Towards a Comprehensive Framework of Monitoring in Young Children:

Shedding Light on the Processes Underlying Monitoring.

Inauguraldissertation

der Philosophisch-humanwissenschaftlichen Fakultät

der Universität Bern

zur Erlangung der Doktorwürde

vorgelegt von

Sophie Wacker

Basel-Stadt (BS)

Begutachtung:

Prof. Dr. Claudia. M. Roebers und Prof. Dr. Mariette van Loon

Selbstverlag, Bern im August 2023

Von der Philosophisch-humanwissenschaftlichen Fakultät der Universität Bern auf Antrag von Prof. Dr. Claudia. M. Roebers und Prof. Dr. Mariette van Loon angenommen.

Bern, den 17.11.2024

Der Dekan Prof. Dr. Elmar Anhalt

To my parents, Therese and Hans

The cumulative dissertation includes the following studies:

Study I

Wacker, S., & Roebers, C. M. (2022). Stop and think: Additional time supports monitoring processes in young children. *Plos One*, 17(9), e0274460. https://doi.org/10.1371/journal.pone.0274460

Study II

Wacker, S., & Roebers, C. M. (2023). Motivating children to (pre) monitor: positive effects on monitoring accuracy?. *Metacognition and Learning*, 1-19. https://doi.org/10.1007/s11409-023-09351-0

Umbrella Paper

Towards a Comprehensive Framework of Monitoring in Young Children: Shedding Light on the Processes Underlying Monitoring

Sophie Wacker

University of Bern

Abstract

Metacognitive monitoring is a fundamental milestone in a child's developmental trajectory. Previous theoretical models of monitoring have predominantly been informed by classical methods such as behavioral measures and observations. Furthermore, to understand monitoring within a broader research context, too little attention has been paid to expand the definition of monitoring. Thus, the main goal of the present research is to propose a comprehensive framework for monitoring and the underlying processes. The foundation of the proposed framework emerges from the studies used in this umbrella paper and these findings are integrated with a theoretical background of processes underlying monitoring. While supporting children in their inhibition skills (study I) resulted in more accurate monitoring, allowing children to differentiate between answer alternatives (study II) negatively affected children's monitoring accuracy. Results from study I indicate that inhibition might be an important underlying process of monitoring. The results from study II suggest that there might be additional processes influencing monitoring such as the high interrelation with memory. Within this dissertation combining the results from our studies expanding them with additional neuronal background of processes underlying monitoring contributes to a new framework of monitoring. Within this framework I propose the dynamic interplay of monitoring with memory and inhibition. This proposed framework might be a first step to better understand why young children often struggle to accurately monitor. More research is needed to shedding light on these dynamic relations between underlying processes and monitoring.

Contents

Introduction

Monitoring describes the ability to introspect and evaluate cognitive processes (Destan et al., 2014; Dunlosky & Metcalfe, 2008). Monitoring and regulating ongoing cognitive functions adaptively is a fundamental milestone in a child's development. These so-called metacognitive functions enable adaptive behavior within various daily tasks (de Bruin & van Gog, 2012; Roebers et al., 2017; van Loon et al., 2014). For example, ongoing surveillance of subjective task performance and comparison with objective performance.

Research in young children suggests that the fine-tuning of monitoring skills is a complex maturational process. A rudimentary form of monitoring emerges at preschool age (Coughlin et al., 2015; Lyons & Ghetti, 2011). When starting school, a child's monitoring ability becomes more sophisticated and shows increasing congruency with actual performance (Kuhn, 2000; Roebers et al., 2019). Nevertheless, at primary school age and beyond, children still have difficulties monitoring accurately. Indeed, research findings still show a discrepancy when monitoring is compared to performance accuracy (Destan & Roebers, 2015; Finn & Metcalfe, 2014). Younger children often struggle to differentiate between correct and incorrect answers and show a strong tendency towards overconfidence (Bryce et al., 2015; Lipko et al., 2009). This discrepancy and overconfidence, also known as miscalibration, can substantially affect further learning (Dunlosky & Rawson, 2012; van Loon et al., 2022). Monitoring skills are needed for various skills in school, such as mathematics and writing (Harris et al., 2010; Muncer et al., 2022). Moreover, the positive effect of monitoring on school performance is evident even when controlling for intelligence, and the longitudinal impact highlights the importance of this construct for a child's development (Ohtani & Hisasaka, 2018).

Miscalibration resulting from inefficient monitoring skills is an important aspect that can help us to understand how to adequately support children in their developmental age (Hacker et al., 2008). Such higher-order cognitive functions are highly relevant for school success and are also important into adulthood (Bakracevic & Licardo, 2010; Dunlosky et al., 2008). Therefore, the development of a comprehensive model for monitoring maturation is crucial.

Traditionally, models of monitoring have predominantly been informed by classical methods in behavioral psychology, such as behavioral measures and observations (Efklides, 2011; McAlpine et al., 1999; Nelson & Narens, 1990). As such, they run the risk of only describing monitoring from one point of view and do no incorporate possibly relevant findings from other fields with different perspectives, such as neurological considerations (Carlén, 2017; Munakata et al., 2008). Within this umbrella paper, I aim to propose a more comprehensive monitoring framework by incorporating both memory and inhibition into the theoretical foundation of monitoring. To achieve this, results from our studies will be combined with information derived from literature in the field of neurology. Joint consideration of these research branches should contribute to gathering a more profound understanding of processes underlying monitoring.

Investigating processes underlying memory from a neurological point of view should contribute to a more comprehensive understanding of monitoring. Until now, less attention has been placed on understanding the neurological activity of monitoring, but it is important to incorporate this aspect as monitoring includes neuronal signaling in its basic components (del Cul et al., 2009; Yeon et al., 2020). In this regard, imaging studies in adults have shown that the prefrontal cortex (PFC) and the anterior cingulate cortex (ACC) can be regarded as neurological correlates of monitoring (De Martino et al., 2013; Fleming et al., 2012; Insabato et al., 2010). Moreover, activity in the ventromedial PFC and in the right rostrolateral PFC is related to the sense of confidence in a selected answer, and its strength of connection also influences the degree to which metacognitive judgments can be communicated (De Martino et al., 2013). It is known that the maturation of these areas continues until late adolescence (Kolb et al., 2012). The PFC in particular is one of the brain regions that matures relatively late (Carlén, 2017; Treffer & Semendeferi, 2021). Understanding the difference in neuronal activity between children and adults is, therefore, fundamental to understanding the maturational monitoring process.

Furthermore, when considering monitoring, it seems essential to address its association with memory. A noteworthy observation is that memory and monitoring show similar neuronal activity, indicating that the two processes are closely connected (Hanks & Summerfield, 2017; Pereira et al., 2020). As in monitoring, the PFC is also essential for memory development and executing sequential operations (Bauer et al., 2013). Memory traces refer to neuronal connections between brain areas as a result of learning (Asok et al., 2019; Robins, 2017). Within cognitive operations, such memory traces represent an underlying origin of information to guide or evaluate subsequent executions. Hence, the same memory traces might be connected to both monitoring and memory processes. Importantly, we can distinguish between brain activity related to retrieval and monitoring processes suggesting that, although closely related, they are not the same processes (Peters et al., 2017; Samaha et al., 2016; Yeung & Summerfeld, 2012;).

The ability to stop and think about an answer, allowing an accumulation of information, is important for monitoring (Kälin & Roebers, 2020). Such inhibitory skills refer to a core component of executive functions and belong to the umbrella term of higher-order cognitive functions (Miyake & Friedman, 2012). Overall, executive functions might serve as a prerequisite for monitoring (Destan & Roebers, 2015; Jiao et al., 2023; Roebers et al., 2012; Roebers, 2017). Interestingly, the importance of inhibiting a prepotent answer has already been documented (Simpson & Riggs, 2007; 2011). The ability to withhold a prepotent answer might benefit monitoring accuracy because it allows the dissipation of feelings of uncertainty. With progressive maturation, executive functions and monitoring are thought to differentiate and follow a distinct pattern (Bryce et al., 2015; Geurten et al., 2016; Roebers, 2017;). These results indicate an interplay between monitoring and inhibition to compensate for inadequately developed neurological areas in young children.

The neurological basis of monitoring supports the inclusion of memory and inhibition into its theoretical framework (Fleming et al., 2010). While measuring monitoring a number of questions arise. What are the processes underlying monitoring? Is monitoring a distinct process of memory or rather an additional accumulation of an underlying memory trace? To what extent can we measure pure monitoring within young children?

To conclude, it is important to gain a profound understanding of the underlying processes of monitoring and its developmental stages, as successful development of monitoring skills is highly relevant for a child's long-term development (Bayard et al., 2021; Tobias & Everson, 2009). Moreover, the existence of a strong bias towards overconfidence at early stages of monitoring development and its interconnection with inhibition and memory is supportive of incorporation of a neurological point of view. The goal of the present umbrella paper is, therefore, to propose a joint framework of monitoring and its underlying mechanism by including state-of-the-art theories addressing monitoring, memory, and inhibition.

Theoretical Background

In this section I will introduce theories from behavioral psychology and neurology that are relevant components of my comprehensive framework of monitoring. The model of metacognition from Nelson and Narens (1990) serves as the foundation, and the significance of memory and memory traces (Standard Model of System Consolidation & Multiple Trace Theory) will be addressed based on the neurological foundation of our studies. Next, theoretical aspects of metamemory will be outlined (Strategic Regulation of Memory Accuracy). Furthermore, in the context of the neurological background, the question of whether a process leading to a decision is the same or represents a distinct process for monitoring is considered. Finally, the role of inhibition will be outlined (prepotent responses, Simpson et al., 2007). This overview aims to present the relevant parts of different theoretical backgrounds to provide a comprehensive yet differentiated understanding of monitoring.

Monitoring

Metacognitive research most often refers to the foundational model of Nelson and Narens (1990). This theoretical model addresses the interplay of monitoring and control processes and their corresponding information flow. The studies in the present dissertation focus solely on monitoring. However, to understand the relationship between monitoring and control processes, and the importance of monitoring itself, it is essential to consider the entire model. The model represents two closely related levels. The meta-level (metacognition) is superordinate to the object-level (cognition) and enables specific regulation in response to information from lower levels. Between these two hierarchically different levels there is an interplay of bottom-up (monitoring) and top-down (control) processes. Therefore, through monitoring, the meta-level is updated from the lower object-level. As a consequence, the metalevel modifies the object level through control processes. This exchange enables us to monitor current behavior and, subsequently, execution of a corresponding adaptation through a feedback loop. This model defines an essential foundation for understanding monitoring. It emphasizes the exchange between bottom-up and top-down processes, which describe the dynamic exchange between higher and lower cognitive functions. As Nelson and Narens (1990) outlined, the meta-level can not only influence the state of the object-level but can also change the processing itself. Furthermore, it highlights the importance of monitoring, which can be seen as a primary function for evaluating the cognitive processes that serve as the foundation for further regulation of subsequent cognitive processes.

Monitoring is, therefore, described as the bottom-up process transferring information between the object- and the meta-level. In my opinion this model neglects the role of hidden processes underlying monitoring – namely memory and inhibition.

Memory

Memory seems to provide the underlying information for further cognitive processing such as monitoring (Klein, 2015; Zlotnik & Vansintjan, 2019). Memory is generally known as a capacity that entails the acquisition, encoding, storing, and retrieving of information. Memory traces are the underlying units that connect and store information within different brain areas (Asok et al., 2019; Robins, 2017). These memory traces can be seen as a mental representation of stored information, feelings, and experience in the brain. As such, without the prior encoding and retrieval of memories, monitoring does not have the required information to produce a sound meta-level response (Flavell, 1971).

Memory consolidation is an essential part in the learning process (Morgado-Bernal, 2011; Squire et al., 2015). Since a paired associate learning task was used in both of our studies, it is essential to shed light on the memory processes behind them. Neuroimaging studies suggest that, as a direct consequence of learning, neurological changes occur in the hippocampus and the neocortex indicating that the moment of learning parallels a recognizable change in neurological patterns (Clopath et al., 2012; Dash et al., 2004). Furthermore, these changes can immediately influence subsequent behavior and evaluation. The crucial process of memory consolidation is the process by which synaptic connections are formed, which influences the subsequent retrieval of what has been learned (Taylor et al., 2013). Initially, mainly the hippocampus is involved. Subsequently, connections with the neocortex are further elaborated. It is assumed that the shift is increasingly based towards the neocortex, where long-term memories are represented over time. Consolidation thus represents the process by which initially short-lived, unstable memory traces are strengthened, resulting in more stable and long-lived connections (McClelland, 2013). During retrieval processes neurons in the hippocampus and the neocortex are activated. This theoretical assumption has been confirmed repeatedly and is known as the Standard Model of System Consolidation Theory (Nadel et al., 2007; Squire et al., 1992).

Hippocampal memory traces are expanded and reinforced through repeated recall and engagement. The Multiple Trace Theory (MTT) postulates that memory traces between the hippocampus and neocortical regions are strengthened during retrieval (Nadel & Moscovitch, 1997). Engaging with the same stimuli triggers re-encoding, resulting in a more robust memory trace. Furthermore, this strengthening enables a more dominant connection, which is more accessible (Clopath et al., 2012). Specifically, repeated exposure produces a more robust connection than other stimuli. Therefore, it is not surprising that a more pronounced trace is associated with increased access speed and might be easier to recall. Through repetition, these memory traces might be strengthened and adapted in a more connected way, leading to a more robust association (Robertson et al., 2004). Several brain areas are activated during memory retrieval, as shown in fMRI studies (Hayes et al., 2004; Kim, 2020). Indeed, increased brain activity in these specific brain regions has recently been linked to lower reaction times, indicative of increased memory strength through repetition and, therefore, easier access (Nadel & Moscovitch, 2001). If memory traces and memory are indeed a prerequisite for monitoring, we next need to investigate the mechanism which translates information between the two processes.

Metamemory

The term metamemory refers to higher-level judgments about memory and is, therefore, the link between memory and monitoring (Bjork, 1994). Metamemory is the knowledge about memory and the degree to which this knowledge is valid (Koriat, 1993; Nelson & Narens, 1990). The schematic model of strategic regulation of memory accuracy offers monitoring processes under consideration of memory (Koriat & Goldsmith, 1996). Within the strategic regulation of the memory accuracy model, monitoring operates as a mechanism to evaluate the correctness of memory. Within this framework, monitoring and memory retrieval processes lead to the, so-called, best candidate answer. Therefore, choosing the best answer does not only rely on monitoring but also retrieval processes connected to long-term memory. This model emphasizes the importance of considering memory and metamemory together. Having highlighted how monitoring and memory can be seen as intertwined processes from a behavioral psychology perspective, I also want to discuss their relationship from the neurological point of view.

Neurological Findings Regarding Monitoring and Memory Processes

Accumulation of Information

Because memory and metamemory are closely linked, the differentiation of these two processes is challenging. Comparing the two, it seems possible that both evoke and leverage the same neuronal activity in the form of a memory trace and mainly differ in how this is further processed. As such, the question arises as to what degree they are distinct cognitive processes (Fleming et al., 2012). Is monitoring simply the additional accumulation of information after memory retrieval (Yu et al., 2015; Kiani et al., 2014)? Or are the underlying processes different?

Following the theory of the race model (Kepecs et al., 2008; Vickers et al. 1970) additional information about a given answer accumulates over time (De Martino et al., 2013; Navajas et al., 2016). Pleskac and Busemeyer (2010) postulated a two-stage dynamic signal detection theory (2DSD) model, adapted from the drift-diffusion model of Ratcliff (1987). The 2DSD model proposes that, when choosing an answer, evidence is accumulated over time. The occurrence of a decision is determined by the point at which enough information has accumulated and reaches an individual boundary. However, further information accumulates after a decision has been made, serving as the underlying cue for rating confidence in the decision itself (Busey et al., 2000; Kepecs et al., 2008; Merkle & Van Zandt., 2006; Vickers, 1979). The fact that, in this post-decision phase, additional information is accumulated, with further cues also having an influence, is in alignment with the observation that the CJ can differ from a previously selected decision (Pleskac et al., 2010; Stone et al., 2022; Van Zandt & Maldonado-Molina, 2004).

Monitoring represents an additional accumulation of information processing, extending beyond decision processing (Cleremans et al., 2017; Navajas et al., 2016; Pasquali et al.,2010; Yu et al., 2015). Neurological studies revealed neuronal activity in the posterior medial frontal cortex during post-decision evidence accumulation (Fleming et al., 2018). Additionally, the impact of post-decision evidence on the subjective confidence report was modulated by specific activity in the anterior prefrontal cortex. Frontal subregions seem to be particularly important in encoding evidence for one of the answer options, updating the accuracy of the selected answer and transforming a confidence feeling about the selected answer (Fleming et al., 2018). However, because selecting an answer and rating a CJ use accumulated information, the question arises as to what extent these processes differ.

Distinct or Same Processes?

Although neurological studies indicated that, during decision and monitoring processes, the same brain regions are activated, monitoring processes appear to evoke additional neuronal activity. Several neurological studies have investigated the neurological underpinnings of confidence (Bang et al., 2018; Li et al., 2021). Neurological regions involved in confidence are the antero-medial prefrontal cortex (del Cul et al 2009; De Martino et al., 2013; Lebreton et al., 2015), anterior prefrontal cortex, rostrolateral prefrontal cortex (De Martino et al., 2013; Fleming & Huijgen, 2012), temporal lobe (Fleming et al. 2010), and the anterior cingulate cortex (Fleming & Huijgen, 2012). Of note, the perigenual anterior cingulate cortex (pgACC) located in the medial prefrontal cortex is strongly activated during decision confidence. Indeed, activity measured in the pgACC differs within a subject depending on the reported explicit confidence (Bang & Fleming, 2018). However, there are also brain regions, such as the intraparietal sulcus (IPS), activated during the decision process. Neurons in the IPS are, therefore, not only active during the decision process but their information is also used for confidence (Kiani et al., 2009).

There are two views regarding the relationship between decision-making and confidence. One line of research suggests that both cognitive processes are based on the same neurological areas (Boldt & Yeung, 2015; Hanks & Summerfield, 2017; Pereira et al., 2020). Another direction postulates that they are two different processes originating in distinct neurological circuits (Peters et al., 2017; Rahnev et al., 2015; Samaha et al., 2016). Yeon et al. (2020) investigated this intertwined process of neurological underpinnings within a fMRI study in adults. Results revealed that, during decision-making and confidence rating, the same neural circuits are activated (see Yeon et al. (2020)) in specific neurological areas. These findings indicate that there is an overlapping neural response during these processes (Stone et al., 2022). The authors argue that this overlap might stem from the fact that decision and confidence represent a judgment. Moreover, this view is supported by animal studies showing that the same underlying neurons predict decision and confidence (Kepecs et al., 2008; Kiani et al., 2009). Additionally, Kiani and Shadlen (2009) detected specific neurons in the parietal cortex which are activated during decision and confidence judgments (CJ). These results indicate that the underlying information for choice and confidence stem from the same neuronal patterns (Maniscalco & Lau, 2012; Meyniel et al. 2015).

Although it can be assumed that the same or similar neurological circuits are used for both decision-making and monitoring, they may still be conceptually different (Busey et al., 2000; Merkle et al., 2006). Some studies showed that additional brain areas are activated while forming a confidence judgment which are not activated during memory retrieval (Chua et al., 2006; 2009; Fleming et al., 2018; Kim et al., 2007; Moritz et al., 2006; Yeon et al., 2020). They also found brain regions that are activated during the decision process but not during confidence judgments. This is congruent with monitoring being a second order cognitive function compared to the first order function of decision making (Yeung et al., 2012). Moreover, the same underlying process can lead to different signal processing. More precisely, this means that the perception of the signal is different, and the underlying information is processed and viewed differently (Maniscalco, 2012; Yeon et al., 2020).

Looking at the same underlying processes from a different perspective might explain why there are additional regions activated during monitoring. Qiu et al. (2018) investigated the underlying processes of metacognition and decision processing within a decision – re-decision paradigm. The authors hypothesized that a different neural system controls the formation of confidence judgment beyond the decision process. Metacognition thus consists of a system that is independent of memory decision-making. They proposed a model where information about confidence simultaneously emerges during the decision process. Thereafter, second-order processes are involved in converting this accumulated information into confidence. Additionally, the authors postulated that monitoring emerges from a feeling of uncertainty rather than from error detection. To conclude, monitoring and the decision process seem to share at least in some parts the same neuronal activity. However, they also are distinct processes because monitoring triggers additional brain activity.

Inhibition and Prepotent Response

To conclude this section, I will discuss the relationship between inhibition, monitoring, and prepotent responses in children as the last piece for my framework. Inhibitory skills are important to prevent a premature answer, allowing enough time to engage in monitoring (Simpson & Riggs, 2007).

