DOI: 10.1111/cdev.14174

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Toddlers' visual exploration during decisions predicts uncertainty monitoring 1 year later

Sarah Leckey[1,2] | **Christopher Gonzales**[1,2] | **Diana Selmeczy**[3] | **Simona Ghetti**[1,2]

[1]Center for Mind and Brain, University of California, Davis, Davis, California, USA

[2]Department of Psychology, University of California, Davis, Davis, California, USA

[3]Department of Psychology, University of Colorado, Colorado Springs, Colorado Springs, Colorado, USA

Correspondence
Sarah Leckey and Simona Ghetti, Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, CA 95616, USA.
Email: ssleckey@ucdavis.edu and sghetti@ucdavis.edu

Funding information
National Science Foundation, Grant/Award Number: BCS1424058; National Institute of Child Health and Human Development, Grant/Award Number: F31HD102153

Abstract
The longitudinal relation between toddlers' behaviors in uncertain situations (e.g., information seeking, hesitation) and preschoolers' uncertainty monitoring was investigated (between 2014 and 2019 in Northern California; Time 1: N = 183, M = 28.99 months, 53% female, 67.8% White; Time 2: N = 159, M = 41.64 months, 52.2% female). Eye movements and response latencies were recorded as children identified a target from two partially occluded (Time 1) or degraded (Time 2) images. Confidence ratings for identifications were collected at Time 2. At Time 1, gaze transitions between response options, but not response latencies and mental state language, predicted Time 2 uncertainty monitoring. Overall, these findings provide the first direct evidence of connections between toddlers' uncertainty behaviors and preschoolers' uncertainty monitoring.

The ability to assess subjective experiences of uncertainty is foundational to humans' ability to explore, learn, and engage in effective decision making (e.g., De Martino et al., 2013; Fandakova et al., 2017; Fleming et al., 2012; Ghetti et al., 2013). Little is known about the emergence and early development of this ability, but recent evidence suggests that infants and toddlers exhibit behavioral responses to uncertain situations, such as looking towards caregivers or known experts when they are faced with unknown or uncertain situations. For example, infants as young as 6 months look longer at their parent when their expectations about known properties of the world are violated (Walden et al., 2007). By 12 months of age, infants look towards a speaker more when they utter a novel label in the presence of two unknown objects compared to one object, suggesting that they implicitly recognize that the referent of the novel label is unclear (Bazhydai et al., 2020; Vaish et al., 2011). Additionally, 16- to 20-month-olds are more likely to turn to their caregivers for help during difficult compared to easy trials (Goupil et al., 2016). Finally, 2-year-olds respond faster (e.g., tapping the location where a target picture is presented) when they identify the target correctly compared to when they make an inaccurate selection (Leckey et al., 2020), and switch their gaze more frequently between response options during more difficult decisions (i.e., when the target and distracter were similar compared to dissimilar).

Overall, these results suggest that infants and toddlers are sensitive to decision difficulty and attempt to gather more information to disambiguate the situation well before they can experience feelings of confidence that are calibrated to their actual decision accuracy and report them on a confidence scale. The latter abilities are clearly demonstrated between 3 and 4 years of age (Coughlin et al., 2015; Hembacher & Ghetti, 2014). However, it is unclear whether these early behaviors demonstrated by toddlers are functionally related to or

Abbreviations: AOI, areas of interest; CFI, comparative fit index; RMSEA, root mean square error of approximation; ToM, theory of mind.

provide the foundation for uncertainty monitoring later in childhood. Thus, the purpose of the current study was to evaluate whether early behavioral responses to uncertain situations predict later emerging uncertainty monitoring abilities (i.e., indicated by higher subjective confidence for accurate compared to inaccurate decisions). For the purposes of this study, we refer to these early behavioral responses as "uncertainty behaviors". In contrast, we refer to the ability to differentiate accurate from inaccurate decisions in subjective confidence reports (i.e., report greater confidence for accurate compared to inaccurate decisions; Lyons & Ghetti, 2011; Roebers et al., 2004) as "uncertainty monitoring".

We examined two uncertainty behaviors as candidates for longitudinal predictors of uncertainty monitoring, namely gaze transitions between response options prior to rendering a decision and response latencies to the decision. Indeed, research in older children and adults indicates that more gaze transitions and longer response latencies are associated with subjective assessments of uncertainty (Folke et al., 2016; Koriat & Ackerman, 2010; Lyons & Ghetti, 2011; Roderer & Roebers, 2014; Zakay & Tuvia, 1998). Overall, behavioral responses that track response accuracy in toddlers (Dautriche et al., 2022; Leckey et al., 2020) also track confidence ratings when children and adults can provide them. The possibility of a developmental continuity between these behavioral responses and later subjective uncertainty leads to the hypothesis that these uncertainty behaviors may be precursors of later uncertainty monitoring. This prediction may align well with current theoretical accounts about early manifestations of metacognition.

Specifically, it has been proposed that a rudimentary form of uncertainty monitoring is already in place during infancy (Goupil & Kouider, 2019; Proust, 2019; Shea et al., 2014). This rudimentary system allows for the automatic monitoring and control of actions, such as information seeking or selection without conscious awareness (Goupil & Kouider, 2019) and is proposed to underlie uncertainty behaviors which emerge during late infancy (Beran et al., 2012; Goupil et al., 2016). As the underlying brain structures and companion abilities develop and mature, conscious uncertainty monitoring that can be explicitly shared with others emerges and continues to be refined throughout childhood. Although to date, this account does not offer explicit predictions as to whether or to what extent a developmental continuity should be expected from the early rudimentary system to later emerging uncertainty monitoring, their developmental relation is plausible. Specifically, individual differences in uncertainty behaviors may reflect variability in the extent to which toddlers seek out information during difficult decisions. The repeated experience of the association between information seeking, hesitation, and decision difficulty may pave the way for gaining awareness of which decisions may be more, or less, uncertain. Thus, we hypothesized that uncertainty behaviors at 2 years of age would predict the capacity to reflect on uncertainty at 3 years of age, suggesting that the uncertainty behaviors observable at 2 years of age could potentially be a manifestation of this early, more automatic, uncertainty monitoring.

This prediction stands in contrast to another extant theoretical account. In this account, early behavioral responses do not reflect a rudimentary form of uncertainty monitoring, but rather the appraisal of the levels of risk associated with potential choices, resulting in the selection of the less risky choice (Carruthers, 2009; Carruthers & Williams, 2019). This process is not considered metacognitive because it does not require any explicit conceptual understanding of uncertainty, risk, or any other mental operation involved in decision making. From this perspective, there is no obvious reason to predict that early uncertainty behaviors would be longitudinal predictors of uncertainty monitoring. Instead, this account would lead to the alternative hypothesis that children may begin to utilize their uncertainty behaviors as cues to their subjective uncertainty states only when they have already developed the capacity to conceptually understand uncertainty and overtly introspect about and report on uncertainty. Results consistent with this alternative would involve significant concurrent associations between uncertainty behaviors and uncertainty monitoring at age 3, but no direct longitudinal relation from uncertainty behaviors at age 2 and uncertainty monitoring at age 3. Although we agree that early uncertainty behaviors are unlikely to be metacognitive in their own right because we do not have direct evidence that infants and young children are aware of their mental states, we propose that experience with these uncertainty behaviors may facilitate the emergence of uncertainty monitoring.

