2020; Oeri et al., 2020, 2024; Wang and Bonawitz, 2022). For example, Wang and Bonawitz (2022) found that preschoolers quit difficult tasks, especially when the likelihood of reward is low, suggesting that they strategically use explore-exploit strategies by considering task difficulty and reward probabilities, when they adjust their persistence. In our own work, Kim et al. (2024) investigated explore-exploit strategies in a novel persistence task that was age-appropriate but challenging to achieve the goal (catching pretend fish in ponds with diminishing rewards). Using latent class analysis, we found that children aged 3–7 used three different strategies when persisting toward a goal: exploration-dominant, exploitation-dominant, and balanced. The ability to balance exploration and exploitation did not emerge until around age 6. The balanced approach was interpreted as the most adaptive strategy, revealed by this more dynamic approach to task analysis as opposed to simply capturing persisting vs. quitting. Incorporating explore-exploit strategies in studying persistence dynamics is promising, and more research is anticipated in this area.

The second question is: *What are the underlying mechanisms that support young children's intensive and broad exploration in their early lives and their shift to more strategic exploration? One* possible mechanism is children's intrinsic motivation to explore. A study by Liquin and Gopnik (2022) supports the idea that children's heightened exploration tendencies are primarily driven by their strong motivation to explore. The authors tested whether the differences in exploration between children and adults were due to differences in their initial beliefs about the environment—assumptions about which options will be rewarding or costly—or motivational differences. Their findings showed no significant differences in initial beliefs between children and adults, indicating that the differences in exploration were derived from motivation. In a follow-up study, when the same hints about the environment were given, both children and adults made similar inferences, further supporting the motivational account.

Another mechanism could be the development of cognitive control, including executive function and metacognition skills, which are essential for problem-solving (Marulis and Nelson, 2021). Exploration is often described as a complex process, as it demands several situational factors that individuals need to take into account prior to exploration, such as ambiguity, expected value of options, and information gains (Lapidow and Bonawitz, 2023; Le Heron et al., 2020). Optimizing exploration requires integrating cognitive processes, such as causal learning (Bonawitz et al., 2012, 2014), reward-based learning (Wittmann et al., 2023), and executive function/metacognition (Badre et al., 2012; Lee and Carlson, 2015; Otto et al., 2013). The protracted development of explore-exploit strategies, with a late shift from predominant exploration to goal-directed decision-making with more exploitation, may be due to the prolonged maturation of executive function and metacognition (O'Leary and Sloutsky, 2017; Roebers, 2017; Zelazo and Carlson, 2012). However, researchers found that even young children (ages 3–4) can show systematic exploration despite immature top-down regulation, which may be possible via bottom-up regulation of broad attention distribution (Blanco and Sloutsky, 2020, 2021). One study even reported no associations between proactive control and strategic exploration adapted to time horizons (Zhuang et al., 2023). In contrast, in the persistence study mentioned earlier, we found that children aged 3–7 with better executive function skills and metacognitive awareness in post-task interviews tended to balance their exploration and exploitation strategies more effectively in the context of diminishing rewards, even after controlling for age (Kim et al., 2024). We reasoned that children who reflected on the task as it unfolded were better able to monitor and control their strong urge to explore novel options. Since persistence aims at achieving a goal, future studies could examine how to foster younger children's adaptive persistence decision-making by helping them reflect upon their performance and learn flexibility in their thinking process, determining when to keep going and when to change their goals or strategies. As current findings are mixed, however, more research is needed to investigate the relations between explore-exploit strategies and cognitive control.

Conclusion

In conclusion, the dynamics of exploration and exploitation throughout child development is a complex interplay between the desire to seek new information and the need to take advantage of known rewards. Children's exploration is influenced by their intrinsic motivation to explore, and they become more balanced in strategy use with age and the development of cognitive control skills. Understanding these developmental trajectories not only deepens our knowledge but also has practical implications for parenting and educational interventions aimed at fostering adaptive learning and decision-making skills. Future research should continue to examine the underlying mechanisms that support children's exploration and drive the transitions with ages and examine how explore-exploit strategies can be applied to real-life situations, ultimately helping children achieve their goals effectively.

Author contributions

SK: Writing – original draft, Writing – review & editing. SC: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* 3, 397–422.

Badre, D., Doll, B. B., Long, N. M., and Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73, 595–607. doi: 10.1016/j.neuron.2011.12.025

Blanco, N. J., and Sloutsky, V. M. (2020). Attentional mechanisms drive systematic exploration in young children. *Cognition* 202:104327. doi: 10.1016/j.cognition.2020.104327

Blanco, N. J., and Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Dev. Sci.* 24:e13026. doi: 10.1111/desc.13026

Blanco, N. J., and Sloutsky, V. M. (2024). Exploration, exploitation, and development: Developmental shifts in decision-making. *Child Dev.* 95, 1287–1298. doi: 10.1111/cdev.14070

Bonawitz, E., Denison, S., Gopnik, A., and Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cogn. Psychol.* 74, 35–65. doi: 10.1016/j.cogpsych.2014.06.003

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

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* 9, 129–136. doi: 10.1016/0040-5809(76)90040-X

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 933–942. doi: 10.1098/rstb.2007.2098

Conley, M. I., and Baskin-Sommers, A. (2023). Development in uncertain contexts: An ecologically informed approach to understanding decision-making during adolescence. *Cogn Affect Behav Neurosci.* 23, 739–745. doi: 10.3758/s13415-023-01067-7

Constantino, S. M., and Daw, N. D. (2015). Learning the opportunity cost of time in a patch- foraging task. *Cogn Affect Behav Neurosci.* 15, 837–853. doi: 10.3758/s13415-015-0350-y

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879. doi: 10.1038/nature04766

de Liaño, B. G. G., Munoz-Garcia, A., Pérez-Hernández, E., and Wolfe, J. M. (2022). Quitting rules in hybrid foraging search: From early childhood to early adulthood. *Cogn. Dev.* 64:101232. doi: 10.1016/j.cogdev.2022.101232

Doan, T., Castro, A., Bonawitz, E., and Denison, S. (2020). "Wow, I did it!": Unexpected success increases preschoolers' exploratory play on a later task. *Cogn. Dev.* 55:100925. doi: 10.1016/j.cogdev.2020.100925

Frankenhuis, W. E., and Gopnik, A. (2023). Early adversity and the development of explore–exploit tradeoffs. *Trends Cogn. Sci.* 27, 616–630. doi: 10.1016/j.tics.2023.04.001

Giron, A. P., Ciranka, S., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., et al. (2023). Developmental changes in exploration resemble stochastic optimization. *Nat. Human Behav.* 7, 1955–1967. doi: 10.1038/s41562-023-01662-1

Gittins, J. C., and Jones, D. M. (1979). A Dynamic Allocation Index for the Discounted Multiarmed Bandit Problem. *Biometrika* 66, 561–565. doi: 10.1093/biomet/66.3.561

Golinkoff, R. M., Hirsh-Pasek, K., and Singer, D. G. (2006). "Why play = learning: A challenge for parents and educators," in Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth, eds. D. G. Singer, R. M. Golinkoff, and K. Hirsh-Pasek (Oxford: Oxford University Press), 3–12.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375:20190502. doi: 10.1098/rstb.2019.0502

Harms, M. B., Xu, Y., Green, C. S., Woodard, K., Wilson, R., and Pollak, S. D. (2024). The structure and development of explore-exploit decision making. *Cogn. Psychol.* 150:101650. doi: 10.1016/j.cogpsych.2024.101650

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., and Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci.* 19, 46–54. doi: 10.1016/j.tics.2014.10.004

Kim, S., Berry, D., and Carlson, S. M. (2024). Should I stay or should I go? Children's persistence in the context of diminishing rewards. *Dev. Sci.* (Minor revision under review).

Kirkpatrick, S., Gelatt Jr, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* 220, 671–680. doi: 10.1126/science.220.4598.671

Lapidow, E., and Bonawitz, E. (2023). What's in the box? Preschoolers consider ambiguity, expected value, and information for future decisions in explore-exploit tasks. *Open Mind* 7, 855–878. doi: 10.1162/opmi_a_00110

Le Heron, C., Kolling, N., Plant, O., Kienast, A., Janska, R., Ang, Y. S., et al. (2020). Dopamine modulates dynamic decision-making during foraging. *J. Neurosci.* 40, 5273–5282. doi: 10.1523/JNEUROSCI.2586-19.2020

Lee, W. S. C., and Carlson, S. M. (2015). Knowing when to be "rational:" Economic decision-making and executive function in preschool children. *Child Dev.* 86, 1434–1448. doi: 10.1111/cdev.12401

Leonard, J. A., Garcia, A., and Schulz, L. E. (2020). How adults' actions, outcomes, and testimony affect preschoolers' persistence. *Child Dev.* 91, 1254–1271. doi: 10.1111/cdev.13305

Leonard, J. A., Lee, Y., and Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science* 357, 1290–1294. doi: 10.1126/science.aan2317

Leonard, J. A., Martinez, D. N., Dashineau, S. C., Park, A. T., and Mackey, A. P. (2021). Children persist less when adults take over. *Child Dev.* 92, 1325–1336. doi: 10.1111/cdev.13492

Liquin, E. G., and Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition* 218:104940. doi: 10.1016/j.cognition.2021.104940

Lloyd, A., McKay, R., Sebastian, C. L., and Balsters, J. H. (2021). Are adolescents more optimal decision-makers in novel environments? Examining the benefits of heightened exploration in a patch foraging paradigm. *Dev. Sci.* 24:e13075. doi: 10.1111/desc.13075

Lloyd, A., Viding, E., McKay, R., and Furl, N. (2023). Understanding patch foraging strategies across development. *Trends Cogni. Sci.* 27, 1085–1098. doi: 10.1016/j.tics.2023.07.004

Lucca, K., Horton, R., and Sommerville, J. A. (2020). Infants rationally decide when and how to deploy effort. *Nat. Human Behav.* 4, 372–379. doi: 10.1038/s41562-019-0814-0

Marulis, L. M., and Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3-5 year olds. *Metacogni. Learn.* 16, 207–231. doi: 10.1007/s11409-020-09244-6

Meder, B., Wu, C. M., Schulz, E., and Ruggeri, A. (2021). Development of directed and random exploration in children. *Dev. Sci.* 24:e13095. doi: 10.1111/desc.13095

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., et al. (2015). Unpacking the exploration-exploitation tradeoff: a synthesis of human and animal literatures. *Decision* 2, 191–215. doi: 10.1037/dec0000033

Nussenbaum, K., and Hartley, C. A. (2019). Reinforcement learning across development: what insights can we draw from a decade of research? *Dev. Cogn. Neurosci.* 40:100733. doi: 10.1016/j.dcn.2019.100733

Oeri, N., Kälin, S., and Buttelmann, D. (2020). The role of executive functions in kindergarteners' persistent and non-persistent behaviour. *Br. J. Dev. Psychol.* 38, 337–343. doi: 10.1111/bjdp.12317

Oeri, N., Kunz, N. T., and Kälin, S. (2024). Task persistence through a dynamic lens: Understanding temporal-behavioral dynamics in kindergarten children. *J. Appl. Dev. Psychol.* 92, 101642. doi: 10.1016/j.appdev.2024.101642

O'Leary, A. P., and Sloutsky, V. M. (2017). Carving metacognition at its joints: PROTRACTED development of component processes. *Child Dev.* 88, 1015–1032. doi: 10.1111/cdev.12644

Otto, A. R., Gershman, S. J., Markman, A. B., and Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761. doi: 10.1177/0956797612463080

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

Rovee-Collier, C. (1997). Dissociations in infant memory: rethinking the development of implicit and explicit memory. *Psychol. Rev.* 104, 467–498. doi: 10.1037/0033-295X.104.3.467

Schulz, E., Wu, C. M., Ruggeri, A., and Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychol. Sci.* 30, 1561–1572. doi: 10.1177/0956797619863663

Schulz, L. E., and Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Dev. Psychol.* 43, 1045–1050. doi: 10.1037/0012-1649.43.4.1045

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., et al. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *J. Exp. Psychol.* 146, 155–164. doi: 10.1037/xge0000250

Speekenbrink, M. (2022). Chasing unknown bandits: Uncertainty guidance in learning and decision making. *Curr. Dir. Psychol. Sci.* 31, 419–427. doi: 10.1177/09637214221105051

Wang, J., and Bonawitz, E. (2022). Children's sensitivity to difficulty and reward probability when deciding to take on a task. *J. Cogn. Dev.* 24, 341–353. doi: 10.1080/15248372.2022.2152032

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., and Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Exp. Psychol. Gen.* 143, 2074–2081. doi: 10.1037/a0038199

Wittmann, M. K., Scheuplein, M., Gibbons, S. G., and Noonan, M. P. (2023). Local and global reward learning in the lateral frontal cortex show differential development during human adolescence. *PLoS Biol.* 21:e3002010. doi: 10.1371/journal.pbio.3002010

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., and Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nat. Human Behav.* 2, 915–924. doi: 10.1038/s41562-018-0467-4

Zajkowski, W. K., Kossut, M., and Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *Elife* 6:e27430. doi: 10.7554/eLife.27430.016

Zelazo, P. D., and Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Dev. Perspect.* 6, 354–360. doi: 10.1111/j.1750-8606.2012.00246.x

Zhuang, W., Niebaum, J., and Munakata, Y. (2023). Changes in adaptation to time horizons across development. *Dev. Psychol.* 59, 1532–1542. doi: 10.1037/dev0001529

frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 23 September 2024
DOI 10.3389/fdpys.2024.1441395

OPEN ACCESS

EDITED BY
Igor Bascandziev,
Harvard University, United States

REVIEWED BY
Maria Theobald,
University of Trier, Germany
Lucas Lörch,
Leibniz Institute for Research and Information in Education (DIPF), Germany

*CORRESPONDENCE
Andrew G. Young
ayoung20@neiu.edu

RECEIVED 31 May 2024
ACCEPTED 30 August 2024
PUBLISHED 23 September 2024

CITATION
Young AG and Shtulman A (2024) Children's cognitive reflection predicts successful interpretations of covariation data.
Front. Dev. Psychol. 2:1441395.
doi: 10.3389/fdpys.2024.1441395

Children's cognitive reflection predicts successful interpretations of covariation data

Andrew G. Young[1]* and Andrew Shtulman[2]

[1]Department of Psychology, Northeastern Illinois University, Chicago, IL, United States, [2]Department of Psychology, Occidental College, Los Angeles, CA, United States

Introduction: Cognitive reflection is the ability and disposition to reflect on one's own thinking, allowing a person to identify and correct judgments grounded in intuition rather than logic. Cognitive reflection strongly predicts school-aged children's understanding of counterintuitive science concepts. Here, we asked whether children's cognitive reflection similarly predicts a domain-general scientific skill: the interpretation of covariation data.

Method: Five- to 12-year-olds (N = 74) completed a children's Cognitive Reflection Test (CRT-D) and measures of executive functioning. They also interpreted covariation data presented in 2 x 2 contingency tables.

Results and discussion: CRT-D performance predicted children's overall accuracy and the strategies they used to evaluate the contingency tables, even after adjusting for their age, set-shifting ability, inhibitory control, and working memory. Thus, the relationship between cognitive reflection and statistical reasoning emerges early in development. These findings suggest cognitive reflection is broadly involved in children's scientific thinking, supporting domain-general data-interpretation skills in addition to domain-specific conceptual knowledge.

KEYWORDS

cognitive reflection, scientific thinking, evidence evaluation, statistical reasoning, data interpretation, development

1 Introduction

Human reasoning and decision-making are often characterized by the coexistence and interaction of fast intuitive processes and more costly deliberative analytic processes (Kahneman, 2011). The Cognitive Reflection Test (CRT; Frederick, 2005) and its variants (e.g., verbal CRT, Sirota et al., 2021) are the most widely used measures of individual differences in analytic vs. intuitive thinking in adults. CRTs are designed to measure the ability and disposition to override an intuitive incorrect response and engage in deliberative reflection to generate a correct alternative response. Consider the famous bat-and-ball item: "*A bat and a ball cost $1.10 in total. The bat costs $1 more than ball. How much does the ball cost?*" A majority of adults provide the intuitively cued response of 10 cents, failing to realize that the bat itself would then cost $1.10. Adults who provide the correct answer of 5 cents are thought to have engaged in analytic reflection, detecting and inhibiting the incorrect intuitive response that first came to mind and effortfully generating a correct response in its place (see also Bago and De Neys, 2019).

Adult performance on the CRT is widely known as an excellent predictor of rational thinking on heuristics-and-biases tasks and normative thinking dispositions (e.g., Frederick, 2005; Toplak et al., 2011). More broadly, adults with greater cognitive reflection tend to prioritize analysis over intuition across many domains. For example, they demonstrate greater conceptual understanding of science (e.g., astronomy and thermodynamics; Shtulman and McCallum, 2014) and are more likely to endorse contested scientific beliefs (e.g., evolution, climate change, and vaccination; Gervais, 2015; Pennycook et al., 2023). They are also better at rejecting empirically unjustifiable claims, including fake news (Pennycook and Rand, 2019), conspiracy theories (Swami et al., 2014), paranormal beliefs (Pennycook et al., 2012), and social stereotypes (Blanchar and Sparkman, 2020).