There is an ongoing debate about the importance of executive functions for metacognition and *vice versa* (Destan & Roebers, 2015; Jiao et al., 2023; Marulis & Nelson, 2021). Both constructs are classified under self-regulatory processes and are essential for school performance. Studies have shown that executive functions might be a prerequisite for monitoring and can be predictive of successfully engaging in metacognition (Kälin & Roebers, 2020; Roebers, 2017). However, their interplay can change with proceeding development. More precisely, at certain stages in development metacognition and EF are intertwined. With advanced differentiation they develop into separate and mostly independent trajectories (Roebers, 2017).

Lately, the importance of inhibition for metacognition was particularly highlighted. It is assumed that children with more differentiated inhibition skills can take more time to think and to monitor ongoing behavior (Kälin & Roebers, 2020). This is consistent with the view of taking time to respond rather than answering prematurely. Especially young children tend to answer impulsively and prematurely. Simpson et al. (2012) proposed the Passive-Dissipation Model, adapted from the stop-signal paradigm from Verbruggen and Logan (2008). The model proposes that an incorrect and premature answer is racing against a correct answer. The authors argue that an effortful computation would provide a correct answer, whereas an incorrect answer might result from a rash computation or stem from a misleading cue. Connecting this model with inhibition highlights the importance of taking enough time to avoid a premature answer. Taken together, inhibition is relevant as a prerequisite for monitoring and enabling enough time to allow the prepotent response to fade away.

Summary of Results

The overarching aim of our studies was to gain a more profound understanding of the processes underlying monitoring. Our intention was to expand the current understanding through an experimental approach and to shed light on hidden processes of monitoring. We investigated possible reasons for why young children often show inaccurate monitoring skills. In the following, I will summarize the main findings from the studies included in this dissertation.

The *'Stop and Think'* Study

Our first study, '*Stop and Think'* - builds on the idea that monitoring and executive functions are interrelated. Previous research revealed that executive functions could be a prerequisite for metacognition. Immature executive functions could partially modulate the low monitoring accuracy observed in young children. We postulated that immature inhibition skills negatively influence monitoring. Choosing an answer prematurely and not taking the time to *'Stop and Think'* hampers successful engagement in monitoring processes.

The PFC and children's inhibition skills go through a maturation process as they age. The start of school seems an additional critical factor in supporting this development. We investigated children in preschool and in second grade. Within an established paired associate learning task, we implemented what we call '*Stop and Think'* before children could choose a confidence judgment. In this modification we did not allow children to select their answer for a fixed period of time. Using this experimental approach, we intended to support children in their inhibitory skills to prevent a rash and often premature answer. We found that our '*Stop and Think'* modification led to better monitoring accuracy, indicating that pausing to think might benefit a child's ability to monitor.

The *'Pre-Monitoring'* Study

Despite the significant improvement in monitoring accuracy achieved through the *'Stop and Think'* approach, we were interested in whether further modification to the method could enhance the effects. We hypothesized that actively engaging in monitoring compared to just '*Stop and Think'* might benefit a child's monitoring accuracy. Based on '*Stop and Think'*, we addressed the following aspects in a second study known as the *'Pre-Monitoring'* Study. We hypothesized that actively engaging with all answer alternatives will positively influence monitoring accuracy. More precisely, just because we enabled children to *'Stop and Think'* in our first study, we could not be certain about what they were actually doing during this phase. Therefore, we experimentally implemented an additional phase where children had to actively engage in *'Pre-Monitoring'*. This phase entailed a differentiated evaluation of confidence for every answer alternative. We asked children to evaluate four answer alternatives to trigger contradicting information, which they could subsequently consider when evaluating their final recognition choice and when giving their confidence judgment on that final choice.

Furthermore, compared to the *'Stop and Think'* Study, we implemented this phase prior to recognition rather than afterwards. The theoretical background of metamemory postulates that memory and metamemory are closely intertwined. Thus, memory-based recognition might strongly influence the following monitoring. Therefore, evaluating confidence before an answer has been selected in recognition might not interfere with monitoring. Within this design, we intended to weaken the influence of memory traces by enabling children to activate pre-monitoring processes before choosing their final response. Contrary to our expectations, the results indicated that being instructed to pre-monitor *increased* and not decreased overconfidence in young children.

Discussion

The aim of the present umbrella paper was to propose a joint framework of monitoring and its underlying mechanism by including state-of-the-art theories of monitoring, memory, and inhibition. The foundation of the proposed framework emerges from the theoretical background of processes underlying monitoring and will be integrated with the findings from the two studies included in this umbrella paper. Whereby the studies focused on investigating monitoring processes in young children with experimental approaches, the theoretical background expands these results from a neurological point of view. Together they contribute towards a more comprehensive understanding of monitoring processes in young children in a broader research context. After introducing my proposed framework, I will return and address the initial questions: (a) what are the processes underlying monitoring? (b) is monitoring a distinct process of memory or rather an additional accumulation of an underlying memory trace? (c) to what extent can we measure pure monitoring within young children?

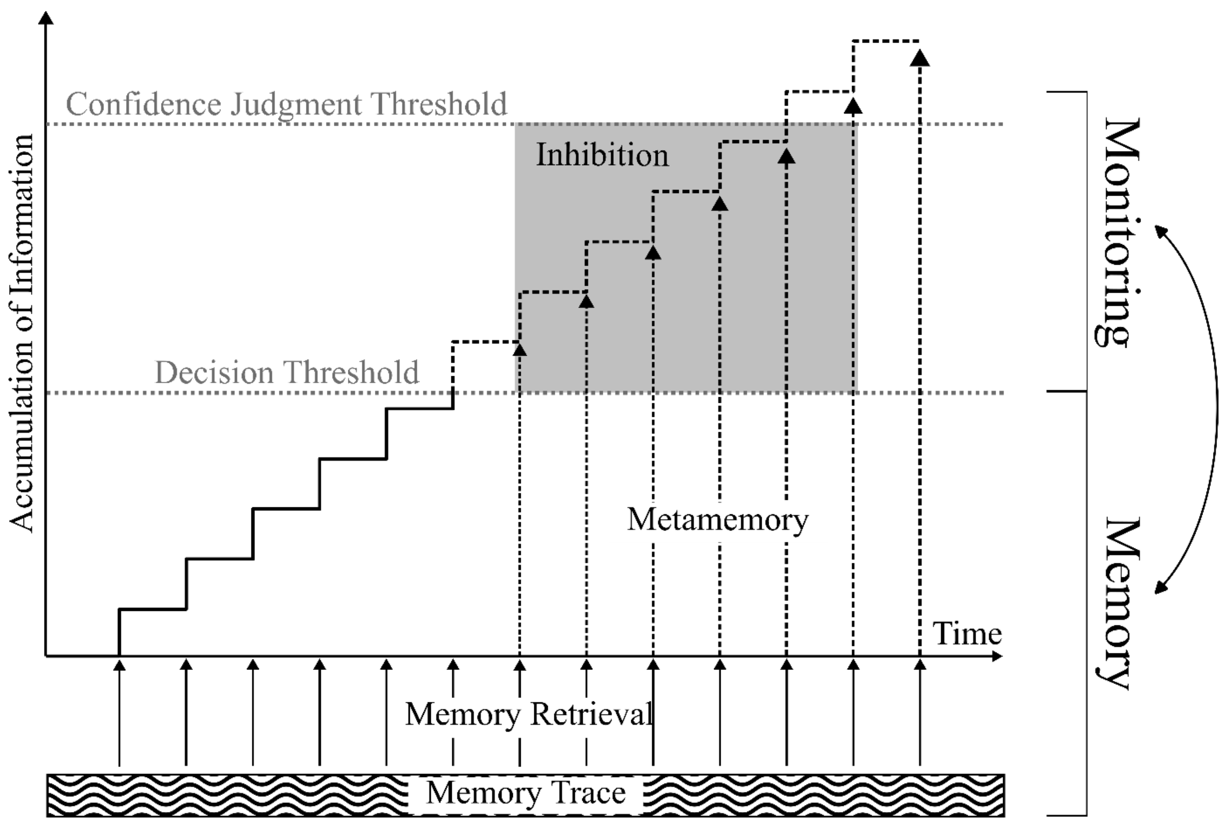
Towards a Joint Model of Monitoring and the Underlying Processes

Including state-of-the-art theories addressing monitoring, memory, and inhibition, I propose a joint framework of monitoring and its underlying processes (see Figure 1). Memory traces serve as an underlying component for the decision-making and monitoring process.

Over time, information is accumulated whereas decision and CJ differ within their threshold. Memory (first-order) and monitoring (second-order) are in constant interaction. The shift from decision-making to monitoring is referred to by the term metamemory, which highlights an intertwined relationship. In addition, inhibition modulates this shift by enabling enough time to switch from first- to second-order functions, to allow further accumulation of information and to engage in monitoring processes.

Figure 1

Joint Framework of Monitoring and the Underlying Processes.



Memory Traces

Memory traces are an underlying neurological structure and can be considered a prerequisite for decision-making and monitoring (Clopath et al., 2012; Robins, 2017). Since our brain can be seen as the center of cognitive processing, the significance of memory for executing cognitive functions such as monitoring is essential (Zlotnik & Vansintjan, 2019). Most importantly, although a memory trace influences subsequent cognitive functions, engaging in cognitive function can also modulate the memory trace (Dash et al., 2004). Hence, when we discuss the role of memory trace in monitoring, it is also essential to consider the opposite direction.

A memory trace contributes to the evaluation process of confidence; however, monitoring evaluation might also affect memory itself (Spellman & Blomell, 2012). As monitoring also induces a retrieval process, this reactivation represents an opportunity to strengthen the underlying memory trace. Due to the attempt to retrieve information from memory, the subsequent processes change and strengthen the existing memory trace and, therefore, change memory itself. Coupling an underlying memory trace with high confidence also affects the strength of this trace (Nadel et al., 2007). Specifically, this means that the underlying information becomes stronger, requiring a more substantial effort to reconsider. This begs the question: why should this answer be reconsidered after a subjectively perceived differential evaluation?

The connection of a specific memory trace was strengthened further due to repeated retrieval of the same stimuli during the tasks set in the first and second study (Clopath et al., 2012). In the second study, evaluating all alternatives and rating one with the highest judgment most likely reinforced the underlying priming of the stimulus and answer option, strengthening the memory trace. Consequently, the subsequent evaluation is based on a strengthened association rather than inducing uncertainty (McClelland, 2013). The previous judgment may influence the corresponding confidence judgment and needs more effort to reconsider a feeling of confidence. Under this point of view, the retrospective confidence judgment represents an additional confirmation of a memory trace.

A strengthened connection might also boost familiarity with the item and increases the time to retrieve this association (Begg et al., 1992). As young children use different cues to guide their behavior, they might get trapped by this strengthened interconnection, increasing the likelihood of being overconfident (Fleming et al., 2018; Talluri et al., 2018). The feeling of the time invested in this evaluation process might be used as a cue for boosting confidence. Regardless of more *'Time to Think'*, as in study I, or a differentiated evaluation, as in study II, both reinforce the subjective time invested and the effort put into evaluation of a specific item. This point of view aligns with the findings from the studies presented in this umbrella paper.

Overall, independent of age, children show a high tendency towards overconfidence in both the first and second studies. Even if the first study revealed small benefits in monitoring accuracy due to a *'Stop and Think'* pause, the tendency towards overconfidence was evident. The same pattern of overconfidence emerged in the second study; however, the results indicated that overconfidence could be more easily boosted than downregulated. This indicated that additional exposure to a *'Pre-Monitoring'* phase boosted overconfidence. These results contribute to the view that overconfidence is a robust phenomenon. As overconfidence also has an evolutionary background and adaptive effects (Johnson & Fowler, 2011) this observation should not necessarily be viewed negatively. Overconfidence can have a positive effect not only in the short term but also in the long term by increasing the probability of success (McKay & Dennett, 2011). Young children still have to learn fundamental skills and, despite numerous failures or setbacks, overconfidence could be a driving force. Nevertheless, overconfident behavior often has negative consequences. Since adults can also be overconfident in dangerous situations, it is essential to consider both sides of the coin (Johnson & Fowler, 2011).

Besides repeated exposure to the same stimuli, visual accentuation can boost confidence and strengthen and stabilize the memory trace (Koriat, 1980; Nadel et al., 2007; Robertson et al., 2004). Compared to the first study, the second study had a more active modification phase by selecting a *'Pre-Monitoring'* judgment for every alternative. Instead of inducing a feeling of uncertainty due to engagement with other alternatives that might also be correct, this *'Pre-Monitoring'* phase evoked a feeling of confidence. This might also influence the subsequent phases and hamper switching to pure monitoring processes.

Accumulation of Information

Results indicate that decision-making and monitoring processes accumulate information over time. Whereas the information emerges from the same baseline, the accumulated information differs between the decision and CJ. Accumulation of information refers to collecting more evidence over time. The threshold represents an individual boundary when enough information is available. Most importantly, between the decision-making and monitoring process, the onset of the thresholds differs based on (or 'from') the timeline. The latter reaches the threshold afterward because it represents a post-decision accumulation after making a decision. The more precise the distinction between these thresholds, the more monitoring emerges independent from a decision-making process. If their boundaries are too closely spaced, monitoring judgment might not be qualitatively different from a decision-making process. The onset of the distinct threshold should have a sufficient time distance from each other to obtain a shift towards a conceptual different accumulation of information.

The *'Stop and Think'* implemented in study I ensures a sufficient distance between these two thresholds. In concrete terms, we have induced a delay to ensure further accumulation of information. Study II does not contain this time window; however, the previous pre-evaluation could also represent a time window allowing an additional accumulation. Since no additional delay was implemented between the subsequent recognition and monitoring phase, the distance between the thresholds tended to converge. Consequently, the boundary is temporally too closely connected to the decision, indicating that the post-decision phase was not a pure monitoring accumulation. Therefore, in the second study, the CJ was more of a confirmation of the decision process, which could strengthen the memory trace and lead to more overconfidence in young children.

Tracking down the mechanisms underlying the *'Stop and Think'* phase or the *'Pre-Monitoring'* phase is challenging. Regardless of where the phase is implemented, it provides an experimental approach for the additional accumulation of information. It represents an accumulation that can affect subsequent recognition and monitoring decisions before recognition. After recognition, it only affects the monitoring judgment. However, additional time could also evoke an invalid accumulation of information and affect the subsequent tasks (Rollwage e al., 2020). Invalid accumulation might result from strengthening a misleading memory trace and collecting more incorrect information.

However, only addressing the time aspect might imply that decision and monitoring differ regarding the time required to finish the accumulation (Stone et al., 2022). Distinguishing these processes only from a time perspective is not exhaustive. Even if the accumulation of information is a necessary process underlying monitoring, inclusion of further neuronal differences between a decision-making process and monitoring should be considered.

Decision-Making and Monitoring Process

I have discussed a shared process underlying monitoring and decision-making, namely memory. Consequently, I addressed that both processes use an accumulation of information emerging from the same memory trace but, at the same time, may differ through subsequent monitoring emerging from post-decision accumulation. Differences between decision-making and monitoring processes will now be addressed.

The decision process can be seen as a first-order execution, whereas monitoring is a second-order function (Yeung et al., 2012). A second-order function is a qualitatively different process because it entails a different way of processing the same underlying information (Fleming et al., 2018; Meyniel et al., 2015; Yeung et al., 2012). Additionally, the goal of an evaluation differs depending on first- *versus* second-order functions. Monitoring depends on the exchange between first- and second-order functions. Within this information flow, monitoring evaluates the first-order information and transforms the output into a feeling of certainty. Neurological studies show that these processes are distinct, and that monitoring activates additional and different neuronal processes than decision-making (De Martino et al., 2013; Grimaldi et al., 2015; Insabato et al., 2010; Qiu et al., 2018).

In the studies presented in this umbrella paper, we have no imaging results and, therefore, can only speculate how our findings would fit into a neurological context. Nevertheless, the first and second studies share some common implications that align with the theoretical assumption mentioned above. We have demonstrated a difference in monitoring accuracy between the age groups, with more sophisticated monitoring skills in the older group as compared to the preschool group (Kuhn, 2000). These findings support that monitoring accuracy is more fine-tuned with age (Lyons & Ghetti, 2011). Maturation of the PFC might contribute to this improvement. Indeed, deficits in second-order functions but not in first-order functions can result from a damaged PFC (Fleming et al., 2014; Rounis et al., 2010). These findings suggest that, although both first- and second-order functions use the same underlying trace, monitoring is another readout of this information (for an overview, see Fleming et al., 2018). Due to age and corresponding maturational processes, older children might be more able to engage in pure monitoring processes than younger children. Moreover, due to neurological readiness, older children might better differentiate between decision-making and monitoring. Younger children might struggle to separate these processes; hence, their CJ is closer to a decision process.

Furthermore, in both studies, children predominantly chose the highest scale point to express their confidence. In the above-discussed section regarding the strengthening of a memory trace, I argued that strengthening could induce overconfidence and might be the reason why a child selects a high CJ. However, another reason could also be that children confirm their selected answer due to a high CJ because the separation between the decision-making and monitoring process is less evident in children than in adults.

Inhibition

Inhibition is a crucial factor influencing the extent to which monitoring processes can engage. Regarding the theoretical background, inhibition seems important in different monitoring steps. First by inhibiting a prepotent response, second by maintaining inhibition for an accumulation of information, and third by modulating the influence of an underlying memory trace (Del Cul et al., 2009; Ratcliff & Starns, 2009; Simpson et al., 2012). The importance of inhibition appears not only on a theoretical but also on a neurological level. Inhibiting a prepotent response might allow the corresponding neuronal monitoring circuits to disseminate neurological signals, inducing a feeling of uncertainty.

Inhibition is another essential key component enabling monitoring processes. Inhibiting a prepotent response could minimize the issue that monitoring relies on the same foundation as the decision-making process (Boldt & Yeung, 2015; Hanks & Summerfield, 2017). Instead, additional neurological correlates of monitoring would have the opportunity to become active and induce feelings based on the spectrum of certainty. Moreover, withholding a hasty response allows for the accumulation of more information which can be used for confidence (Simpson et al., 2007). Furthermore, when new information through accumulation fails to update the current state of the decision, this might negatively influence further cognitive processes; for example, cognitive flexibility and reconsideration of previously learned items (Fleming et al., 2018). In contrast, it is surprising that research addressing the relationship between executive function and metacognition is only partially consistent. Some studies point to a clear link, while others cannot show a connection. Even though the importance of inhibition is evident, there is a need for future research to better understand the reciprocal process between executive function and metacognition (Bryce et al., 2014; Destan et al., 2015). Besides the acquisition method, different executive function and metacognitive methods should be compared to integrate different aspects of each construct.

In our first study, we hypothesized that allowing children more *'Time to Think'* would benefit their monitoring accuracy. This was tested by implementing a pause between recognition and the monitoring judgment (Kälin et al., 2020). The pause should prevent a prepotent response enabling an accumulation of information that goes beyond the decision process by allowing a feeling of uncertainty to spread (Stone et al., 2022; Van Zandt & Maldonado-Molina, 2004). Furthermore, we showed children an animation to aid their engagement in the monitoring processes. With this delay, we tried to disentangle the intertwined processes of memory and metamemory (Baranski & Petrusic, 1998; Bjork, 1994). This approach resulted in more accurate monitoring in our experimental conditions. Since the *'Stop and Think'* condition improved their monitoring accuracy, this could indicate that the support of inhibitory abilities led to more successful monitoring.

The results indicate that our modification might contribute to disentangling these processes. However, a question remains about whether the *'Stop*', the *'Think'*, or the joint *'Stop and Think'* condition results in the biggest improvement (Baranski & Petrusic, 1998; Stone et al., 2022). The *'Stop'* condition allows for the inhibition of a prepotent response and additional accumulation of information, whereas the *'Think'* condition leads to a stronger activation and engagement of monitoring. My expectation would be that the joint *'Stop and Think'* condition would show the biggest improvement, potentially beyond the sum of its parts. Stopping alone does not guarantee engagement in monitoring and only thinking does not prevent prepotent responses, especially in young children. However, further research is needed to confirm this link.

What Are the Processes Underlying Monitoring?

Within this umbrella paper, I highlighted the role of memory, which I consider one of the most important underlying processes. Monitoring is a probability or degree representing a memory trace (Bang et al., 2018) and can be seen as access to memory evaluating how the memory trace can be trusted (Li et al., 2021). The high interrelation between memory and monitoring underscores that they should be considered in tandem.

Furthermore, the information accumulation process should also be considered when understanding monitoring (Kepecs et al., 2008; Merkle & Van Zandt., 2006). Even if the accumulated information emerges from a memory trace, collecting evidence over time should also be incorporated into a monitoring framework (Asok et al., 2019). However, given the theoretical background of metamemory, monitoring should be able to have access to memory but, at the same time, not become entrapped in a (mis)leading memory trace (Rounis et al., 2010; Fleming et al., 2010) which would result in more time being required to accumulate additional information.