We recognize that there are additional accounts surrounding the factors underlying the emergence of uncertainty monitoring during the preschool years, including the roles of early awareness of mental states and conceptual understanding of the mind. Around 2 years of age, toddlers begin to show initial, explicit access to their feelings of ignorance (e.g., Harris et al., 2017; Marazita & Merriman, 2004) as signaled by their use of the verbal expression "I don't know" or gestures such as shrugging (Leckey et al., 2020). This early access to primitive mental states and capacity to communicate may support young children's ability to identify and label confidence states above and beyond behavioral signals (Goupil & Kouider, 2019; Paulus et al., 2013).

With regard to conceptual understanding of the mind, toddlers' developing theory of mind (ToM) abilities may also promote uncertainty monitoring. ToM, which is the understanding of mental states in oneself and others, and how these mental states may change depending on situations (Wellman & Liu, 2004), develops across early childhood (Wellman, 2011). This ability may help children gain an understanding of how mental processes are deployed during a task, and how decisions may be inaccurate, which would then allow them to subjectively assess their own performance (Perner, 2000). These ideas have led researchers to propose that a ToM develops before uncertainty monitoring, as the latter requires a representation upon oneself (Carruthers, 2009). Others argue that uncertainty monitoring must develop first, because in order to attribute mental states onto others, one must first be able to understand them within oneself (Goldman, 2006). With either hypothesis, indicators of ToM, such as understanding of false belief (Carruthers, 2009), would be related to uncertainty monitoring. However, previous studies testing this hypothesis in the context of the same perceptual task used here failed to find an association. Specifically, Coughlin et al. (2015) found that 3- to 5-year-olds who exhibited better false belief understanding did not show better uncertainty monitoring; however, they were more likely to ask for help on difficult trials (See also Baer et al., 2021). This result suggests that understanding of false belief may be more important to evaluate how to act on states of uncertainty, rather than computing one's own uncertainty assessments, especially when others' states of knowledge are to be considered. Nevertheless, we deemed it important to assess the association between explicit false belief understanding and uncertainty monitoring in the context of the present study. In this research, an alternative model with these variables was also evaluated in order to assess whether individual differences in the extent toddlers generally use mental state language and preschoolers developing ToM abilities may additionally contribute to uncertainty monitoring.

The present study

We conducted a pre-registered longitudinal study to examine how individual differences in uncertainty behaviors at the age of 2, including gaze transitions between response options and response latencies (Leckey et al., 2020), predicted uncertainty monitoring abilities measured through subjective un/certainty ratings at the age of 3. At the first assessment point (Time 1), 2-year-olds completed a perceptual identification task in which they were asked to identify a target from partially occluded images (Figure 1a). These same participants returned about 1 year later to complete their second assessment (Time 2). At that time, they completed a conceptually similar, but age-appropriate perceptual identification

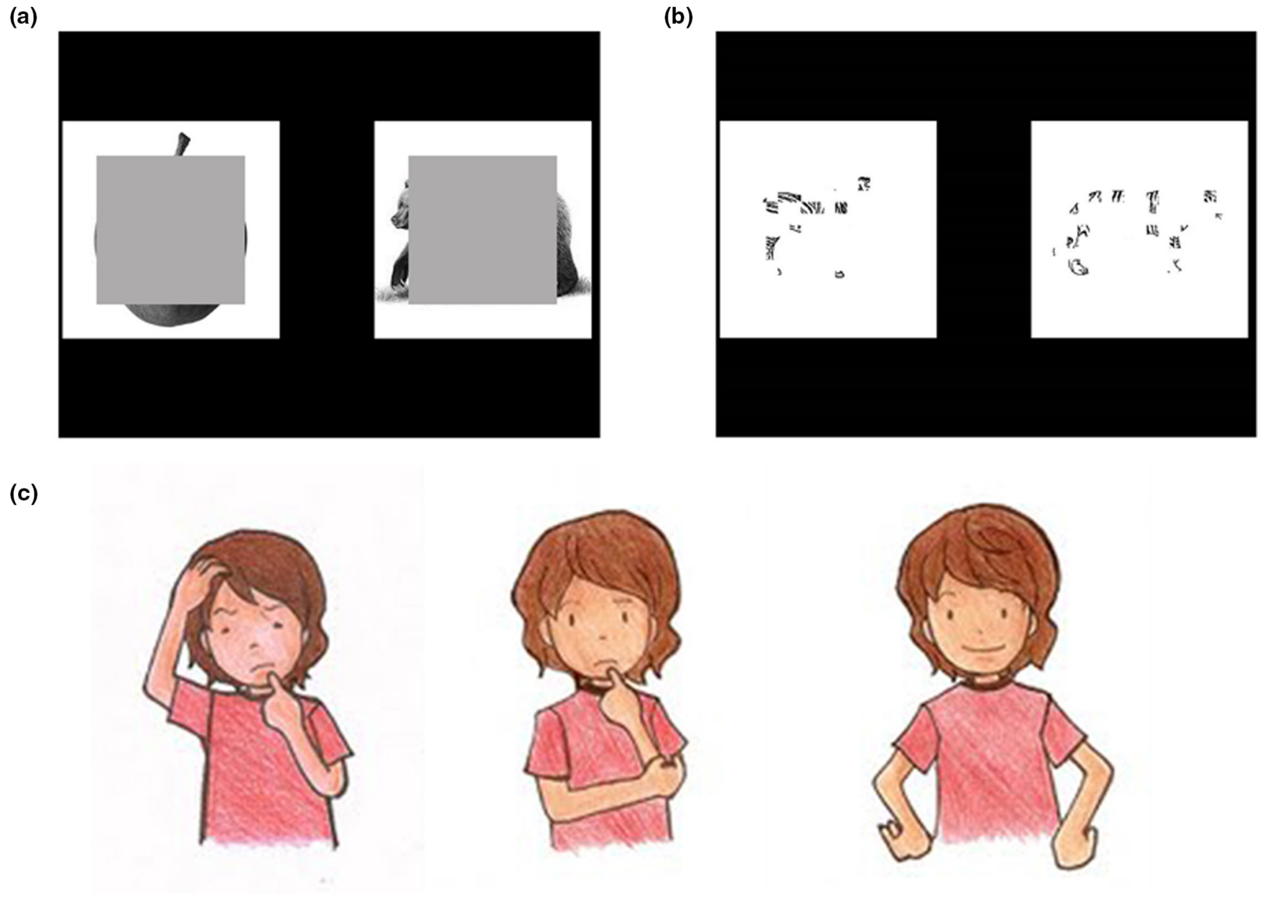

FIGURE 1 Perceptual task stimuli. (a) Example of Time 1 perceptual task stimuli. (b) Example of Time 2 perceptual task stimuli. (c) Example of pictorial confidence rating scale for Time 2.

task (e.g., Coughlin et al., 2015; Figure 1b), which required children to select a target from two response options depicting visually degraded objects and to rate their confidence for each identification (Figure 1c). To the extent that gaze transitions and response latencies (at either Time 1 or Time 2) are related to uncertainty monitoring, we predicted that more gaze transitions and longer response latencies would predict better uncertainty monitoring. A model testing the additional hypotheses concerning early mental state language and ToM discussed earlier was also assessed. We predicted that if mental state language and ToM are related to uncertainty monitoring, then more mental state language use (at either Time 1 or Time 2) and better ToM at Time 2 would predict better uncertainty monitoring at Time 2.