Recent studies using a verbal CRT for elementary-school-aged children, the Cognitive Reflection Test–Developmental Version (CRT-D), have found cognitive reflection to be a similarly powerful predictor of children's thinking and reasoning (Shtulman and Young, 2023). Performance on the CRT-D predicts rational thinking on heuristics-and-biases tasks and normative thinking dispositions in children from the U.S. (Young et al., 2018) as well as China (Gong et al., 2021). Furthermore, the CRT-D predicts children's understanding of counterintuitive concepts in biology, physics, and mathematics, as well as their ability to learn from instruction targeting these concepts (Young and Shtulman, 2020a,b; Young et al., 2022).

The above evidence suggests cognitive reflection supports the development of domain-specific scientific knowledge. However, domain-general scientific skills and practices (e.g., data interpretation, experimentation, and argumentation) are also fundamental to the development of scientific thinking (Zimmerman, 2007; Shtulman and Walker, 2020; NGSS Lead States, 2013). The present study investigates whether cognitive reflection predicts children's successful interpretations of covariation data.

Interpreting covariation data is a critical skill, as both children and adults need to draw conclusions from and update beliefs in response to data encountered in their everyday lives. By the end of the preschool years, children are able to interpret and revise their beliefs based on simple patterns of covariation data (Koerber et al., 2005; Schulz et al., 2007). However, many children and adults have great difficulty interpreting more complex patterns of covariation. When presented with covariation data in a 2 x 2 contingency table like the one shown in Figure 1, children and adults commonly generate inaccurate judgments and use non-optimal strategies that often neglect parts of the data (Shaklee and Mims, 1981; Shaklee and Paszek, 1985; Saffran et al., 2019; Osterhaus et al., 2019). For example, Saffran et al. (2016) found that 2nd and 4th graders justified their covariation judgments by mentioning only two-cells of a contingency table (e.g., *A* and *B*, but not *C* and *D*) on ∼33% of trials, and mentioned the normative comparison of ratios (i.e., conditional probabilities) on only ∼3% of trials.

Prior research has not directly examined whether cognitive reflection predicts children's interpretations of covariation data presented in 2 x 2 contingency tables. However, cognitive reflection does predict reasoning on several related tasks. In adults, cognitive reflection is positively associated with accurate interpretations of covariation data that are presented sequentially (e.g., Saltor et al., 2023). Stanovich, Toplak, and colleagues have also found that cognitive reflection predicts adolescent and adult performance on composite measures of scientific thinking that include items

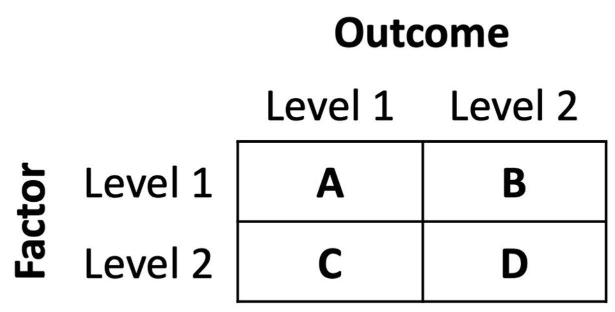


FIGURE 1
Example of a 2 x 2 contingency table with labeled cells.

on covariation detection in 2 x 2 contingency tables, though they do not report correlations with covariation items specifically (Stanovich et al., 2016; Toplak and Stanovich, 2024).

Obersteiner et al. (2015) have suggested children's invalid strategy use on 2 x 2 contingency tables arises from two common intuitive biases: base-rate bias (i.e., ignoring the base rate at which some effect occurs) and whole number bias (i.e., focusing on whole number components of fractions rather than the overall ratios). Children with greater cognitive reflection are less likely to exhibit both of these biases. For example, CRT-D performance predicts normative reasoning on base-rate sensitivity and denominator neglect/ratio bias tasks (e.g., Gong et al., 2021; Young and Shtulman, 2020a). Furthermore, middle school students' CRT-D performance positively predicts their mature number sense, including perceiving fractions as numbers (rather than separate numerators and denominators) and rich conceptual understandings of rational and whole numbers (Kirkland et al., 2024).

Finally, considering multiple hypotheses and focusing on disconfirmation (rather than confirmation) are both thought to improve correct interpretation of contingency tables (e.g., Osterhaus et al., 2019). Cognitive reflection facilitates children's reasoning about possibilities (Shtulman et al., 2023), and might similarly facilitate children's reasoning about multiple hypotheses. Additionally, adults who rely on counterexamples to solve reasoning problems tend to have higher CRT scores and more accurate covariation judgments (Béghin and Markovits, 2022; Thompson and Markovits, 2021). These multiple lines of evidence suggest children who exhibit greater cognitive reflection should be more successful in interpreting covariation data than those who exhibit less.

In this study we measured school-aged children's performance on the CRT-D and explicit judgments of covariation data presented in 2 x 2 contingency tables. We adopted our stimuli and procedure from Saffran et al. (2016). That is, we presented data in a grounded context (i.e., plant foods and plant growth) using symmetrical tables that compared two potential causes rather than the presence and absence of one candidate cause (i.e., Food A vs. Food B, rather than Food A vs. No Food). Both contextual grounding and symmetry of variables support children's and adults' successful interpretations of covariation data (Osterhaus et al., 2019; Saffran et al., 2016). We considered children's covariation judgment accuracy and strategy use. Prior research has usually examined children's and adults' strategies for interpreting covariation data by eliciting verbal explanations and justifications (e.g., Saffran et al., 2016, 2019) or by evaluating patterns of correct responding across items (e.g., Shaklee and Paszek, 1985; Osterhaus et al., 2019). We used patterns of responding, including specific errors, to assess children's strategies.

We also measured children's executive functions, including set-shifting, inhibitory control, and working memory. Inhibitory control processes have been hypothesized to support children's covariation judgments (e.g., reducing base-rate and whole number biases; Obersteiner et al., 2015). Similarly, limited working memory capacity might contribute to children's use of strategies that neglect parts of a data table (Saffran et al., 2019). However, prior research has not directly examined children's executive functions and their evaluations of 2 x 2 contingency tables. Measuring executive functions also allowed us to further examine the predictive utility of cognitive reflection. Research with children and adults suggests the predictive strength of cognitive reflection is largely independent of executive functions (e.g., Toplak et al., 2011; Young and Shtulman, 2020a), but this may not be the case for interpreting covariation data. Thus, we asked whether the CRT-D is a useful predictor of covariation judgment accuracy and strategy use after adjusting for children's age and executive functions.

2 Method

2.1 Participants

Our participants were 74 children in kindergarten through 6th grade. Their mean age was 7 years and 5 months, and they were approximately balanced for gender (42 female, 32 male). Children were recruited from public playgrounds in Southern California. The present data is a subset of 86 children reported on in Gorman (1986) investigation of fake news detection. Eight children from this larger dataset did not complete the covariation judgment task and were excluded from the present analyses. Additionally, four children who had not yet entered kindergarten were excluded, as the covariation judgment task we used has not been used with preschoolers in prior research.

2.2 Measures and materials

2.2.1 Cognitive reflection test—developmental version

Children completed the 9 item CRT-D (Young and Shtulman, 2020a) as a measure of cognitive reflection. The test consists of brainteasers designed to elicit intuitive, yet incorrect, responses that children can correct upon further reflection. An example item is "If you're running a race and you pass the person in second place, what place are you in?" The lure response is first, but the correct answer is second, as you have not passed the person in first. We used the number of correct responses as children's score, with higher scores indicating greater cognitive reflection.

2.2.2 Executive function tasks

2.2.2.1 Verbal fluency

Children completed two verbal fluency tasks as measures of endogenous set-shifting (Munakata et al., 2012). They named as many animals as they could in 1 min and as many foods as they could in 1 min (without repetition). To be successful, children had to recognize the need to switch subcategories when they had exhausted exemplars from the current subcategory (e.g., breakfast foods) and also decide what new subcategory to switch to (e.g., desserts, fruits, or snacks) without external cues. Children's responses were audio-recorded and transcribed. Children's performance on the animal and food tasks were similar ($Mean_{\text{Animal}} = 12.6$, $SD_{\text{Animal}} = 5.5$ vs. $Mean_{\text{Food}} = 11.9$, $SD_{\text{Food}} = 5.6$) and highly correlated, $r_{(38)} = 0.73$. We used the mean number of items across the animal and food tasks as children's verbal fluency score. In cases where we did not have data for both verbal fluency tasks (e.g., due to recording errors or attrition), we scored their performance on a single verbal fluency task.

2.2.2.2 Toolbox flanker inhibitory control and attention test

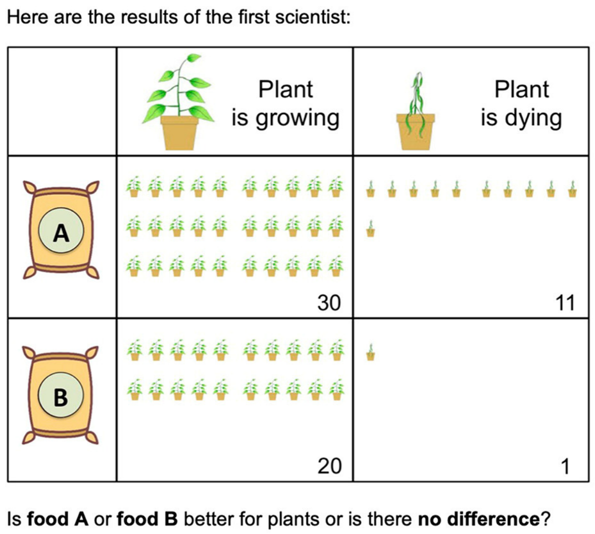
Children completed the tablet-based Flanker Test from the NIH Toolbox Cognition Battery (Zelazo et al., 2013; iPad Version 1.11). The test measures both attention and inhibitory control, requiring children to indicate the left-right orientation of a middle stimulus while inhibiting attention to four flanking stimuli. Scoring of the Toolbox Flanker is based on both accuracy and reaction time (Zelazo et al., 2013). We used uncorrected standardized scores (*Mean* = 100, *SD* = 15), which reflect overall level of performance relative to the entire NIH Toolbox normative sample, regardless of age or other demographic factors. Higher scores indicate better performance.

2.2.2.3 Backward digit span

Children completed a backward digit span task that required maintenance and manipulation of items in working memory (Alloway et al., 2009). The experimenter read a sequence of numbers at a pace of one per second. Children were then asked to repeat the numbers in reverse order. Children were given a practice trial of 3 digits and then test trials starting at 2 digits, increasing by 1 digit after every 2 trials. The task ended when children failed both trials of a given length or at the conclusion of the 8-digit trials. We used the highest span with at least one correct trial as children's score (Alloway et al., 2009). Scores could range from 1 to 8 (a score of 1 was assigned if children failed both 2-digit trials).

2.2.3 Covariation judgment materials

We used six 2 × 2 contingency table items from Saffran et al. (2016). Items were presented in the context of a story about scientists developing different plant foods to improve plant growth. The rows of the tables were labeled with illustrations indicating "Food A" and "Food B," the two levels of the independent variable. The columns of the tables were labeled with illustrations indicating the "plant is growing" and the "plant is dying," the two levels of the dependent variable. Cell frequencies were depicted with illustrations and numbers. Figure 2 shows an example item



Item 1		Item 2		Item 3	
30	30	6	7	8	8
27	10	10	29	16	16

Item 4		Item 5		Item 6	
15	10	30	11	20	10
16	30	20	1	10	5

FIGURE 2
Sample item **(top)** and cell frequencies for the six covariation data tables **(bottom)**.

as presented to children and the cell frequencies of the six items. The normative strategy for solving such problems is to compare conditional probabilities. In this strategy, also called *comparison of ratios*, a solver compares the proportion of cases with growing plants that received Food A to the proportion of cases with growing plants that received Food B [e.g., A/(A+B) vs. C/(C+D)].

As described in Saffran et al. (2016), items were designed with a number of characteristics in mind. There were two items depicting no relationship (Items 3 and 6), two items favoring Food A (Items 2 and 4), and two items favoring Food B (Items 1 and 5). For the majority of items, attention to only the first row (*A vs. B* strategy) or only the first column (*A vs. C* strategy) would yield incorrect judgments. The structure of relationships between cell values also varied across items. Three items included the same cell value twice (Items 1, 3, 6), one item had cell values that were simple multiples (Item 6), and other items had less salient relationships (e.g., 16 is about half of 30 in Item 4). Finally, the difference between the cells in the two rows and the cells in the two columns was the same for Items 3 and 5, but not the other items.

2.3 Procedure

Children completed the study on-site with the consent of their guardians. Trained research assistants worked one-on-one with the children to complete the tasks at tables adjacent to the playgrounds. Depending on the measure, research assistants either read the items aloud or displayed them on an iPad, and the children responded verbally or via touch screen. Children completed the tasks in the following order: animal verbal fluency, Flanker, CRT-D, covariation judgment (described below), backward digit span, food verbal fluency. Children also completed a short (<5 min) fake news detection activity between the CRT-D and covariation judgment tasks; the results for this activity are presented in Gorman (1986) and will not be considered here. Most children completed the entire study session in 20–25 min.

2.3.1 Covariation judgment procedure

The procedure and script of our covariation task closely followed the symmetrical condition of Experiment 1 in Saffran et al. (2016). A researcher first introduced the covariation judgment task by providing a grounded context with the following story:

> In this game you are going to think about some scientists that are trying to invent foods that help plants grow. Each scientist has invented two different plant foods and wants to figure out if one food is better for helping plants grow, or if there is no difference between the foods. So each scientist did an experiment. The scientists gave one of their foods to one set of plants, and then gave their other food to a different set of plants. After a few weeks, the scientists checked to see if the plants grew well or died.

After the introduction, the researcher explained the structure of a sample table that contained no data:

> Each scientist wrote down what they saw in a table like this. This row will show the plants that got food A (researcher pointed across 1st row). This row will show the plants that got food B (researcher pointed across 2nd row). This column will show the plants that grew well (researcher pointed down 1st column). This column will show the plants that died (researcher pointed down 2nd column).

The researcher then asked two comprehension questions to make sure children understood the meaning and structure of the table ("Before we move on, can you show me which box will have plants that got food A and are growing? Can you show me which box will have plants that got food B and are dying?). Children who failed these questions received a second explanation of the sample table and answered the comprehension questions again. After passing the comprehension questions, children were presented with the six contingency tables one at a time in random order. For each table children were asked to make a judgment about which plant food was better based on the results of the scientist's experiment (e.g., "Here are the results of the first scientist. Is food A or food B better for plants or is there no difference?"). Children's six judgments were scored for accuracy (e.g., responding Food B for

Item 1). The McDonald's ω total of the measure was 0.77 (Zinbarg et al., 2005).

2.4 Coding

We coded children's response patterns across the items for the use of six strategies observed in prior studies (see Saffran et al., 2016). Strategies relying on all four cells of the tables included the normative *comparison of ratios* strategy [e.g., A/(A+B) vs. C/(C+D)] and the *comparison of differences* strategy [e.g., (A-C) vs. (B-D)]. Strategies relying on just two cells of the tables included *A vs. B*, *C vs. D*, *A vs. C,* and *B vs. D*. Table 1 shows the expected response patterns of these strategies across the 6 items.

We coded children's strategy use according to the following scheme. First, we coded children as using a given strategy if their response patterns exactly matched the expected pattern generated by that strategy across all six items. Seventeen children met this criterion. Next, we coded the remaining children as using a given strategy if their response patterns matched the expected pattern generated by that strategy on five of six items. Twenty-one additional children matched a strategy on five of six items. Previous studies have employed similar, less stringent criteria when coding children's strategies for interpreting 2 x 2 contingency tables (e.g., Shaklee and Paszek, 1985), allowing for noise or distraction. Finally, we coded the 36 children whose response patterns did not match any strategy on at least five items as using other/mixed strategies.

A potential concern with coding matches on five of six items is the possibility of ties (i.e., matching more than one strategy). The probability of matching more than one strategy on five of six items is 0.0069 (see Supplementary material). Thus, the opportunity for ties, given the coding scheme and strategy particulars, was quite low. Three children did have response patterns that matched two strategies on five of six items (see Table 1). Two children matched on the *comparison of ratios* and *comparison of differences* strategies (both 4-cell strategies). One child matched on *C vs. D* and *B vs. D* (both 2-cell strategies).

In the following strategy use analyses, we compared children in terms of the larger 4-cell, 2-cell, and mixed/other strategy categories. We classified the two children matching on *comparison of ratios* and *comparison of differences* in the larger 4-cell strategy category and the child that matched on *C vs. D* and *B vs. D* in the larger 2-cell strategy category.

3 Results

3.1 Descriptive statistics and missing data

Table 2 presents summary statistics of and bivariate Pearson correlations among our variables of interest. As can be seen in Table 2, our primary data set contained missing verbal fluency, Flanker, and backward digit span data for a number of children. Three children were missing verbal fluency data due to audio-recording failures. Thirty-one children were missing Flanker data due to experimenter or software errors. Eleven children were missing backward digit span data due to drop out or parent interruption. One possible reason for data loss was our setting (playgrounds), which may have introduced additional distractions and interruptions compared to more typical lab settings. Overall, 36 of 74 children provided incomplete data, resulting in missing values for 10.1% of the primary data set. We ran a Hawkins test for data missing completely at random (MCAR) with the R package *MissMech* (Jamshidian et al., 2014), which revealed insufficient evidence to reject the assumption that data were MCAR ($p = 0.499$). To increase statistical power, reduce bias, and account for the uncertainty induced by these missing data, we generated 50 imputed data sets via predictive mean matching using the R package *mice* (van Buuren and Groothuis-Oudshoorn, 2011; for multiple imputation details, see: https://osf.io/t37hn/). We used these multiply imputed data for all following inferential analyses.