Inhibition is a critical component influencing the degree to which pure monitoring can emerge as a further process underlying monitoring (Baranski & Petrusic, 1998; Kälin et al., 2020; Simpson et al., 2007).

The proposed framework of processes underlying monitoring is not exhaustive, and there might be further components modulating monitoring. However, including memory and inhibition is a first step to shed light on hidden monitoring processes (Munakata et al., 2008).

Most importantly, I want to highlight that monitoring is not memory, nor will I argue that memory is more relevant. When looking at the theoretical background, one might suppose that memory traces are so dominant that they will have a strong influence, overshadowing monitoring. Nevertheless, this is different from where my line of argumentation is leading. The overarching goal is not to completely disentangle monitoring from memory. Monitoring will always depend on memory as it is a judgment built on information from memory (Bjork, 1994; Koriat & Goldsmith, 1996). However, retrieving information from memory does not imply that monitoring is not a distinct process. Moreover, pure monitoring entails a part of memory, and sophisticated monitoring skills are distinct and characterized by modulating the strength of the corresponding memory trace. Developing fine-tuned monitoring skills is a highly demanding maturation and can positively influence performance. When these advantages are evident, monitoring can be a powerful tool for a child's development and beyond into adulthood.

Is Monitoring a Distinct Process of Memory or Rather an Additional Accumulation of an Underlying Memory Trace?

Even though the decision-making and monitoring processes can be based on the same underlying memory trace, they are conceptually different. Monitoring can evaluate additional information beyond the decision-making process and represents a different perspective of the same underlying information (Fleming et al., 2014; Florent et al., 2015; Meyniel et al., 2015). However, clearly disentangling these processes is more complex than expected (Meyniel et al., 2015; Spellman et al., 2008). Neurological studies in adults support the view that decision-making and monitoring activate different neuronal patterns. As our studies only included young children, and we have no neurological data supporting this view, we can only discuss the findings from a theoretical point of view.

Nevertheless, the findings indicate that there might be a shift towards more distinct processes throughout development. This shift mainly emerges from the maturation of specific neuronal regions such as PFC (De Martino et al., 2013; Schneider, 2010). This point of view might explain why young children show less accurate monitoring skills and more overconfidence than adults. The findings suggest that separation between these two processes might be difficult, especially at a young age, and children are more prone to select a hasty response. The experience and exposure to engaging with higher-order cognitions in the school context encourage this differentiation.

To What Extent Can We Measure Pure Monitoring Within Young Children?

Provocatively, the above-mentioned findings challenge whether we can measure pure monitoring in young children. Moreover, this question seems more complex than anticipated and demands a more profound understanding of monitoring (Spellman et al., 2008).

Regarding the results of our studies, a CJ cannot be considered an isolated and memory-independent retrieval process. To capture pure monitoring, the CJ should not only rely on the same memory trace *per se* but on a distinct evaluation of confidence regarding the chosen answer.

Not only is there a high interrelation between memory and metamemory but the methodological complexity is also challenging. Our experimental approaches have tried to investigate different methodological aspects. At the same time, our data showed how sensitive such evaluations can be, as well as how they can provoke effects in precisely the opposite direction than expected. It is difficult to separate how these processes are kept apart within the successive evaluation of confidence after a recognition phase. The separation of whether a CJ is driven by a feeling of confidence rather than by an additional confirmation of the memory trace blurs into one another. This consideration indicates that the difficulty of clearly separating these processes leads children to have issues in distinguishing these processes separately. Consequently, the successive task phases cannot capture pure monitoring and may strengthen a cascade of information, with reinforcement from long-term memory resulting in over-optimism. Whether miscalibration emerges from monitoring or is mainly out of inappropriate translation into probabilistic terms, remains open (Koriat, 1980).

Even if the development of theories and a more profound theoretical background is of great importance, sufficient time should be taken for developing new tasks to measure monitoring. Nevertheless, as supported by our experimental designs, just as it is essential to identify what can support children in their monitoring, it is also important to know what does not benefit them. This fundamental research is extremely relevant to better understanding the construct of monitoring and a child's development. The more precise and defined our findings are, the more precise we can design appropriate interventions and research projects to support children in their everyday life. In the literature, a few interventions to improve monitoring have already been suggested and could provide important insights; for example, through feedback (Lipko et al., 2012; O'Leary & Sloutsky, 2017; Oudman et al., 2022; van Loon & Roebers, 2021). The proposed framework of processes underlying monitoring might enrich such intervention studies by drawing attention to these processes and their corresponding challenges.

To conclude, capturing pure monitoring will always be a challenge due to a high interrelation with memory. It may be that, with age, capturing pure monitoring becomes increasingly accessible. However, even then, it will never be completely separable from memory.

Prospects

Monitoring skills are fine-tuned with age. The increasing maturation of neurological correlates for monitoring, starting with school, is an additional driving force (Schneider, 2010). The school context further stimulates development and subsequently supports the child to engage with second-order cognitive function. In our studies, we included children from preschool and second grade. Expanding the age range by including older children and adults would enable further tracking from the developmental perspective (Miller, 2013). Moreover, identifying an age group that would benefit from a pre-monitoring phase would support the assumption that a shift towards pure monitoring processes becomes easier with age.

As most studies have investigated neurological methods in adults only (del Cul et al 2009; De Martino et al., 2013; Lebreton et al., 2015), applying these methods to younger age groups would be essential. Investigating monitoring tasks with additional neurological methods would allow us to get a better understanding of brain regions activated by specific task sequences. We can then track to what extent a child's neuronal responses differ between separate task sequences. Furthermore, comparing children with inaccurate monitoring skills with those with more sophisticated skills would allow us to identify different neurological patterns. One would anticipate additional activated monitoring areas within older age groups compared to the decision-making process.

Furthermore, it would be interesting to adapt our experimental approaches to investigate what kind of support would benefit monitoring accuracy in children. An open question remains about whether and how improvements in monitoring accuracy could be achieved. Theoretically, it might be possible that a delay encourages further processing at the base-level task rather than a more significant deployment of monitoring processes. The delay would, therefore, refer more to the memory than the monitoring process. Structuring the delay to guide the focus more strongly to monitoring could address this concern; for example, through a red audio instruction. From the methodological point of view, another major issue could be caused by children undergoing numerous and repetitive trials (Nadel & Moscovitch, 2001; Robertson et al., 2004). Children might progressively lose interest and completely stop engaging with the later iteration of the task. Hence, as the task progresses, the *'Stop and Think'* phase might no longer have the desired effect. Introducing a more active *'Stop and Think'* phase might be the first step to overcoming this issue.

Additionally, instead of a between-subject study design, I would suggest to include the *'Stop and Think'* as a within factor. More precisely, presenting items with and without the *'Stop and Think'* in a randomized order. Using this experimental approach, we could, therefore, evaluate whether a specific effect is only evident when items are combined with a *'Stop and Think'*. If monitoring accuracy with items accompanied by a *'Stop and Think'* is more accurate than with items without, we can assume that *'Stop and Think'* does indeed have a direct positive effect.

In order to further address the concept evaluated in the second study, it would also be compelling to implement the *'Pre-Monitoring'* phase after recognition. Children would choose an answer in recognition and, before they can select a mostly overconfident judgment in the CJ phase, they would undergo a preliminary monitoring phase. Within this timeline, a delay would have been integrated that is more in line with study one – but potentially more active. Instead of only stopping in this phase we can also control that they actively engage in thinking.

Evaluating answer alternatives after the decision would increase the probability that the memory trace from the decision-making process is strengthened and would be contrary to our original idea. However, as the results indicate that, regardless of the time point, the underlying memory trace interferes with monitoring this should not affect monitoring accuracy. Instead, especially because memory traces seem to be an underlying and dominant process of monitoring, it would be more relevant to activate monitoring rather than trying to reduce the memory trace.

The more important aspect seems to be in supporting children in distinguishing these processes and achieving the transition to higher cognitive functions. Within this point of view, enabling children to differentiate between alternatives after recognition but before rating a CJ might let them reconsider their prior selected answer and prepare them to slowly shift towards monitoring processes.

Furthermore, including implicit measures could expand our understanding of monitoring (Ackerman & Koriat, 2011). This could involve not only evaluating how long children take for monitoring judgment but also tracking of their eye movements. Tracking eye movements are a direct indicator of cognitive processing by the brain (Gottlieb et al., 2013). Leckey et al. (2020) investigated gaze switching between answer alternatives and found response latencies as a cue for confidence. Gaze switching was longer when items were difficult, indicating longer reaction times for evaluating more demanding tasks. Information seeking is linked to subjective confidence (Desender et al., 2018). A future study including eye tracking might open new insights into the underlying processes. Evaluating how long children looked at the answer options, how they compared different alternatives, or if they looked at the alternatives within the monitoring process, could be an enriching method helping to understand information processes and monitoring (De Martino et al., 2013; Gottlieb et al., 2013).

At this point, it is important to emphasize that our primary intention was not to force children to become progressively better at monitoring. Instead, we intended to gain a better understanding of the processes underlying monitoring, in order to obtain a more comprehensive picture of a child's development. Each child shows an individual development trajectory. It is not the child who has to change but rather the environment must accept individual trajectories. However, a more profound understanding of monitoring can be used to understand a child's development beyond the expected norm. Consequently, future research should address results not only on a group but also on an individual level. For example, in our second study, some children did benefit from a *'Pre-Monitoring'* judgment. However, these children were in the minority; hence, their results were overshadowed by other subgroups. The use of latent profile analysis could shed light on the same data from a different perspective. For example, the EF and MC relationship could be more comprehensively investigated individually, including different measures of EF and MC to obtain a broader representation of the constructs. Individual profiles in executive functions could be compared to monitoring outcomes. Within these individual EF profiles, there might be differences in monitoring.

Conclusion

The present dissertation contributes to a more profound understanding of monitoring in young children. Whereas the studies used in this dissertation shed light on the processes underlying monitoring using an experimental approach, the umbrella paper expands these results from a different point of view. Specifically, the integration of the underlying neurological perspective, and the significance of memory and inhibition, contribute to a differentiated framework of monitoring (Munakata et al., 2008). Monitoring refers to a subjective estimation of the likelihood that a chosen answer might be correct (Fleming, 2017).

Even if monitoring and memory seem to overlap more often than expected, they can still be separated on a neurological level (Grimaldi et al. 2015). Monitoring will always depend on memory as it depends on information from a memory trace. However, while monitoring requires memory, sophisticated monitoring skills are characterized by modulating the strength of the corresponding memory trace.

References

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology*: Applied, *17*(4), 406–417. https://doi.org/10.1037/a0025129

Asok, A., Leroy, F., Rayman, J. B., & Kandel, E. R. (2019). Molecular mechanisms of the memory trace. *Trends in Neurosciences*, *42*(1), 14-22. https://doi.org/10.1016/j.tins.2018.10.005

Bakracevic Vukman, K., & Licardo, M. (2010). How cognitive, metacognitive, motivational and emotional self-regulation infuence school performance in adolescence and early adulthood. *Educational Studies*, 36(3), 259–268. https://doi.org/10.1080/03055690903180376

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *115*(23), 6082-6087. https://doi.org/10.1073/pnas.1800795115

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology*: *Human Perception and Performance*, *24*(3), 929-945. https://psycnet.apa.org/doi/10.1037/0096-1523.24.3.929

Bayard, N. S., van Loon, M. H., Steiner, M., & Roebers, C. M. (2021). Developmental Improvements and Persisting Difficulties in Children's Metacognitive Monitoring and Control Skills: Cross-Sectional and Longitudinal Perspectives. *Child Development*, *92*(3), 1118–1136. https://doi.org/10.1111/cdev.13486

Begg, I. M, Anas, A., & Farinacci, S. (1992). Dissociation of process in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121, 446-458. https://psycnet.apa.org/doi/10.1037/0096-3445.121.4.446

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing About Knowing* (pp. 185–205). MIT Press.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478-3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bryce, D., Whitebread, D., & Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning*, *10*(2), 181–198. https://doi.org/10.1007/s11409-014-9120-4

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26 – 4.

Carlén, M. (2017). What constitutes the prefrontal cortex?. *Science*, *358*(6362), 478-482. https://doi.org/10.1126/science.aan8868

Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, *29*(4), 1150-1160. https://doi.org/10.1016/j.neuroimage.2005.09.058

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, *21*(9), 1751-1765. https://doi.org/10.1162/jocn.2009.21123

Cleremans, A. (2017). As Above, So Below: Tangled Loops between Consciousness and the Unconscious. *Before Consciousness: In Search of the Fundamentals of Mind*, 83-105.

Clopath, C. (2012). Synaptic consolidation: an approach to long-term learning. *Cognitive Neurodynamics*, *6*, 251-257.

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science*, *18*(6), 957–971. https://doi.org/10.1111/desc.12271

Dash, P. K., Hebert, A. E., & Runyan, J. D. (2004). A unified theory for systems and cellular memory consolidation. *Brain Research Reviews*, *45*(1), 30-37. https://doi.org/10.1016/j.brainresrev.2004.02.001

de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

del Cul, A., Dehaene, S., Reyes, P., Bravo, E., Slachevsky, A., 2009. Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain 132*(9), 2531–2540. https://doi.org/10.1093/brain/awp111

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105-110. http://dx.doi.org/10.1038/nn.3279

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761-778. https://doi.org/10.1177/0956797617744771

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

Destan, N., & Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfdence? *Metacognition and Learning*, *10*(3), 347–374. https://doi.org/10.1007/s11409-014-9133-z

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunlosky, J., & Rawson, K. A. (2012). Overconfdence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational psychologist*, *46*(1), 6-25. https://doi.org/10.1080/00461520.2011.538645

Finn, B., & Metcalfe, J. (2014). Overconfdence in children's multi-trial judgments of learning. *Learning and Instruction*, 32, 1–9. https://doi.org/10.1016/j.learninstruc.2014.01.001

Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, *2*(4), 421-453. https://doi.org/10.1016/0010-0285(71)90025-9

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1338–1349. https://doi.org/10.1007/978-3-642-45190-4_11

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, *32*(18), 6117-6125. https://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811-2822. https://doi.org/10.1093/brain/awu221

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617-624.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541-1543. https://doi.org/10.1126/science.1191883

Geurten, M., & Willems, S. (2016). Metacognition in early childhood: Fertile ground to understand memory development?. *Child Development Perspectives*, *10*(4), 263-268. https://doi.org/10.1111/cdep.12201

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585-593. https://doi.org/10.1016/j.tics.2013.09.001

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, *55*, 88-97. https://doi.org/10.1016/j.neubiorev.2015.04.006

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 429–456). Erlbaum Publishers.

Hanks, T. D., & Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron*, *93*(1), 15-31. http://dx.doi.org/10.1016/j.neuron.2016.12.003

Harris, K. R., Santangelo, T., & Graham, S. (2010). Metacognition and strategies instruction in writing. *Metacognition, Strategy use, and Instruction*, 226-256.

Hayes, S. M., Ryan, L., Schnyer, D. M., & Nadel, L. (2004). An fMRI study of episodic memory: retrieval of object, spatial, and temporal information. *Behavioral Neuroscience*, *118*(5), 885. https://doi.org/10.1037%2F0735-7044.118.5.885

Insabato, A., Pannunzi, M., Rolls, E. T., & Deco, G. (2010). Confidence-related decision making. *Journal of Neurophysiology*, *104*(1), 539-547. https://doi.org/10.1152/jn.01068.2009

Jiao, X., Zhang, A., & Bu, X. (2023). Relating metacognition and executive functions to early mathematical and language skills in children aged 4–5 years. *Metacognition and Learning*, 1-16.

Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317-320. https://doi.org/10.1038/nature10384

Kälin, S., & Roebers, C. M. (2020). Time-based measures of monitoring in association with executive functions in kindergarten children. *Zeitschrift für Psychologie*, *228*(4), 244-253. https://doi.org/10.1027/2151-2604/a000422

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227-231.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329-1342. http://dx.doi.org/10.1016/j.neuron.2014.12.015

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759-764. https://doi.org/10.1126/science.1169405

Kim, H. (2020). An integrative model of network activity during episodic memory retrieval and a meta-analysis of fMRI studies on source memory retrieval. *Brain Research*, 1747, 147049. https://doi.org/10.1016/j.brainres.2020.147049

Klein, S. B. (2015). What memory is. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(1), 1-38. https://doi.org/10.1002/wcs.1333

Kolb, B., Mychasiuk, R., Muhammad, A., Li, Y., Frost, D. O., & Gibb, R. (2012). Experience and the developing prefrontal cortex. *Proceedings of the National Academy of Sciences*, *109*(supplement_2), 17186-17193. https://doi.org/10.1073/pnas.1121251109

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609-639. https://psycnet.apa.org/doi/10.1037/0033-295X.100.4.609

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. Journal of Experimental Psychology: *Human Learning and Memory*, *6*(2), 107. https://psycnet.apa.org/doi/10.1037/0278-7393.6.2.107

Kuhn D. (2000). Metacognitive development. *Current Directions in Psychological Science*, *9*(5), 178–181. https://doi.org/10.1111/1467-8721.00088

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159-1167.

Leckey, S., Selmeczy, D., Kazemi, A., Johnson, E. G., Hembacher, E., & Ghetti, S. (2020). Response latencies and eye gaze provide insight on how toddlers gather evidence under uncertainty. *Nature Human Behaviour*, *4*(9), 928–936.

Li, H. H., Sprague, T. C., Yoo, A. H., Ma, W. J., & Curtis, C. E. (2021). Joint representation of working memory and uncertainty in human cortex. *Neuron*, *109*(22), 3699-3712. https://doi.org/10.1016/j.neuron.2021.08.022

Lipko, A. R., Dunlosky, J., Lipowski, S. L., & Merriman, W. E. (2012). Young children are not underconfdent with practice: The beneft of ignoring a fallible memory heuristic. *Journal of Cognition and Development*, *13*(2), 174–188. https://doi.org/10.1080/15248372.2011.577760

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfdence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, *103*(2), 152–166. https://doi.org/10.1016/j.jecp.2008.10.002

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*(6), 1778-1787.

https://doi.org/10.1111/j.1467-8624.2011.01649.x

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422– 430. http://dx.doi.org/10.1016/j .concog.2011.09.021

Marulis, L. M., & Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacognition and Learning*, *16*(1), 207-231.

McAlpine, L., Weston, C., Beauchamp, C., Wiseman, C., & Beauchamp, J. (1999). Building a metacognitive model of reflection. *Higher Education*, 37, 105-131.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190. https://psycnet.apa.org/doi/10.1037/a0033812

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, *32*(6), 493-510. https://doi.org/10.1017/S0140525X09990975

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*(1), 78-92. http://dx.doi.org/10.1016/j.neuron.2015.09.039

Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, *135*(3), 391– 408. https://psycnet.apa.org/doi/10.1037/0096-3445.135.3.391

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

Miller, P. H. (2013). The history of memory development research: remembering our roots. *The Wiley Handbook on the Development of Children's Memory*, 1, 15-40. https://doi.org/10.1002/9781118597705.ch2

Morgado-Bernal, I. (2011). Learning and memory consolidation: linking molecular and behavioral data. *Neuroscience*, 176, 12-19. https://doi.org/10.1016/j.neuroscience.2010.12.056

Moritz, S., Gläscher, J., Sommer, T., Büchel, C., & Braus, D. F. (2006). Neural correlates of memory confidence. *Neuroimage*, *33*(4), 1188-1193. https://doi.org/10.1016/j.neuroimage.2006.08.003

Munakata, Y., Stedron, J. M., Chatham, C. H., & Kharitonova, M. (2008). 22 Neural Network Models of Cognitive Development. *Handbook of Developmental Cognitive Neuroscience*, 367.