METHODS

Participants

Time 1 included 183 toddlers aged 25–34 months (M=28.99 months, 97 female). Time 2 occurred on average 12.61 months (SD=3.39) later and included 159 children from Time 1, now preschoolers, aged 36–60 months (M=41.64 months, 83 female). Reasons for participants not returning for Time 2 included not responding to recruitment calls (6), moved out of area (3), no longer interested in participating (2), not returning after scheduling sessions (2), or not being recruited for Time 2 (11).

Parents reported their children's race as White (N=124), Asian (N=5), African American (N=5), Pacific Islander (N=1), multi-racial (N=37), and unreported (N=11). Parents of 31 children identified them as Hispanic. Families' household incomes were less than \$15,000 (N=4), \$15,000–\$25,000 (N=4), \$25,000–\$40,000 (N=16), \$40,000–\$60,000 (N=32), \$60,000–\$90,000 (N=34), more than \$90,000 (N=89) and unreported (N=4). The children were recruited from a Northern California community between 2014 and 2019, from a database of families contacted from birth records, who had expressed interest in participating in child development studies. None of the children had a history of developmental or speech delays. This experiment was approved by the Institutional Review Board of the University of California, Davis, and informed consent was obtained from all parents.

Materials and procedure

Time 1 measures

Perceptual task

Toddlers were assessed on two versions of a perceptual identification task (Figure 1a). Each version had 20 trials. This task included a manipulation of decision difficulty. On each trial, toddlers saw two occluded images side by side. The stimulus pairs were either perceptually and semantically similar (e.g., an elephant and a bear) or dissimilar (e.g., a cow and broccoli) to one another. Toddlers were asked to select the target image by touching it (touchscreen version) or pointing to it (eye-tracker version). Stimuli included 40 images of common objects and animals (Yassa et al., 2011) and were selected based on age-of-acquisition norms (Kuperman et al., 2012). Results from this task have been reported separately (Leckey et al., 2020); therefore, the decision difficulty manipulation is not discussed further and we collapsed across decision difficulty for the purpose of the current analyses. Here, we discuss how we collected and processed data for the purpose of the longitudinal follow-up.

The eye-tracker version was administered on a Tobii T-120 17-in eye-tracker monitor. Children sat on their parents' laps approximately 60 cm from the monitor. All stimuli were 10 cm × 10 cm (visual angle 9.53) with 4.45 cm (visual angle 4.24) between them. Tobii Studio's standard infant calibration was used; a cartoon cat was presented at 5 points on the screen accompanied by a sound effect. The experiment proceeded when children's gaze was captured at all 5 points. Default Tobii fixation filter settings were used for eye movement data reduction (velocity threshold: 35 pixels per sample; distance threshold: 35 pixels; minimum fixation duration: 83 ms). Parents wore dark sunglasses to ensure that their eye movements were not unintentionally recorded and were instructed to refrain from speaking to their child during the task and hold their child to prevent them from leaning forward or moving excessively. Toddlers were introduced to the task by being told, "Now we are going to find some things that are hiding behind boxes!" and were instructed to respond by pointing to the appropriate image on the screen. For each trial, the experimenter asked, "Can you find the (*target*)?" on a blank screen and then pressed a button to present the trial. As soon as the child pointed to a response, the experimenter keyed in their response, advancing the task to a blank screen with a fixation cross in the middle. If the child refused to respond, the experimenter keyed in a separate code and moved on to the next trial.

The touchscreen version was identical to the eye-tracker version, except that children sat in front of the monitor on a child-sized chair and were instructed to respond by touching the appropriate image on an upright touchscreen monitor. As soon as the child touched an image, the task advanced to a blank screen before starting the next trial. As in the eye-tracker version, if a child refused to respond, the experimenter keyed in a separate code and moved on to the next trial. The eye-tracker version of the perceptual discrimination task was always completed before the touchscreen version due to piloting revealing a tendency for toddlers to lean forward during the eye-tracker version if they completed the touchscreen version first, which interfered with accurate eye movement data collection.

Not all participants contributed data in both versions of the perceptual task due to some toddlers being uncooperative in one (27) or both versions (8), not returning for all sessions (3), computer issues (2), or they completed an advanced pilot version of the perceptual task at Time 1 (6). Therefore, 149 toddlers contributed eye-tracker data and 157 toddlers contributed touchscreen data.

The variables obtained from this perceptual task included accuracy, response latencies, and gaze transitions. Response latencies were obtained from the touchscreen version of the task and were measured as the time from when the stimulus was shown on the screen to the time when the participant touched the screen for their response. The final response latency variable was an average of response latencies from all valid trials. Gaze transitions were obtained from the eye-tracker version of the task. In order to calculate gaze transitions, we used Tobii Studio software to create areas of interest (AOIs). These AOIs encompassed individual square images surrounding the target and distractor stimuli respectively, so that each trial had a target AOI and a distractor AOI. Our gaze transitions variable was calculated by counting how many times the children's gaze transitioned from one stimulus to the other during a trial. We defined a transition as a fixation on an AOI that was preceded by a fixation to the other AOI, including instances in which there were fixations on other (non-AOI) areas of the screen in between fixations to AOIs. The first fixation to an AOI in a trial was not counted as a transition, so the minimum number of transitions in a trial was zero. The final gaze transition variable was an average of gaze transitions from all valid trials. The accuracy variable was an average of the accuracy scores from the touchscreen and eye-tracker versions of the task. To calculate this average, we first calculated the participants average accuracy for both versions separately, and then took an average of those two scores. We considered this approach acceptable because there were no significant differences between participants scores on the different versions of the task ($t(136)=-1.63$, $p=.11$, $d=0.25$) and there was a significant correlation between the two, $r=.34$, $p<.001$. For the participants who did not complete both versions of the task, we predicted missing values by utilizing a linear regression model predicting either touchscreen accuracy from eye-tracker accuracy or vice versa. The resulting accuracy average was thus obtained from two scores across all participants.

Before calculating the accuracy and response latency variables from the touchscreen version, we removed trials for which parents reported that the child did not know the word for the object they were being requested to find. This resulted in 114 trials (3.63%) across 60 toddlers being eliminated from analyses. Then, we removed trials for which children did not provide an answer. This resulted in 25 trials (0.80%) across 17 toddlers being eliminated. Next, we removed any trials with response latencies less than 700 ms in duration. These trials were likely to be responses produced before processing the stimuli or trials in which the toddler was inattentive. Previous research has indicated that the average time for 5-year-olds to make a motor response is around 750 ms (Droit-Volet, 2010), so with our sample, it was decided that 700 ms was a reasonable reaction time cutoff. This criterion resulted in 9 trials (0.29%) across 8 toddlers being eliminated from analyses. Finally, we also removed trials where the z-scored response latencies across each individual participant were ±3 SDs. This resulted in 76 trials (2.42%) across 76 toddlers being eliminated from analyses. Once these eliminations were made average response latencies and accuracy for each participant were calculated.