3.2 Judgment accuracy

We fit a Bayesian multilevel logistic regression model to examine the predictive utility of cognitive reflection, along with age and executive functions, for children's judgment accuracy (for an analogous Frequentist analysis, see the Supplementary material). We modeled children's judgment accuracy as repeated binomial trials (i.e., correct vs. incorrect across the 6 covariation items) with CRT-D, age, verbal fluency, Flanker, and backward digit span as predictors. The model also included by-participant random intercepts. Predictor variables were scaled to mean = 0 and $SD = 1$. We used weakly informative priors for all regression parameters, including Normal ($\mu = 0, \sigma = 2.5$) for beta coefficients. We used the brm_multiple() function from the R package *brms* to fit the model to each of the 50 imputed data sets and produce a final pooled model by combining the posterior distributions from each imputed fit (Bürkner, 2017). We report median posterior point-estimates and 89% Credible Intervals (CrI) from this pooled model distribution. Graphical posterior predictive checks, R_{hat} values, and effective posterior sample size (ESS) values were satisfactory (Muth et al., 2018).

Finally, we used projective predictive variable selection (via the R package *projpred*; Piironen et al., 2023) to examine the importance of model predictors for out-of-sample predictive performance (i.e., how well a model should predict a new child's judgment accuracy). This method uses posterior information from a reference model to find smaller candidate models whose predictive distributions closely match the reference predictive distribution (Piironen and Vehtari, 2017). The method begins with a forward search through the model space, starting from an empty model (intercept-only), and at each step adding the variable that minimizes the predictive discrepancy to the reference model. Next, Bayesian leave-one-out cross-validation (Vehtari et al., 2017) and a decision criterion are used to determine the final size of the submodel. We selected the smallest submodel within 1 standard error of the predictive performance of the reference model, using expected log pointwise predictive density (elpd) as the measure of predictive performance. This approach allowed us to assess the relative importance of the CRT–D as a predictor in comparison to other measured variables. As an example, if projective predictive variable selection suggested a model with age, CRT-D, and Flanker as predictors, we could conclude that age is the most important

TABLE 1 Expected response patterns for strategies.

	Response					
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
4-cell strategies						
Comparison of ratios	B	A	ND	A	B	ND
Comparison of differences	B	A	ND	A	ND	A
2-cell strategies						
A vs. B	ND	B	ND	A	A	A
C vs. D	B	A	ND	A	B	B
A vs. C	A	B	B	B	A	A
B vs. D	B	A	A	A	B	B

A = "Food A", B = "Food B", ND = "No Difference."

TABLE 2 Variable summary statistics and bivariate correlations.

			Correlations				
Variable	N	Mean (SD)	1.	2.	3.	4.	5.
1. Correct covariation judgments (out of 6)	74	2.46 (1.64)	—				
2. Age (in years)	74	7.38 (1.85)	0.46***	—			
3. CRT-D (out of 9)	74	2.96 (2.25)	0.50***	0.52***	—		
4. Verbal fluency	71	12.11 (5.26)	0.07	0.04	−0.04	—	
5. Flanker	43	82.00 (18.31)	0.25	0.62***	0.09	−0.05	—
6. Backward digit span	63	3.13 (1.34)	0.11	0.43***	0.24	−0.14	0.40*

*p < 0.05.
***p < 0.001.

predictor, followed by CRT-D and then Flanker. Further, we could conclude that any variables not selected add no additional predictive information. For further details on the models and projective predictive variable selection, see: https://osf.io/t37hn/.

Figure 3A shows the relationship between CRT-D and children's judgment accuracy. A 1 SD increase in CRT-D predicted a 1.65 increase in the odds of a correct judgment, 89% CrI [1.23, 2.22]. Figure 3B and Supplementary Table 1 display the parameter estimates from the model. Children's age also predicted judgment accuracy. A 1 SD increase in age predicted a 1.48 increase in the odds of a correct judgment, 89% CI [1.02, 2.18]. Projective predictive variable selection suggested a submodel with CRT-D and no other predictors. The model with CRT-D as the only predictor had similar out-of-sample predictive performance to the full model, Δ elpd = 0.64, SE = 2.59. Overall, these results indicate that CRT-D performance predicted children's correct interpretations of covariation data over and above their age and executive functions. Furthermore, CRT-D performance is the single best predictor of children's covariation judgement accuracy among the measured variables.

Children's verbal fluency, Flanker, and backward digit span did not predict judgment accuracy in the model that included CRT-D. These measures also did not predict judgment accuracy when considered independently and modeled as single predictors (see Supplementary Figure 1).

3.3 Strategy use

Table 3 summarizes children's coded strategy use, including mean judgment accuracies by strategy group. Consistent with previous research using the same materials (Saffran et al., 2016, 2019), more children used a two-cell strategy than a four-cell strategy, with *A vs. C* being the most common. Similarly, descriptive results show that overall judgment accuracy was weakly connected to strategy categories. For example, children that used a two-cell *C vs. D* or *B vs. D* strategy judged more items correctly than children that used a two-cell *A vs. B* or *A vs. C* strategies, and children that used a more sophisticated four-cell *comparison of differences* strategy.

We fit a Bayesian multinomial logistic regression model to examine the relationships between cognitive reflection, age, and executive functions on children's strategy use (for an analogous Frequentist analysis, see Supplementary material). We modeled children's strategy use with two-cell strategies as the reference category and CRT-D, age, verbal fluency, Flanker, and backward digit span as predictors. Similar to our accuracy model, we used scaled predicter variables (mean = 0, *SD* = 1) and weakly informative priors [e.g., Normal (μ = 0, σ = 2.5) for beta coefficients]. We generated a pooled model from fits to the 50 imputed data sets and performed projective predictive variable selection. Graphical posterior predictive checks, R_{hat} values,

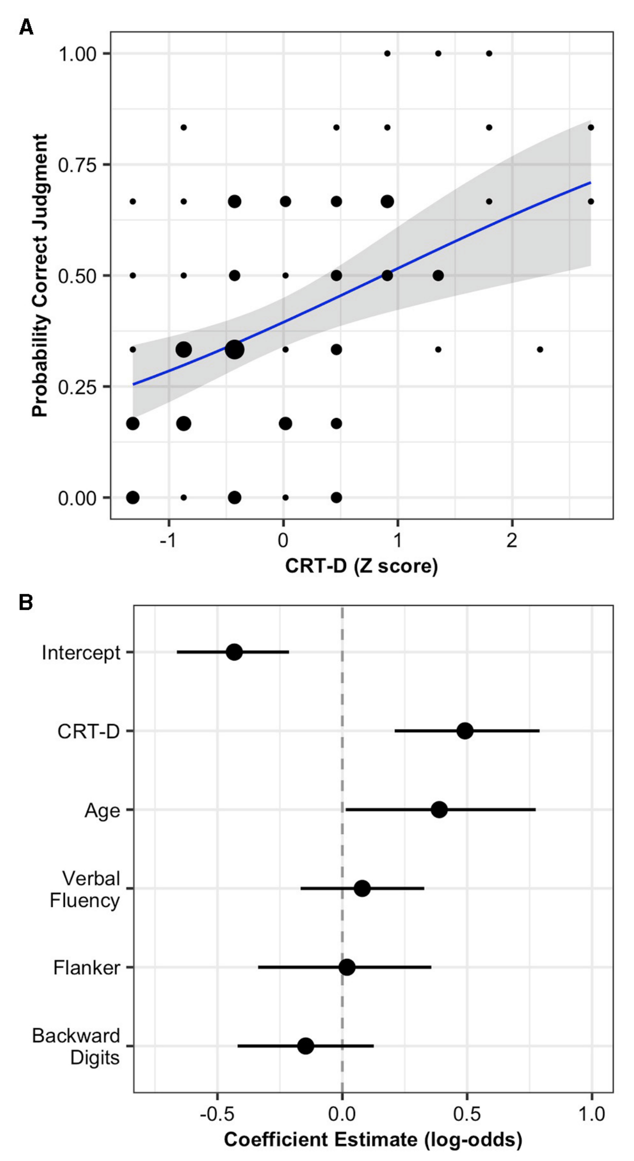


FIGURE 3
(A) Estimated probability of correct covariation judgment by CRT-D score. Ribbons represent 89% Credible Intervals. Points represent children's raw percent of correct judgments with size proportional to number of children. **(B)** Accuracy model coefficient estimates with 89% Credible Intervals.

TABLE 3 Summary of children's coded strategy use.

Strategy category	# of children	Mean correct covariation judgments (out of 6)
4-cell strategies	13	4.54
Comparison of ratios	4	5.75
Comparison of differences	7	3.71
Tie: comparison of ratios and comparison of differences	2	5
2-cell strategies	25	1.64
A vs. B	2	2.5
C vs. D	2	5
A vs. C	16	0.0375
B vs. D	4	4
Tie: C vs. D and B vs. D	1	4
Other/mixed strategies	36	2.28

Other/mixed children did not match an expected strategy response pattern on at least 5 of 6 items.

and ESS values were satisfactory. For details, see: https://osf.io/t37hn/.

Figure 4A shows the relationships between CRT-D and children's strategy use. A 1 *SD* increase in CRT-D predicted an 8.02 increase in the odds of using a four-cell strategy over a two-cell strategy, 89% CrI [3.12, 22.03]. Additionally, a 1 *SD* increase in CRT-D predicted a 2.29 increase in the odds of using a mixed/other strategy over a two-cell strategy, 89% CrI [1.06, 5.06]. Figure 4B and Supplementary Table 4 display the parameter estimates from the strategy use model. Projective predictive variable selection suggested a submodel with CRT-D and no other predictors. The model with CRT-D as the only predictor had similar out-of-sample predictive performance to the full model, Δ elpd =0.59, $SE = 3.74$. Overall, these results indicate that CRT-D performance predicted children's strategy use over and above their age and executive functions. Furthermore, CRT-D performance is the single best predictor of children's strategy use among the variables measured.

Children's age, verbal fluency, Flanker, and backward digit span did not predict strategy use in the model that included CRT-D. When considered independently and modeled as single predictors (see Supplementary Figure 2), only children's age predicted using a four-cell strategy over a two-cell strategy, OR = 2.28, 89% CrI = [1.27, 4.26]. However, children's age (OR = 0.58, 89% CrI = [0.35, 0.95]), Flanker (OR = 0.52, 89% CrI = [0.28, 0.90]), and backward digit span (OR = 0.60, 89% CrI = [0.37, 0.94]) independently predicted using a mixed/other strategy relative to a two-cell strategy. These effects suggest that with increasing age, inhibitory control, and working memory, children were more likely to use a two-cell strategy over a mixed/other strategy. In contrast, children with greater cognitive reflection were more likely to use a mixed/other strategy over a two-cell strategy (i.e., in the combined model).

4 Discussion

The present study examined whether cognitive reflection predicts school-aged children's interpretations of covariation data. In line with prior research, a majority of children in the present study had difficulty interpreting covariation data presented in 2 x 2 contingency tables and used sub-optimal strategies that neglected parts of the data (e.g., Shaklee and Mims, 1981; Shaklee and Paszek, 1985; Saffran et al., 2016). However, we found children with greater CRT-D scores generated more accurate judgments and were more likely to use sophisticated four-cell strategies than children with lower CRT-D scores. Cognitive reflection predicted correct interpretations and strategy use even after adjusting

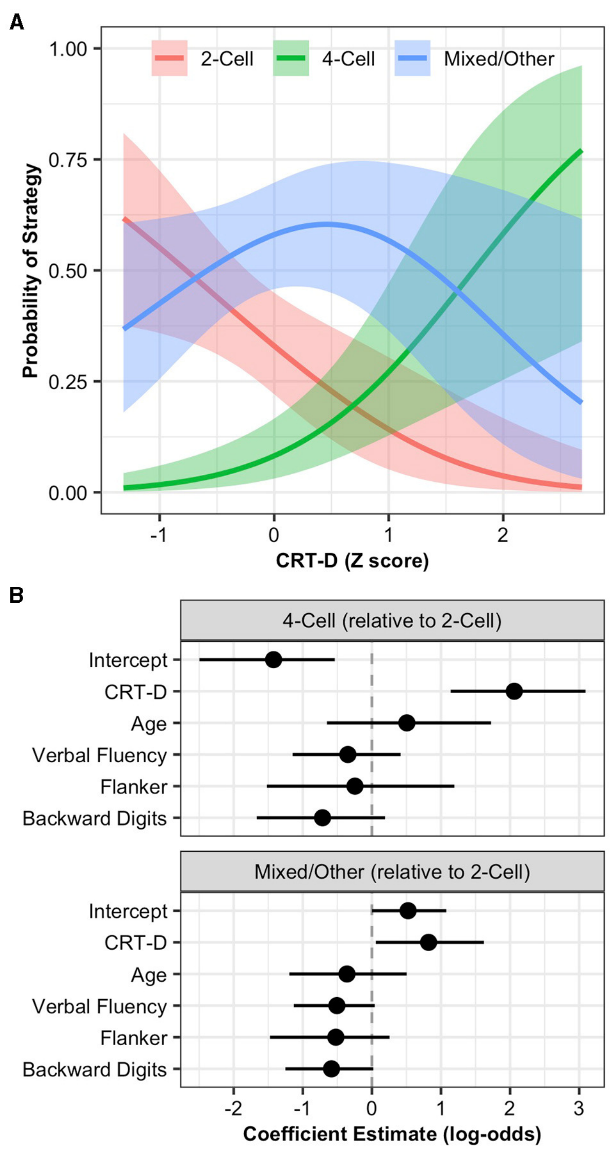


FIGURE 4
(A) Estimated strategy probabilities by CRT-D score. Ribbons represent 89% Credible Intervals. **(B)** Strategy use model coefficient estimates with 89% Credible Intervals.

for children's age, set-shifting, inhibitory control, and working memory. Moreover, if we wanted to predict a new school-aged child's accuracy or strategy use in the present task, their CRT-D score is the first and only measure we should collect. Age and the executive functioning measures did not provide any additional out-of-sample predictive value.

Our findings are consistent with prior research demonstrating cognitive reflection predicts covariation judgment accuracy in adolescents and adults (e.g., Saltor et al., 2023; Stanovich et al., 2016; Toplak and Stanovich, 2024). Why does cognitive reflection facilitate children's correct interpretations of covariation data? One possibility is that cognitive reflection helps children override intuitive base-rate and whole number biases (e.g., Young and Shtulman, 2020a; Gong et al., 2021; Kirkland et al., 2024), which are thought to drive inadequate strategies on 2 x 2 contingency tables (Obersteiner et al., 2015). Similarly, overall mathematical ability supports successful interpretations of covariation data in adults (Osterhaus et al., 2019). Children's cognitive reflection is positively associated with greater math ability in several domains, including understanding the equal sign (Young and Shtulman, 2020a), mature number sense (Kirkland et al., 2024), and use of the distributive property (Clerjuste et al., 2024). It may be that more reflective children in the present study were more likely to have the requisite mathematical skills to correctly judge 2 x 2 contingency tables, even after adjusting for age. Future research should directly measure children's base-rate bias, whole number bias, and general mathematical ability to better understand the relationship between cognitive reflection and interpretations of covariation data.

Another explanation is that cognitive reflection facilitates children's modal cognition and greater consideration of possibilities (Shtulman et al., 2023, 2024). Children with greater cognitive reflection may have approached the data tables by entertaining multiple hypotheses (e.g., Food A is better vs. Food B is better vs. Food A and B are similar), thereby increasing their focus on disconfirming hypotheses and considering more data cells (Ackerman and Thompson, 2017; Osterhaus et al., 2019). In contrast, less reflective children may have focused on confirmatory testing of fewer hypotheses (e.g., Food A is better, so compare *A* vs. *C*). Similarly, adults with greater cognitive reflection are more likely to rely on counterexamples when solving reasoning problems (Thompson and Markovits, 2021). A greater consideration of counterexamples may have led children to focus on disconfirmation in the present study. Future studies are needed to investigate the role of cognitive reflection in children's hypothesis testing. Comparing more and less reflective children in open-ended experimentation or causal learning tasks would be a fruitful approach to exploring how cognitive reflection influences children's navigation of hypothesis spaces and their strategies for testing those hypotheses.

Children with greater cognitive reflection likely had a metacognitive advantage in the present task. Xu et al.'s (2022) Meta-Reasoning framework suggests metacognitive monitoring and control are integral to our reasoning and problem-solving processes. In adults, the CRT has been used to study several meta-reasoning processes. For example, more reflective adults have superior conflict detection (i.e., sensitivity to conflict between intuitive judgments and logical principles; Šrol and De Neys, 2021), better meta-reasoning discrimination (i.e., deciding whether an answer is likely correct and should be reported or withheld; Strudwicke et al., 2023), and are less likely to overestimate their performance relative to less reflective individuals (Mata et al., 2013). Xu et al. (2022) argue that failure on the CRT is essentially a metacognitive failure associated with the Feeling of Rightness. Individuals with a strong Feeling of Rightness are less likely to reconsider, change, or spend additional time thinking about an initial intuitive response. In contrast, a weak Feeling of Rightness should trigger deliberation and a greater probability of changing answers.