Muncer, G., Higham, P. A., Gosling, C. J., Cortese, S., Wood-Downie, H., & Hadwin, J. A. (2022). A meta-analysis investigating the association between metacognition and math performance in adolescence. *Educational Psychology Review*, *34*(1), 301-334.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217-227. https://doi.org/10.1016/S0959-4388(97)80010-4

Nadel, L., Winocur, G., Ryan, L., & Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. *Debates in Neuroscience*, 1, 55-66. https://doi.org/10.1007/s11559-007-9003-9

Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in confidence. *Current Opinion in Behavioral Sciences*, 11, 55– 60. http://dx.doi.org/10.1016/j.cobeha.2016.05.005

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125-141. https://doi.org/10.1016/S0079-7421(08)60053-5

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning*, 13, 179-212. https://doi.org/10.1007/s11409-018-9183-8

O'Leary, A. P., & Sloutsky, V. M. (2017). Carving metacognition at its joints: Protracted development of component processes. *Child Development*, *88*(3), 1015–1032. https://doi.org/10.1111/cdev.12644

Oudman, S., van de Pol, J., & van Gog, T. (2022). Efects of self-scoring their math problem solutions on primary school students' monitoring and regulation. *Metacognition and Learning*, *17*(1), 213–239. https://doi.org/10.1007/s11409-021-09281-9

Pasquali, A., Timmermans, B., Cleeremans, A., 2010. Know thyself: metacognitive networks and measures of consciousness. *Cognition*, *117*(2), 182–190. https://doi.org/10.1016/j.cognition.2010.08.010

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S. et al. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, *117*(15), 8382-8390. https://doi.org/10.1073/pnas.1918335117

Peters, M. A., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M. et al. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*, *1*(7), 0139.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864. https://psycnet.apa.org/doi/10.1037/a0019737

Qiu, L., Su, J., Ni, Y., Bai, Y., Zhang, X., Li, X., & Wan, X. (2018). The neural system of metacognition accompanying decision-making in the prefrontal cortex. *PLoS Biology*, *16*(4), e2004037. https://doi.org/10.1371/journal.pbio.2004037

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664-1680. https://doi.org/10.1177/0956797615595037

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59. https://psycnet.apa.org/doi/10.1037/0033-295X.85.2.59

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological review*, *116*(1), 59. https://psycnet.apa.org/doi/10.1037/a0014086

Robertson, E. M., Pascual-Leone, A., & Miall, R. C. (2004). Current concepts in procedural consolidation. *Nature Reviews Neuroscience*, *5*(7), 576-582.

Robins, S. K. (2017). Memory traces. *The Routledge Handbook of Philosophy of Memory*, 1, (pp. 76-87). Guilford Press.

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

Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, 7, 151-173.

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N. S., & van Loon, M. H. (2019). The role of children's metacognitive experiences for cue utilization and monitoring accuracy: A longitudinal study. *Developmental Psychology*, *55*(10), 2077. https://doi.org/10.1037/dev0000776

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confdence drives a neural confrmation bias. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-16278-6.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165-175. https://doi.org/10.1080/17588921003632529

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Frontiers in Psychology*, 7, 851. https://doi.org/10.3389/fpsyg.2016.00851

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114-121. https://doi.org/10.1093/oxfordhb/9780199336746.013.1

Schneider, W. (2010). Metacognition and memory development in childhood and adolescence. In H. S. Waters & W. Schneider (Eds.), *Metacognition, Strategy Use, and Instruction* (pp. 54–84). New York, NY: Guilford Press.

Schneider, W., & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied Metacognition* (pp. 224–257). Cambridge University Press.

Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions?. *Developmental Psychology*, *43*(2), 417. https://doi.org/10.1037/0012-1649.43.2.417

Simpson, A., & Riggs, K. J. (2011). Under what conditions do children have difficulty in inhibiting imitation? Evidence for the importance of planning specific responses. *Journal of Experimental Child Psychology*, *109*(4), 512-524. https://doi.org/10.1016/j.jecp.2011.02.015

Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory processes: How response prepotency is created and overcome. *Developmental Science*, *15*(1), 62-73. https://doi.org/10.1111/j.1467-7687.2011.01105.x

Spellman, B. A., Bloomfield, A., & Bjork, R. A. (2008). Measuring memory and metamemory: Theoretical and statistical problems with assessing learning. *Handbook of Metamemory and Memory*, (pp. 95-101). Oxford University Press.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, 120, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195-231. https://doi.org/10.1037/0033-295X.99.2.195

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766.

Stone, C., Mattingley, J. B., & Rangelov, D. (2022). On second thoughts: changes of mind in decision-making. *Trends in Cognitive Sciences, 26*(5), 419-431. https://doi.org/10.1016/j.tics.2022.02.004

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confrmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, *28*(19), 3128–3135. https://doi.org/10.1016/j.cub.2018.07.052

Teffer, K., & Semendeferi, K. (2012). Human prefrontal cortex: evolution, development, and pathology. *Progress in Brain Research*, 195, 191-218. https://doi.org/10.1016/B978-0-444-53860-4.00009-X

Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. *In Handbook of Metacognition in Education* (pp. 107-127). Routledge.

van Loon, M. H., Bayard, N. S., Steiner, M., & Roebers, C. M. (2022). The accuracy and annual rank-order stability of elementary school children's self-monitoring judgments. *Journal of Applied Developmental Psychology*, 80, 101419. https://doi.org/10.1016/j.appdev.2022.101419

van Loon, M. H., & Roebers, C. M. (2021). Using feedback to support children when monitoring and controlling their learning. In D. Moraitou & P. Metallidou (Eds.), *Trends and Prospects in Metacognition Research Across the Life Span,* (pp. 161 -184). Springer. https://doi.org/10.1007/978-3-030-51673-4_8

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1147–1166. https://doi.org/10.1037/0278-7393.30.6.1147

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418-424. https://doi.org/10.1016/j.tics.2008.07.005

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37-58. https://doi.org/10.1080/00140137008931117

Vickers D. (1979). Decision processes in visual perception. New York, NY: Academic Press.

Yeon, J., Shekhar, M., & Rahnev, D. (2020). Overlapping and unique neural circuits are activated during perceptual decision making and confidence. *Scientific Reports*, *10*(1), 20761.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310-1321. doi:10.1098/rstb.2011.0416

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, 144, 489 –510. http://dx.doi.org/10.1037/xge0000062

Zlotnik, G., & Vansintjan, A. (2019). Memory: An extended definition. *Frontiers in Psychology*, 10, 2523. https://doi.org/10.3389/fpsyg.2019.02523

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes über die Universität vom 5. September 1996 und Artikel 69 des Universitätsstatuts vom 7. Juni 2011 zum Entzug des Doktortitels berechtigt ist. Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die Dissertation zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.

Bern, August 2023 Sophie Wacker

Acknowledgments

First and foremost, I would like to thank Claudia Roebers for giving me the opportunity of doing a Ph.D. I sincerely appreciated our interesting discussions, your guidance through the past three years, and the very happy moments dancing in the office when a paper was accepted. Your impressive knowledge and joy in research are a very powerful inspiration for me. Furthermore, I would like to thank Mariëtte van Loon, who agreed to be my second supervisor. I would also like to express a special thank you to Niamh and Florian for giving me feedback on my manuscript and to Chris for proofreading it. Additionally, I am very grateful to be part of such a wonderful and inspiring team - thank you for your support! Finally, I would like to express the warmest thank you to my mom, Therese, my partner Luzi, my family, and my friends for being part of this journey. From the bottom of my heart, I am deeply grateful for all the support and love I have received.

Appendix A: Study I

Stop and think: Additional time supports monitoring processes in young children

Sophie Wacker & Claudia M. Roebers

University of Bern

RESEARCH ARTICLE

Stop and think: Additional time supports monitoring processes in young children

Sophie Wacker*, **Claudia M. Roebers**

Department of Psychology, University of Bern, Bern, Switzerland

* sophie.wacker@unibe.ch

Citation: Wacker S, Roebers CM (2022) Stop and think: Additional time supports monitoring processes in young children. PLoS ONE 17(9): e0274460. https://doi.org/10.1371/journal.pone.0274460

Editor: Micah B. Goldwater, University of Sydney, AUSTRALIA

Received: February 22, 2022

Accepted: August 30, 2022

Published: September 15, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0274460

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Abstract

When children evaluate their certainty, monitoring is often inaccurate. Even though young children struggle to estimate their confidence, existing research shows that monitoring skills are developing earlier than expected. Using a paired associates learning task with integrated monitoring, we implemented a time window to—"Stop and Think"—before children generated their answers and evaluated their confidence in the chosen response. Results show that kindergarten and second grade children in the—"Stop and Think"—condition have higher monitoring accuracy than the control group. Implementing a time window thus seems to support children in their evaluation of different certainty levels. Relating individual differences in independently measured inhibitory control skills revealed a correlation between monitoring and inhibition for kindergarteners.

Introduction

Metacognition research consistently reveals that young children show inaccurate monitoring skills. That is, they are overly optimistic when evaluating their performance. Accurate monitoring is important for a wide range of cognitive domains, including academic achievement [1]. Although often overoptimistic, young children have been found to be able to accurately monitor their performance (for example, in everyday life asking back when ambiguous information is provided or in play situations hesitating when executing an ambiguous demand [2–4]. The current approach tests the possibility that children's inaccurate monitoring is—at least in part—due to young children not taking enough time to engage in monitoring processes actively. We will explore this question in two ways. For one, we will experimentally induce a time window during which children are asked to monitor and compare different responses regarding their likelihood of being the correct answer. For another, we will independently quantify participants' inhibitory control skills and relate them to their monitoring ability.

Monitoring is a fundamental part of metacognition [5, 6], describing an individual's capability to reflect and supervise cognitive processes [4]. There are several methods to measure different monitoring aspects. Monitoring processes can be measured before a memory test and are called prospective judgments. These include, for example, judgments of learning or feelings of knowing [7]. The present study focused on retrospective monitoring processes, measured after a memory test, and described as confidence judgments. Children, adolescents, and adults experience and report their confidence on different levels, ranging from very

Competing interests: The authors have declared that no competing interests exist.

unsure to very sure. Thereby, kindergarten and primary school children have more difficulties estimating their performance accurately than older individuals [3, 8–10]. That is, they report being really sure, often independent of their response's accuracy. This overconfidence is partly due to imprecise monitoring skills [11–14]. Even for incorrect answers, young children often give high confidence judgments, suggesting their ability to reflect on certainty is far from fully developed at this age. During the primary school years, monitoring becomes more sophisticated and differentiated and shows an increasing congruency with actual performance [15, 16].

The theoretical background of the present approach is a broader conceptualization of higher-order cognitive self-regulation, entailing metacognition and executive functions [17, 18]. Self-regulatory skills are increasingly recognized to embrace executive functions (EF) and monitoring [19–23]. Executive functions are top-down regulating processes and include updating, shifting, and inhibition aspects [24]. Especially inhibitory control skills are needed for many everyday tasks, including learning and monitoring [25]. In younger children, inhibitory control and metacognition are assumed to be connected interactively. For example, recent results reported by Kälin and Roebers [21] uncovered an association between monitoring and executive functions. The authors showed that better inhibition was related to longer latencies when giving confidence judgments in a paired associates learning and recognition paradigm. This suggests that children with better inhibitory skills took more time to report their confidence in the selected answer. Most interestingly, these longer latencies resulted in more accurate monitoring judgments. In other words, children's monitoring accuracy was better when they took longer to generate a monitoring judgment—according to the concept of "more time to think".

With increasing age and experience, monitoring and EF are thought to differentiate and follow distinct developmental trajectories [18]. Based on their findings, Roebers et al. [22] postulated that well-developed EF are necessary to develop metacognitive skills. Deficits in monitoring processes could result from immature executive functions because a certain level of EF skills is needed to perform metacognition successfully. Therefore, the association might be stronger in younger than older children. This assumption is corroborated by findings showing that in 5- compared to 7-years-olds, monitoring is more closely related to inhibitory control skills [19]. Most recent longitudinal research addressing the interrelation of those constructs revealed that EF at an early age predicts self-regulated learning one year later but not vice versa [26]. The present study included kindergarten children and second graders to confirm previous findings: we aimed to further explore that inhibition may indeed be more critical for younger compared to older children's monitoring accuracy, with the assumption that school attendance and academic tasks gradually train children's monitoring skills [4].

From a neuropsychological perspective, inhibition might serve as a monitoring prerequisite. To engage in monitoring, the responsible neural networks need time to loop signals from the anterior cingulate cortex (ACC) to frontal structures [27]. Prefrontal structures, especially the ACC, are considered a neurological correlate for monitoring and cognitive processing [28]. Therefore, neurological monitoring signals need time for transmission [29]. Consequently, immature inhibitory skills may not provide enough time for these signals to be strengthened and passed on to related neural structures [30]. Only if an individual takes enough time to process information, monitoring aspects can come into effect [17, 31, 32]. In other words, if individuals can inhibit their prepotent responses, hesitate and ask themselves: Am I really sure about my answer?, this should benefit their monitoring accuracy [33].

However, for these processes, one must develop and experience a feeling of uncertainty. The engagement with uncertainty (carefully evaluating the own levels of certainty) may trigger metacognitive processing and can result in better performance (due to a more differentiated and conscious evaluation) [17]. Indirect evidence supporting the view that time may play a crucial role for monitoring accuracy stems from research on the delayed JOL-effect. Delayed compared to immediate judgments-of-learning are typically more accurate, both in adults and in children [34, 35]. In cognitive tasks, for example, in a memory test, experimentally inducing a delayed response by providing additional time before responding has repeatedly been found to be an efficient means to increase the accuracy of children's responses [25, 36–38]. Simpson et al. [39] showed that if a child must wait a set time to generate an answer, this answer was more likely to be correct than answering immediately. Poor task performance can result from a prepotent response. With additional time, reflective processing may result in better performance.

None of these studies have yet applied this concept to monitoring. We will build on these findings and explore the extent to which a "Stop and Think" instruction may positively affect children's monitoring accuracy. Additionally, research focusing on the accuracy of a confidence judgment based on the prior answer showed that information processing does not end after the decision is made [40]. On the contrary, the accumulation of further information processing evolves during the interval between an answer and the corresponding confidence judgment. This accumulation may also profit from more time which is in line with our assumption. Inhibiting the prepotent response and allowing neurological signals to strengthen [41, 42] may also allow the accumulation of additional information, which may be guided by monitoring processes.

To our knowledge, no study tried to explore the influence of increased time to monitoring on children's monitoring accuracy. In an experimental setting, we implemented a delay during which the child should "Stop and Think". In the present study children solved a paired associate learning task. After studying several item pairs, subjects had to choose one out of four answer alternatives that matched the corresponding stimulus picture (recognition phase). The "Stop and Think" delay was inserted after the recognition and before the subsequent monitoring. Afterwards, children had to select a confidence judgment by rating how sure they were that they chose the correct item pair. We hypothesized that being "forced" to take more time to monitor and prevent fast and thus undifferentiated monitoring judgments would positively affect children's monitoring skills temporarily. More time until monitoring judgments are given may allow the individual to pause and reflect on the ongoing cognitive and metacognitive processes, ideally leading to better monitoring. We expected small benefits from additional time against the background of the above-mentioned findings [19, 43]. To evaluate the impact of additional time on different aspects of monitoring accuracy, we analyzed a relative (monitoring discrimination, i.e., the difference in confidence between correct and incorrect responses [44]) and an absolute score (i.e., overconfidence, the deviation of certainty from performance). We did not expect any effect on recognition as the delay was only inserted after participants had chosen an alternative.

From an individual differences perspective and in parallel to the theoretical background outlined above on the relation between inhibition and monitoring, inhibition might be a candidate factor contributing to high confidence in children, irrespective of performance [19]. One might expect that better inhibition allows the child to hesitate instead of jumping on an answer and reporting high confidence, and to reflect on the likelihood of different alternatives to be correct and thus to monitor more accurately. However, more research is needed to understand the relation between monitoring and inhibition. Despite intensive research on metacognition and its development, relatively little attention has been paid to individual differences within homogenous age groups.

The preschool and kindergarten age represents a critical time window for executive function development [45, 46]. In cognitive tasks requiring inhibitory control, findings show that younger compared to older children benefit more from a delay [36, 47]. These results indicate that younger children need more support in inhibiting their impulsive behavior. Because of developmental maturation and still relatively immature inhibition functions [19, 48], we hypothesize that kindergarten children would benefit more from a "Stop and Think!" instruction compared to second graders.

To address the role of individual differences for monitoring beyond our experimental manipulation, we also assessed inhibitory control skills independently from metacognition. This allows us to explore the relationship between inhibition and monitoring accuracy in the control condition in which no "Stop and Think!" instruction and no delayed monitoring judgments were induced. We expected that individual differences in inhibition would be weakly but positively related to monitoring accuracy in younger but not necessarily in older (school) children in the control condition. In addition, we examined the relationship between inhibition and monitoring accuracy in the experimental group (only children with the "Stop and Think" instruction). Thus, we investigated whether there are individual differences regarding the extent to which a delay can contribute to improving monitoring accuracy. For example, children with poor inhibition might benefit more from a delay than children with already sophisticated inhibitory control skills.

Methods

Participants

Data stems from N = 393 children from rural and urban areas in a mid-European country. For the analysis, we recruited a sample of N = 202 (44.6% female) kindergartners between 4–6 years of age (M = 73.6 months, SD = 7.4 months) and N = 191 second graders (45.5% female) between 7–9 years of age (M = 94.2 months, SD = 7.1 months). Participants represent a sample of middle-class families mostly of Caucasian descent. The Ethics Committee of the Faculty of Human Sciences at the University of Bern approved ethical consent for the study (Approval No: 2002–100005). The parents or legal guardians of all participating children signed an informed consent. Further, all children were asked verbally to participate prior to the testing. They were further explained that they could terminate the task at any time. No child ever did. Data is entirely anonymous. Due to technical problems, we excluded N = 2 participants. Additionally, during one test session, N = 20 children had to quarantine due to COVID-19. These children were also excluded from the analyses reported below because they did not solve the paired associate learning task. Due to the current restrictions, there was no opportunity to retest them. We had to exclude N = 4 participants with an accuracy of 0% or 100% in the recognition block for the paired association task as they did not generate complete monitoring data. Restrictions due to COVID -19 and because several children had to be in quarantine, we could not examine inhibition data of all children. Therefore, the Heart and Flower task analysis is limited to an N = 330.

Procedure and measures

Children performed two different computer-based tasks, running on tablets (Samsung Galaxy S6). During the study, trained investigators were present. Test sessions took place in a group setting in children's schools, with each participant listening to the pre-recorded instructions through headphones. Children solved a paired associates learning task with integrated monitoring (30 to 40 min.). In this task, children were to log in their answers by touching predefined areas on the screen with their index fingers. The children solved a paired association learning task encased in a cover story of two children to assess the monitoring aspect. Following a familiarization phase, the task was composed of 3 phases. In the first, the learning phase, participants learned different numbers of item pairs (kindergarteners: 16 items, second graders: 22 items). Each item pair was presented for 4s. After the learning phase, participants solved a filler task for about 1 minute, followed by a recognition phase. Participants were shown one constituent from an item pair and had to choose one out of four possible answers as being the matching item. There was no time limit for choosing the matching item. After choosing an answer in the recognition phase, participants were immediately asked to provide a confidence judgment (CJ) for their selection in the final monitoring phase. Participants had to indicate their certainty on a 7-point Likert scale (adapted from [27]).

Children were randomly assigned to either the control group (CG; they solved the task as described above) or the experimental group (EG). Participants in the EG had to wait a set time before choosing a CJ. Research suggests that the diffusion of neurological signals to the prefrontal cortex needs about 200–250 ms [20, 30]. Other studies found that implementing a delay of 4 seconds leads to a performance improvement [36]. Therefore, we chose an interval representing a reasonable pause allowing enough time for diffusion and time for reflection. For this purpose, we implemented a fixed delay of 8s before participants could choose a CJ. Throughout this 8s delay, an animation was implemented, during which the pictures became gradually transparent and smaller. At the same time, one out of the two protagonists appeared with a big speech bubble containing the thermometer. This sequence represented the protagonists showing that they are taking time to think about their answers and their certainty following a pattern:—"Stop and Think"—All other procedures did not differ from the control group.

We generated 12 item pairs with medium difficulty (index with .57) for the kindergarteners and 15 items (index with .60) for the second graders. Item pairs for kindergarteners with a very high (index below .32, N = 2) and very low (index above .77, N = 2) difficulty and for second graders correspondingly (index below .32, N = 5; index above .77, N = 2), served as anchor items and were not used for the analysis. To address relative aspects of monitoring, we calculated a discrimination score to quantify the ability to discriminate between CJ for correct and CJ for incorrect answers [49, 50]. Additionally, we used the bias index for absolute aspects of monitoring [50]. The bias index maps to a continuous range between underestimation (-1), accurate estimation (0), and overestimation (+1).

In another session (15 min.), with a minimum delay of one week, each child solved the Hearts and Flowers task capturing inhibition and cognitive flexibility [51, 52]. For this task, two external response buttons were connected to the computer and placed on the right and left sides of the screen. In the congruent condition (heart block; N = 24 trials), a heart appeared on the right or the left side of the screen. Children had to press the button on the same side where the heart appeared. In the subsequent incongruent condition (flower block; N = 36 trials), children were to press the button on the opposite side of where the flower appeared. In the final mixed block, congruent (heart) and incongruent (flower) trials were combined and appeared in pseudorandomized order (N = 60 trials). The presentation of the stimuli was during 2500 ms, followed by an interstimulus interval of 500 ms. *Dependent Variables*. We calculated the Rate Correct Score (RCS) for every block [53], reflecting the amount of correctly solved items per second. For the Hearts and Flowers task, we excluded (N = 34) participants because overall accuracy was lower than .50 (below change level). Reaction times under 200 ms were excluded as they typically represent reflexes or second corrective responses to the previous trial. Our primary interest lay in the RCS of the flower block, which is considered to represent mainly inhibition [51].