Before calculating the accuracy and gaze transition variables for the eye-tracker version, similar to the touchscreen version, we eliminated trials for which the word was not known. This resulted in 120 trials (4.06%) across 57 toddlers being removed from analysis. Next, we removed trials for which children did not provide an answer. This resulted in 46 trials (1.56%) across 22 toddlers being eliminated from analyses. Finally, we also removed trials for which the eye-tracker did not measure any look time to either AOI, target or distractor. This criterion resulted in the exclusion of 367 trials (12.41%) across 84 toddlers being eliminated from analyses. Once these eliminations were made accuracy and average gaze transitions for each participant were calculated.

Use of ignorance expression

To assess toddlers use of ignorance expressions during the laboratory visit, we had parents report on their child's use of mental state language sentences using an adapted version of the Cognition portion of the Internal States Language Questionnaire (Bretherton & Beeghly, 1982). The sentences included "I think", "I know", "I guess", "I'm sure" and "I don't know". Parents rated how often their child used each of these phrases on a five-point scale from 1 ("never/not yet") to 5 ("often"). Although we collected information on all of these phrases with this questionnaire, we chose to only use the single item that assessed the frequency of use of "I don't know" because this phrase has been shown to emerge earlier than other mental state language phrases, has a more direct behavioral expression (e.g., shrugging; Harris et al., 2017), and has been used in a previous paper with 2-year-olds (Leckey et al., 2020).

General vocabulary

To determine whether any findings concerning mental state language were driven by general language ability, we conducted a supplemental path model which included a measure of general vocabulary. Time 1 general vocabulary was assessed with MacArthur-Bates Communicative Development Inventory-II language questionnaire

(Fenson et al., 2000). This is a questionnaire filled out by the parent which asks about their toddlers' vocabulary, grammar, semantics, pragmatics, and comprehension. For this study, we used the toddlers' vocabulary score, resulting from parents checking off which of 100 words is expressed by their toddlers. The vocabulary score was calculated by taking the average of the 100 words that the parents checked off.

Time 1 procedure

Toddlers participated in 3 sessions, spaced about a week apart and they received a book after each session for their participation. Before data collection for each session, the experimenter played with the child outside of the testing room for about 5min in order to build rapport and increase comfort with the laboratory environment. During Session 1, parents completed demographic and general vocabulary questionnaires and reported on the ignorance language question. During Session 2, we administered the eye-tracker version of the perceptual discrimination task, and during Session 3, we administered the touchscreen version of the task. In addition, participants completed other measures which fall beyond the scope of the current research.

Time 2 measures

Perceptual task

Participants completed a perceptual discrimination task (Figure 1b), which differed from that used at Time 1 in that images were degraded as in previous studies with this age group (Coughlin et al., 2015). Moreover, preschoolers were asked to provide confidence ratings (touchscreen version only; Figure 1c). Stimuli included 108 line drawings of objects and animals (Cycowicz et al., 1997), that are typically known to children as young as 3years of age based on previous studies (e.g., Lyons & Ghetti, 2011) and age-of-acquisition norms (Morrison et al., 1997). An automated program was used to degrade the line drawings by removing a specific number of pixels from each image. Twenty percent of pixels were removed and the percentage of pixels removed were the same for each drawing. Two sets of 40 drawings were created for counter balancing, 8 images were used for practice trials, and 20 images were used for confidence practice.

The touchscreen version began with training the preschoolers on choosing a target from the two degraded images. They completed four practice trials where they were asked to choose an image and were given feedback on the accuracy of their choice. When they were correct, the experimenter would give positive feedback, (e.g., “Great job! We chose this drawing because we saw that the gorilla was in there!”) and when they were incorrect, the experimenter would correct them (e.g., “Oops, the gorilla is actually in here. But that's okay because it's just practice.”). After these practice trials, the experimenter would train the preschoolers on using the confidence scale. The confidence scale was a 3-point pictorial scale with pictures of a child displaying facial and body expressions of low, moderate, or high confidence at each point of the scale (Figure 1c). The same child image was used for each level and could be interpreted as being female or male. The experimenter first verbally described each level of the uncertainty scale. Low-confidence was described as being “not so sure about your answer”, moderate confidence as “kind of sure about your answer”, and high confidence as “really sure about your answer”. After, the preschoolers completed 10 practice trials where they were asked to find an object or image from two degraded line drawings and then reported their confidence for that answer by selecting one of the confidence scale's images. The experimenter would provide feedback on their confidence scale selection; giving positive feedback when the selection seemed to match the outward expression of uncertainty (e.g., “Great job! You were really sure about that one, so you touched this face.”), and prompting them to touch another image when their selection did not match the outward expression of uncertainty (e.g., “Hmm, it seemed like you were not so sure about that one. Remember, if you're not so sure about your answer, this is the face you touch.”). After the 10 practice trials, the experimenter verified the preschoolers understanding of each point on the confidence scale and stressed the importance of using all three pictures during the rest of the task. Preschoolers then completed 20 test trials, where no feedback was given. They were asked to first identify a degraded image and immediately after making their choice, they reported their confidence in their choice.

The eye-tracking version required the preschoolers to complete the same 20 trials as the touchscreen version (i.e., same item pairs). However, no training on the task was done and they were not asked to report their confidence levels. In addition, to ensure less movement for accurate eye movement data, the preschoolers were given a soft-tipped “wand” in order to point to their choice and the experimenter would key in their response. In both tasks, for any trial where the preschooler refused to submit an answer, the experimenter would press a separate key and that trial was taken out of analysis. Unlike Time 1, the touchscreen version was always completed before the eye-tracker version. This was because the confidence scale was completed during the touchscreen version and we were concerned their confidence ratings would be impacted if the children made their confidence choices during the second round of seeing the images.

Similar to Time 1, not all participants at Time 2 contributed data in both perceptual task versions due to being uncooperative in one (9) or both versions (1). Therefore, 153 preschoolers contributed eye-tracking data and 154 preschoolers contributed touchscreen data.

Similar to Time 1, response latencies and accuracy were obtained from the touchscreen version and gaze transitions and accuracy were obtained from the eye-tracker version. In addition, average confidence and uncertainty monitoring variables were obtained from the touchscreen version. Similar to Time 1, there were no statistically significant differences between the accuracy scores in the touchscreen version compared to the eye-tracker version ($t(148)=0.69$, $p=.493$, $d=0.08$), and the two scores were significantly correlated ($r=.40$, $p<.001$), therefore, we again collapsed across the two scores. Average confidence was calculated by taking the average of all confidence ratings across all valid trials for each participant. The uncertainty monitoring variable was calculated by subtracting average confidence for inaccurate trials from average confidence for the accurate trials. Data processing and calculations were the same as Time 1 for all variables. For the touchscreen variables, no trials were eliminated due to preschoolers not providing an answer or for responses being below the 700 ms cutoff. Fifty trials (1.63%) across 50 preschoolers were eliminated due to our outlier cutoff of 3 SD above or below the individuals mean. For the eye-tracker variables, there were 36 trials (1.18%) across 17 preschoolers which were eliminated due to preschoolers not providing an answer. Additionally, there were 270 trials (8.82%) across 85 preschoolers eliminated from analysis due to no looking time measured towards the target or the distractor.

Mental state language

Similar to Time 1, parents were asked to report on their child's use of mental state language sentences using the adapted version of the Cognition portion of the Internal States Language Questionnaire (Bretherton & Beeghly, 1982). However, at Time 2, we utilized the entire subscale in our analysis and the mental state language score was the average across the 5 ratings.