Although the present study was not designed to examine children's meta-reasoning, children's use of mixed/other strategies may be an indicator of meta-reasoning. In particular, children's age, inhibitory control, and working memory predicted an increased use of two-cell strategies over mixed/other strategies. However, cognitive reflection predicted greater use of mixed/other strategies relative to two-cell strategies. One interpretation of these puzzling results is that more reflective children were more metacognitively aware of the inadequacy of their strategies across items and attempted to compensate by using multiple strategies throughout the task (as opposed to using multiple strategies by happenstance, as children with lower executive function skills may have done). To test this possibility, future studies should employee methods from the meta-reasoning and problem-solving literatures, such as eliciting confidence ratings and verbal justifications of strategies. If children's meta-reasoning is indexed by cognitive reflection, we should expect children with greater CRT-D scores to show a stronger correspondence between confidence ratings and strategy variability.

Executive functions did not support children's interpretations of 2 x 2 contingency tables in the present task. Set-shifting, inhibitory control, and working memory (measured via verbal fluency, Flanker, and backward digit span tasks, respectively) did not predict judgment accuracy or strategy-use after adjusting for cognitive reflection. When considered in isolation, both inhibitory control and working memory predicted increased use of two-cell strategies (relative to mixed/other strategies). These results are surprising given prior suggestions that inhibitory control and working memory might support the use of four-cell strategies (Obersteiner et al., 2015; Saffran et al., 2019). To our knowledge, prior research has not directly measured children's executive functions and judgments of 2 x 2 contingency tables. It is an open question whether these results would generalize to different measures of executive functions (e.g., visuospatial working memory rather than verbal working memory) or older samples with more requisite math knowledge to execute four-cell strategies.

Additionally, while cognitive reflection draws on executive function skills (e.g., inhibiting an intuitive response, shifting to an alternative response, and holding the question and possible responses in mind), it also requires the metacognitive ability to engage, coordinate, and sustain these skills on one's own (Simonovic et al., 2023; Shtulman and Young, 2023). Thus, the pattern of cognitive reflection predicting reasoning above and beyond executive functioning in children and adults (e.g., Young and Shtulman, 2020a; Shtulman et al., 2023; Toplak et al., 2011) may be the rule rather than the exception.

This study suggests that cognitive reflection may be broadly involved in children's scientific thinking. Prior research has shown that cognitive reflection supports children's domain-specific scientific knowledge (Young and Shtulman, 2020a,b). The present data highlight that cognitive reflection similarly supports children's data interpretation, a domain-general scientific skill. Further research might explore the role of cognitive reflection in children's evidence and data evaluation in other contexts, including interpretation of ambiguous, disconfirming, or confounded data (Cook et al., 2011; Schulz and Bonawitz, 2007; Theobald et al., 2024). Future work should also explore the role of cognitive reflection in other scientific skills and practices. Given children's performance in the present task, we might expect cognitive reflection to support hypothesis testing, falsification, and experimentation skills more generally.

Furthermore, given influential social models of rationality, we might expect cognitive reflection to support reasoning from disagreement (Young et al., 2012; Langenhoff et al., 2023), collaboration (Shtulman and Young, 2021), and argumentation (Mercier and Sperber, 2011). Research has already begun to explore some of these avenues. For example, Nissel and Woolley (2024) demonstrated that cognitive reflection predicted children's preference for arguments supported by statistical visualizations over anecdotal evidence. We anticipate cognitive reflection will be implicated in many domain-general scientific skills and practices.

Our findings have potential implications for education. Prior studies suggest more reflective children tend to learn more from instruction on counterintuitive science and mathematics concepts (Young and Shtulman, 2020b; Young et al., 2022). Children with greater cognitive reflection might similarly learn more from instruction on how to evaluate 2 x 2 contingency tables and other statistical reasoning topics, where performance is often undermined by inaccurate intuitions. If so, children's CRT-D performance might be used to target children who are ready for instruction or in need of additional or alternative instruction. Recent research has also found modest success in enhancing adult cognitive reflection via intervention (e.g., Isler and Yilmaz, 2023; Simonovic et al., 2023). It remains an open question whether children's cognitive reflection can be substantively improved via targeted instruction and training. Success in enhancing children's cognitive reflection might yield downstream effects, such as improving interpretation of 2 x 2 contingency tables and facilitating science learning more broadly.

To conclude, we have shown that cognitive reflection is a strong and unique predictor of elementary-school-aged children's correct interpretation of covariation data and the strategies they use to evaluate the 2 x 2 contingency tables. Indeed, the CRT-D was the single best out-of-sample predictor of children's judgment accuracy and strategy use, outperforming age, set-shifting, inhibitory control, and working memory. These data highlight cognitive reflection as a critical variable in children's data-interpretation skills and contribute to a growing literature demonstrating that cognitive reflection is broadly involved in children's developing scientific thinking.

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 at: Open Science Framework, https://osf.io/t37hn/.

Ethics statement

The studies involving humans were approved by Occidental College's 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

AY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. AS: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, 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 James S. McDonnell Foundation through an Understanding Human Cognition Scholar Award to Andrew Shtulman.

Acknowledgments

We would like to thank Kidspace Children's Museum, Julieta Alas, Connor Allison, Taleen Berberian, Yukimi Hiroshima, Ellen McDermott, Phoebe Patterson, Khetsi Pratt, and Emma Ragan for their assistance with data collection. We also extend our gratitude to Andrea Saffran and Martha Alibali for sharing their materials and engaging in helpful discussions about their task.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdpys.2024.1441395/full#supplementary-material

References

Ackerman, R., and Thompson, V. A. (2017). Meta-reasoning: monitoring and control of thinking and reasoning. *Trends Cogn. Sci.* 21, 607–617. doi: 10.1016/j.tics.2017.05.004

Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Dev.* 80, 606–621. doi: 10.1111/j.1467-8624.2009.01282.x

Bago, B., and De Neys, W. (2019). The smart System 1: evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Think. Reason.* 25, 257–299. doi: 10.1080/13546783.2018.1507949

Béghin, G., and Markovits, H. (2022). Reasoning strategies and prior knowledge effects in contingency learning. *Mem. Cogn.* 50, 1269–1283. doi: 10.3758/s13421-022-01319-w

Blanchar, J. C., and Sparkman, D. J. (2020). Individual differences in miserly thinking predict endorsement of racial/ethnic stereotypes. *Soc. Cogn.* 38, 405–421. doi: 10.1521/soco.2020.38.5.405

Bürkner, P. C. (2017). brms: an R package for Bayesian multilevel models using Stan. *J. Statist. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Clerjuste, S. N., Guang, C., Miller-Cotto, D., and McNeil, N. M. (2024). Unpacking the challenges and predictors of elementary-middle school students' use of the distributive property. *J. Exp. Child Psychol.* 244:105922. doi: 10.1016/j.jecp.2024.105922

Cook, C., Goodman, N. D., and Schulz, L. E. (2011). Where science starts: spontaneous experiments in preschoolers' exploratory play. *Cognition* 120, 341–349. doi: 10.1016/j.cognition.2011.03.003

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

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

Gong, T., Young, A. G., and Shtulman, A. (2021). The development of cognitive reflection in China. *Cogn. Sci.* 45:e12966. doi: 10.1111/cogs.12966

Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *Br. J. Psychol.* 77, 85–96. doi: 10.1111/j.2044-8295.1986.tb01984.x

Isler, O., and Yilmaz, O. (2023). How to activate intuitive and reflective thinking in behavior research? A comprehensive examination of experimental techniques. *Behav. Res. Methods* 55, 3679–3698. doi: 10.3758/s13428-022-01984-4

Jamshidian, M., Jalal, S., and Jansen, C. (2014). MissMech: an R Package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *J. Statist. Softw.* 56, 1–31. doi: 10.18637/jss.v056.i06

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kirkland, P. K., Guang, C., Cheng, Y., and McNeil, N. M. (2024). Mature number sense predicts middle school students' mathematics achievement. *J. Educ. Psychol.* 2024:880. doi: 10.1037/edu0000880

Koerber, S., Sodian, B., Thoermer, C., and Nett, U. (2005). Scientific reasoning in young children: preschoolers' ability to evaluate covariation evidence. *Swiss J. Psychol.* 64, 141–152. doi: 10.1024/1421-0185.64.3.141

Langenhoff, A. F., Engelmann, J. M., and Srinivasan, M. (2023). Children's developing ability to adjust their beliefs reasonably in light of disagreement. *Child Dev.* 94, 44–59. doi: 10.1111/cdev.13838

Mata, A., Ferreira, M. B., and Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: a dual-process perspective on overconfidence. *J. Personal. Soc. Psychol.* 105:353. doi: 10.1037/a0033640

Mercier, H., and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behav. Brain Sci.* 34, 57–74. doi: 10.1017/S0140525X10000968

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

Muth, C., Oravecz, Z., and Gabry, J. (2018). User-friendly Bayesian regression modeling: a tutorial with rstanarm and shinystan. *Quantit. Methods Psychol.* 14, 99–119. doi: 10.20982/tqmp.14.2.p,099

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

Nissel, J., and Woolley, J. D. (2024). Anecdata: children's and adults' evaluation of anecdotal and statistical evidence. *Front. Dev. Psychol.* 2:1324704. doi: 10.3389/fdpys.2024.1324704

Obersteiner, A., Bernhard, M., and Reiss, K. (2015). Primary school children's strategies in solving contingency table problems: the role of intuition and inhibition. *ZDM* 47, 825–836. doi: 10.1007/s11858-015-0681-8

Osterhaus, C., Magee, J., Saffran, A., and Alibali, M. W. (2019). Supporting successful interpretations of covariation data: beneficial effects of variable symmetry and problem context. *Quart. J. Exp. Psychol.* 72, 994–1004. doi: 10.1177/1747021818775909

Pennycook, G., Bago, B., and McPhetres, J. (2023). Science beliefs, political ideology, and cognitive sophistication. *J. Exp. Psychol.* 152, 80–97. doi: 10.1037/xge0001267

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., and Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition* 123, 335–346. doi: 10.1016/j.cognition.2012.03.003

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

Piironen, J., Paasiniemi, M., Catalina, A., Weber, F., and Vehtari, A. (2023). *projpred: Projection Predictive Feature Selection. R Package Version 2.8.0.* Available at: https://mc-stan.org/projpred/

Piironen, J., and Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Statist. Comput*. 27, 711–735. doi: 10.1007/s11222-016-9649-y

Saffran, A., Barchfeld, P., Alibali, M. W., Reiss, K., and Sodian, B. (2019). Children's interpretations of covariation data: explanations reveal understanding of relevant comparisons. *Learn. Instr*. 59, 13–20. doi: 10.1016/j.learninstruc.2018.09.003

Saffran, A., Barchfeld, P., Sodian, B., and Alibali, M. W. (2016). Children's and adults' interpretation of covariation data: does symmetry of variables matter? *Dev. Psychol.* 52:1530. doi: 10.1037/dev0000203

Saltor, J., Barberia, I., and Rodr?guez- Ferreiro, J. (2023). Thinking disposition, thinking style, and susceptibility to causal illusion predict fake news discriminability. *Appl. Cogn. Psychol.* 37, 360–368. doi: 10.1002/acp.4008

Schulz, L. E., and Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Dev. Psychol.* 43, 1045–1050. doi: 10.1037/0012-1649.43.4.1045

Schulz, L. E., Bonawitz, E. B., and Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Dev. Psychol.* 43, 1124–1139. doi: 10.1037/0012-1649.43.5.1124

Shaklee, H., and Mims, M. (1981). Development of rule use in judgments of covariation between events. *Child Dev.* 1981, 317–325. doi: 10.2307/1129245

Shaklee, H., and Paszek, D. (1985). Covariation judgment: systematic rule use in middle childhood. *Child Dev.* 1985, 1229–1240. doi: 10.2307/1130238

Shtulman, A., Goulding, B., and Friedman, O. (2024). Improbable but possible: training children to accept the possibility of unusual events. *Dev. Psychol.* 60, 17–27. doi: 10.1037/dev0001670

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. doi: 10.1016/j.jecp.2023.105727

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 Walker, C. (2020). Developing an understanding of science. *Ann. Rev. Dev. Psychol.* 2, 111–132. doi: 10.1146/annurev-devpsych-060320-092346

Shtulman, A., and Young, A. G. (2021). Learning evolution by collaboration. *BioScience* 71, 1091–1102. doi: 10.1093/biosci/biab089

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

Simonovic, B., Vione, K., Stupple, E., and Doherty, A. (2023). It is not what you think it is how you think: a critical thinking intervention enhances argumentation, analytic thinking and metacognitive sensitivity. *Think. Skills Creat.* 49:101362. doi: 10.1016/j.tsc.2023.101362

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., and Marshall, A. C. (2021). Measuring cognitive reflection without maths: development and validation of the verbal cognitive reflection test. *J. Behav. Decision Mak.* 34, 322–343. doi: 10.1002/bdm.2213

Šrol, J., and De Neys, W. (2021). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Think. Reason.* 27, 38–68. doi: 10.1080/13546783.2019.1708793

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

Strudwicke, H. W., Bodner, G. E., Williamson, P., and Arnold, M. M. (2023). Open-minded and reflective thinking predicts reasoning and meta-reasoning: evidence from a ratio-bias conflict task. *Think. Reason.* 2023, 1–27. doi: 10.1080/13546783.2023.2259548

Swami, V., Voracek, M., Stieger, S., Tran, U. S., and Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition* 133, 572–585. doi: 10.1016/j.cognition.2014.08.006

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

Thompson, V. A., and Markovits, H. (2021). Reasoning strategy vs. cognitive capacity as predictors of individual differences in reasoning performance. *Cognition* 217:104866. doi: 10.1016/j.cognition.2021.104866

Toplak, M. E., and Stanovich, K. E. (2024). Measuring rational thinking in adolescents: the assessment of rational thinking for youth (ART-Y). *J. Behav. Decision Mak.* 37:e2381. doi: 10.1002/bdm.2381

Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Mem. Cogn.* 39, 1275–1289. doi: 10.3758/s13421-011-0104-1

van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *J. Statist. Softw*. 45, 1–67. doi: 10.18637/jss.v045.i03

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statist. Comput*. 27, 1413–1432. doi: 10.1007/s11222-016-9696-4

Xu, S., Shtulman, A., and Young, A. G. (2022). "Can children detect fake news?," in *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, eds. J. Culbertson, A. Perfors, H. Rabagliati, and V. Ramenzoni (Toronto, ON: Cognitive Science Society), 2988–2993.

Young, A. G., Alibali, M. W., and Kalish, C. W. (2012). Disagreement and causal learning: others' hypotheses affect children's evaluations of evidence. *Developmental Psychology* 48, 1242. doi: 10.1037/a0027540

Young, A. G., Powers, A., Pilgrim, L., and Shtulman, A. (2018). "Developing a cognitive reflection test for school-age children," in *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, eds. T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish (Austin, TX: Cognitive Science Society), 1232–1237.

Young, A. G., Rodriguez-Cruz, J., Castaneda, J., Villacres, M., Macksey, S., and Church, R. B. (2022). "Individual differences in children's mathematics learning from instructional gestures [Conference presentation abstract]," in *2022 Meeting of the Cognitive Development Society*. Madison, WI. Available at: https://cogdevsoc.org/wp-content/uploads/2022/04/CDS2022AbstractBook.pdf (accessed May 1, 2024).

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.* 11:532088. doi: 10.3389/fpsyg.2020.01247

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., and Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): measuring executive function and attention. *Monogr. Soc. Res. Child Dev*. 78, 16–33. doi: 10.1111/mono.12032

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Dev. Rev.* 27, 172–223. doi: 10.1016/j.dr.2006.12.001

Zinbarg, R. E., Revelle, W., Yovel, I., and Li, W. (2005). Cronbach's alpha, Revelle's beta, McDonald's omega: their relations with each and two alternative conceptualizations of reliability. *Psychometrika* 70, 123–133. doi: 10.1007/s11336-003-0974-7

frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fdpys.2024.1454717

OPEN ACCESS

EDITED BY
Loren Marulis,
Connecticut College, United States

REVIEWED BY
Shiyi Chen,
University of Idaho, United States
Jérôme Clerc,
Université Grenoble Alpes, France
Marion Leclercq,
University of Lille, France

*CORRESPONDENCE
Florian Jonas Buehler
✉ florianjbuehler@gmail.com

RECEIVED 25 June 2024
ACCEPTED 15 October 2024
PUBLISHED 06 November 2024

CITATION
Buehler FJ and Oeri N (2024) Sneaky Snake: assessing metacognitive behavior in 5 to 6 year-olds with an unsolvable task.
Front. Dev. Psychol. 2:1454717.
doi: 10.3389/fdpys.2024.1454717

Sneaky Snake: assessing metacognitive behavior in 5 to 6 year-olds with an unsolvable task

Florian Jonas Buehler* and Niamh Oeri

Department of Developmental Psychology, University of Bern, Bern, Switzerland

In the present study, we developed an unsolvable behavioral metacognitive task for kindergarten children. The task was designed to gain insight into how children's metacognitive processes, measured as monitoring (e.g., checking the plan) and control behavior (e.g., seeking a piece), operate in a problem-solving task that mimics real-life scenarios. Five to six-year-old kindergarten children ($N = 72$) were asked to build a wooden snake according to a plan. The middle piece of the snake (fourth out of seven pieces) was missing, making the task unsolvable. Other than expected, metacognitive behavior was not related to teacher ratings of metacognitive self-regulation. However, we found age differences. Children in kindergarten year two ($M = 5.85$ years old) showed more control behavior than children in kindergarten year one ($M = 5.05$ years old). Surprisingly, we did not find age differences in monitoring behavior. Lastly, we found that metacognitive behavior differed between the solvable part (before the missing piece is reached) and the unsolvable part (after the missing piece is reached). Children showed more monitoring and less control behavior in the solvable part than in the unsolvable part. The current study contributes to the metacognitive research methodology by capturing children's metacognitive processes in action using an ecological-valid, unsolvable behavioral task.