Statistical analysis

Our study follows a 2 (control vs. experimental group) x 2 (kindergarteners vs. second graders) between-subject design. We used Scipy, Numpy, Pandas, and StatsModel, running on Python for data analysis, and Seaborn and Matplotlib for data visualization. Our dependent variables were examined by between-subject analysis of variances (ANOVAs) concerning monitoring. With partial eta squared (η_p^2), we estimated the effect sizes. To explore the relationship between inhibition and monitoring accuracy, we evaluated correlations analysis and reported their corresponding coefficients (*r*).

Results

Preliminary analysis

We conducted a between-subject ANOVA to rule out that an improvement in monitoring accuracy may be driven by primary differences in performance accuracy between the CG and EG. A significant main effect of age ($F(1,392) = 5.08$, $p = .025$, $\eta_p^2 = .013$) revealed higher performance accuracy scores for second graders ($M = .57$, $SD = .18$) -corresponding to 13 out of 22 correctly solved items—than for kindergartners ($M = .53$, $SD = .19$)—corresponding to correctly solving 9 out of 16 items. Thus, there was a well-balanced database including an about equal number of correct and incorrect answers and their confidence judgment for the monitoring analyses reported below. The main effect of the condition ($F(1,392) = 3.08$, $p = .08$, $\eta_p^2 = .008$) and the interaction ($F(1,392) = .121$, $p = .728$, $\eta_p^2 = .00$) did not reach significance. Therefore, we can assume that an improvement in monitoring accuracy observed in the EG is not an artifact of better performance accuracy.

As a preliminary analysis to evaluate the performance in the inhibition measure, we calculated a between-subject ANOVA. The main effect of age ($F(1,325) = 91.03$, $p < .001$, $\eta_p^2 = .219$) was significant, with higher correctly solved items per second for second graders ($M = .45$, $SD = .12$) than kindergarteners ($M = .33$, $SD = .11$). The main effect of the condition ($F(1,325) = 3.34$, $p = .068$, $\eta_p^2 = .01$) and the interaction ($F(1,325) = .68$, $p = .41$, $\eta_p^2 = .002$) did not reach significance. These results indicate that performance in inhibitory control skills was comparable across the CG and EG.

Monitoring

To address relative monitoring accuracy, we evaluated the discrimination score. This score taps children's ability to metacognitively discriminate in their confidence judgments between correctly and incorrectly recognized item pairs by giving substantially higher CJ for correct than for incorrect recognition. Results of the between-subject ANOVA revealed a significant main effect of age ($F(1,389) = 14.43$, $p < .001$, $\eta_p^2 = .036$), with higher discrimination scores for second graders ($M = 1.50$ $SD = 2.53$) compared to kindergarteners ($M = .48$, $SD = 2.65$). In addition, a significant main effect of condition was identified ($F(1,389) = 4.23$, $p = .04$, $\eta_p^2 = .011$), due to participants in the EG ($M = 1.2$, $SD = 2.6$) achieving better discrimination between correct and incorrect items compared to CG ($M = .6$, $SD = 2.6$), that is, achieving more accurate monitoring (see Fig 1). The interaction did not reach significance ($F(1,389) = .04$, $p = .842$, $\eta_p^2 = .00$), thus the effect of the delay was similar in the two age groups.

As the literature offers ample evidence for young children‘s performance overestimation, we were also interested in an absolute score of monitoring, the bias index. This score can range from underestimation (negative values), perfect estimation (values around zero) to overestimation (positive values). The ANOVA with age and experimental condition as between-subject factor revealed a significant main effect of age ($F(1,389) = 13.71$, $p < .001$, $\eta_p^2 = .034$). Kindergartners ($M = .29$, $SD = .26$) show a stronger overconfidence compared to second graders ($M = .19$, $SD = .255$). A main effect of condition ($F(1,389) = 5.39$, $p = .021$, $\eta_p^2 = .014$) was also found, with participants in the EG ($M = .21$, $SD = .26$) showing less overconfidence compared to the CG ($M = .28$, $SD = .26$) (see Fig 2). Contrary to our hypothesis, this effect was

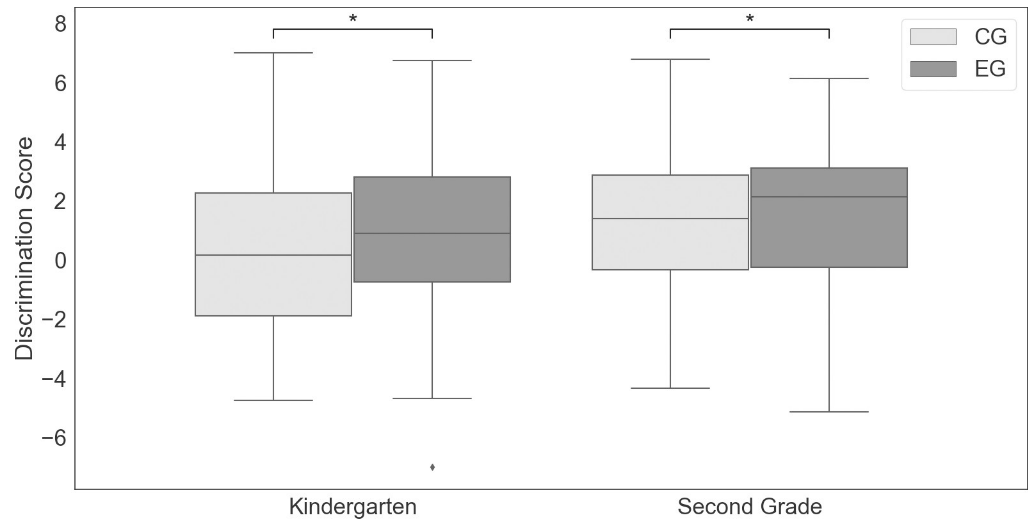


Fig 1. Distribution of the discrimination score separated for condition and age. Note. Boxplot for the dependent variable discrimination score, separated for Age (Kindergartners vs. Second Graders), and Condition (Control Group (CG) vs. Experimental Group (EG)). Whiskers represent 1.5 * interquartile range.

https://doi.org/10.1371/journal.pone.0274460.g001

about equally strong in both age groups as the interaction did not reach significance ($F(1,389)$ = .382 p = .537, η_p^2 = .001).

Individual differences in inhibition

We will report the correlations separately for the two conditions. In the control group, we will explore whether individual differences in inhibition are related to monitoring accuracy (discrimination and bias index) independently from our—"Stop and Think"–manipulation. The reported results are therefore only based on participants in the CG. For this analysis, we related individual differences of the Rate Correct Score within the flower block of the Hearts and

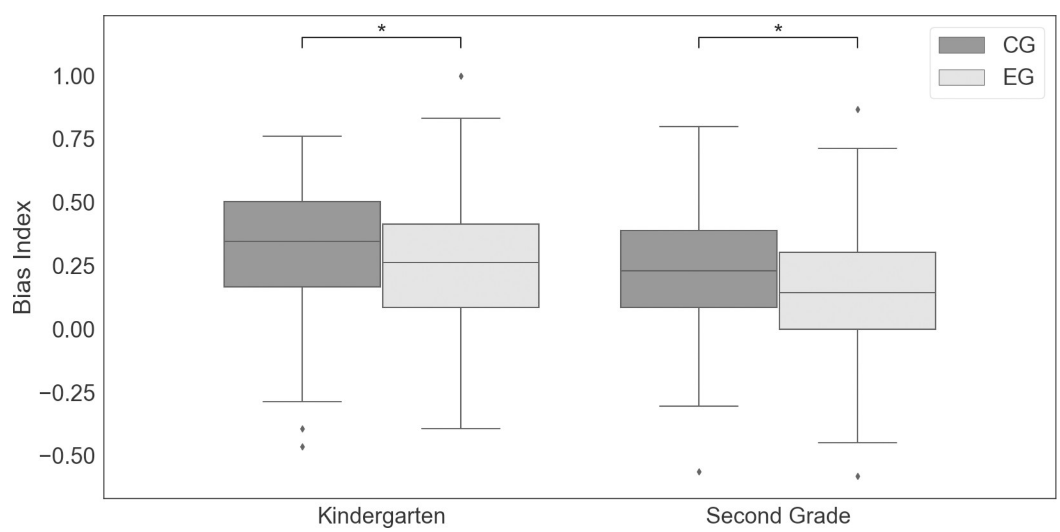


Fig 2. Distribution of the bias index separated for condition and age. Note. Boxplot for the dependent variable bias index, separated for Age (Kindergartners vs. Second Graders), and Condition (Control Group (CG) vs. Experimental Group (EG)). Whiskers represent 1.5 * interquartile range.

https://doi.org/10.1371/journal.pone.0274460.g002

Flowers task to monitoring skills (discriminations score and bias index). For kindergarteners, correlational analysis revealed a significant positive correlation between the discrimination score and the inhibition RCS (r = .225, p = .034, n = 88). Higher values in the discrimination score (representing more accurate monitoring) were related to better performance in the inhibition RCS (more correctly solved items per second within the flower block). This correlation only represents a small effect. Regarding the bias index and the inhibition RCS, no significant correlation was observed (r = -.144, p = .178, n = 88). Regarding second graders, no significant correlations were observed, neither for the discrimination score (r = .20, p = .104, n = 67) nor for the bias index (r = -.061, p = .626, n = 67).

In the experimental condition, by correlating inhibition with our monitoring measures, we will address whether inserting the delay between recognition and monitoring has a differential effect on participants depending on their inhibitory control skills. Correlational analysis addressing the discrimination score revealed no significant correlation for kindergarteners (r = .169, p = .108, n = 91) and second graders (r = .133, p = .233, n = 82). Additionally, no significant correlation was found for the bias index for the kindergarteners (r = -.119, p = .261, n = 91) as well as for second graders (r = -.075, p = .502, n = 82).

Discussion

The present study sheds light on young children's difficulties to accurately monitor memory performance by realizing an experimental approach and addressing individual differences in monitoring to inhibition. For one, we induced a delay between recognition and reporting confidence, and for another, we related performance in inhibition (measured with the Heart and Flower task) to our monitoring measures.

Monitoring

In line with previous research, our results confirmed that second graders and kindergarteners already show indications of emerging monitoring skills [10, 11, 54]. In absolute terms, children were able to discriminate substantially between correct and incorrect responses, but their evaluation of incorrect item pairs was still highly overoptimistic [12]. This pattern of results underlines the still undifferentiated monitoring skills in young children [10, 55, 56].

For the relative (discrimination score) and the absolute (bias index) measure of monitoring accuracy, findings pointed into the same direction. Second graders showed a more sophisticated discrimination between CJ for correct and CJ for incorrect items and less overconfidence than kindergartners. Thus, of the age differences reported above concerning discrimination and overconfidence fit nicely into the existing literature [56–59].

As to our experimental manipulation, our results suggested that waiting and reflecting on certainty and uncertainty for the selected answer (participants in the "Stop and Think" condition) did indeed lead to better monitoring discrimination. Moreover, children who were forced to wait and reflect also showed less overconfidence. Implementing a time window thus seemed to support children in their evaluation of confidence and led to more accurate monitoring. Our findings indicate that a brief pause where the child can "Stop and Think" can improve not only performance (as was shown in previous studies: [38, 39, 60]), but also monitoring accuracy. Especially children with difficulties inhibiting a prepotent response may benefit from more time [25]. Further, giving time to enhance monitoring accuracy is also in line with recent findings [40] indicating that information processing is not terminated when a decision is made. More information seems to accumulate between a memory decision and the corresponding monitoring judgment, supporting the idea that additional time may lead to more accurate evaluations due to the accumulation of information supervised by monitoring processes. Even though our experimentally induced manipulation cannot be seen as a congruent and identical method compared to research focusing on a delay, results appear to indicate that the underlying processes are related to each other. In addition to the advantages over "more time to think" known from previous research, the present study discovered an impact on two conceptually different monitoring measures. It is of particular interest that the benefits were not limited to just one aspect of monitoring; instead, our findings might hint at the possibility that monitoring processes overall were affected. This is promising for future research.

Nevertheless, the present study revealed only small effects on the "Stop and Think" manipulation. Perhaps, implementing an extended time window is insufficient to reduce this overoptimistic behavior entirely. Therefore, the possible negative side effects resulting from overconfidence [61], such as ending a learning phase too early or not investing enough time in tasks with increased demands, cannot be completely overcome with such an approach [62, 63]. By giving children more time to consider their answers and their confidence, we could only enable one aspect: providing time for transmission and allowing the individual to be prepared at least on a neurological level. This delay may not be sufficient to fully profit from this neurological readiness; an individual must experience the benefits of monitoring in an environment in which advantages can emerge (for example, asking back when unsure to avoid errors).

Contrary to our assumption, kindergarteners did not disproportionally benefit more from our manipulation. It may be possible that second graders' inhibition skills are at this time not as far developed as monitoring processes may require. Research based on neurological studies indicates that inhibitory control skills' maturation continues until adulthood [64], and the ability to inhibit a prepotent response evolves until adolescence [65]. Therefore, it seems reasonable that second graders equally benefited from our induced support [66]. The spectrum of an individual benefiting from such a "Stop and Think" may be much broader than expected.

When it came to our individual differences approach and our attempt to better understand the role of inhibition for monitoring, our results revealed a significant positive correlation between inhibition and the monitoring discrimination score only for kindergarteners. This relation may indicate that better inhibition skills can indeed be associated with more accurate monitoring skills, especially at an earlier age. However, these findings were not confirmed considering the bias index. Therefore, the relation between inhibition and monitoring accuracy seems still not fully understood. Although other research [21] suggests that accurate monitoring may result from better EF, our results do not reflect the strong interrelation we had expected. The results only displayed a significant correlation for kindergarteners, but the correlations from both age groups were very close in their r values. The insignificant correlation within the second graders may be due to a reduction in the statistical power because of the somewhat smaller sample size in the older age group. Therefore, this insignificant correlation should be interpreted with caution [67]. Results from the present study question strong assumptions that accurate monitoring can be supported through a certain level of inhibition skills. Our findings indicate that inhibition is a necessary but not a sufficient prerequisite for accurate monitoring in children. In her review, Roebers [18] noted that methodical differences are likely to contribute to the typically weak connection between monitoring and inhibitory skills. Capturing in detail and with different methods subcomponents of both constructs may enable to compare different subcomponents. For example, Kälin et al. [21] found a relation between inhibition and implicit, but not explicit measures of monitoring. The association between metacognition and EF could thus vary as a function of inhibition and monitoring measures. In fact, a meta-analysis showed that comparing different tasks for measuring inhibition is specific to a given age range [68]. The utilization of a specific measure must be adapted to a precise age range of interest because, over time, the behavioral manifestation of inhibitory skills changes. This finding highlights the complexity of choosing the right measure for the correct age range when relating monitoring and inhibition to each other. Recent results suggest that metacognitive skills are far more present in young children (3 to 4 years old) than previously expected and that EF and metacognition are related to each other, underscoring the importance of pursuing research in this direction [26, 43]. Correlational analysis within the experimental group revealed no significant relation between inhibition and monitoring accuracy. These findings suggest that our manipulation does not affect participants differently depending on their inhibitory control abilities.

Implication

Even though the results yielded only small effects, they shed light on yet not fully understood monitoring processes. Given the theoretical background [18, 69] and the effects of the present study, the evidence supports the idea that additional time for monitoring may indeed result in more accurate monitoring. Also, from a neurological perspective, we would expect that neural signals generated from the ACC transferred to frontal regions need time for transmission [27]. In other words, the metacognitive neurological signal then has time to strengthen and influence monitoring processes. Accurate monitoring is highly relevant for everyday life situations. Not only young children but even adults benefit from more sophisticated monitoring skills. Observing, reviewing, and evaluating the ongoing cognitive processes are essential in the school setting, higher education, and following career [70, 71]. To the best of our knowledge, the present study is the first that tried to increase the time window for children's monitoring, facilitating a transmission for metacognition within an experimental design.

Limitation

No study is perfect and this one is no exception. Naturally, we cannot be sure that children actively engaged in monitoring during the delay. It is possible that despite our—"Stop and Think"—instruction, the processing of metacognitive signals was not increased. Perhaps for some children, the process of profoundly thinking about their answers can only be achieved if they, for example, have an intrinsic willingness. For future research, an implemented reminder during the animation may trigger cognitive activation for monitoring [72]. Additionally, comparing a delay without any instructions and, therefore, simply allowing more time to reflect in an unguided way would lead to a differentiated understanding of the hidden processes. Based on our and previous findings, the willingness together with additional time to reflect are essential [36, 38].

Conclusion

The present results indicate that giving young children more time to—"Stop and Think"—can improve monitoring accuracy and reduce overconfidence. Additionally, the outcomes suggest that this time window during which children take time to process and generate their answers and evaluate their confidence in the chosen answer can be strengthened with external support. With a more profound understanding of the underlying processes and how they can be supported, we may help facilitate learning activities for students and support teachers in the school setting.

Supporting information

S1 File. Data monitoring.
(XLSX)

S2 File. Data inhibition.
(XLSX)

Author Contributions

Conceptualization: Sophie Wacker, Claudia M. Roebers.

Formal analysis: Sophie Wacker.

Methodology: Sophie Wacker, Claudia M. Roebers.

Project administration: Sophie Wacker, Claudia M. Roebers.

Supervision: Claudia M. Roebers.

Writing – original draft: Sophie Wacker.

Writing – review & editing: Sophie Wacker.

References

1. Bellon E, Fias W, De Smedt B. Metacognition across domains: Is the association between arithmetic and metacognitive monitoring domain-specific? PLoS One. 2020; 15(3), e0229932. https://doi.org/10.1371/journal.pone.0229932 PMID: 32163453
2. Hembacher E, Ghetti S. Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. Psychological Science. 2014; 25(9):1768–1776. https://doi.org/10.1177/0956797614542273 PMID: 25015686
3. Lyons KE, Ghetti S. I don't want to pick! introspection on uncertainty supports early strategic behavior. Child Development. 2013; 84(2):726–736. https://doi.org/10.1111/cdev.12004 PMID: 23278486
4. Schneider W, Löffler E. The development of metacognitive knowledge in children and adolescents. In Dunlosky J. & Tauber S. K. (Eds.), The Oxford handbook of metamemory. 2015; 491–518. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.10
5. Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist. 1979; 34(10):906–911. https://doi.org/10.1037/0003-066x.34.10.906
6. Jameson KA, Narens L, Goldfarb K, Nelson TO. The influence of near-threshold priming on metamemory and recall. Acta Psychologica. 1990; 73(1):55–68. https://doi.org/10.1016/0001-6918(90)90058-n PMID: 2316387
7. Nelson TO. Metamemory: A theoretical framework and new findings. The Psychology of learning and motivation. 1990; 26, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5
8. Coughlin C, Hembacher E, Lyons KE, Ghetti S. Introspection on uncertainty and judicious help-seeking during the preschool years. Developmental Science. 2015; 18(6):957–971. https://doi.org/10.1111/desc.12271 PMID: 25483236
9. Geurten M, Willems S. Metacognition in early childhood: Fertile ground to understand memory development? Child Development Perspectives. 2016; 10(4):263–268. https://doi.org/10.1111/cdep.12201
10. Lyons KE, Ghetti S. The development of uncertainty monitoring in early childhood. Child Development. 2011; 82(6):1778–1787. https://doi.org/10.1111/j.1467-8624.2011.01649.x PMID: 21954871
11. Bayard NS, van Loon MH, Steiner M, Roebers CM. Developmental improvements and persisting difficulties in children's metacognitive monitoring and control skills: Cross-sectional and longitudinal perspectives. Child Development. 2021; 92(3):1118–1136. https://doi.org/10.1111/cdev.13486 PMID: 33529372
12. Finn B, Metcalfe J. Overconfidence in children's multi-trial judgments of learning. Learning and Instruction. 2014; 32:1–9. https://doi.org/10.1016/j.learninstruc.2014.01.001$
13. Lipko AR, Dunlosky J, Merriman WE. Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. Journal of Experimental Child Psychology. 2009; 103(2):152–166. https://doi.org/10.1016/j.jecp.2008.10.002 PMID: 19058813
14. Shin HE, Bjorklund DF, Beck EF. The adaptive nature of children's overestimation in a strategic memory task. Cognitive Development. 2007; 22(2):197–212. https://doi.org/10.1016/j.cogdev.2006.10.001
15. Kuhn D. Metacognitive development. Current Directions in Psychological Science. 2000; 9(5):178–181. https://doi.org/10.1111/1467-8721.00088

16. Roderer T, Roebers CM. Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. Metacognition and Learning. 2010; 5(3):229–250. https://doi.org/10.1007/s11409-010-9059-z

17. Lyons KE, Zelazo PD. Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. Advances in child development and behavior. 2011; 40:379–412. https://doi.org/10.1016/b978-0-12-386491-8.00010-4 PMID: 21887967