Theory of mind

Preschoolers completed two versions of the explicit false belief task from Wellman and Liu (2004). This task was completed on the same touchscreen monitor as the perceptual task. Preschoolers were introduced to a character (version 1: Scott; version 2: Sally) looking for their lost item (version 1: mittens; version 2: sunglasses), which was hidden in one of two locations. The child was told that the character's item was really in one of the locations (version 1: in a backpack; version 2: in a purse) but the character thought the item was in the second location (version 1: in the closet; version 2: in the bedroom). After hearing the story, children were asked where the character would search for the item. Children would either point to their answer on the screen or answer verbally and the experimenter would key in their response. If the child answered correctly (chose the place where the character thought the object was) they were awarded a score of 1. If they answered incorrectly (chose the place the object actually was) they were awarded a score of 0. The scores from both versions of the task were averaged for a final ToM score which could take on values of 0, 0.5, or 1.

Time 2 procedure

Participation at Time 2 was completed across three sessions. During session 1, preschoolers completed both versions (touchscreen and eye-tracker) of the perceptual task and one version of the false belief task. During session 2, parents completed the mental state language questionnaire and during session 3, preschoolers completed the second version of the false belief task. Similar to Time 1, participants completed other measures that were beyond the scope of the current study.

Analytical approach

The analyses reported in the manuscript were preregistered (Open Science Framework; https://osf.io/t642d) and are therefore confirmatory analyses. Additional analyses reported in the Supporting Information were also pre-registered (Supporting Information Results 1 and 2; https://osf.io/ahde8/) and were also confirmatory analyses.

Path models

To test the main hypotheses, we utilized path models. Path models rely on multiple regression methods to examine how multiple independent variables may predict a dependent variable (Lleras, 2005). Unlike a simple multiple regression model, path models allow researchers to examine both direct and indirect effects on a dependent variable, by requiring specification of relationships among all variables included in the model. In our path model diagrams, variables in boxes represent observed variables. Single-direction arrows depict the direction of the proposed effect, with the tip of the arrow pointing towards the variable that is being influenced by the other variable. Correlations between variables are depicted by double-headed arrows.

With our path models, we examined both concurrent and longitudinal relations between Time 1 and Time 2 behaviors and Time 2 uncertainty monitoring, while accounting for age, accuracy across tasks, and Time 2 levels of average confidence, gaze transitions, response latencies, and mental state language. If early behavioral responses are developmental antecedents of later uncertainty monitoring, we should expect a significant direct path from Time 1 variables of gaze transitions and response latencies to Time 2 uncertainty monitoring, with more gaze transitions and longer response latencies predicting better uncertainty monitoring. However, if children learn to use behavioral responses as cues to uncertainty monitoring only when they are already able to provide subjective assessments of uncertainty, then no direct longitudinal path between Time 1 behaviors and Time 2 uncertainty monitoring is expected and only Time 2 behavioral responses would be significant predictors.

We tested our models using the RStudio package lavaan, which uses full information maximum likelihood to utilize all available data while accounting for patterns of missing data (Rosseel, 2012). We determined fit of our models by a non-significant chi-square difference test, a comparative fit index (CFI) of .9 or greater, and a root mean square error of approximation (RMSEA) of .06 or less. Finally, response latencies were transformed from milliseconds to seconds and age variables were transformed from months to years in order to reduce the order of magnitude of standard errors to be similar to the other predictors in our models.

RESULTS

Descriptive statistics for all variables can be found in Table 1 and correlations between all variables in the path models can be found in Table S1.

Preliminary analyses

Time 1 results were analyzed elsewhere (Leckey et al., 2020). As for Time 2 results, we examined the preschoolers' accuracy scores to confirm that the task was appropriate for this age group. Overall, accuracy scores at Time 2 were lower than accuracy scores at Time 1, $t(143) = 15.15$, $p < .001$, $d = 1.62$, suggesting that the task was harder for the participants at Time 2. However, accuracy scores were similar for the eye-tracker version ($M = 0.58$, SD = 0.14) and the touchscreen version ($M = 0.59$, SD = 0.11) and they were above chance in both the eye-tracker version, $t(152) = 7.44$, $p < .001$, $d = 0.60$, and the touchscreen version, $t(153) = 10.16$, $p < .001$, $d = 0.82$ (i.e., accuracy higher than 0.5). These results confirm that participants were able to complete the perceptual task successfully overall at Time 2.

TABLE 1 Means and standard deviations of study variables.

Variable	*N*	*M*	SD
Time 1 variables			
Age	183	28.99	1.70
Gaze transitions	149	1.38	0.66
Response latencies	157	4169.68	1973.68
Accuracy	169	0.77	0.12
"I don't know" ratings	177	3.60	1.49
Time 2 variables			
Age	159	41.64	3.88
Gaze transitions	153	1.01	0.48
Response latencies	154	4108.62	1808.37
Accuracy	158	0.59	0.10
Mental state language	145	3.69	0.76
Theory of mind	160	0.59	0.37
Average confidence	154	1.44	0.46
Uncertainty monitoring	154	0.01	0.32

Before conducting our path models, we also examined whether there were differences in Time 1 variables between the children who completed both time points compared to the children who only completed Time 1. We found that there were no significant differences in Time 1 age, general vocabulary, response latencies, gaze transitions, accuracy, or "I don't know" ratings ($ps \geq .249$).

Finally, we additionally examined the zero-order correlations between predictors in our path models (Table S1). Correlations were not consistent across all measures at both Time points. Gaze transitions and response latencies at Time 1 were significantly correlated ($r = .20$, $p = .022$), whereas, the correlation between gaze transitions and response latencies at Time 2 failed to achieve conventional levels of statistical significance ($r = .14$, $p = .093$), although correlations at both time points did not significantly differ in magnitude according to the Steiger z test ($z = .51$, $p = .61$). In addition, gaze transitions and accuracy were not correlated at Time 1 ($r = .01$, $p = .917$), but were significantly correlated at Time 2 ($r = .22$, $p = .007$).

Gaze transitions, response latencies, and uncertainty monitoring

To investigate how Time 1 average gaze transitions and response latencies predicted uncertainty monitoring at Time 2, we entered these Time 1 predictors in a path model along with the same predictors assessed at Time 2 (Figure 2). Mean accuracy, mean confidence, and age were also included as covariates to ensure that longitudinal relations did not depend on overall differences in the accuracy on the task, tendency to report high confidence, or variability in age. This model allowed us to examine the paths from Time 1 variables accounting for Time 2 variables and relevant covariates. Our model was not saturated and had acceptable levels of absolute and relative model fit as indicated by a non-significant chi-square test (robust $\chi^2(13) = 18.99$, $p = .123$), the RMSEA = .05, and CFI = .87.