KEYWORDS

metacognition, monitoring, control, kindergarten, metacognitive behavior, unsolvable task

1 Introduction

Metacognitive processes are typically assessed verbally. Metacognitive processes describe the ability to monitor and control ongoing cognitive processes (Nelson and Narens, 1990). In metacognitive tasks that capture monitoring and control, children are often asked to evaluate their study progress (i.e., judgments of learning) to rate their confidence in a given answer (i.e., confidence judgments) or to decide which materials they would like to study again (i.e., restudy selections) (e.g., Baer et al., 2021; Bayard et al., 2021; Destan et al., 2014). For example, children are asked, "How sure are you that your answer is correct?" However, verbal metacognitive assessment requires language skills and conscious metacognitive awareness that may not be sufficiently developed, especially in young children. Thus, using verbal assessments to estimate metacognition may be misleading as children's metacognition in everyday situations differs from such verbal judgments. Behavioral metacognitive tasks have been proposed to address this shortcoming. In behavioral metacognitive tasks, children assemble train tracks (Bryce and Whitebread, 2012) or puzzles (Marulis and Nelson, 2021) according to a plan. Such problem-solving tasks allow observing spontaneously occurring metacognitive behavior, such as checking the plan or correcting a mistake. These tasks may be closer related to real-life scenarios because they are almost identical to typical children's games (e.g., assembling train tracks or puzzles). Although these behavioral tasks provide an opportunity to observe metacognitive behavior, our understanding of non-verbal metacognitive behaviors and developmental variation is currently limited due to different task constraints. The current study aims to address these shortcomings by introducing an unsolvable behavioral metacognition task. By providing a task with high ecological validity and systematically analyzing different monitoring and control behaviors, the study will contribute to a fine-grained understanding of metacognitive processes in 5–6-year-old kindergarten children.

Most research on metacognitive processes is based on the influential framework by Nelson and Narens (1990). Developed initially for metamemory, the framework distinguishes between metacognitive monitoring and control, which can be applied to other domains such as problem-solving (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). Monitoring is a bottom-up process to accumulate task information (e.g., evaluating task difficulty). Control is a top-down process initiating actions at the task level. For example, evaluating a task as highly challenging is a monitoring process. Consequently, seeking help based on the evaluation is a typical control process. Both processes are closely related and crucial for children's self-regulated learning and academic achievement (Dunlosky and Metcalfe, 2009; Roebers, 2017).

Metacognitive monitoring and control develop from an early age. Research using perceptual tasks shows that from the age of 3, children can monitor their performance by reporting higher confidence in correct than incorrect trials (e.g., identifying degraded pictures; Coughlin et al., 2015; Gonzales et al., 2021; Lyons and Ghetti, 2011). From age 4, children seem to be able to monitor their performance in memory tasks (e.g., remembering picture pairs; Destan et al., 2014; Hembacher and Ghetti, 2014). From the age of 5, children show signs of metacognitive control as they are more likely to withdraw an incorrect answer than a correct answer (Bayard et al., 2021; Destan et al., 2014; Destan and Roebers, 2015; Kim et al., 2021). However, despite these impressive findings, it is important to acknowledge that these tasks require well-developed language skills and conscious metacognitive awareness. As mentioned above, these skills may not be fully developed in kindergarten children, yielding biased results for children with low language skills and/or lower metacognitive awareness. Non-verbal metacognitive tasks, therefore, provide an opportunity to analyze metacognitive processes independent of a child's language skills. For instance, behavioral tasks allow us to observe children's spontaneously occurring metacognitive processes without explicitly asking about them (e.g., "How sure are you that your answer is correct?"). In the following, we will refer to behavioral observations of metacognitive processes (e.g., monitoring and control) as metacognitive behavior. Thereby, it is important to note that previous studies (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021) have used the term metacognitive skills to describe metacognitive behavior. In the present study, we used the term metacognitive behavior to emphasize the behavioral and non-verbal aspects of behavioral assessments of metacognition. Studies focusing on metacognitive behavior (Bryce and Whitebread, 2012; Marulis and Nelson, 2021) are scarce but reveal similar developmental patterns: From the age of 3, children show monitoring (e.g., checking the construction) and control behavior (e.g., clearing space) when building three-dimensional puzzles according to a plan (Marulis and Nelson, 2021).

By simulating real-life scenarios in metacognitive tasks, we can gain insight into how metacognitive processes operate in everyday situations. Bryce and Whitebread (2012) introduced a problem-solving task in which children (5–7 years) were asked to assemble train tracks according to a model. The task allows one to observe metacognitive monitoring (e.g., checking the construction, checking the model) and control processes (e.g., clearing space, stating a plan). The results showed quantitative and qualitative differences in monitoring and control behaviors between 5- and 7-year-olds. Furthermore, metacognitive behavior was related to teacher ratings of children's metacognition [CHILD questionnaire by Whitebread et al. (2009)], suggesting convergent validity for the developed problem-solving task. Results confirmed the age-sensitivity of the task. Age differences indicated reliable age discrimination for both metacognitive processes, monitoring (e.g., checking the model) and control (e.g., sorting materials). Similarly, in the Wedgits© task (Marulis and Nelson, 2021), 3 to 5-year-olds had to assemble three-dimensional puzzles according to a plan. Metacognitive behavior was coded similarly to the train track task by Bryce and Whitebread (2012). However, in their analyses, the authors focused on aggregated scores of monitoring and control and did not distinguish between different types of monitoring or control behaviors. Results showed that metacognitive monitoring and control can be reliably observed at the age of 3. Overall, both studies suggest that metacognitive behavior in real-life play situations can be reliably observed at a very early age.

In addition to the benefits of observing metacognitive processes in real-life scenarios, behavioral metacognitive tasks have two further advantages. First, observing metacognition in behavioral tasks allows us to capture metacognitive behavior not only quantitatively but also qualitatively. Most standardized tasks provide quantitative, aggregated mean-based estimates for metacognitive processes. For example, typical memory tasks (Destan et al., 2014) and picture identification tasks (Lyons and Ghetti, 2011) yield aggregated (mean-) scores for metacognitive monitoring or control. While these tasks have provided insights into children's metacognitive development (see for an overview Roebers, 2017), the mean-based approach fails to capture different types of metacognitive monitoring and control processes involved in a task. Behavioral tasks, however, allow us to capture both quantifiable indexes and the opportunity to analyze the quality of the behavior. Thus, assessing metacognition through behavioral tasks not only provides insight into how often a behavioral strategy is displayed but also provides a more detailed understanding of the type of metacognitive behavior children display when faced with a challenge.

Second, behavioral tasks allow us to observe successful metacognitive performance as well as unsuccessful metacognitive performance, also known as metacognitive failure (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). In an unsolvable behavioral task, two types of metacognitive failure can be observed: Failure of metacognitive monitoring and failure of metacognitive control. In our approach, monitoring failure occurs when participants mistakenly assemble the wrong piece without realizing the error. An incorrectly assembled piece suggests a failure in monitoring, such as failing to gather correct information about the piece. Furthermore, we conceptualize metacognitive control failure as any form of off-task behavior. Off-task is defined as any behavior that does not serve task completion constructively (Oeri and

Roebers, 2021). When showing off-task behavior, children fail to maintain goal-directed control behavior, such as seeking a piece. Especially when a child is asked to work independently on a task without any adult scaffolding, metacognitive control failure in terms of off-task behavior is likely to occur. Observing metacognitive failures, such as making mistakes and off-task behavior, provides insights into different aspects of the task that might be particularly challenging for children. Thus, observing successful metacognition (i.e., monitoring and control) and metacognitive failure (i.e., making mistakes and off-task) within the same task provides a comprehensive understanding of how metacognitive processes play out and which aspects might be especially challenging to exert metacognition successfully.

Despite these exciting advantages of behavioral tasks, methodological challenges currently limit our understanding of metacognitive behavior in more detail. A common challenge in any behavioral task is the intertwined effects of ability, age, and previous experience. More precisely, task difficulty can impact participants' performance, potentially leading to biased results if it varies significantly between participants. Thus, keeping task difficulty constant across participants is essential to capture the skills of interest reliably (e.g., Dunlosky et al., 2016). Bryce and Whitebread (2012) addressed the issue by introducing two different age-dependent train track tasks, an easy and a more difficult one. Even though performance between age groups was matched for task difficulty, such an approach does not control for ability differences within the age groups. Depending on previous experience with train tracks, task difficulties could still vary largely within the respective age groups. In the Wedgit task, difficulty was held constant by giving children increasingly complex puzzles until they could not complete them within 4 min (Marulis and Nelson, 2021). While such an approach ensures that metacognitive performance is assessed at the individual threshold of maximal performance, it may impact motivation and tiredness, as some children need to complete many more trials than others to achieve their maximum. Another less time-consuming approach is to make the task unsolvable. Despite the fact that previous experiences may influence motivation and potential strategies for approaching the task, the unsolvable nature of the task keeps task difficulty constant across the participants without requiring them to complete numerous trials below their performance threshold.

Second, to analyze different metacognitive strategies the monitoring and control strategies must be observed at a minimal frequency. The train track and the Wedgits task report an average of 8–11 monitoring and control behaviors per minute. However, for the train track task, the most frequent monitoring behavior (i.e., "checking own construction") was observed on average twice per minute, and the most frequent control behavior (i.e., "clearing space") was observed 0.5 times per minute. Furthermore, behaviors shown by less than 25% of the children were excluded from the micro-level analysis behaviors due to the limited range of scores. The low frequencies of target behaviors restrict the reliability of metacognitive behavior estimates, making it challenging to capture potential developmental shifts. A possibility to address this issue would be by introducing more diverse features of the target and distractor items. More specifically, using items that vary in color and symbols forces the participants to monitor and control their behavior more diligently, yielding more possibilities to observe monitoring and control behavior.

Lastly, previous behavioral tasks have focused solely on the metacognitive behavior's frequency. Although this provides important information on how often monitoring and control behavior can be observed, it does not give any information on how long participants engage in the respective behavior. Especially when trying to solve a problem, persisting with a behavior increases the chance for the behavior to be successful. For example, searching for the train track takes a minimal amount of time. If child A searches for a train track for 1 s and child B searches for 10 s, the likelihood of success is higher in child B, but the frequency score would be identical in both children. Although duration is by no means a guarantee for success, it does enhance the chance of being successful. Thus, including the duration of behavior in the coding and the behavior analysis may be a potential route to understanding behavioral patterns in more detail. Combining the frequency and duration may provide insight into different effective and non-effective patterns of monitoring and control when solving a problem.

Building on the foundation of Bryce and Whitebread's (2012) and Marulis and Nelson's (2021) behavioral tasks, we developed an unsolvable behavioral metacognition task (The Sneaky Snake) to observe monitoring and control behavior in kindergarten children. Similar to Bryce and Whitebread (2012), children (4–5 years) had to assemble a snake using wooden pieces according to a model. Different from Bryce and Whitebread (2012), the wooden pieces were colored (green, blue, yellow) and had different symbols on them (dots, triangles, squares). The fact that the snake pieces varied in color, size, shape, and symbols increased the need for a thorough inspection, verification, and reassessment. By increasing the complexity of the target and distractor features, we aimed to observe more metacognitive behavior per minute than in previous tasks and, potentially, more fine-grained developmental differences in the variety of observed metacognitive behaviors. The task was designed to be unsolvable. More specifically, the middle piece of the snake was missing. There were enough distractor pieces to ensure participants did not realize the piece was missing. Additionally, we verified that the children did not know that the piece was missing: We coded whether children had systematically tried all distractor pieces. If this was the case, we excluded the child ($n = 1$). Only by trying every distractor piece could one infer that the fourth piece was actually missing. Through the task's unsolvable nature, we aimed to keep task difficulty constant across participants. It prevents ceiling effects as no participant can fully complete the task (e.g., Dunlosky et al., 2016).

When developing a new task, it is also important to include established measures to examine the validity of the task. We evaluated the task's convergent validity by comparing the observed metacognitive behavior (monitoring and control) with the BRIEF-P Plan/Organize scale (German version: Daseking and Petermann, 2013). The BRIEF-P is a teacher questionnaire on children's self-regulation problems in the classroom. The scale includes items on metacognitive processes and executive functions. Recent studies suggest metacognition is closely related to self-regulation skills, such as executive functions (Bryce et al., 2015; Roebers, 2017; Marulis et al., 2020; Marulis and Nelson, 2021). For instance, Marulis and Nelson (2021) found a relationship between metacognitive behavior and executive functions. Therefore, the BRIEF-P is a suitable tool for assessing convergent validity with more general self-regulation skills in the classroom context. We also explored the relationship between the observed metacognitive behavior and a verbal assessment of children's metacognition in a ball-throwing task (Schneider, 1998). In the ball-throwing task, children had to estimate how many out of 10 balls they could successfully throw into a basket. This allows us to estimate convergent validity with a classical verbal assessment of metacognitive processes (i.e., What do you think how many of these 10 balls will you hit into the basket?).

The hypotheses were the following: (1) We expect the observed monitoring and control behaviors in 5–6-year-old kindergarten children to be negatively related to teacher ratings of self-regulation problems (BRIEF-P Plan/organize scale). More precisely, we expect more frequent and longer monitoring and control behavior to be related to fewer problems reported in the Plan and Organize scale. (2) We expect more frequent monitoring and control behavior in older than younger children. (3) We expect that older children spend more time (longer durations) with monitoring and control behaviors than younger children. (4) Without any *a priori* expectation of the change in metacognitive behavior, we exploratory compared metacognitive behavior before reaching the missing piece (solvable interval) with metacognitive behavior after the missing piece (unsolvable interval).

2 Method

2.1 Participants

In the preregistered study,[1] we relied on a random subsample from a larger study (https://doi.org/10.17605/OSF.IO/JYCV7) on children's self-regulated learning ($N=193$). The target sample size for the present study was $N=66$ children and is based on previous studies (Bryce and Whitebread, 2012; Marulis and Nelson, 2021). We included six additional children to account for potential dropouts, resulting in $N=72$ children (47% female). Children were recruited from different public kindergartens. Seventy-four percent of the children in the sample had at least one parent with a university degree, indicating a high socioeconomic background. Moreover, 67% of the sample were native speakers, which reflects the number of native children in Swiss schools (Federal Statistical Office, 2024). A majority of children had Swiss parents (59%), 13% had parents from other European countries, 4% had African parents, 4% had Asian parents, and for 19%, we do not have any information on the ethnic background. Children in the first kindergarten year ($n=37$) were $M=5.05$ ($SD=0.33$) years old, and children in the second kindergarten ($n=35$) were $M=5.85$ ($SD=0.46$) years old.

2.2 Procedure

Before testing, parents gave informed written consent, and children gave verbal assent. Ethical approval for the study was obtained from the faculty ethics committee (Approval No. 2023-07-01). We collected data from September 2023 until December 2023. Six trained experimenters individually tested participants. The child's parents were not present during testing. Testing took place in a quiet room at the kindergarten. Among all administered tasks, task order was counterbalanced.

2.3 Sneaky Snake

The task measures metacognitive behavior in an unsolvable task. The task was adapted from Bryce and Whitebread (2012). The participants had to assemble a colored wooden snake according to a model. The model (picture) and a box with the target and distractor pieces were placed on a mat (Figure 1). The snake (test trial version) consisted of seven target pieces. Overall, 38 additional distractor pieces were placed in a box. To increase task difficulty and elicit different task behaviors, several distractors were used: The snake pieces differed in colors (green, blue, yellow), shape (short and long bent pieces, four different sizes of straight pieces), and symbols (dots, triangles, squares). First, the participants completed a practice trial (a snake with three pieces) to ensure they understood the task. During the practice trial, participants received feedback from the experimenter. After a successful practice trial, children were asked to assemble another snake (test trial). Children were instructed to start building the snake from the head. They were also instructed to build the snake on the mat. The test trial was unsolvable. The picture model of the test trial consisted of seven pieces, but the fourth piece of the snake was missing. For the test trial, the experimenter left the room and returned after 5 min or whenever the child ended the task. Children's behavior was videorecorded. We piloted the task several times to determine the optimal number and qualities (colors, shapes, sizes, and symbols) of target and distractor pieces. Intensive piloting was necessary to ensure that the task was suitable for 5–6-year-olds yet challenging enough to observe metacognitive behavior. We excluded 10 children from the analyses because they were not able to successfully complete the practice trial, decided to interrupt the task (i.e., going to the bathroom), were interrupted by a third person, or understood that the task was unsolvable (children who had systematically tried all the distractor pieces, $n=1$). Excluding children who understood the task is unsolvable is relevant to keeping the task demands constant across participants.