18. Roebers CM. Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. Developmental Review. 2017; 45:31–51. https://doi.org/10.1016/j.dr.2017.04.001

19. Bryce D, Whitebread D, Szűcs D. The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. Metacognition and Learning. 2015; 10 (2):181–198. https://doi.org/10.1007/s11409-014-9120-4

20. Fernandez-Duque D, Baird JA, Posner MI. Executive Attention and metacognitive regulation. Consciousness and Cognition. 2000; 9(2):288–307. https://doi.org/10.1006/ccog.2000.0447 PMID: 10924249

21. Kälin S, Roebers CM. Time-based measures of monitoring in association with executive functions in kindergarten children. Journal of Psychology. 2020; 228(4):244–253. https://doi.org/10.1027/2151-2604/a000422

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

23. Spiess MA, Meier B, Roebers CM. Prospective memory, executive functions, and metacognition are already differentiated in young elementary school children. Swiss Journal of Psychology. 2015; 74 (4):229–241. https://doi.org/10.1024/1421-0185/a000165

24. Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science. 2012; 21(1):8–14. https://doi.org/10.1177/0963721411429458 PMID: 22773897

25. Simpson A, Riggs KJ. Under what conditions do young children have difficulty inhibiting manual actions? Developmental Psychology. 2007; 43(2):417–428. https://doi.org/10.1037/0012-1649.43.2.417 PMID: 17352549

26. Davis H, Valcan DS, Pino-Pasternak D. The relationship between executive functioning and self-regulated learning in Australian children. British Journal of Developmental Psychology. 2021; 39(4):625–652. https://doi.org/10.1111/bjdp.12391 PMID: 34241911

27. Koriat A, Shitzer-Reichert R. Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In Chambres P., Izaute M., & Marescaux P.-J. (Eds.), Metacognition: Process, function, and use. New York, NY: Kluwer. 2002: pp. 1–17. https://doi.org/10.1037/0278-7393.32.5.1133

28. Marsh R, Zhu H, Schultz RT, Quackenbush G, Royal J, Skudlarski P, et al. A developmental fMRI study of self-regulatory control. Human Brain Mapping. 2006; 27(11):848–863. https://doi.org/10.1002/hbm.20225 PMID: 16421886

29. Fleming SM, Dolan RJ. The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2012; 367:1338–1349. https://doi.org/10.1098/rstb.2011.0417 PMID: 22492751

30. Wessel JR. Error awareness and the error-related negativity: Evaluating the first decade of evidence. Frontiers in Human Neuroscience. 2012; 6:1–16. https://doi.org/10.3389/fnhum.2012.00088

31. Espinet S D, Anderson JE, Zelazo PD. N2 amplitude as a neural marker of executive function in young children: An ERP study of children who switch versus perseverate on the Dimensional Change Card Sort. Developmental Cognitive Neuroscience. 2012; 2:49–S58. https://doi.org/10.1016/j.dcn.2011.12.002 PMID: 22682910

32. Espinet SD, Anderson JE, Zelazo PD. Reflection training improves executive function in preschool-age children: Behavioral and neural effects. Developmental Cognitive Neuroscience. 2013; 4:3–15. https://doi.org/10.1016/j.dcn.2012.11.009 PMID: 23280362

33. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. Relating introspective accuracy to individual differences in brain structure. Science. 2010; 329:1541–1543. https://doi.org/10.1126/science.1191883 PMID: 20847276

34. Nelson TO, Dunlosky J. "When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". Psychological Science. 1991; 2(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

35. Schneider W, Visé M, Lockl K, Nelson TO. Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. Cognitive Development. 2000; 15(2), 115–134. https://doi.org/10.1016/S0885-2014(00)00024-1

36. Beck SR, Carroll DJ, Brunsdon VEA, Gryg CK. Supporting children's counterfactual thinking with alternative modes of responding. Journal of Experimental Child Psychology. 2011; 108(1):190–202. https://doi.org/10.1016/j.jecp.2010.07.009 PMID: 20817168

37. Carroll DJ, Blakey E, Simpson A. Can we boost preschoolers' inhibitory performance just by changing the way they respond? Child Development. 2021; 92(6):2205–2212. https://doi.org/10.1111/cdev.13617 PMID: 34214196

38. Ling DS, Wong CD, Diamond A. Do children need reminders on the day-night task, or simply some way to prevent them from responding too quickly? Cognitive Development. 2016; 37:67–72. https://doi.org/10.1016/j.cogdev.2015.10.003 PMID: 26949287

39. Simpson A, Riggs KJ, Beck SR, Gorniak SL, Wu Y, Abbott D, et al. Refining the understanding of inhibitory processes: How response prepotency is created and overcome. Developmental Science. 2012; 15(1):62–73. https://doi.org/10.1111/j.1467-7687.2011.01105.x PMID: 22251293

40. Navajas J, Bahrami B, Latham PE. Post-decisional accounts of biases in confidence. Current Opinion in Behavioral Sciences. 2016; 11, 55–60. https://doi.org/10.1016/j.cobeha.2016.05.005

41. Dougherty M. A review of neural networks applied to transport. Transportation Research Part C: Emerging Technologies. 1995; 3(4), 247–260. https://doi.org/10.1016/0968-090X(95)00009-8

42. Stevenson IH, Rebesco JM, Miller LE, Körding KP. Inferring functional connections between neurons. Current opinion in neurobiology. 2008; 18(6), 582–588. https://doi.org/10.1016/j.conb.2008.11.005 PMID: 19081241

43. Marulis LM, Nelson LJ. Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. Metacognition and Learning. 2021; 16(1):207–231. https://doi.org/10.1007/s11409-020-09244-6

44. Dunlosky J, Thiede KW. Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. Learning and Instruction. 2013; 24(1):58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

45. Moriguchi Y, Hiraki K. Prefrontal cortex and executive function in young children: A review of NIRS studies. Frontiers in Human Neuroscience. 2013; 7:867. https://doi.org/10.3389/fnhum.2013.00867 PMID: 24381551

46. Zhong J, Rifkin-Graboi A, Ta AT, Yap KL, Chuang KH, Meaney MJ, et al. Functional networks in parallel with cortical development associate with executive functions in children. Cerebral Cortex. 2014; 24(7):1937–1947. https://doi.org/10.1093/cercor/bht051 PMID: 23448875

47. Rivière J, Brisson J, Aubertin E. The interaction between impulsivity, inhibitory control and scale errors in toddlers. European Journal of Developmental Psychology. 2020; 17(2):231–245. https://doi.org/10.1080/17405629.2019.1567324

48. Lo YH, Liang WK, Lee HW, Wang CH, Tzeng OJL, Hung DL, et al. The neural development of response inhibition in 5- and 6-year-old preschoolers: An ERP and EEG study. Developmental Neuropsychology. 2013; 38(5):301–316. https://doi.org/10.1080/87565641.2013.801980 PMID: 23862634

49. Fleming SM, Lau HC. How to measure metacognition. Frontiers in Human Neuroscience. 2014; 8:443. https://doi.org/10.3389/fnhum.2014.00443 PMID: 25076880

50. Schraw G. A conceptual analysis of five measures of metacognitive monitoring. Metacognition and Learning. 2009; 4(1):33–45. https://doi.org/10.1007/s11409-008-9031-3

51. Davidson MC, Amso D, Anderson LC, Diamond A. Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia. 2006; 44(11):2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006 PMID: 16580701

52. Diamond A, Barnett WS, Thomas J, Munro S. The early years: Preschool program improves cognitive control. Science. 2007; 318(5855):1387–1388. https://doi.org/10.1126/science.1151148

53. Woltz DJ, Was CA. Availability of related long-term memory during and after attention focus in working memory. Memory & Cognition. 2006; 34(3):668–684. https://doi.org/10.3758/bf03193587

54. Destan N, Hembacher E, Ghetti S, Roebers CM. Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. Journal of Experimental Child Psychology. 2014; 126:213–228. https://doi.org/10.1016/j.jecp.2014.04.001 PMID: 24945686

55. Koriat A, Goldsmith M, Schneider W, Nakash-Dura M. The credibility of children's testimony: Can children control the accuracy of their memory reports? Journal of Experimental Child Psychology. 2001; 79(4):405–437. https://doi.org/10.1006/jecp.2000.2612 PMID: 11511131

56. Roebers CM. Confidence judgments in children's and adults' event recall and suggestibility. Developmental Psychology. 2002; 38(6):1052–1067. https://doi.org/10.1037//0012-1649.38.6.1052 PMID: 12428714

57. Ghetti S, Lyons KE, Lazzarin F, Cornoldi C. The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. Journal of Experimental Child Psychology. 2008; 99(3):157–181. https://doi.org/10.1016/j.jecp.2007.11.001 PMID: 18191139

58. O'Leary AP, Sloutsky VM. Carving metacognition at its joints: Protracted development of component processes. Child Development. 2017; 88(3): 1015–1032. https://doi.org/10.1111/cdev.12644 PMID: 27759890

59. Wall JL, Thompson CA, Dunlosky J, Merriman WE. Children can accurately monitor and control their number-line estimation performance. Developmental Psychology. 2016; 52(10):1493–1502. https://doi.org/10.1037/dev0000180 PMID: 27548391

60. Diamond A, Kirkham N, Amso D. Conditions under which young children can hold two rules in mind and inhibit a prepotent response. Developmental Psychology. 2002; 38(3):352–362. https://doi.org/10.1037/0012-1649.38.3.352 PMID: 12005379

61. Destan N, Roebers CM. What are the metacognitive costs of young children's overconfidence? Metacognition and Learning. 2015; 10(3):347–374. https://doi.org/10.1007/s11409-014-9133-z

62. Dunlosky J, Rawson KA. Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. Learning and Instruction. 2012; 22(4):271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

63. Metcalfe J, Finn B. Evidence that judgments of learning are causally related to study choice. Psychonomic Bulletin and Review. 2008; 15(1):174–179. https://doi.org/10.3758/pbr.15.1.174 PMID: 18605499

64. Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: What have we learned about cognitive development? Trends in Cognitive Sciences. 2005; 9(3):104–110. https://doi.org/10.1016/j.tics.2005.01.011 PMID: 15737818

65. Shing YL, Lindenberger U, Diamond A, Li SC, Davidson MC. Memory maintenance and inhibitory control differentiate from early childhood to adolescence. Developmental Neuropsychology. 2010; 35 (6):679–697. https://doi.org/10.1080/87565641.2010.508546 PMID: 21038160

66. Bryce D, Szucs D, Soltész F, Whitebread D. The development of inhibitory control: An averaged and single-trial lateralized readiness potential study. NeuroImage. 2011; 57(3):671–685. https://doi.org/10.1016/j.neuroimage.2010.12.006 PMID: 21146618

67. Makin TR, de Xivry JJO. Science Forum: Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. Elife. 2019; 8, e48175. https://doi.org/10.7554/eLife.48175

68. Petersen IT, Hoyniak CP, McQuillan ME, Bates JE, Staples AD. Measuring the development of inhibitory control: The challenge of heterotypic continuity. Developmental Review. 2016; 40:25–71. https://doi.org/10.1016/j.dr.2016.02.001 PMID: 27346906

69. Marulis LM, Baker ST, Whitebread D. Integrating metacognition and executive function to enhance young children's perception of and agency in their learning. Early Childhood Research Quarterly. 2020; 50:46–54. https://doi.org/10.1016/j.ecresq.2018.12.017

70. Hertzog C, Dunlosky J. Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. Current directions in psychological scienc. 2011; 20(3), 167–173. https://doi.org/10.1177/0963721411409026 PMID: 24478539

71. Kane S, Lear M, Dube CM. Reflections on the role of metacognition in student reading and learning at higher education level. Africa Education Review. 2014; 11(4), 512–525. http://dx.doi.org/10.1080/18146627.2014.935001

72. Barker JE, Munakata Y. Time isn't of the essence: Activating goals rather than imposing delays improves inhibitory control in children. Psychological Science. 2015; 26(12):1898–1908. https://doi.org/10.1177/0956797615604625 PMID: 26546078

Appendix B: Study II

Motivating children to (pre) monitor: positive effects on monitoring accuracy?

Sophie Wacker & Claudia M. Roebers

University of Bern

Metacognition and Learning
https://doi.org/10.1007/s11409-023-09351-0

Motivating children to (pre)monitor: positive effects on monitoring accuracy?

Sophie Wacker[1] · Claudia M. Roebers[1]

Received: 27 January 2023 / Accepted: 14 June 2023

Abstract
When young children evaluate their confidence, their monitoring is often overoptimistic, that is, inaccurate. The present study investigated a potential underlying mechanism for kindergarteners' and second graders' overconfidence within a paired associates learning paradigm. We implemented a pre-monitoring phase motivating children to differentially evaluate their confidence for each alternative *before* children could choose an answer in the subsequent recognition phase. For one, we intended to weaken the influence of one single and prepotently selected memory trace. For another, we motivated and enabled children to evaluate all four answer alternatives concerning their certainty before evaluating their final recognition choice by giving a confidence judgment. We compared monitoring discrimination and monitoring bias with a control condition whose task sequence did not include a pre-monitoring judgment. Contrary to our expectations, the pattern of results indicated that being instructed to pre-monitor did *increase* and not decrease overconfidence in young children. The present results will be discussed against the background of memory-metamemory interaction, confirmation bias, and methodological issues.

Keywords Cognitive development · Metacognition · Monitoring · Overconfidence · Children

Introduction

One of the most critical milestones in a child's development is reaching the ability to monitor and regulate cognitive functions. So-called metacognitive processes allow to monitor behavior and enable adaptive information processing under varying task demands (Conn et al., 2018; Wang et al., 1993). In metacognitive research, one distinguishes between declarative (i.e., knowledge about learning strategies and one's memory; Flavell, 1971) and procedural metacognition (monitoring and control processes; Schneider & Löffler, 2016). Thereby both, accurate monitoring (i.e., ongoing

✉ Sophie Wacker
sophie.wacker@unibe.ch

Claudia M. Roebers
claudia.roebers@unibe.ch

[1] Department of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

surveillance of subjective task performance and comparison with objective performance) and efficient control processes (i.e., processes involved in the regulation of one's one cognitive performance: re-study selections, allocation of study time) are essential for school achievement (Dunlosky & Metcalfe, 2008; de Bruin & van Gog, 2012; van Loon et al., 2014), and lifelong learning (Bakracevic Vukman & Licardo, 2010). First signs of emerging metacognition are already observable at preschool age (Coughlin et al., 2015; Geurten & Bastin, 2019; Lipowski et al., 2013; Lyons & Ghetti, 2013). Further and significant improvements are then consistently found, especially during the school years (Lyons & Zelazo 2011; Schneider & Löffler, 2016). Thus, a critical age window for metacognitive development occurs at a young age building the basis for a protracted development into adolescence.

Although early signs of emerging metacognitive skills have repeatedly been documented, monitoring processes of kindergarten and young primary school children still need to get fine-tuned and more differentiated, and thus, their monitoring accuracy is consistently lower compared to adults (Schneider, 2014). Specifically, young children seem to have difficulties reliably and metacognitively differentiating between correct and incorrect performance in their monitoring judgments (Destan & Roebers, 2015; Finn & Metcalfe, 2014; Lipko et al., 2009), with a strong tendency to a positive bias, also called "overconfidence". Some authors have argued that one reason for their overoptimistic monitoring may be that young children do not take enough time to monitor alternative answers and their final responses but rather jump too quickly on a positive (very certain) monitoring judgment (Bryce et al., 2015; Roebers, 2017). This overconfidence, also labelled as miscalibration, severely and negatively impacts subsequent control decisions and hampers efficient adjustments within learning processes (e.g., be uncertain and then go back and double-check; Dunlosky & Rawson, 2012; Hacker et al., 2008; van Loon et al., 2022). The question thus arises whether young children are generally incapable of monitoring accurately or whether, under certain circumstances, they can be supported to monitor more accurately. This is the focus of the present approach.

Because overconfidence is a well-known phenomenon in metacognitive research including kindergarten and young primary school children, some studies have tried to improve and support monitoring processes in young children by applying different experimental designs, for example, inducing feedback (Lipko et al., 2012; O'Leary & Sloutsky, 2017; Oudman et al., 2022; van Loon & Roebers, 2021). However, the evidence as to whether overconfidence can be corrected is very inconsistent (Bol et al., 2005; Nietfeld et al., 2005; Saenz et al., 2019). Even more importantly, the underlying processes contributing to children's inaccurate monitoring are yet to be understood.

The theoretical background of our experimental manipulation is the interplay of memory retrieval, memory monitoring, and strategic decisions about memory performance as outlined by Bjork (1994) and Koriat and Goldsmith (1996). Within these frameworks, memory and metamemory are no longer considered separated, independent constructs, but rather to be closely intertwined. Monitoring and control processes uniformly rely on retrieving information from long-term memory (Bjork, 1994; Koriat, 2015). According to the model of strategic regulation of memory accuracy (Koriat & Goldsmith, 1996), information from long-term memory is retrieved and, at the same time, monitoring processes are triggered to choose the best candidate answer (the answer with the highest/most positive monitoring judgment). These fast and early retrieval processes from long-term memory coupled with early monitoring processes are likely to regulate or influence later monitoring processes, an interaction that has been neglected when addressing children's developing monitoring skills. By asking children to report on their confidence *before* selecting a memory answer, we sought to step into this memory-monitoring interplay to gain insights into their very early monitoring processes.

Against the background of such strategic regulation processes based on ongoing, early monitoring processes that are intertwined with memory retrieval, children's tendency to quickly jump on an answer seems fatal, or at least, be contributing to overconfidence. That is, when quickly selecting an answer, more detailed differentiation of different degrees of confidence and thorough comparisons among the answer alternatives (or, in the case of free recall, a further search for an answer with a higher likelihood to be correct) rarely occurs (Barnes et al., 1999). Indirect evidence for the assumption that children's overconfidence stems – at least in part – from their reluctance to engage in (pre)monitoring processes stems from studies in which children were highly motivated to answer correctly (e.g., receiving gold coins for every correct answer). With such a paradigm, young children were found to be quite capable of monitoring accurately (Roebers & Fernandez, 2002; Koriat & Ackerman, 2010). These results suggest that efficient monitoring abilities are already present at a young age but not yet automatically used.

Furthermore, the first and possibly impulsive answer is not always the most suitable. Resisting a prepotent response (Simpson et al., 2012) and considering additional information to evaluate the most suitable answer may be terminated prematurely (Nelson, 1990). In doing so, children might be neglecting contradicting evidence, which then contributes to a confirmation bias, increasing confidence (Koriat et al., 1980). Investing time and effort in comparing the different degrees of confidence of the different answer alternatives might not seem relevant to young children. In other words, overconfidence might be reduced by enabling and powerfully motivating children to address and consider contradictory information. For this reason, our design prompted children to actively consider information against and in favor for the chosen answer, as well as for the three other alternatives (Koriat et al., 1980). In other words, we prompted monitoring processes at time point during the task sequence in which no definitive answer had yet been selected.

The present study

The present study implemented a pre-monitoring phase that motivated children to evaluate their confidence not only in one but in all answer alternatives *before* choosing an answer (recognition). Participants solved a paired associate learning task with a pre-monitoring (only the experimental condition), a recognition, and a post-monitoring phase. Before choosing an alternative in the recognition phase, participants in the experimental condition had to give a monitoring judgment for every answer alternative. After choosing one alternative in the recognition phase, participants then rated their confidence in the chosen alternative, as is done in classical studies using confidence judgments (Destan et al., 2014; Dunlosky & Metcalfe, 2008). We included kindergarten and second graders because some metacognitive monitoring abilities are already developed at kindergarten age (Coughlin et al., 2015), and the transition to school age represents an important time window for further fine-tuning of monitoring abilities (Schneider, 2014).

We hypothesized that participating in the experimental condition (including pre-monitoring judgments) would improve monitoring accuracy compared to the control condition (no pre-monitoring judgments). More precisely, we expected the pre-monitoring phase to encourage a comprehensive evaluation of all four alternatives (Koriat & Goldsmith, 1996; Navajas et al., 2016). This in turn, should lead to more accurate monitoring in the following recognition phase. This was expected, because at the time of the pre-monitoring judgments, no answer had yet been selected possibly reducing the memory retrieval-monitoring coupling. For another, participants had to judge each alternative actively to determine the best candidate answer. Thus, participants had to invest time in searching for contradicting information rather than immediately choosing one answer and ignoring the alternatives.