Before examining longitudinal predictions, we note that average gaze transitions and average response latencies were significantly associated in the Time 1 task, $\beta = .29$, SE = .19, $p = .044$ and trending the same

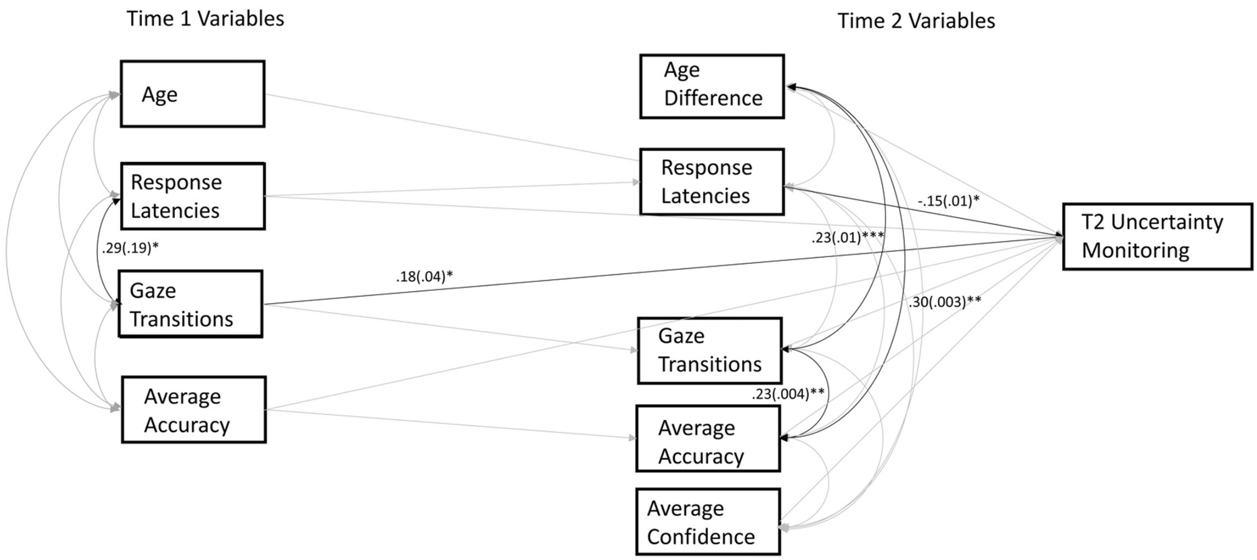

FIGURE 2 Path model examining the relations among uncertainty behaviors at Time 1 and Time 2 and their ability to predict uncertainty monitoring at Time 2. Non-significant paths are gray. Not pictured: Significant covariances between Time 1 average accuracy and age difference, $\beta = -.19$, SE = .003, $p = .039$ and between Time 1 response latencies and age difference, $\beta = .21$, SE = .06, $p = .028$. *$p < .05$, **$p < .01$, ***$p \leq .001$.

direction in the Time 2 task, $\beta = .13$, SE = .06, $p = .057$, suggesting that these variables functioned similarly at each time point as far as capturing the time taken by visual exploration prior to committing to a response. As for our longitudinal tests, results revealed a significant path between Time 1 gaze transitions and Time 2 uncertainty monitoring, $\beta = .18$, SE = .04, $p = .035$, indicating that toddlers who switched gaze more frequently during the task at Time 1 showed better uncertainty monitoring (i.e., higher confidence ratings reported for accurate trials compared to inaccurate trials) at Time 2. There was also a significant path between Time 2 response latencies and Time 2 uncertainty monitoring, $\beta = -.15$, SE = .01, $p = .028$, indicating, contrary to our prediction, that children with faster response latencies during the task at Time 2 showed better uncertainty monitoring (see Supporting Information Results 1 for a pre-registered complementary analysis using multilevel models suggesting this result is driven by accurate trials and Supporting Information Results 2 for a pre-registered path model utilizing drift diffusion parameters instead of average response latencies).

The path between Time 1 response latencies and Time 2 uncertainty monitoring and the path between Time 2 gaze transitions and Time 2 uncertainty monitoring were not significant ($\beta = -.08$, SE = .01, $p = .259$; $\beta = .03$, SE = .05, $p = .712$). Thus, our results were consistent with our main hypothesis that Time 1 gaze transitions would predict Time 2 uncertainty monitoring. However, this was not the case with regard to response latencies and our alternative hypothesis that these behaviors would only be associated concurrently with uncertainty monitoring.

Evaluating the role of mental state language and ToM

To explore the role of mental state language and ToM, we examined Time 1 parental report of the frequency of use of the "I don't know" phase, and Time 2 parental report of mental state language use and the false belief task as predictors of Time 2 uncertainty monitoring (Figure S1). The same variables and covariates as our main model were also included. Additionally, we added paths from Time 1 response latencies and gaze transitions to Time 2 mental state language and ToM. Similar to our main model, this model had good fit, with a non-significant chi-square test (robust $\chi^2(22) = 30.44$, $p = .108$), RMSEA = .05, and CFI = .87. In addition to the significant paths seen in the main model, there was also a significant path from Time 1 "I don't know" ratings to Time 2 mental state language, $\beta = .30$, SE = .04, $p < .001$, such that more frequent use of "I don't know" at Time 1 predicted higher use of mental state language at Time 2. However, the paths from Time 1 "I don't know" ratings, Time 2 mental state language, and Time 2 ToM to Time 2 uncertainty monitoring were not significant ($\beta = -.07$, SE = .02, $p = .392$; $\beta = .06$, SE = .04, $p = .498$; $\beta = -.05$, SE = .07, $p = .568$). Finally, the paths from Time 1 response latencies and gaze transitions to Time 2 mental state language and ToM were also not significant ($\beta = -.04$, SE = .03, $p = .643$; $\beta = .08$, SE = .09, $p = .311$; $\beta = -.02$, SE = .02, $p = .883$; $\beta = -.002$, SE = .05, $p = .982$). Thus, neither mental state language nor ToM abilities seemed to be related to the emergence of uncertainty monitoring. Similar results were found when a measure of overall vocabulary at Time 1 was accounted for in the model (Supporting Information Results 3), although model fit overall was poor which motivated our retention of a simpler model without general vocabulary.

DISCUSSION

Young children amass considerable knowledge in a short period of time, availing themselves with relentless questioning (Chouinard et al., 2007) and remarkable ability to identify reliable sources of information (Pasquini et al., 2007) and to respond to feedback (Leonard et al., 2020). Against this backdrop, the ability to subjectively evaluate one's level of uncertainty (Ghetti et al., 2013; Koriat & Goldsmith, 1996) is critical to direct learning efforts and monitor current knowledge states. Recent research has suggested that infants and toddlers exhibit uncertainty behaviors such as turning to adults for help, or transitioning their gaze between response options during difficult decisions (Goupil et al., 2016; Goupil & Kouider, 2016; Leckey et al., 2020), which may provide the foundation for later emerging uncertainty monitoring.

Using a longitudinal design, we found that gaze transitions between response options at 2 years of age positively predicted uncertainty monitoring 1 year later. Previous research has proposed that gaze transitions allow one to gather more information about response choices before making a decision when uncertain (Folke et al., 2016; Leckey et al., 2020). Our longitudinal relation suggests that toddlers who were more likely to visually explore and compare response options prior to committing a response may have more opportunities over time to recognize differences in states of knowledge leading to gaining better ability to discriminate these states with subjective experiences of uncertainty. Future studies should include experimental manipulations encouraging repeated visual exploration of response options to pinpoint the exact mechanism through which early opportunities for gathering and weighing visual evidence may promote uncertainty monitoring over time. This longitudinal relation persisted after controlling for concurrent gaze transitions at age 3, which did not predict concurrent uncertainty monitoring, but were related to accuracy. Therefore, although gaze transitions may be implicit indictors of how well children gather accurate evidence before making decisions even at age 3, children at this age may not have direct access yet, as adults do (Folke et al., 2016), to the correspondence between trial-level gaze transitions and uncertainty and thus to the use of this behavior to assess subjective uncertainty.