2.4 Coding scheme for metacognitive behavior

The foundation of the developed coding scheme is based on the coding scheme for metacognitive behavior by Bryce and Whitebread (2012). The coding scheme developed by Bryce and Whitebread has four distinct features that suggest it is a valuable tool for observing metacognitive behavior: (1) It is based on Nelson and Narens' (1990) theoretical framework. (2) It is age sensitive. (3) It shows convergent validity with teacher ratings of children's metacognition. (4) It has been successfully applied to a slightly different problem-solving task by Marulis and Nelson (2021). Based on the metacognitive behaviors described by Bryce and Whitebread (2012) and Nelson and Narens' theoretical framework (1990), we adapted the coding scheme for the Sneaky Snake task. We coded monitoring, control, making mistakes, and off-task behaviors. To differentiate between monitoring and control behaviors, we relied on Nelson and Narens' framework (1990), describing monitoring as a bottom-up process accumulating task

1 https://doi.org/10.17605/OSF.IO/ZX86H

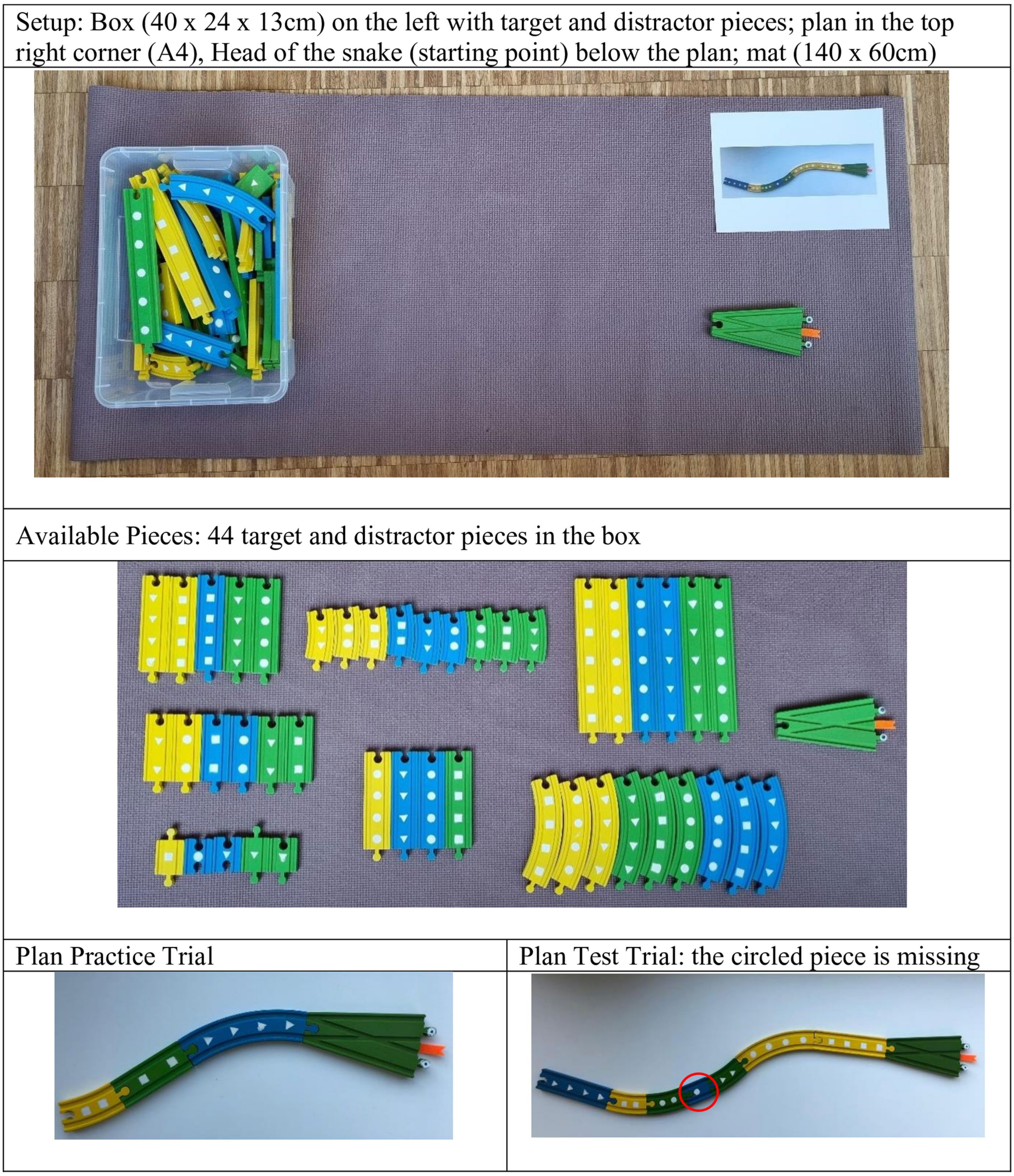


FIGURE 1
The Sneaky Snake task.

information (e.g., studying the plan) and control as a top-down process initiating actions at the task level (e.g., seeking a piece in the box). Making mistakes and off-task behavior describe two types of metacognitive failure. Making mistakes describes a monitoring failure (e.g., building in an incorrect piece) and occurs when a person may not have accumulated enough task information to make an accurate decision. Off-task behavior describes a control failure (e.g., walking away from the task) and occurs when a person fails to maintain goal-directed behavior at the task level. To examine the validity of the coding scheme, i.e., if these behaviors were observable, we randomly selected and coded 10 videos. After these 10 initial codings, we had to revise the coding scheme because some behaviors did not occur as we theoretically assumed. For example, because of the low occurrences of verbalizations in the Sneaky Snake we reduced the number of verbalization categories to monitoring and control. The final coding scheme is displayed in Table 1.

After the first full coding round, we had to drop four behaviors ("Comparing a single piece with the plan," "Comparing pieces,"

TABLE 1 Sneaky Snake coding scheme.

Behavior	Example	Occurrences per minute *M* (*SD*)	Seconds per minute *M* (*SD*)
Monitoring behavior			
Checking plan	Child glances back to the plan while seeking in the box for a piece.	4.2 (1.47)	11.25 (3.51)
Inspecting a piece	Child takes a closer look at a snake piece by counting the number of symbols on the piece.	2.69 (1.25)	4.76 (2.59)
Monitoring verbalization	"This is a difficult task" or "1, 2, 3, 4 [child counts squares on a piece]"	0.18 (0.54)	0.65 (2.16)
Comparing a single piece with the plan	Child puts a snake piece next to the plan and glances back and forth between the piece and the plan.	–	–
Comparing pieces	Child compares a blue curve with dots with a blue curve with squares.	–	–
Checking own construction	Child checks their construction by overviewing the built snake.	–	–
Sum score monitoring	–	7.08 (2.6)	16.67 (6.17)
Control behavior			
Seeking	Child seeks in the box for a piece.	4.44 (1.21)	24.78 (7.51)
Adjustments	Child replaces a piece in the snake to correct an error.	0.13 (0.17)	0.66 (1.21)
Control verbalization	"A yellow curve with four squares [child repeats what they are seeking]"	0.05 (0.32)	0.22 (1.37)
Grouping pieces	The child groups yellow pieces in one place.	–	–
Empty the box	Child empties the box.	–	–
Sum score control	–	4.63 (1.29)	25.66 (7.25)
Monitoring failure			
Mistakes	Child builds in an incorrect piece.	0.58 (0.73)	5.15 (5.6)
Control failure			
On-task-off-task	Child builds a snake that is unrelated to the task and the plan.	0.75 (0.75)	3.8 (4.12)
Off-task	Child walks away and does not interact with the task anymore.	0.3 (0.26)	4.26 (4.65)
Sum score off-task	–	1.06 (0.8)	8.05 (6.63)

All behaviors were coded in the first round of coding. In the second coding round, behaviors in bold were maintained, whereas all other behaviors were dropped because of low frequencies and reliabilities.

"Checking own construction," and "Grouping pieces") because they were not uniquely identifiable and reliably distinguishable from other behaviors. It was challenging to distinguish between "comparing a piece with the plan," "checking the original plan," and "inspecting an object." For example, a child looks back and forth between a piece and the plan while holding the piece in their hand. This behavior could be coded as a single instance of "comparing the piece with the plan" or as two instances of "inspecting the piece" and "checking the plan." To decrease ambiguity in the coding process, we decided to focus on fewer behaviors but clearly identifiable and reliably observable behaviors. Finally, we excluded "emptying the box" because it was not observed. Table 1 shows an overview of the coded behaviors, including their occurrences and durations. A more detailed version of the coding scheme is available here: https://osf.io/3dtnv/.

Because the total test time varied between subjects, we divided the duration and occurrences per behavior through the total minutes spent on the task. If children were briefly disturbed by a third party (e.g., another child running in the test room), the disturbance time was deducted from the total time. (No behaviors were rated when the child was disturbed; $n = 8$; disturbance time: $M = 3.39$ s; $range = 1$–6.13 s). We also computed aggregated scores of monitoring, control, and off-task behavior. We summed all individual behaviors contributing to monitoring, control, or off-task behavior. Monitoring failures consisted of a single measurement based on mistakes. See Table 1 for the mean scores.

We double-coded 28 (39%) of the videos. Interrater reliability for monitoring (*ICC* occurrence/min = 0.85; *ICC* duration/min = 0.59), control (*ICC* occurrence/min = 0.94; *ICC* duration/min = 0.95), making mistakes (*ICC* occurrence/min = 0.93; *ICC* duration/min = 0.96), and off-task behavior (*ICC* occurrence/min = 0.93; *ICC* duration/min = 0.97) was excellent. We transcribed all verbalizations during the coding process. Two independent raters categorized the transcriptions as monitoring or control behaviors and solved disagreements by discussion.

2.5 Self-regulation skills

To assess children's self-regulation skills in the classroom, teachers filled out two subscales, Plan/Organize and Emotional Control of the Behavior rating inventory of executive function-preschool version (Brief-P German version; Daseking and Petermann, 2013). We computed normed T-scores separated by age and gender for the Plan/Organize and Emotional Control scales. Normed mean scores are presented in Table 2. A lower T-score indicates fewer problems reported in the Plan/Organize and Emotional Control scales.

2.6 Ball-throwing

The task based on Schneider (1998) measures overconfidence. Participants were asked to throw 10 balls into a basket from a 120 cm distance. Participants started with a practice trial (10 balls). After the practice trial, they were asked to predict how many balls they would successfully throw into the box in the test trial. In the test trial, children threw 10 balls again.

We calculated metacognitive accuracy based on children's prediction [0–10] and test trial scores [0–10]. The accuracy score indicates the absolute difference between predicted and scored balls. A score closer to 0 indicates higher accuracy. Mean scores can be found in Table 2.

2.7 Statistical analysis

Shapiro Wilk tests revealed non-normal distributions of metacognitive behavior. Therefore, we relied on Spearman correlations and Mann–Whitney U Tests for group comparisons and Cohen's d as effect sizes. For data analysis, we used R (R Core Team, 2021). We computed Spearman correlations with the purrr package [version 1.0.2], MANOVAs with the manova() function of base R, and Mann–Whitney U Tests with the wilcox.test() function of base R. The R code for data analysis was developed with the support of OpenAI's GPT-4 model (OpenAI, 2024 version). The dataset and R script are available here: https://osf.io/3dtnv/.

3 Results

3.1 Descriptives

The final sample consisted of $n = 62$ children. Children spent $M = 247.84$ s ($SD = 67.48$) on the Sneaky Snake task. Regarding the solvable and the unsolvable parts, most children ($n = 55/62$) reached the unsolvable part of the task: They assembled the first three snake pieces correctly and started looking for the fourth missing piece. On average, participants worked on the task for $M = 94.03$ s ($SD = 47.45$) during the solvable part and for $M = 156.21$ s ($SD = 68.78$) during the unsolvable part. Table 1 reports the mean occurrences and duration of all observed behaviors in the Sneaky Snake task. Single categories dominated monitoring and control behavior. The most prevalent monitoring behavior was checking the plan (occurrences/min $M = 4.2$; duration/min $M = 11.25$), followed by inspecting a piece (occurrences/min $M = 2.69$; duration/min $M = 4.76$). The most prevalent control behavior was seeking (occurrences/min $M = 4.44$; duration/min $M = 24.78$), followed by adjustments (occurrences/min $M = 0.13$; duration/min $M = 0.66$). Therefore, we relied on sum scores of monitoring, control, and off-task behavior for the analyses. Making mistakes consisted of a single score. Interestingly, inspections of occurrences and duration of behaviors revealed slightly different patterns. While the most frequent behavior (occurrences/min) was monitoring, the longest (duration/min) behavior was control.

As indicated in Table 2, normed t-scores (normed for age and gender) on the BRIEF-P Plan/Organize scale and the Emotional Control scale were normally distributed, indicated by mean scores close to 50 and standard deviations close to 10. Moreover, in the ball-throwing task, children overestimated their performance. They predicted to score more balls than they did, which is typical for this age group (Schneider, 1998; Xia et al., 2023).

3.2 Validating the metacognitive behavior codings

To evaluate convergent validity we correlated metacognitive behavior in the Sneaky Snake task with a validated teacher questionnaire of self-regulated behavior (BRIEF-P). We expected negative correlations between the observed monitoring and control behaviors (occurrences and duration of the behaviors) and the BRIEF-P Plan/Organize scale. However, while controlling for age, we found no relation between monitoring or control behavior and the Plan/Organize scale for either of the variables, occurrence or duration. Also, off-task behavior and making mistakes were unrelated to the Plan/Organize scale.

Next, we explored the relationship between the observed monitoring and control behaviors and teacher ratings of self-regulated behavior in the BRIEF-P Emotional Control scale. While controlling for age, the results showed a negative correlation between monitoring occurrences and the Emotional Control scale ($\rho = -0.27$, $p = 0.042$), indicating that more monitoring is associated with fewer emotional control issues reported in the Emotional Control scale. Moreover, results showed a trend toward a negative correlation between control duration and the Emotional Control

TABLE 2 Mean scores BRIEF-P and Ball-throwing.

Scale	*M* (*SD*)	Range
BRIEF-P		
Plan/organize [t]	50.2 (10.31)	39–74
Emotional control [t]	47.45 (9.72)	40–76
Ball-throwing		
Prediction	7.1 (2.59)	2–10
Performance	5.03 (2.24)	0–10
Metacognitive accuracy	2.77 (2.43)	0–9

t = age and gender normed t-scores; Metacognitive accuracy = Performance - Prediction.

TABLE 3 Sneaky Snake mean scores for first and second kindergarten year.

Score	5-year-olds (Kindergarten 1) $n = 31$		6-year-olds (Kindergarten 2) $n = 31$	
	M (*SD*)	Range	*M* (*SD*)	Range
Monitoring				
Occurrences [occ/min]	6.71 (1.52)	3.8–10.4	7.44 (3.34)	1.6–17.11
Duration [s/min]	16.39 (4.58)	6.05–28.63	16.94 (7.5)	3.13–47.12
Control				
Occurrences [occ/min]	4.59 (1.29)	2.4–7.2	4.67 (1.32)	3–8.82
Duration [s/min]	23.05 (5.47)	14.98–37.58	28.27 (7.92)	9.9–43.66
Mistakes				
Occurrences [occ/min]	0.78 (0.92)	0–4.2	0.38 (0.38)	0–1.34
Duration [s/min]	7.00 (6.71)	0–24.14	3.3 (3.4)	0–10.95
Off-task				
Occurrences [occ/min]	1.24 (0.93)	0–3.6	0.87 (0.61)	0–2.75
Duration [s/min]	9.29 (6.48)	0–28.86	6.81 (6.66)	0–31.54

Significant differences between kindergarten 1 and 2 are bold ($p < 0.05$).

scale ($\rho = -0.25$, $p = 0.054$), indicating that longer control behavior is associated with fewer problems in regulating and controlling emotions. All other behaviors (monitoring duration, control occurrences, making mistake occurrences and duration, and off-task occurrences and duration) were unrelated to the Emotional Control scale.

Lastly, we evaluated the convergent validity between metacognitive behavior and metacognitive verbal performance prediction. Correlations controlled for age revealed no relation between monitoring or control behavior and metacognitive accuracy in the ball-throwing task for either of the variables, occurrence or duration. Also, off-task behavior and making mistakes were unrelated to the performance prediction in the ball-throwing task.

3.3 Age differences in metacognitive behavior

We compared 5 and 6-year-olds on occurrences and duration of monitoring, control, making mistakes, and off-task behavior. A MANOVA revealed a trend toward a significant age difference (*Pillai's trace* = 0.23, F (8, 53) = 1.99, $p = 0.066$). Following up with Mann–Whitney U Tests revealed that 5-year-olds showed less control behavior (duration: $U = 265$, $p = 0.002$; $d = -0.77$), a trend to more off-task behavior (duration: $U = 619.5$, $p = 0.051$; $d = 0.38$), and more mistakes (duration: $U = 630.5$, $p = 0.034$; $d = 0.69$) than 6-year-olds. Mann–Whitney U Tests for all other age comparisons (monitoring occurrences and duration, control occurrences, making mistake occurrences, off-task occurrences) were not significant. See Table 3 for means and Figure 2 for boxplots. In summary, we partially confirmed our hypothesis regarding age: Older children show more control and less off-task behavior and made fewer mistakes. However, we did not find an age difference in monitoring behavior.

3.4 Metacognitive behavior in the solvable and unsolvable task

To explore how the task's unsolvable nature affected children's behavior, we compared the behavior during the solvable part of the task (i.e., before reaching the missing piece) to the behavior during the unsolvable part of the task (i.e., after reaching the missing piece). Therefore, we compared the 60 s before the missing piece (solvable interval) with the 60 s after the missing piece (unsolvable interval). Wilcoxon signed-rank tests revealed that compared to the unsolvable interval in the solvable interval children exhibited more monitoring behavior (occurrences: $W = 983.5$, $p = 0.003$; $d = 0.43$; duration: $W = 1'049$, $p = 0.02$; $d = 0.28$), less control behavior (duration: $W = 372$, $p = 0.001$; $d = -0.48$), made more mistakes (duration: $W = 560$, $p < 0.001$; $d = 0.53$), and less off-task behavior (duration: $W = 150$, $p = 0.02$; $d = -0.34$). Wilcoxon signed-rank tests for all other task solvability comparisons (control occurrences, mistake occurrences, and off-task occurrences) were not significant. See Table 4 for mean scores and Figure 3 for boxplots. In summary, children show more monitoring, less control, and less off-task behavior in the solvable than unsolvable interval.