Method

Participants

The sample for the present study consisted of N = 330 children from urban and rural regions in the vicinity of a university town [further information withheld for blinded review]. The study included preschool children and second grade students. We excluded participants with less than 25% correct recognition (preschool children N = 18, second grade students N = 7). Recognition scores under 25% correct may be based on chance level as there were four alternatives. This exclusion led to a total sample of N = 305, with N = 143 kindergartners (49.7% female) between 5-6 years of age ($M = 5.9$, $SD = .47$), and N = 162 second grade students (53.1% female) between 7-8 years of age ($M = 7.8$, $SD = .36$). The Ethics Committee of the Faculty of Human Sciences of the University had approved the study (Approval No: 2002-100005). Only children who had written parental consent were allowed to participate in the study. Additionally, participation was voluntary for the children, and they were informed that they could terminate the tasks at any time without giving any explanation. No child ever did. Data were treated entirely anonymously.

Measures and Procedures

Participants solved a computer-based task running on tablets (Samsung Galaxy S7) using the app software *Ionic*. All instructions during the task were given via headphones. The task had been extensively piloted regarding understandability and usability for children from both age groups. During the testing phase, children were closely supervised to assist in case of technical problems. Children solved a paired associate learning task to capture monitoring processes (about 30 minutes). The task was accompanied by a cover story about two children who are taking pictures of animals (e.g., bird) that are engaged in actions toward different objects (e.g., bread). Participants were instructed to memorize the associated pairs to help the protagonists find the associated pictures that belong together later on.

Participants were randomly assigned to one of two conditions (control vs. experimental condition). Being part of the control condition entailed the following structure. After a familiarization phase, preschool children had to study 20, and second grade students 26 item pairs, respectively. Each pair was presented for four seconds. After a one-minute filler task to ensure standardized delay and to prevent memory strategies, the stimulus picture (the animal) was presented with four alternatives (recognition phase; four different objects, i.e., feed, environment). A monitoring phase followed this recognition phase (Fig. 1). Now, children had to evaluate their confidence by reporting their degree of confidence on a classical 7- Likert scale which was illustrated with a thermometer (Koriat & Shitzer-Reichert, 2002) ranging from blue (very unsure) to red (very sure). Confidence judgments were converted into values from 1 to 7. Prior to the testing phase children were introduced to the

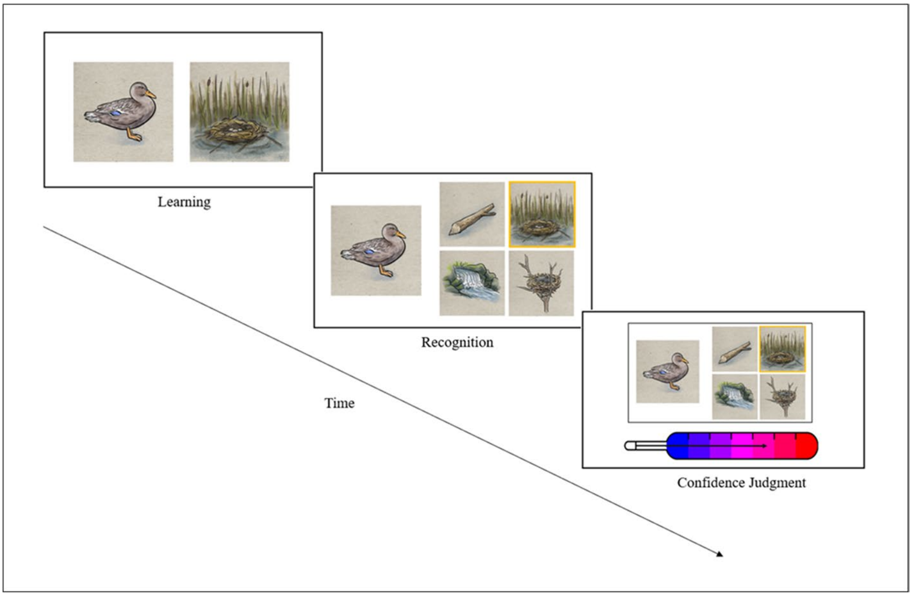


Fig. 1 Schematic Illustration of the Task Procedure

thermometer and had to solve different examples correctly and give appropriate confidence judgments before proceeding to start the testing phase.

The procedure of the experimental condition followed the same sequence except for an implemented "pre-monitoring phase" between learning and recognition. After the filler task, participants saw the stimulus picture and the four alternatives. Underneath every alternative, a 7-Likert scale appeared (Fig. 2). The Likert scale was visualized with seven stars of increasing size (ranging from a small star on the left to a large star on the right). For every alternative, participants were told to choose a star. An interactive animation introduced the stars. The smallest star was introduced as follows: "If you think the pictures do not belong together, you can choose a small star." The stars in the middle section were visualized on a continuum from a small to a large star. For these stars, the instruction: "If you think it could be the right picture, you can choose stars in the middle" was given. In the end, it was explained that choosing all seven stars including the largest star would represent a maximum of confidence introduced with the following explanation: "If you are convinced that these images belong together, you can choose all the stars up to the largest star". A practice trial accompanied every introduction. Using the abovementioned questions, children were asked which star they would choose. When they chose the correct star, they proceeded to the next question until they answered every practice question with the intended star. If participants chose an inappropriate star, they were corrected by the experimenter. When the exercise was solved 100 percent correctly, the practice trials were terminated, and the program progressed to the actual task. These pre-monitoring judgments (stars) were converted into values from 1 to 7.

To evaluate monitoring accuracy, we included two different measures. For quantifying relative monitoring accuracy, a discrimination score was calculated for each child, indicating how reliably children can metacognitively differentiate between correct and incorrect recognition with their confidence judgment (difference score; Dunlosky & Thiede, 2013;

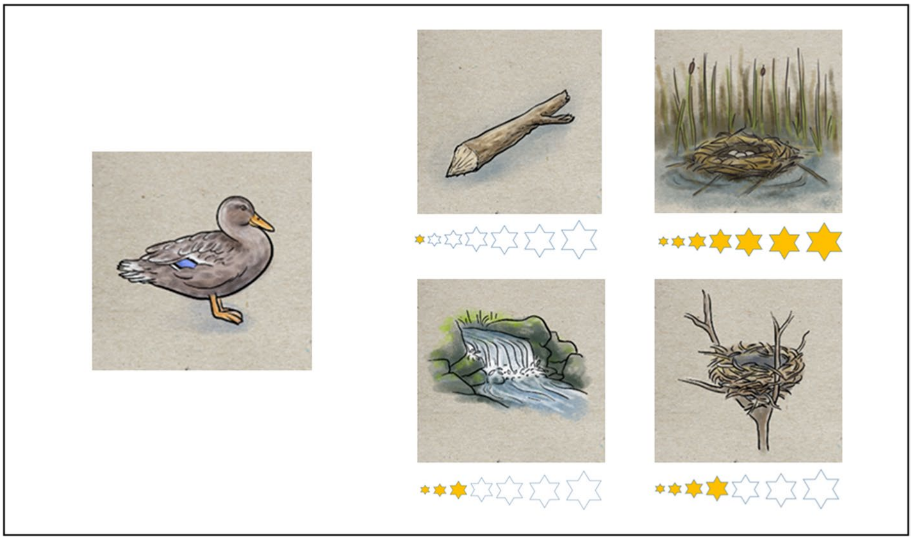


Fig. 2 Pre-Monitoring Phase for Experimental Condition

Fleming & Lau, 2014; Schraw, 2009). Thereby, higher confidence judgments for correct than for incorrect responses indicate more accurate monitoring. Further, we included a measure of absolute monitoring accuracy, a bias index, reflecting under- to overconfidence (Schraw, 2009). Values for absolute monitoring accuracy can vary between +1 (overestimation) and –1 (underestimation). The closer the value is to zero, the more accurately the confidence rating matches performance.

Statistical Analysis

Data was analyzed with Python using Scipy, Numpy, Pandas, and StatsModel as packages. We used Seaborn and Matplotlib for data visualization. Our hypotheses regarding different monitoring aspects were analyzed by running between-subject analyses of variances (ANOVAs). As an estimation for effect sizes, we used partial eta squared (η_p^2).

Results

Performance

Mean performance for kindergarteners was 9 out of 20 correctly solved items (mean proportion correct: 45%, $SD = .14$), and 14 out of 26 correctly solved items for second graders (mean proportion correct: 53%, $SD = .15$). This represents an approximately equal number of correctly and incorrectly solved items across both age groups and thus provides a balanced database for evaluating monitoring processes.

Because monitoring measures are intertwined with performance, we wanted to rule out that differences in monitoring between the two conditions (control vs experimental condition) may be due to differences in performance. Therefore, we first analyzed performance by means of a between-subject ANOVA. As expected, there was a significant main effect of age (F (1, 301) = 26.8, $p < .001$, $\eta_p^2 = .08$), with higher performance for second graders ($M = .53$, $SD = .15$) than for kindergarteners ($M = .45$, $SD = .14$). Furthermore, there was no significant main effect of condition (F (1, 301) = 1.07, $p = .30$, $\eta_p^2 = .01$) and no interaction (F (1, 301) = .757, $p = .39$, $\eta_p^2 = .00$).

Monitoring

To capture metacognitive discrimination (relative monitoring accuracy), we computed a difference score by subtracting confidence for correctly from incorrectly solved items. In young children higher values then indicate better discrimination between correctly and incorrectly solved items. The ANOVA with age and condition as between-subject factors revealed a significant main effect of age (F (1, 301) = 34.50, $p < .001$, $\eta_p^2 = .10$). Second graders ($M = .83$, $SD = 1.9$) outperformed kindergartners ($M = -.40$, $SD = 1.8$) in terms of metacognitive discrimination. The main effect of condition (F (1, 301) = 2.64, $p = .105$, $\eta_p^2 = .01$) and the interaction (F (1, 301) = 1.40, $p = .24$, $\eta_p^2 = .01$) did not reach significance.

Thus, contrary to our assumption, when comparing relative monitoring accuracy between the control and experimental condition, this difference was not significant. The descriptive data even suggested that participants in the control condition reached a better discrimination score. Although non-significant, it appears with respect to relative monitoring accuracy that engaging in pre-monitoring (providing star ratings) before choosing an answer may even negatively influence monitoring accuracy in children.

Turning to the bias index, this score indicates an absolute monitoring accuracy measure and ranges from underestimation to overestimation. The ANOVA with age and condition as between-subject factors revealed a main effect of age (F (1, 301) = 39.77, $p < .001$, $\eta_p^2 = .12$). In line with our assumption, older children ($M = .27$, $SD = .19$) show less overconfidence than younger children ($M = .42$, $SD = .20$). Again, contrary to our expectations, results revealed a significant main effect of condition (F (1, 301) = 14.74, $p < .001$, $\eta_p^2 = .05$), with *less* overconfidence in the control condition ($M = .30$, $SD = .21$) than in the experimental condition ($M = .38$, $SD = .19$). In other words, participating in the experimental condition and evaluating the four alternatives before selecting the best candidate answer by giving pre-monitoring judgments seemed to *increase* rather than decrease overconfidence in young children, independent of age as the non-significant interaction revealed (F (1, 301) = 2.4, $p = .12$, $\eta_p^2 = .01$).

Post Hoc Analysis: Searching for Reasons for Unexpected Results

As the results so far did not confirm our assumptions regarding the effects of our implemented pre-monitoring judgments, we ran some post hoc analysis to shed light on the implemented pre-monitoring judgments. Firstly, we looked at the distribution of pre-monitoring judgments within an item (*differentiation within alternatives*, see below). In doing so, we analyzed whether participants discriminated between the four alternatives (give different numbers of stars within one item indicative of active monitoring). Secondly, we evaluated the relation of the pre-monitoring judgments with subsequent recognition. More precisely, we analyzed which alternative obtained the highest pre-monitoring judgment, and then addressed whether this alternative was selected in the recognition phase (*relation pre-monitoring judgment and answer in recognition*, see below). Thirdly, we compared the pre-monitoring judgment with the monitoring judgment given *after* recognition (*consistency between pre-monitoring judgment and confidence judgment*, see below).

Differentiation within Alternatives Results revealed that only N = 7 children did not differentiate between the alternatives by giving every alternative the same pre-monitoring judgment. Because all other participants had variability in their scoring pattern (scoring different pre-monitoring judgments within four alternatives), we are tempted to believe that the vast majority of children followed our instructions and seriously monitored differentially; that is, rated lower, or higher levels based on their evaluations for the four alternatives.

Relation Pre-Monitoring Judgment and Answer in Recognition Descriptive statistics of both age groups revealed that out of 100% (N = 3049) registered pre-monitoring judgments, 88.3% (N = 2692) of the alternatives having received the highest level were also selected in the recognition phase. Only 11.7% (N = 357) of the alternatives that had received the highest pre-monitoring judgments were not selected in the recognition phase. These results suggest that there is a very strong relation between pre-monitoring and recognition (see Table 1). This pattern of result remained the same when we analyzed the two age groups separately.

Additionally, for kindergarteners and second graders, pre-monitoring judgments were generally high (see Table 1). Overall, participants favored one alternative and gave it the highest pre-monitoring judgments to express the outcome of their pre-monitoring. This trend confirms young children's tendency to choose rather high confidence judgments (in any case, even in the face of contradicting information).

Consistency between Pre-Monitoring Judgment and Confidence Judgment To evaluate the consistency of the two different monitoring measures, we looked at the correspondence between pre-monitoring and confidence judgment, as a function of the correctness of recognition. By comparing the levels of both monitoring measures, we evaluated if participants tended to rate higher, lower or the same level of confidence on both monitoring measures but on the two different scales. Because the pattern of results revealed no difference between kindergarteners and second graders, we report them collectively. A descriptive overview of all judgments of both age groups can be found in Table 2.

High consistency would mean that reporting a very low pre-monitoring judgment leads to a high likelihood of choosing a very low confidence judgment in the retrospective confidence judgment (see marked diagonal in Table 2). The diagonal of agreements adds up to 68.1% of all monitoring judgments. We further split the judgment range into two groups

Table 1 Pre-Monitoring Judgments (Kindergarteners and Second Graders) as a Function of Match in the Recognition Task in Percent

Pre-Monitoring Judgments							
	1	2	3	4	5	6	7
Recognition Match	0.0	1.2	1.4	3.1	5.8	6.8	81.7
Recognition no Match	0.0	6.2	4.8	5.0	10.4	9.5	64.2

Recognition match = answer rated with the highest pre-monitoring judgment was also selected in recognition. Recognition no match = answer rated with highest pre-monitoring judgment was not selected in recognition

Table 2 Pre-Monitoring and Confidence Judgments (Kindergarteners and Second Graders)

Pre-Monitoring Judgments								
		1	2	3	4	5	6	7
Confidence Judgments	1	0.0	0.1	0.1	0.3	0.1	0.0	2.1
	2	0.0	0.1	0.1	0.2	0.1	0.0	0.7
	3	0.0	0.0	0.1	0.5	0.3	0.3	0.9
	4	0.0	0.1	0.4	0.6	0.8	0.8	4.4
	5	0.0	0.2	0.2	0.5	1.3	1.3	3.9
	6	0.0	0.1	0.2	0.5	0.9	2.2	5.9
	7	0.0	0.6	0.4	0.7	2.3	2.3	63.9

This table displays values in percent under the assumption that the answer with the highest star was also selected in recognition

(1- 6 and 7). Regarding the judgments' range from 1 to 6, results suggest that only 4.2% of these monitoring judgments showed consistency. However, for the highest rating (very sure in the pre-monitoring judgment and very sure in the confidence judgment), results showed a total agreement of 63.9%, thus very high consistency across judgments. This distribution clearly shows that most retrospective confidence judgments are built on a very high pre-monitoring judgment and a very high confidence rating that matches the pre-monitoring judgment. Thus, most participants reported very high confidence in both monitoring measures (Fig. 3). To address whether the high confidence was primarily observed concerning correct answers (then, high confidence seems justified), we ran the same analyses for correct and incorrect recognition, respectively. Interestingly, the pattern of results turned out to be equivalent.

As reported above, most confidence judgments (68.1%) were in agreement with the pre-monitoring judgment. The remaining 31.9% were either rated higher or lower in the subsequent confidence judgment. Out of these remaining 31.9% a total of 9.3% of judgments mirrored higher confidence compared to the previous pre-monitoring judgment indicating a tendency towards a strengthening of the subjective confidence. But, 22.6% of judgments were found to mirror *lower* confidence compared to the previous pre-monitoring judgment, indicating a decrease in overconfidence as we had initially expected. Furthermore, we conducted a Pearson Chi-Squared test. There was a significant relationship between the ratings of pre-monitoring and confidence judgments, $X^2(30) = 550.20$, $p < .001$, $N = 2692$. The relationship indicates a moderate effect ($CC = .412$, $p < .001$).

Discussion

Overconfidence is a well-known challenge in metacognitive research in children. Although some studies have already attempted to improve children's monitoring, results were inconsistent, and underlying reasons for children's overconfidence still need to be explored. We implemented a pre-monitoring phase enabling children to monitor possible answers before making a final decision in the recognition phase. With this experimental approach, we intended to motivate a differentiated evaluation of each alternative thereby preventing that just one, hastily selected answer, builds up a dominating memory trace. We hypothesized

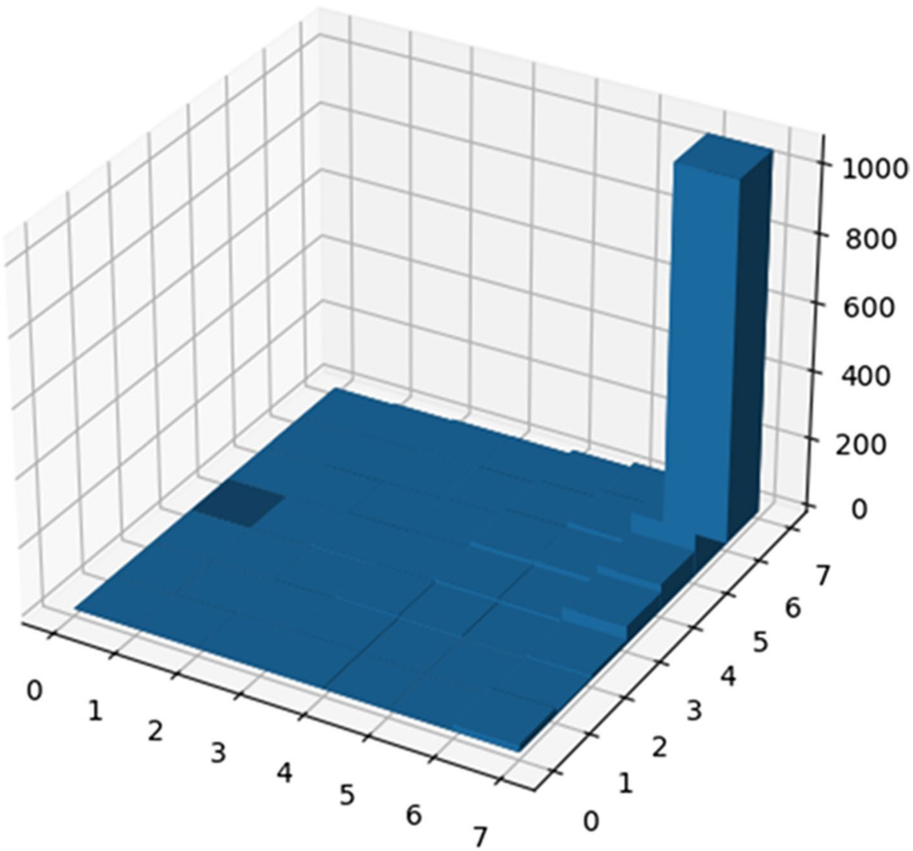


Fig. 3 Association between Pre-Monitoring and Confidence Judgements. *Note.* 3D Plot displays pre-monitoring judgments on the left (y-axis) and confidence judgments on the right (x-axis) for kindergarteners and second graders. The number of judgments is mapped on the z-axis

that participating in the pre-monitoring phase would enable a deeper and more differentiated monitoring, considering all four alternatives (in the sense of the regulation of memory accuracy of Koriat & Goldsmith, 1996), which might positively influence young children's monitoring accuracy.

In line with previous research, our results revealed that kindergarteners and second graders were able to metacognitively discriminate between correct and incorrect answers (Lyons & Ghetti, 2011). As expected, both age groups showed a high tendency towards overconfidence (Destan et al., 2014; Finn & Metcalfe, 2014; van Loon et al., 2017). Furthermore, our results align with studies showing that the ability to monitor adequately increases with age (Schneider & Löffler, 2016).

To address the effect of the experimental condition, we investigated two measures of monitoring accuracy. The findings pointed in the same direction for the relative (discrimination score) and the absolute (bias index) measure of monitoring accuracy. Contrary to our expectations, participating in the experimental condition decreased rather than increased monitoring accuracy in both age groups. More precisely, pre-monitoring was not helpful for children when asked to discriminate between correct and incorrect answers.

Pre-monitoring seemed even to increase overconfidence in both age groups. Given the inconsistent findings in studies addressing interventions to improve monitoring, the present results might not be entirely surprising (O'Leary & Sloutsky, 2017; Oudman et al., 2022). Nevertheless, our study confirms that when trying to improve monitoring accuracy in children, researchers are facing severe methodological and theoretical challenges (Azevedo, 2020; Schraw & Moshman, 1995).