Unlike gaze transitions, response latencies at age 2 did not predict later uncertainty monitoring. Given that Leckey et al. (2020) found that response latencies in 2-year-olds distinguished between correct and incorrect responses, failure to detect a longitudinal relation suggests that hesitation alone may not promote the capacity to introspect on uncertainty, and implies, instead, that how toddlers garner more information through visual inspection while computing the decision, may be more critical early in development.

In contrast to our predictions, shorter response latencies at age 3 concurrently predicted better uncertainty monitoring. We had predicted that longer response latencies would predict better uncertainty monitoring, which is consistent with the idea that hesitation may be a behavioral indicator of uncertainty monitoring as found in preschoolers (Lyons & Ghetti, 2011), older children (Koriat & Ackerman, 2010), and adults (Robinson et al., 1997; Zakay & Tuvia, 1998). The opposite direction of this relation suggests, instead, that 3-year-olds who respond more efficiently in the task may have better uncertainty monitoring. The supplemental analysis utilizing a multilevel model to examine children's responses at the trial level (Supporting Information Results 1) provided results consistent with this possibility (i.e., 3-year-olds who responded more quickly on accurate decisions reported higher confidence on those decisions compared to 3-year-olds who responded more slowly). In other words, the relation between response latency and confidence was not about slower responses providing a cue to uncertainty and helping to discriminate between accurate and inaccurate decisions, but rather being more confident when effective and efficient decisions were made.

The lack of longitudinal relation for gaze transitions and response latencies suggests little continuity of these behaviors during this developmental period, but could potentially be due to task differences. The fact that gaze transitions and responses latencies are related to each other at each time point is consistent with this possibility. We did our best to design the tasks to be comparable between time points; however, we elected to use a simpler task involving occluded images at Time 1 because of a concern that degraded images would be too difficult for toddlers. It is possible that if toddlers and preschoolers completed the exact same task at both time points that we would find a relation between gaze transitions and response latencies at both time points. Future research should examine this possibility.

We tested an additional exploratory model based on extant proposals that early use of mental state language and ToM may contribute to the emergence of uncertainty monitoring (Sodian et al., 2012). It has been suggested that toddlers use of verbal expressions of ignorance, such as "I don't know," may increase their understanding of confidence states. Similarly, in order to achieve the ability to assess when one's cognitive operation yielded a correct versus incorrect decision, it has been theorized that one must understand that mental states can explain decisions (ToM; Flavell, 2000; Goldman, 2006). Early use of expressions of ignorance predicted later use of mental state language consistent with the idea that there is a continuity in the extent to which children access and communicate their states of knowledge. Therefore, we examined whether toddlers use of “I don't know” and preschoolers' ToM ability predicted preschoolers' uncertainty monitoring. However, neither toddlers' parental reported use of “I don't know” nor preschoolers' ToM ability were significant predictors of uncertainty monitoring at age 3. This result suggests that overall use of mental state language and ToM may not be critical to support the emergence of capacity to experience and report uncertainty on discrete perceptual decisions.

Although our failure to detect an association between explicit false belief and uncertainty monitoring is consistent with other studies (Baer et al., 2021; Coughlin et al., 2015), several proposals exist surrounding the connections between the abilities to understand mental states in oneself and others, which would predict associations between these two abilities (Carruthers, 2009). This position is not ubiquitous and the possibility that ToM and uncertainty monitoring are separate systems resulting in no correlations among them has been proposed (Carruthers, 2009). Thus, it is possible that an explicit understanding of the mental states of others is not necessary to compute confidence ratings that are calibrated to the accuracy of one's, immediately preceding, cognitive decision (Nichols & Stich, 2003). However, other studies have found predicted associations between uncertainty monitoring and ToM, such as a recent study which revealed impaired uncertainty monitoring and ToM abilities in children with autism spectrum disorder compared to typically developing children and, across both groups, uncertainty monitoring was associated with ToM (Nicholson et al., 2021). It is possible that testing context in our current study may not have been optimal to detect an association because children were not asked to act on the basis of their uncertainty (e.g., seeking help from an adult; Coughlin et al., 2015). Therefore, future research may gain from examining associations between other behaviors stemming from uncertainty assessments in order to obtain a more comprehensive picture of the relation between uncertainty monitoring and ToM.

The lack of a relation between mental state language and uncertainty monitoring may also suggest that there is a difference between a parent's perception of their child's use of mental states and the child's experience of uncertainty in a laboratory task. Differences between laboratory tasks and survey measures have been shown in past research in other fields, including self-regulation (Eisenberg et al., 2019). For example, Eisenberg et al. (2019) showed that although there are a variety of laboratory tasks and surveys used to measure self-regulation and they are thought to measure the same construct, these laboratory tasks and survey measures actually do not correlate with one another. Thus, it is possible that the use of survey measures for mental state language can explain its failure to predict a laboratory measure. We also recognize that our measures of mental state language in toddlers and in preschoolers may not have been fully comprehensive, given that our toddler measure was a single item. We chose a single-item assessment because “I don't know” has been shown to be one of the first mental state language phrases spoken by 2-year-olds and we wished to use a similar assessment as prior literature (Leckey et al., 2020). Although some researchers argue that single-item scales can be just as reliable as multi-item scales (Bergkvist, 2015; Loo, 2002), future studies may examine the relation between uncertainty monitoring and additional uncertainty terms used in everyday language (Meder et al., 2022), for both the toddler and preschooler age groups.

We note that even though past research has shown that 3-year-old preschoolers are able to monitor their uncertainty (Coughlin et al., 2015: Lyons & Ghetti, 2013), as a group, our participants did not show reliably higher confidence for accurate compared to inaccurate answers (Table 1). One reason for this discrepancy between past research and ours may be that our sample was younger and more diverse than the samples included in previous research. The ability to monitor uncertainty is still emerging at 3 years of age, with likely large individual variation, which we observed. We were reassured by several findings that suggest that these participants were approaching our tasks as expected, including that these same children exhibited reliable responses to decision difficulty at age 2 (Leckey et al., 2020) and age 3 in their eye movements and response latencies (Supporting Information Results 4). Nevertheless, future research should replicate our findings with a slightly older sample to ensure a sample with better overall uncertainty monitoring scores.

Finally, future research should examine potential motivations behind perceptual tasks, like the one used in the current study, and their links to uncertainty monitoring. For example, both of the tasks utilized in this study were introduced as a “finding game”, and the blocked or degraded nature of the task was created to be engaging and interesting for the children in order to encourage them to try their best while also eliciting some uncertainty about the answer. Future research could examine whether this engagement is driven by the children visually inspecting the stimuli for the sake of resolving this uncertainty versus the children trying to fulfill a desire to know what the answer is.

In summary, we showed that gaze transitions between response options at age 2 predict preschoolers' uncertainty monitoring abilities, whereas response latencies and mental state language did not. In addition, we found that concurrent behaviors of gaze transitions, mental state language, and ToM did not predict preschoolers' uncertainty monitoring abilities, however, response latencies did. These results underscore the importance of early visual exploration during decision making as a marker of evidence seeking and future capacity to introspect on uncertainty.