4 Discussion

Simulations of real-life scenarios in behavioral metacognitive tasks can provide insights into how young children's emerging metacognitive processes operate in everyday situations. In the present study, we developed an unsolvable behavioral task with high ecological validity to improve current behavioral methods to capture metacognitive monitoring and control behaviors in 5–6-year-olds. Building on existing behavioral metacognitive tasks (Bryce and Whitebread, 2012; Marulis and Nelson, 2021), three features were modified: First, to hold task difficulty constant across all participants, the task was designed to be unsolvable. Second, three distractors, shape, color, and size, were used to increase the frequency of different metacognitive strategies.

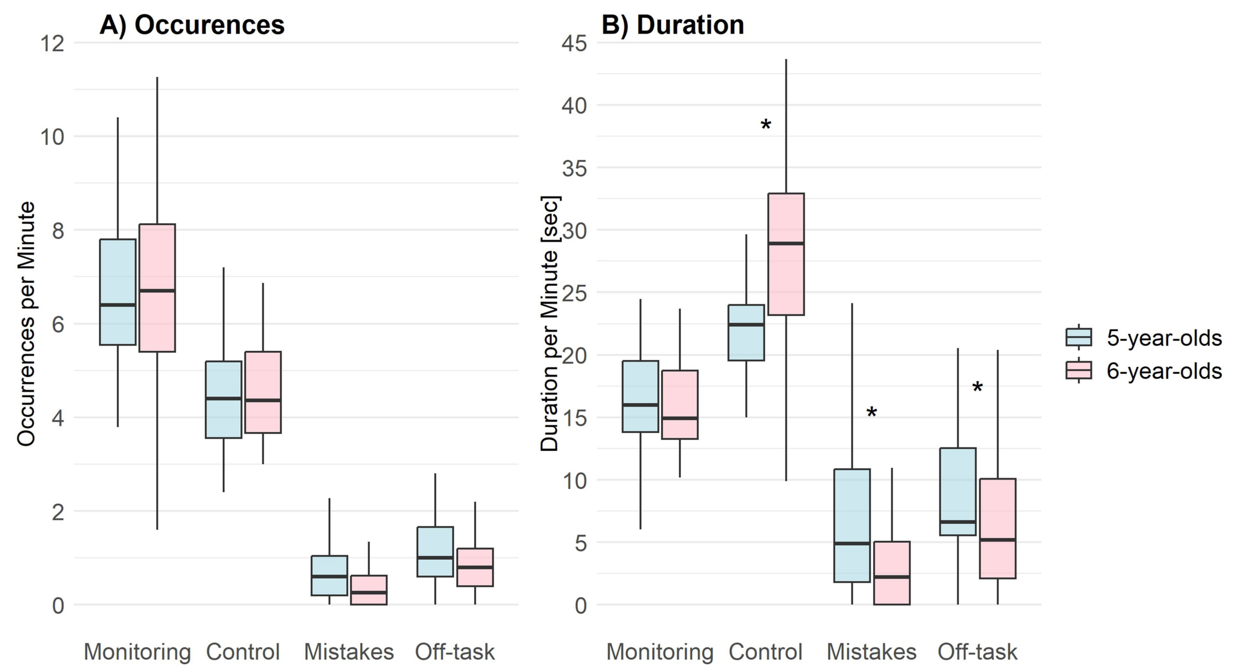


FIGURE 2
Boxplots for (A) occurrences and (B) duration of metacognitive behavior for 5 and 6-year-olds. *Significant differences between 5 (kindergarten 1) and 6-year-olds (kindergarten 2) at $p < 0.05$.

TABLE 4 Mean scores for behavior in the solvable and unsolvable task part.

Score	Solvable *n* = 55		Unsolvable *n* = 55	
	M (*SD*)	Range	*M* (*SD*)	Range
Monitoring				
Occurrences [occ/min]	9.29 (4.21)	2–18	7.19 (3.92)	0–20
Duration [s/min]	19.27 (8.64)	5.32–42.18	15.96 (9.53)	0–60
Control				
Occurrences [occ/min]	5.25 (1.75)	2–9	4.84 (2.14)	0–10
Duration [s/min]	22.3 (8.92)	8.81–48.53	29.9 (13.05)	0–50.31
Mistakes				
Occurrences [occ/min]	1.06 (1.03)	0–4	0.87 (1.2)	0–6
Duration [s/min]	8.85 (10.04)	0–39.88	3.12 (6.38)	0–28.15
Off-task				
Occurrences [occ/min]	0.51 (1.12)	0–7	0.75 (1.14)	0–5
Duration [s/min]	1.53 (4.8)	0–31.61	4.37 (7.48)	0–40.92

Scores for the solvable part are based on the 60 s before children reached the unsolvable piece, and scores for the unsolvable part are based on the 60 s after children reached the unsolvable piece. Significant differences between the solvable and unsolvable task part are bold ($p < 0.05$).

Third, to understand metacognitive processes more comprehensively, in addition to observing the frequency of the observable behaviors, the duration of the behaviors was coded, too.

The results for the Sneaky Snake task can be summarized as follows: Overall, the analysis showed that the two most frequently observed metacognitive behaviors were "seeking" (i.e., metacognitive control behavior) and "checking the plan" (i.e., metacognitive monitoring behavior). "Checking the plan" was shown twice as often than the next frequent behavior, "inspecting a piece." The difference for metacognitive control was even more pronounced: "Seeking" was shown four times more than the next frequent behavior, "adjusting." These differences are also reflected in the duration of how long the behaviors were shown.

Other than expected, the correlation analysis showed that the observed metacognitive behaviors were not related to the teacher's estimations of children's metacognitive regulation skills measured with the BRIEF-P Plan/Organize scale. Moreover, metacognitive

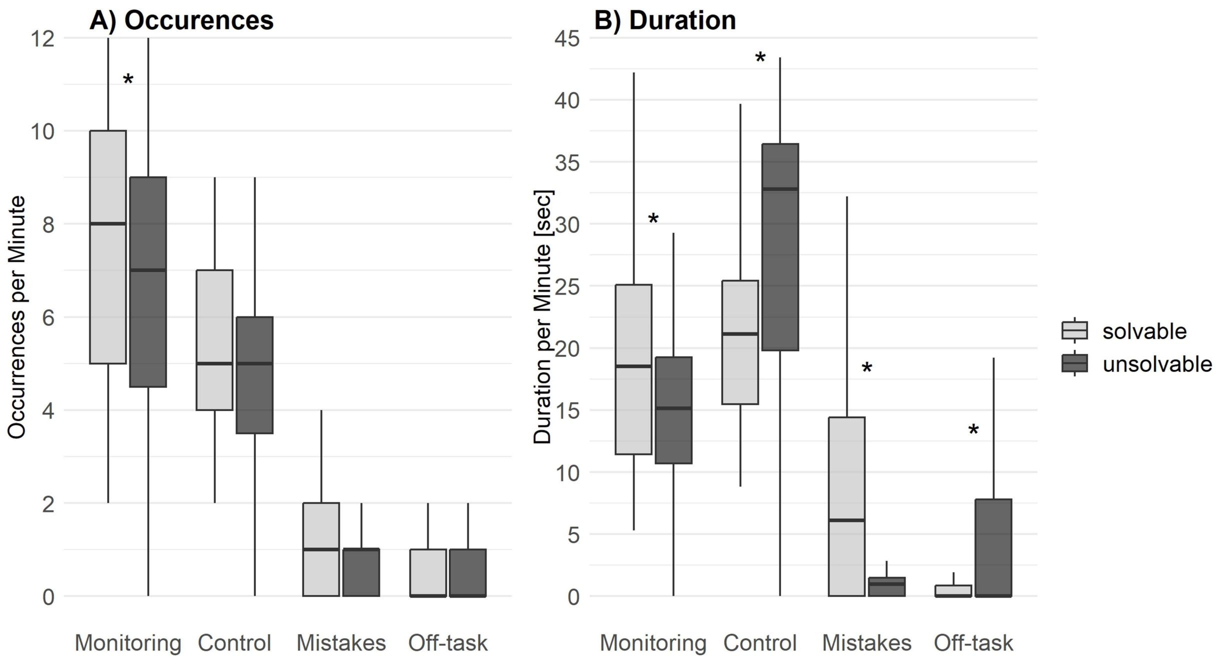


FIGURE 3
Boxplots for (A) occurrences and (B) duration of metacognitive behavior in the solvable and unsolvable part. *Significant differences in metacognitive behavior between the solvable and unsolvable part at $p < 0.05$.

accuracy in a ball-throwing task (i.e., a classical verbal assessment of metacognition) was not related to metacognitive behavior. Expected age differences were found for control behavior (i.e., seeking behavior) and off-task behavior but not for monitoring behavior or making mistakes. Comparing the solvable part of the task to the unsolvable part showed that while the task was solvable, children showed more monitoring but less control behavior. Once the task was unsolvable, children tended to show more off-task behavior.

4.1 Evaluating the unsolvable metacognitive task

The Sneaky Snake task was developed to address existing constraints in capturing behavioral metacognitive processes in young children in a real-life, familiar play context. The task consists of wooden train track pieces. These wooden train tracks are a common toy in kindergartens and children's homes in Switzerland. However, to ensure that all children were familiar with the task, we included an extensive practice trial with feedback. The practice trial was without any time limit, so every child could take as much time as they needed to get familiar with the task.

We included the BRIEF-P questionnaire (Daseking and Petermann, 2013) to validate the behavioral task. Other than expected, we found no relationship between metacognitive behavior in the Sneaky Snake task and teacher ratings of metacognitive regulation for either of the two variables, occurrences or duration. These findings indicate that the behavior children show when asked to work independently on a problem-solving task does not match the teacher's estimation of how well a child is able to plan and organize their behaviors to pursue a goal in the classroom context. While these findings are different from what we expected, it might be that the regulation demands during the Sneaky Snake tasks differ substantially from the metacognitive regulation demands in the classroom context. Thus, it might be that the two measurements focus on different aspects of metacognition. More specifically, the Sneaky Snake captures metacognitive behavior in a single play session, whereas in a typical classroom setting, children must monitor and control their behavior in the presence of many other children or when working in a group setting with peers. Given that metacognition has been suggested to be domain-specific (e.g., Baer et al., 2021; van Loon et al., 2024), and these two measurements capture different metacognitive aspects, this may explain the lack of correlations between the BRIEF-P and the Sneaky Snake.

Moreover, the plan/organize scale includes only one item that captures the child's ability to work on a difficult task; the remaining nine questions capture more classroom situations demanding to follow classroom rules. In addition to the plan/organize scale, the BRIEF-P also allows the computation of an emergent metacognition index as described in the BRIEF-P manual. This index combines the plan/organize and working memory scale. The working memory scale captures additional metacognitive components (e.g., difficulty staying on task or following multi-step instructions). However, the working memory scale was not assessed in the present study, and consequently, the metacognitive index could not be computed. It might be that the metacognitive index would reveal relationships with metacognitive behaviors that were not apparent with the plan/organize scale. Using a different questionnaire, Bryce and Whitebread (2012) found positive relations between metacognitive behavior and teacher rating with the CHILD questionnaire (Whitebread et al., 2009). The different findings may be explained by a slightly different focus of the two questionnaires: Whereas the CHILD questionnaire assesses adaptive metacognitive skills in the classroom (e.g., uses previously taught strategies), the BRIEF-P is a clinical scale focusing on metacognitive regulation problems (e.g., does not complete tasks, even after receiving hints). It may be that the CHILD questionnaire captures metacognitive skills more closely aligned with the Sneaky Snake task than the BRIEF-P.

Future research could also consider a questionnaire more closely related to metacognitive behavior reflected in the Sneaky Snake, such as the CHILD questionnaire. Moreover, a questionnaire on children's metacognitive knowledge about the Sneaky Snake could be valuable for gaining further insights into children's metacognitive awareness. For instance, Marulis and Nelson (2021) interviewed children after the Wedgits© Puzzle with the Metacognitive Knowledge Interview (e.g., Do you think you did a good or not so good job on the puzzles?; see Marulis et al., 2016). By relating the child's answers during the interview to their behavior, the Metacognitive Knowledge Interview may provide insight into the child's level of consciousness during the Sneaky Snake task.

Interestingly, metacognitive monitoring and control were related to teacher-rated emotional control skills. We found that children who more frequently monitored their behavior and spent more time with control actions (mainly seeking the snake pieces) were also better able to control their emotions. Thus, it seems that the ability to control one's own emotions may be crucial to maintaining metacognitive monitoring and control when facing an unsolvable task.

Lastly, we also explored the relationship between metacognitive accuracy in the ball-throwing task and metacognitive behavior in the Sneaky Snake. The results showed that metacognitive accuracy in the ball-throwing task was not related to metacognitive monitoring and control behavior in the Sneaky Snake task. As mentioned previously, such performance predictions as an indicator of metacognitive accuracy are an explicit and verbal assessment of children's metacognition (e.g., Xia et al., 2023). The present finding suggests that verbal assessment of metacognition and metacognitive behavioral processes reflect distinct metacognitive processes that, especially in early development, may operate more independently. The distinct measurement domains might also explain the zero correlations. Furthermore, the Sneaky Snake task is a problem-solving task, whereas the ball-throwing task is a motor task. As suggested above, metacognition in young children may be domain-specific (Baer et al., 2021; van Loon et al., 2024), explaining why metacognition in the Sneaky Snake task and metacognition in the ball-throwing task were unrelated. Future research could assess behavioral and verbal metacognition in the same task to further investigate whether behavioral and verbal metacognitive processes develop independently or develop differently between domains.

Comparing the Sneaky Snake task to existing behavioral tasks such as the Train track (Bryce and Whitebread, 2012) and the Wedgits© task (Marulis and Nelson, 2021) shows that increasing the number of distractors does not necessarily increase the frequency of the metacognitive monitoring and control behaviors. Similar to the train track study in the present study, not all behaviors were observed at a minimum frequency to be analyzed. In fact, three monitoring behaviors and two control behaviors were so rarely shown that we had to exclude them from the analysis. To address the limited frequency issue, developing more complex behavioral tasks involving multiple subsequent steps may be a way to elicit more diverse metacognitive behaviors in the participants.

4.2 Metacognitive processes captured through behavioral observation

As expected, we found age differences in several metacognitive indices (i.e., control, off-task, mistakes). Older children showed more control behavior than younger children. More specifically, when older children showed metacognitive control, they tended to show the behavior for longer periods but not necessarily more frequently. Contrary to our expectations and different than Bryce and Whitebread's (2012) findings, we did not find age differences in monitoring behavior. The age range between the investigated groups might explain the different findings. Bryce and Whitebread (2012) compared 5-year-olds with 7-year-olds, whereas we compared 5-year-olds with 5 years and 9-months-olds. Age differences in monitoring are likely more pronounced when comparing groups of children with a more significant age difference. Finally, we also found shorter periods of off-task behavior and mistakes in older children than in younger ones. These findings align with the literature suggesting that metacognitive failure decreases with age (Bryce and Whitebread, 2012). Overall, the present findings suggest that in kindergarten, differences in metacognitive behavior occur primarily in the duration, not the frequency of metacognitive behavior. More specifically, older children may spend more time with goal-directed control behavior, which may be related to making fewer mistakes and showing less off-task behavior. The differing pattern of results for occurrences and duration emphasizes the importance of including both measurements in future studies.

Furthermore, the present findings are also somewhat contradictory to longitudinal studies with verbal metacognitive assessments. Whereas Bayard et al. (2021) and Gonzales et al. (2021) found more pronounced developmental improvements in metacognitive monitoring than control, we found age differences in metacognitive control but not in metacognitive monitoring. This opposing result pattern between the present results and verbal metacognition assessments emphasizes the need to further understand how language skills and metacognitive awareness may influence these verbal judgments. One way to address this knowledge gap is through longitudinal research, which includes both measurement approaches: behavioral metacognitive task and verbal metacognitive assessments. Through such an approach, we would be able to disentangle how language and metacognitive awareness might be driving developmental trajectories of monitoring and control.

The unique feature of the Sneaky Snake task is that it is unsolvable. The main aim of designing an unsolvable task was to hold task difficulty constant for all participants. Most children (89%) reached the unsolvable part of the task, indicating that task difficulty was indeed comparable between children. However, the fact that the task consists of a solvable part and then becomes unsolvable allows us to examine an increase in metacognitive regulation demands. When the participant reaches the unsolvable part of the task, the regulation demands increase significantly as no moment of success facilitates metacognitive regulation and motivation to complete the task. The change in regulation demands was mirrored nicely in all four observed behaviors: Comparing the behaviors shown in the solvable part to the unsolvable part showed that while monitoring behavior and making mistakes decreased from the solvable to the unsolvable task part, control and off-task behavior increased. These behavioral changes may reflect the changing task demands from the solvable to the unsolvable part. While monitoring one's progress when building the snake is crucial, the same amount of progress cannot be made in the unsolvable part when searching for the missing piece. Searching for the missing piece requires a high maintenance of goal-directed behavior when facing difficulty. The higher metacognitive demands in the unsolvable part may also explain the increase in off-task behavior. Finally, fewer mistakes in the unsolvable part may result from the seeking behavior; while children were searching for the next piece, they did not place any pieces, consequently lowering the risk of making mistakes. Overall, in terms of ecological validity, the shift in the task from solvable to unsolvable mimics real-life situations quite accurately. In class and more generally when learning something new, most children are faced with the situation that initially, when starting the task, they can complete the first part but then are confronted with difficulty. Maintaining this edge is where learning eventually happens. It is also precisely at this threshold and beyond where metacognitive skills are most needed to accomplish a goal successfully. The current version of the task has yet to be improved. However, examining metacognitive processes at the threshold from solvable to unsolvable, as well as when the task is unsolvable, may be interesting for future research to gain a more detailed understanding of metacognitive processes in action.