We did some post hoc tests to evaluate what children really did during the pre-monitoring phase and whether there was consistency across the different phases of our task. Analysis revealed that overall, participants seriously discriminated between every answer alternative. Providing different pre-monitoring judgments for each alternative, we can assume that participants did not make fast and superficial decisions and simply moved on to the next item prematurely without at least considering the other alternatives (Simpson et al., 2007). Even if this pre-monitoring phase seemed to trigger a differentiated comparison as we had anticipated, the consequence differed from what we expected as it increased rather than decreased overconfidence. Children might – already at this stage – have trapped in the confirmation bias (Nickerson, 1998; Peters, 2022).

Results indicated that the alternative rated with the highest pre-monitoring judgment was, in most cases, also chosen in recognition. This connection suggests that choosing a high pre-monitoring judgment may indicate high certainty already and therefore serves as a cue for again selecting this alternative as the final answer. The high consistency between the pre-monitoring, recognition choice, and the confidence judgment supports this interpretation. Moreover, participants also chose the same scale level to express their confidence.

Reasons for Overconfidence

Memory and Monitoring

The attempt to separate monitoring processes from memory seems more difficult than expected (Bjork, 1994; Koriat & Goldsmith, 1996; Roebers & Schneider, 2005). This finding raises the question of to what extent it is possible to consider and investigate monitoring and memory as separate processes (Koriat, 2007). The dynamic association between memory and metamemory seems to be underestimated and even more critical than assumed (Busey et al., 2000; Bjork, 1994; Vernon & Usher, 2003). Pre-monitoring judgments may already be related to memory processes (Dinsmore & Parkinson, 2013; Ferreira et al., 2019; Vernon & Usher, 2003). These considerations raise the question of whether a confidence judgment can be considered an isolated and memory-independent process. Are confidence judgments really a representation of confidence or rather a feeling of ease of accessibility from memory? This assumption might be supported by research addressing the delayed JOL effect. Immediate JOLs are less accurate than delayed JOLs by showing less agreement with subsequent performance (Nelson & Dunlosky, 1991). That is, delayed judgments of learning are more likely to match subsequent recognition than immediate judgments of learning. The pronounced agreement between delayed judgments and recognition might be due to the retrieval of similar memory traces.

Cues and Heuristics

Deciding under uncertainty may trigger some rules or heuristics to choose an alternative that seems to be the best (Tversky & Kahneman, 1973). It has already been shown that even young children use cues as a source of information to guide their memory decisions

(Geurten & Willems, 2016; Koriat, 1997; Koriat et al., 2009; Roebers et al., 2019). Using invalid cues, for example, such as how easily accessible the learned item pairs can be retrieved (Dunslosky & Nelson, 1992; Koriat, 2007; Koriat & Levy-Sadot, 2001; Mazzoni & Nelson, 1995) might have influenced monitoring accuracy. The more accessible such information can be retrieved, the higher the feeling of confidence (Bastin et al., 2019; Benjamin, 2005; Metcalfe & Finn, 2008; Unkelbach & Stahl, 2009) and might, therefore, also guide monitoring processes (Hertzog et al., 2003; Kelley & Lindsay, 1993). Furthermore, the evaluation of pre-monitoring can act as an indirect mediator between cue and target. Due to the selection of a high pre-monitoring judgment, the connection between cue and target is further reinforced (Atiya et al., 2019). Thus, this implicit reinforcement can be associated with higher confidence (Pyc & Rawson, 2010; Whittlesea & LeBoe, 2003). Repeated exposure to the same item throughout the task may have strengthened the influence of these cues. However, invalid cue utilization might then be a reason for the observed overconfidence in young children.

Evidence Accumulation and Selective Attention

Navajas et al. (2016) suggested that even if a decision has been made, there is still further accumulation of information that can work in favor for or against one's answer. Several studies have shown that a hasty answer often suffers from inaccuracy (Simpson et al., 2007). With additional time to perform a differentiated analysis, including monitoring, further information is considered before choosing the best candidate answer (Atiya et al., 2019; Wacker & Roebers, 2022). Current research addressed evidence accumulation under uncertainty with a perceptual discrimination paradigm in toddlers (Leckey et al., 2020). Including eye movement measures, such as gaze switching, shed light on the implicit processing of evidence accumulation. Results revealed that toddlers accumulate more information through gaze switching when uncertain. Therefore, depending on item difficulty, a pre-monitoring phase might lead to more gaze switching between difficult items and slower evidence accumulation resulting in lower confidence. Therefore, a differentiated evaluation of all four alternatives and a subsequent monitoring phase, as in our study design, should provide a valuable opportunity (Murphy et al., 2015). However, results indicated that instead of accumulating further information, children might feel more confident through a misleading accumulation of invalid information (Talluri et al., 2018; Rollwage et al., 2020). In addition, the time invested (i.e., to exclude other alternatives with lower certainty) could strengthen the feeling of accomplishing a differentiated process. Consequently, the preferred answer with the highest certainty rating would gain a subjective feeling of confidence.

Furthermore, selective activation of attention to a particular alternative influences subsequent cognitive activation (Nadel et al., 2012). The selective attention may have contributed to boosting overconfidence. Moreover, after choosing an alternative, the selected answer was highlighted with a yellow frame. This visual accentuation might have additionally increased selective attention to that one particular alternative.

A Closer Look at the Scale Level of Both Monitoring Measures

We can distinguish three different patterns of association between the pre-monitoring and the corresponding confidence judgments. In the first pattern, children enhanced their confidence judgment compared to the prior pre-monitoring judgment indicating a tendency towards increasing overconfidence through pre-monitoring. In the second pattern, participants chose the same confidence level for the pre-monitoring and the subsequent confidence judgment - this association was most frequent. However, as our results revealed, for both monitoring measures, most answers were already scored on the highest scale level. Therefore, the first pattern of association (tendency towards enhancing confidence through the task) may be underrepresented because of that obvious ceiling effect. Because children mainly chose already the highest score in their pre-monitoring judgment, their confidence judgment could not be rated higher – even if children would have wanted to. Contrarily, the possibility of downregulating their confidence would have been possible, as shown in the third pattern of association. Such a downregulation becomes visible when a confidence judgment is lower than the prior pre-monitoring judgment. This downregulation indicates that a differentiated evaluation of answer alternatives might trigger doubts or lead children to give their confidence second thoughts. As a result, children might reduce their over-optimistic estimation and rate a more fine-tuned and accurate monitoring judgment. Results from the latter pattern, that is, a downregulation of confidence, are still promising. In line with our assumptions, in around one-fifth of the item judgments there was a downregulation of confidence. In other words, in 20% of the items, a differentiated evaluation might indeed have led to more accurate monitoring in our participants. Besides the strong pattern of overconfidence, this smaller proportion of downregulating confidence seems promising. Of course, it raises the question of to what extent we can support children to increase the proportion of downregulation.

From a theoretical point of view, the attempt to downregulate confidence might be only relevant for incorrect answers because being sure concerning a correct answer is reasonable and well justified. However, results revealed that regardless of correctness, most confidence judgments indicated the same overoptimistic confidence rating. Thus, it is not that children should no longer be confident in correct answers per se, but children might benefit in general from a more fine-tuned monitoring.

Implications

Based on the present results, the question arises as to what extent we can measure pure monitoring. These findings may, therefore, also influence research on different monitoring measures. The high accordance of pre-monitoring judgment and recognition underscores the assumption that monitoring judgments such as JOL and FOK are predictive of later retrieval (Hart, 1965). However, this prediction may be influenced by the strength of the memory traces (Koriat, 2007). The best candidate answer may be selected based on retrieval processes and indirectly triggers searching information confirming this alternative (Koriat & Goldsmith, 1996; Vernon & Usher, 2003). Having previously engaged in a differentiated evaluation for choosing the best candidate answer may foster a subjective feeling of high confidence in the decision, which is then – in most cases – maintained. Overall, children most often choose the highest point on the scales to express their confidence. The tendency of favoring high confidence levels in both of our scales (pre-monitoring and confidence judgments) confirms that overconfidence is a robust phenomenon. Including a broader age range in future research would be interesting. The present study revealed indications that participants sometimes downregulated their confidence judgments after the pre-monitoring phase. Including different age groups could – in this context - shed light on whether there are further subgroups or a possible age range where participants may systematically benefit from pre-monitoring.

Limitations

The selection of our stimulus pictures with animals and the corresponding food or environment of the animal might have induced a particularly stronger bias towards high confidence compared to other paired-associate learning tasks. Children at this age are often in contact with animals (i.e., books, etc.). Consequently, the familiarity of the pictures might have reinforced a strong feeling of familiarity at this age. The influence of this cue, including "ease of processing" might have reinforced the confirmation bias. As even young children use mnemonic cues to evaluate their confidence (Geurten et al., 2018; Koriat, 2007; Roebers et al., 2019), these may have further increased confidence. Therefore, using less familiar stimuli, may reduce familiarity and the concomitant confirmation bias (Destan & Roebers, 2015; Roderer & Roebers, 2010).

Furthermore, even if we evaluated our items and their corresponding item difficulty, there might be some noise within the alternatives. For example, some alternatives might associate more or less strongly with the stimulus picture, but this should be unsystematic across children. Of course, some items were easier or more difficult, also depending on the answer alternatives, but such variability in item difficulty is important to have in a task. Moreover, continuously changing confidence is also part of daily life experiences. However, the present approach contained a larger number of items than other studies including these age groups, and in our pilot study, we carefully evaluated different answer alternatives. Together, we have reason to believe that the task appropriately mapped both age-related differences and individual differences, within age-groups.

We used the same scales as in the retrospective confidence judgments for the pre-monitoring phase. However, they might trigger different processes because one phase is completed before and the other after recognition. Nevertheless, it is important to note that our focus was laid on the retrospective confidence judgments, thereby investigating whether a pre-monitoring phase can positively affect monitoring compared to the control condition. Only within our post hoc analysis we compared the two measures, however, not with the point of view that both arise on the base of the same cue but rather to investigate on which scale level participants expressed their confidence.

Conclusion

In sum, the present study offers new insights into the association between memory and metamemory, showing that the relationship seems far more complex than previously expected. Therefore, the question arises of how we can capture monitoring processes more independent from prepotent memory traces. Results indicated that the influence of prepotent memory traces might have additionally increased overconfidence in young children. The association between memory and metamemory seems underestimated and more critical than assumed. More research is needed to understand the underlying mechanisms of overconfidence in young children.

Author's contribution Conceptualization: Sophie Wacker, Claudia M. Roebers.
Methodology: Sophie Wacker, Claudia M. Roebers.
Formal analysis: Sophie Wacker.
Project administration: Sophie Wacker, Claudia M. Roebers.
Supervision: Claudia M. Roebers.
Writing – original draft: Sophie Wacker.
Writing – review & editing: Sophie Wacker.

Funding Open access funding provided by University of Bern

Data availability Data are available on request from the author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Ethical approval based on the APA's Ethical Principles of Psychologists and Code of Conduct (2002) was obtained from the Faculty's Ethics Committee (Faculty of Human Sciences, University of Bern; Approval No. 2020-10-00005).

Informed consent Written consent (by caregivers of participants, data was anonymous).

References

Atiya, N. A., Rañó, I., Prasad, G., & Wong-Lin, K. (2019). A neural circuit model of decision uncertainty and change-of-mind. *Nature Communications, 10*(1), 1–12. https://doi.org/10.1038/s41467-019-10316-8

Azevedo, R. (2020). Reflections on the field of metacognition: Issues, challenges, and opportunities. *Metacognition and Learning, 15*(2), 91–98. https://doi.org/10.1007/s11409-020-09231-x

Bakracevic Vukman, K., & Licardo, M. (2010). How cognitive, metacognitive, motivational and emotional self-regulation influence school performance in adolescence and early adulthood. *Educational Studies, 36*(3), 259–268. https://doi.org/10.1080/03055690903180376

Barnes, A. E., Nelson, T. O., & Dunlosky, J. (1999). 10 An Integrative System of Metamemory Components Involved in Retrieval. Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application, (pp. 287-313). Cambridge, MA: MIT Press.

Bastin, C., Besson, G., Simon, J., Delhaye, E., Geurten, M., Willems, S., & Salmon, E. (2019). An integrative memory model of recollection and familiarity to understand memory deficits. *Behavioral and Brain Sciences, 42*, e281. https://doi.org/10.1017/S0140525X19000621

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remembering" and a caution about purportedly nonparametric measures. *Memory & Cognition, 33*(2), 261–269. https://doi.org/10.3758/BF03195315

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing About Knowing* (pp. 185–205). MIT Press.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Bryce, D., Whitebread, D., & Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning, 10*(2), 181–198. https://doi.org/10.1007/s11409-014-9120-4

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*(1), 26–48. https://doi.org/10.3758/BF03210724

Conn, M., M'Bale, K., & Josyula, D. (2018). Multi-level metacognition for adaptive behavior. *Biologically Inspired Cognitive Architectures, 26*, 174–183. https://doi.org/10.1016/j.bica.2018.10.006

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science, 18*(6), 957–971. https://doi.org/10.1111/desc.12271

de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Destan, N., & Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition and Learning*, *10*(3), 347–374. https://doi.org/10.1007/s11409-014-9133-z

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

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. https://doi.org/10.1016/j.learninstruc.2012.06.001

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374–380. https://doi.org/10.3758/BF03210921

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Ferreira, C. S., Charest, I., & Wimber, M. (2019). Retrieval aids the creation of a generalized memory trace and strengthens episode-unique information. *NeuroImage, 201*, 115996. https://doi.org/10.1016/j.neuroimage.2019.07.009

Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1–9. https://doi.org/10.1016/j.learninstruc.2014.01.001

Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology, 2*(4), 421–453. https://doi.org/10.1016/0010-0285(71)90025-9

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Geurten, M., & Bastin, C. (2019). Behaviors speak louder than explicit reports: Implicit metacognition in 2.5-year-old children. *Developmental Science, 22*(2), e12742. https://doi.org/10.1111/desc.12742

Geurten, M., & Willems, S. (2016). Metacognition in early childhood: Fertile ground to understand memory development? *Child Development Perspectives, 10*(4), 263–268. https://doi.org/10.1111/cdep.12201

Geurten, M., Meulemans, T., & Willems, S. (2018). A closer look at children's metacognitive skills: The case of the distinctiveness heuristic. *Journal of Experimental Child Psychology, 172*, 130–148. https://doi.org/10.1016/j.jecp.2018.03.007

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 429–456). Erlbaum Publishers.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208–216. https://doi.org/10.1037/h0022263

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of memory and language, 32*(1), 1–24. https://doi.org/10.1006/jmla.1993.1001

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.012

Koriat, A. (2015). Knowing by doing: When metacognitive monitoring follows metacognitive control. In D. S. Lindsay (Ed.) & C. M. Kelley (Trans.) & A. P. Yonelinas, H. L. Roediger II (Eds.), Remembering:

Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby (pp. 185–197). Psychology Press.
Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13*(3), 441–453. https://doi.org/10.1111/j.1467-7687.2009.00907.x
Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490
Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 34–53. https://doi.org/10.1037/0278-7393.27.1.34
Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive Judgments and their Accuracy. In P. Chambres, M. Izaute, & P. J. Marescaux (Eds.), *Metacognition*. Springer. https://doi.org/10.1007/978-1-4615-1099-4_1
Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107
Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The easily learned, easily remembered heuristic in children. *Cognitive Development, 24*(2), 169–182. https://doi.org/10.1016/j.cogdev.2009.01.001
Leckey, S., Selmeczy, D., Kazemi, A., Johnson, E. G., Hembacher, E., & Ghetti, S. (2020). Response latencies and eye gaze provide insight on how toddlers gather evidence under uncertainty. *Nature Human Behaviour, 4*(9), 928–936.
Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology, 103*(2), 152–166. https://doi.org/10.1016/j.jecp.2008.10.002
Lipko, A. R., Dunlosky, J., Lipowski, S. L., & Merriman, W. E. (2012). Young children are not underconfident with practice: The benefit of ignoring a fallible memory heuristic. *Journal of Cognition and Development, 13*(2), 174–188. https://doi.org/10.1080/15248372.2011.577760
Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology, 49*(8), 1505–1516. https://doi.org/10.1037/a0030614
Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*(6), 1778–1787. https://doi.org/10.1111/j.1467-8624.2011.01649.x
Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development, 84*(2), 726–736. https://doi.org/10.1111/cdev.12004
Lyons, K. E., & Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. *Advances in Child Development and Behavior, 40*, 379–412. https://doi.org/10.1016/B978-0-12-386491-8.00010-4
Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1263–1274. https://doi.org/10.1037/0278-7393.21.5.1263
Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084–1097. https://doi.org/10.1037/a0012580
Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *Elife, 4*, e11946. https://doi.org/10.7554/eLife.11946
Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews, 36*(7), 1640–1645. https://doi.org/10.1016/j.neubiorev.2012.03.001
Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in confidence. *Current Opinion in Behavioral Sciences, 11*, 55–60. https://doi.org/10.1016/j.cobeha.2016.05.005
Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5
Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2*(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175
Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. *The Journal of Experimental Education, 74*(1), 7–28 http://www.jstor.org/stable/20157410

O'Leary, A. P., & Sloutsky, V. M. (2017). Carving metacognition at its joints: Protracted development of component processes. *Child Development, 88*(3), 1015–1032. https://doi.org/10.1111/cdev.12644

Oudman, S., van de Pol, J., & van Gog, T. (2022). Effects of self-scoring their math problem solutions on primary school students' monitoring and regulation. *Metacognition and Learning, 17*(1), 213–239. https://doi.org/10.1007/s11409-021-09281-9

Peters, U. (2022). What is the function of confirmation bias? *Erkenntnis, 87*(3), 1351–1376. https://doi.org/10.1007/s10670-020-00252-1

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 335–335. https://doi.org/10.1126/science.1191465

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

Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development*, *3*(4), 415–443. https://doi.org/10.1080/15248372.2002.9669676

Roderer, T., & Roebers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. *Metacognition and Learning*, *5*(3), 229–250. https://doi.org/10.1007/s11409-010-9059-z

Roebers, C. M., & Schneider, W. (2005). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology*, *91*(1), 24–44. https://doi.org/10.1016/j.jecp.2005.01.001

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N. S., & van Loon, M. H. (2019). The role of children's metacognitive experiences for cue utilization and monitoring accuracy: A longitudinal study. *Developmental Psychology*, *55*(10), 2077. https://doi.org/10.1037/dev0000776

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications, 11*(1), 1–11. https://doi.org/10.1038/s41467-020-16278-6

Saenz, G. D., Geraci, L., & Tirso, R. (2019). Improving metacognition: A comparison of interventions. *Applied Cognitive Psychology, 33*(5), 918–929. https://doi.org/10.1002/acp.3556

Schneider, W. (2014). *Memory development from early childhood through emerging adulthood*. Springer.

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 491–518). Oxford University Press.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371. https://doi.org/10.1007/BF02212307

Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology, 43*(2), 417–428. https://doi.org/10.1037/0012-1649.43.2.417

Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory processes: How response prepotency is created and overcome. *Developmental Science, 15*(1), 62–73.

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology, 28*(19), 3128–3135. https://doi.org/10.1016/j.cub.2018.07.052

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*(1), 22–38. https://doi.org/10.1016/j.concog.2008.09.006

van Loon, M. H., & Roebers, C. M. (2021). Using feedback to support children when monitoring and controlling their learning. In D. Moraitou & P. Metallidou (Eds.), *Trends and prospects in metacognition research across the life span*. Springer. https://doi.org/10.1007/978-3-030-51673-4_8

van Loon, M. H., de Bruin, A. B., van Gog, T., van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta psychologica, 151*, 143–154. https://doi.org/10.1016/j.actpsy.2014.06.007

van Loon, M., de Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? Developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology*, 158, 77–94. https://doi.org/10.1016/j.jecp.2017.01.008

van Loon, M. H., Bayard, N. S., Steiner, M., & Roebers, C. M. (2022). The accuracy and annual rank-order stability of elementary school children's self-monitoring judgments. *Journal of Applied Developmental Psychology*, 80, 101419. https://doi.org/10.1016/j.appdev.2022.101419

Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre-and postretrieval mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 339–346. https://doi.org/10.1037/0278-7393.29.3.339

Wacker, S., & Roebers, C. M. (2022). Stop and think: Additional time supports monitoring processes in young children. *PLoS ONE*, *17*(9), e0274460. https://doi.org/10.1371/journal.pone.0274460

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of educational research, 63*(3), 249–294. https://doi.org/10.3102/00346543063003249

Whittlesea, B. W., & LeBoe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49*(1), 62–79. https://doi.org/10.1016/S0749-596X(03)00009-3