FUNDING INFORMATION

This work was supported by a grant from the National Science Foundation (NSF; BCS1424058) to SG. Any opinions, findings, and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the NSF. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. SL was supported by National Institute of Child Health and Human Development Grant F31HD102153.

DATA AVAILABILITY STATEMENT

The data and code necessary to reproduce the analyses presented here are publicly accessible, as are the materials necessary to attempt to replicate the findings. Analyses were also pre-registered. Data, code, materials, and the pre-registration for this research are available at the following URL https://osf.io/td2be/.

ORCID

Sarah Leckey https://orcid.org/0000-0002-9008-7749
Christopher Gonzales https://orcid.org/0000-0003-4312-5914
Diana Selmeczy https://orcid.org/0000-0001-7963-9170

REFERENCES

Baer, C., Malik, P., & Odic, D. (2021). Are children's judgments of another's accuracy linked to their metacognitive confidence judgments? *Metacognition and Learning*, *16*, 485–516. https://doi.org/10.1007/s11409-021-09263-x

Bazhydai, M., Westermann, G., & Parise, E. (2020). "I don't know but I know who to ask": 12-month-olds actively seek information from knowledgeable adults. *Developmental Science*, *23*, e12938. https://doi.org/10.1111/desc.12938

Beran, M. J., Brandl, J. L., Perner, J., & Proust, J. (2012). On the nature, evolution, development, and epistemology of metacognition: Introductory thoughts. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 1–18). Oxford University Press.

Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. *Marketing Letters*, *26*, 245–255. https://doi.org/10.1007/s11002-014-9325-y

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, *18*, 906–921. https://doi.org/10.1037/0012-1649.18.6.906

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*, 121–182. https://doi.org/10.1017/S0140525X09000545

Carruthers, P., & Williams, D. M. (2019). Comparative metacognition. *Animal Behavior and Cognition*, *6*, 278–288. https://doi.org/10.26451/abc.06.04.08.2019

Chouinard, M. M., Harris, P. L., & Maratsos, M. P. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*, i–129.

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

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237. https://doi.org/10.1006/jecp.1996.2356

Dautriche, I., Goupil, L., Smith, K., & Rabagliati, H. (2022). Two-year-olds' eye movements reflect confidence in their understanding of words. *Psychological Science*, *33*, 1842–1856. https://doi.org/10.1177/09567976221105208

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

Droit-Volet, S. (2010). Stop using time reproduction tasks in a comparative perspective without further analyses of the role of the motor response: The example of children. *European Journal of Cognitive Psychology*, *22*, 130–148. https://doi.org/10.1080/09541440902738900

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*, 2319–2332. https://doi.org/10.1038/s41467-019-10301-1

Fandakova, Y., Selmeczy, D., Leckey, S., Grimm, K. J., Wendelken, C., Bunge, S. A., & Ghetti, S. (2017). Changes in ventromedial prefrontal and insular cortex support the development of metamemory from childhood into adolescence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7582–7587. https://doi.org/10.1073/pnas.1703079114

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Reply: Measuring variability in early child language: Don't shoot the messenger. *Child Development*, *71*, 323–328. https://doi.org/10.1111/1467-8624.00147

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, *24*, 15–23. https://doi.org/10.1080/016502500383421

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

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*, 1–8. https://doi.org/10.1038/s41562-016-0002

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

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press on Demand.

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

Goupil, L., & Kouider, S. (2019). Developing a reflective mind: From core metacognition to explicit self-reflection. *Current Directions in Psychological Science*, *28*, 403–408. https://doi.org/10.1177/0963721419848672

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3492–3496. https://doi.org/10.1073/pnas.1515129113

Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017). Young children communicate their ignorance and ask questions. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7884–7891. https://doi.org/10.1073/pnas.1620745114

Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, *25*, 1768–1776. https://doi.org/10.1177/0956797614542273

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, *13*, 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*, 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. https://doi.org/10.3758/s13428-012-0210-4

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*, 928–936. https://doi.org/10.1038/s41562-020-0913-y

Leonard, J. A., Garcia, A., & Schulz, L. E. (2020). How adults' actions, outcomes, and testimony affect preschoolers' persistence. *Child Development*, *91*, 1254–1271. https://doi.org/10.1111/cdev.13305

Lleras, C. (2005). Path analysis. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (Vol. *3*, pp. 25–30). Elsevier.

Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of Managerial Psychology*, *17*, 68–75. https://doi.org/10.1108/02683940210415933

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*, 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*, 726–736. https://doi.org/10.1111/cdev.12004

Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language*, *51*, 458–472. https://doi.org/10.1016/j.jml.2004.06.008

Meder, B., Mayrhofer, R., & Ruggeri, A. (2022). Developmental trajectories in the understanding of everyday uncertainty terms. *Topics in Cognitive Science*, *14*, 258–281. https://doi.org/10.1111/tops.12564

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*, 528–559. https://doi.org/10.1080/027249897392017

Nichols, S., & Stich, S. (2003). How to read your own mind: A cognitive theory of self-consciousness. In Q. Smith, & A. Jokic (Eds.), *Consciousness: New philosophical essays* (pp. 157–200). Oxford: Clarendon Press.

Nicholson, T., Williams, D. M., Lind, S. E., Grainger, C., & Carruthers, P. (2021). Linking metacognition and mindreading: Evidence from autism and dual-task investigations. *Journal of Experimental Psychology: General*, *150*, 206–220. https://doi.org/10.1037/xge0000878

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Paulus, M., Proust, J., & Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: An eye-tracking and pupillometric study. *Frontiers in Psychology*, *4*, 145. https://doi.org/10.3389/fpsyg.2013.00145

Perner, J. (2000). Memory and theory of mind. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 297–312). Oxford University Press.

Proust, J. (2019). From comparative studies to interdisciplinary research on metacognition. *Animal Behavior and Cognition*, *6*, 309–328. https://doi.org/10.26451/abc.06.04.10.2019

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, *82*, 416–425.

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

Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). "It's magic!" the effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology*, *87*, 320–335. https://doi.org/10.1016/j.jecp.2004.01.004

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*, 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Sodian, B., Thoermer, C., Kristen, S., & Perst, H. (2012). A review on the development of metacognition in young children. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 119–133). Oxford University Press.

Vaish, A., Demir, Ö. E., & Baldwin, D. (2011). Thirteen-and 18-month-old infants recognize when they need referential information. *Social Development*, *20*, 431–449. https://doi.org/10.1111/j.1467-9507.2010.00601.x

Walden, T., Kim, G., McCoy, C., & Karrass, J. (2007). Do you believe in magic? Infants' social looking during violations of expectations. *Developmental Science*, *10*, 654–663. https://doi.org/10.1111/j.1467-7687.2007.00607.x

Wellman, H. M. (2011). Developing a theory of mind. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 258–284). Wiley-Blackwell.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*, 968–979. https://doi.org/10.1002/hipo.20808

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica*, *98*, 103–115. https://doi.org/10.1016/S0001-6918(97)00037-1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Leckey, S., Gonzales, C., Selmeczy, D., & Ghetti, S. (2024). Toddlers' visual exploration during decisions predicts uncertainty monitoring 1 year later. *Child Development*, *00*, 1–13. https://doi.org/10.1111/cdev.14174