4.3 Limitations

Even though our aim was to develop a task to address constraints in existing behavioral tasks, with our adaptions, we could not reliably observe all behaviors as planned in the first version of the coding scheme (see Table 1). Especially "glancing behavior" was difficult to distinguish. For instance, when a child puts a piece next to the plan, it was difficult to distinguish whether the child solely glanced at the plan ("checking the plan") or actively glanced back and forth between the piece and the plan ("comparing a single piece with the plan"). Therefore, we combined some of the categories in the second version of the coding scheme. For instance, we coded "checking the plan" and "comparing a single piece with the plan" as the same behavior ("checking the plan"). To address this issue in future studies, the task size should be increased. For example, instead of a picture of the snake, a same-size snake model could be used. The snake model should be placed further away from the building mat to allow for a more precise distinction between checking the model or comparing the piece with the model.

Motivation has an essential impact on any human behavior including metacognition (e.g., Efklides, 2011; Marulis and Nelson, 2021; Zimmerman and Moylan, 2009). Therefore, it's possible that motivation affects the four observed behaviors (i.e., monitoring, control, mistakes, and off-task) to different degrees and that the demands on motivation even increased during the unsolvable part of the task. Unfortunately, in the current design, it is not possible to disentangle motivation from the observed behaviors. Further research using different incentives could investigate how motivation is related to monitoring, control, mistakes, and off-task behavior.

Furthermore, the focus of the BRIEF-P questionnaire made it difficult to validate the task. The zero relation between the metacognitive behavior and the teacher ratings limit our understanding of the extent to which the task effectively captures metacognitive behavior in 5–6-year-old kindergarten children. However, the BRIEF-P working memory scale may be interesting to include in future studies. Combining the working memory scale with the plan/organize scale would allow computing an emergent metacognition index, which might be more closely related to metacognitive behavior. Finally, the present cross-sectional study limits our understanding of developmental differences in metacognitive behavior. Longitudinal designs are required to understand developmental differences in more detail.

4.4 Conclusion

We investigated young children's metacognitive behavior in an unsolvable task. The task was designed to gain insight into how children's metacognitive processes operate in a problem-solving task that mimics real-life scenarios (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). Similar to previous studies (e.g., Bryce and Whitebread, 2012; see for reviews Roebers, 2017; Xia et al., 2023), we found age differences in metacognition. Older children showed longer control behavior than younger children. Furthermore, results suggest differing metacognitive behaviors depending on whether a task is solvable or unsolvable. We observed more monitoring and less control behaviors in the solvable than unsolvable part of the task. Although the task still needs further improvement, the unsolvable nature of the task assesses metacognitive processes at a crucial threshold: Most learning happens at the edge between solvable and unsolvable, similar to what Vygotsky described as the zone of proximal development (Vygotsky, 1978). The nature of the Sneaky Snake task allows us to capture metacognitive processes precisely at this edge, potentially providing insight into metacognitive processes during a crucial moment in the learning process. The current study contributes to the research methodology to capture metacognitive processes in action by introducing an unsolvable behavioral metacognitive task.

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 below: https://osf.io/3dtnv/.

Ethics statement

The study involving humans was approved by the Ethikkommission der Phil.-hum. Fakultät University of Bern. 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

FB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. NO: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was supported by the following grant: Swiss National Science Foundation, Grant/Award Number: 1001C_197336.

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

Baer, C., Ghetti, S., and Odic, D. (2021). Perceptual and memory metacognition in children. Proceedings of the annual meeting of the cognitive science perceptual and memory metacognition in children 43, 3061–3067.

Bayard, N. S., van Loon, M. H., Steiner, M., and Roebers, C. M. (2021). Developmental improvements and persisting difficulties in children's metacognitive monitoring and control skills: cross-sectional and longitudinal perspectives. *Child Dev.* 92, 1118–1136. doi: 10.1111/cdev.13486

Bryce, D., and Whitebread, D. (2012). The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving. *Metacogn. Learn.* 7, 197–217. doi: 10.1007/s11409-012-9091-2

Bryce, D., Whitebread, D., and Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacogn. Learn.* 10, 181–198. doi: 10.1007/s11409-014-9120-4

Coughlin, C., Hembacher, E., Lyons, K. E., and Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev. Sci.* 18, 957–971. doi: 10.1111/desc.12271

Daseking, M., and Petermann, F. (2013). BRIEF-P-Verhaltensinventar zur Beurteilung exekutiver Funktionen für das Kindergartenalter [German adaptation of the behavior rating inventory of executive function® - preschool version (BRIEF®-P)]. Bern: HUBER.

Destan, N., Hembacher, E., Ghetti, S., and Roebers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5-to 7-year-old children. *J. Exp. Child Psychol.* 126, 213–228. doi: 10.1016/j.jecp.2014.04.001

Destan, N., and Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacogn. Learn.* 10, 347–374. doi: 10.1007/s11409-014-9133-z

Dunlosky, J., and Metcalfe, J. (2009). Metacognition: a textbook for cognitive, educational, life span & applied psychology. Thousand Oaks, CA: Sage Publications.

Dunlosky, J., Mueller, M. L., and Thiede, K. W. (2016). "Methodology for investigating human metamemory: problems and pitfalls" in The Oxford handbook of metamemory. eds. J. Dunlosky and S. K. Tauber (New York, NY: Oxford University Press).

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* 46, 6–25. doi: 10.1080/00461520.2011.538645

Federal Statistical Office. (2024). *Compulsory education*. Available at: https://www.bfs.admin.ch/bfs/en/home/statistics/education-science/pupils-students/compulsory.html.

Gonzales, C. R., Merculief, A., McClelland, M. M., and Ghetti, S. (2021). The development of uncertainty monitoring during kindergarten: change and longitudinal relations with executive function and vocabulary in children from low-income backgrounds. *Child Dev.* 93, 524–539. doi: 10.1111/cdev.13714

Hembacher, E., and Ghetti, S. (2014). Don't look at my answer: subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychol. Sci.* 25, 1768–1776. doi: 10.1177/0956797614542273

Kim, S., Senju, A., Sodian, B., Paulus, M., Itakura, S., Okuno, A., et al. (2021). Memory monitoring and control in Japanese and German preschoolers. *Mem. Cogn.* 51, 708–717. doi: 10.3758/s13421-021-01263-1

Lyons, K. E., and Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Dev.* 82, 1778–1787. doi: 10.1111/j.1467-8624.2011.01649.x

Marulis, L. M., Baker, S. T., and Whitebread, D. (2020). Integrating metacognition and executive function to enhance young children's perception of and agency in their learning. *Early Child Res. Q.* 50, 46–54. doi: 10.1016/J.ECRESQ.2018.12.017

Marulis, L. M., and Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacogn. Learn.* 16, 207–231. doi: 10.1007/s11409-020-09244-6

Marulis, L. M., Palincsar, A. S., Berhenke, A. L., and Whitebread, D. (2016). Assessing metacognitive knowledge in 3–5 year olds: the development of a metacognitive knowledge interview (McKI). *Metacogn. Learn.* 11, 339–368. doi: 10.1007/s11409-016-9157-7

Nelson, T. O., and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motivat. Adv. Res. Theory* 26, 125–173. doi: 10.1016/S0079-7421(08)60053-5

Oeri, N., and Roebers, C. M. (2021). Task persistence in kindergarten children: disentangling age from schooling effects. *Br. J. Dev. Psychol.* 39, 217–230. doi: 10.1111/BJDP.12358

OpenAI. (2024). ChatGPT (October 2024 version) [Large language model]. Available at: https://chat.openai.com/.

R Core Team (2021). R: a language and environment for statistical computing [computer software]. Available at: https://www.r-project.org/ (Accessed October 21, 2024).

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

Schneider, W. (1998). Performance prediction in young children: effects of skill, metacognition and wishful thinking. *Dev. Sci.* 1, 291–297. doi: 10.1111/1467-7687.00044

van Loon, M., Orth, U., and Roebers, C. (2024). The structure of metacognition in middle childhood: evidence for a unitary metacognition-for-memory factor. *J. Exp. Child Psychol.* 241:105857. doi: 10.1016/J.JECP.2023.105857

Vygotsky, L. S. (1978). "Interaction between learning and development" in Readings on the development of children. eds. M. Gauvain and M. Cole (New York, US: Scientific American Books), 34–40.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., et al. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacogn. Learn.* 4, 63–85. doi: 10.1007/s11409-008-9033-1

Xia, M., Poorthuis, A. M. G., and Thomaes, S. (2023). Why do young children overestimate their task performance? A cross-cultural experiment. *J. Exp. Child Psychol.* 226:105551. doi: 10.1016/j.jecp.2022.105551

Zimmerman, B. J., and Moylan, A. R. (2009). "Self-regulation: where metacognition and motivation intersect" in Handbook of metacognition in education. eds. B. J. Zimmerman and A. R. Moylan (New York, US: Routledge).

frontiers | Frontiers in Developmental Psychology

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

Check for updates

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

 10.3389/fdpys.2025.1496651

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

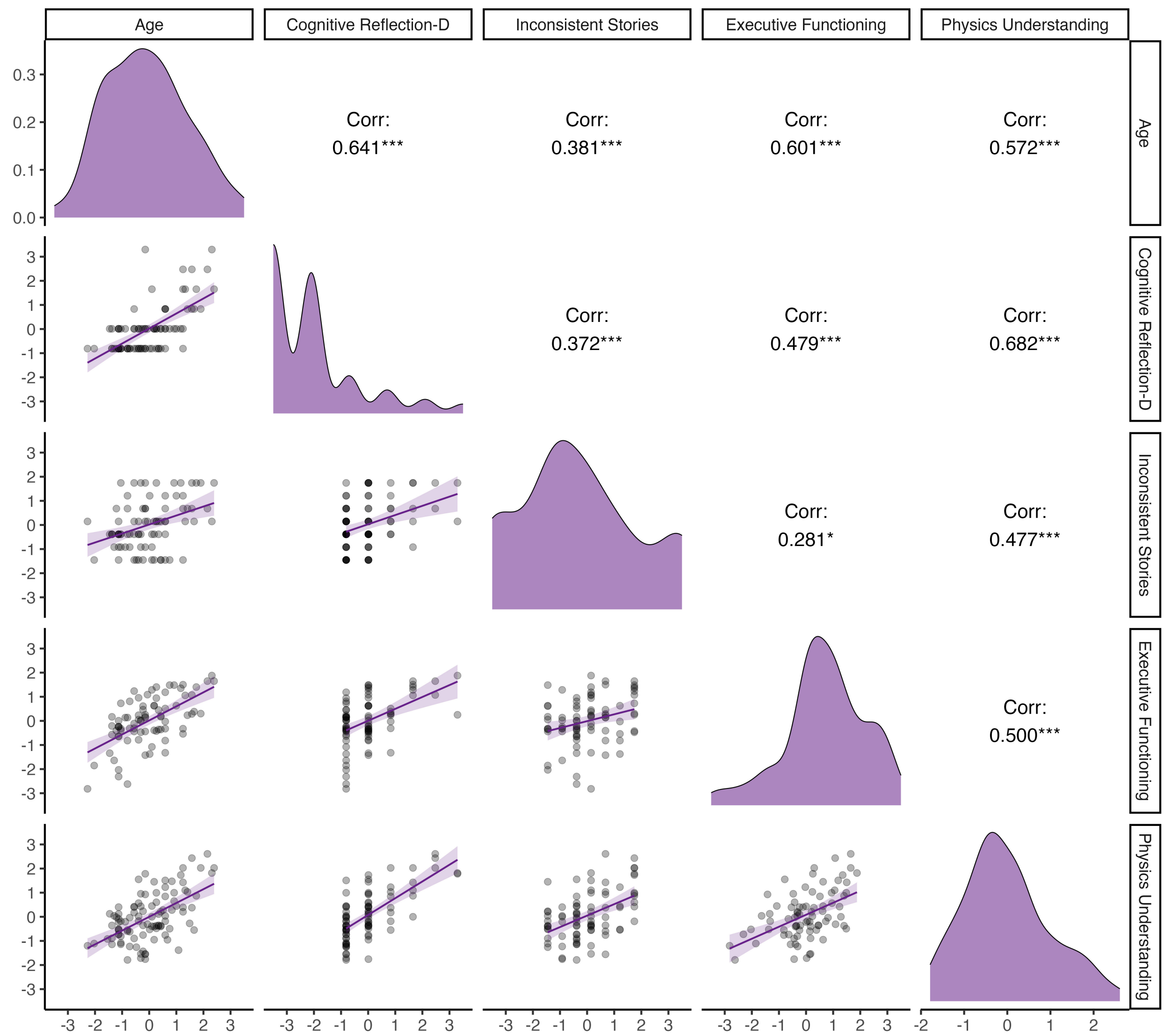


FIGURE 1
Bivariate correlations between Physics Understanding and the domain general predictor variables ($n = 81–100$). * < 0.05, ** < 0.01, *** < 0.001. The distributions of all variables are on the diagonal. Below the diagonal are the bivariate scatterplots. All plotted variables represent z-scored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

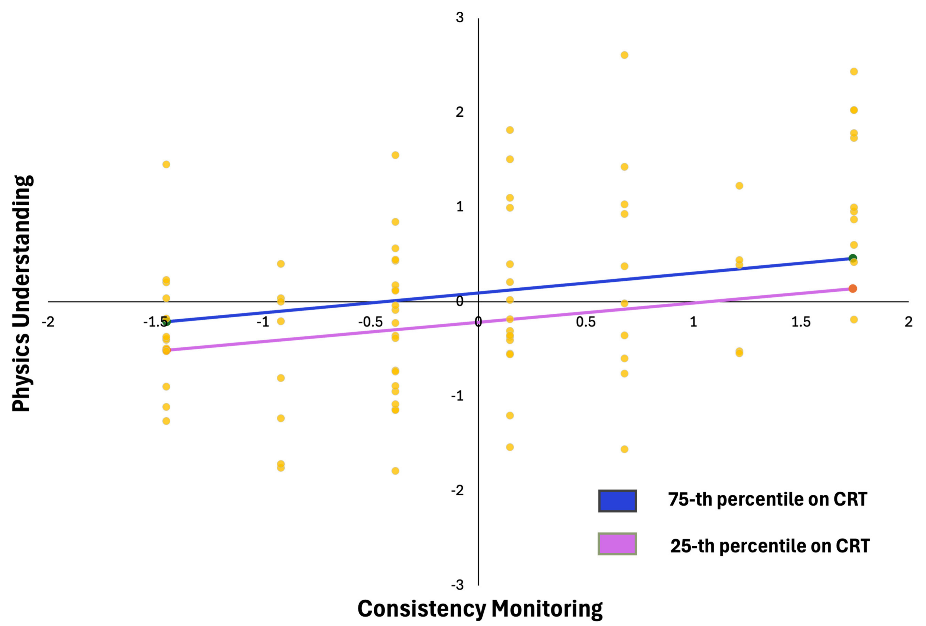


FIGURE 2
Consistency monitoring is positively correlated with Physics Understanding when controlling for Age, Executive Functioning, and Cognitive Reflection. Fitted lines of prototypical children who scored at the 25th percentile (bottom line) and the 75th percentile (top line) on Cognitive Reflection—Developmental. 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

Frontiers in
Developmental Psychology

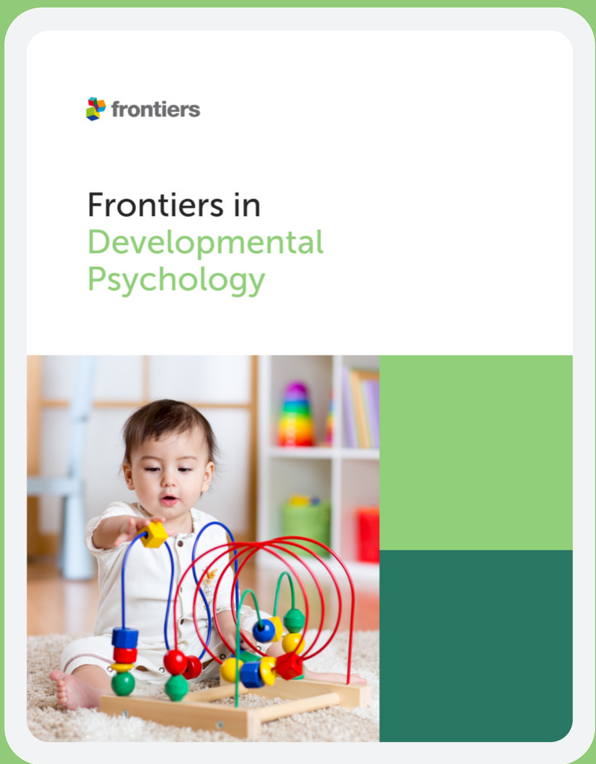
Explores human development and adaptation across the human lifespan from prenatal to old age

Advances our understanding of the cognitive, social, and emotional development of humans and the effects these internal processes have on education, culture and identity

---

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact