Developmental Psychology

Children Flexibly Compare Their Confidence Within and Across Perceptual Domains

Carolyn Baer and Darko Odic

Online First Publication, September 10, 2020. http://dx.doi.org/10.1037/dev0001100

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

http://dx.doi.org/10.1037/dev0001100

BRIEF REPORT

Children Flexibly Compare Their Confidence Within and Across Perceptual Domains

Carolyn Baer and Darko Odic
University of British Columbia

How does a person make decisions across perceptual boundaries? Here, we test the account that confidence constitutes a common currency for perceptual decisions even in childhood by examining whether confidence can be compared across distinct perceptual dimensions. We conducted a strict test of domain-generality in confidence reasoning by asking 6- to 7-year-olds to compare their confidence in 2 decisions, either from the same perceptual dimension (e.g., number vs. number) or from two different perceptual dimensions (e.g., area vs. emotion). Not only could children compare their confidence across and within domains but there were no differences in their abilities to make within- and across-domain comparisons. Our findings support the idea that confidence is represented in a common format even in childhood, which could provide an account for perceptual integration in childhood that doesn't necessitate the use of language.

Keywords: certainty, confidence, perceptual decision-making, representations, domain-generality

Supplemental materials: http://dx.doi.org/10.1037/dev0001100.supp

How does a person make decisions across perceptual boundaries? Perceptual systems interpret the world using distinct units: Representing the number of objects in a set is distinct from representing the emotion expressed on a face. These perceptual units are not interchangeable—one does not typically think that one number is "angrier" than another—and yet, we make decisions that compare and integrate these percepts. For instance, a person's behavior may greatly differ when approached by a large, angry group of people compared with a small, happy group. This leaves the mind with the challenge of comparing apples to oranges: How can one make sensible decisions across distinct perceptual sources?

Some theories advocate for a single common currency that facilitates these decisions: *subjective confidence* (see de Gardelle & Mamassian, 2014; Shea & Frith, 2019). In recent behavioral work, adult observers flexibly decided whether they were more sure of a visual decision or an independent auditory decision (de Gardelle, Le Corre, & Mamassian, 2016). In a similar vein, neurophysiological work shows common neural circuits for confidence decisions across perceptual boundaries, mostly in the prefrontal cortex (PFC; for a review see Rouault, McWilliams, Allen, & Fleming, 2018). Whereas this work remains inconclusive about confidence as a common currency for decisions outside perception (e.g., memory, executive functioning; see Mazancieux, Fleming, Souchay, & Moulin, 2020; Rouault et al., 2018), subjective confidence could facilitate decision-making across perceptual boundaries.

However, there is conflicting developmental evidence for this account. In one study, 5- to 8-year-olds' betting behavior (a nonverbal confidence judgment) was uncorrelated in number and emotional expression decisions (Vo, Li, Kornell, Pouget, & Cantlon, 2014), as were explicit confidence judgments in math and memory decisions until age 11 in another (Geurten, Meulemans, & Lemaire, 2018). These two studies suggest that children may not represent confidence as a common currency across distinct perceptual and cognitive domains. In contrast, one study with 6- to 9-year-olds found correlations between number, surface area, and emotion perception certainty judgments (Baer, Gill, & Odic, 2018). There, children were shown two possible trials—one associated with higher certainty, one with lower—and prospectively decided which would maximize their accuracy. Despite no correlation in the three perceptual accuracies, children's ability to select the higher certainty trial correlated across domains.

Carolyn Baer and Darko Odic, Department of Psychology, University of British Columbia.

This work was funded by the Social Sciences and Humanities Research Council of Canada through an Insight Development Grant to Darko Odic and a Canada Graduate Scholarship to Carolyn Baer. Thanks to Katrina Zelko, Deveena Basi, and Emilie Kniefel for their help with data collection and to the schools and families for their support. Our study design, hypotheses, and analysis plan were preregistered and are available at http://aspredicted.org/ei5ux.pdf and stimuli and data available at https://osf.io/74dcv/.

Correspondence concerning this article should be addressed to Carolyn Baer, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada. E-mail: cebaer@psych.ubc.ca

There are, naturally, key differences between these studies that could make them difficult to compare at face value. For one, Geurten et al. (2018) examined math and memory decisions, which may not share confidence as a common currency compared to perceptual decisions (Rouault et al., 2018). Second, Baer et al. (2018) asked children to select one of two questions to solve, which, although capitalizing on children's desire to succeed, potentially also tapped into children's desire to challenge themselves. Baer et al. also used prospective judgments, whereas Vo et al. (2014) used retrospective judgments, which may rely on different metacognitive processing (see Pouget, Drugowitsch, & Kepecs, 2016, but note that Geurten et al., 2018, found similar performance between these judgment types).

Even more important, Geurten et al. (2018) and Vo et al. (2014) used measures of confidence *calibration* to measure individual differences. That is, children indicated high or low confidence for every trial individually, and responses were matched with their accuracy on those trials through φ and γ correlations and by plotting receiver operating characteristic curves to statistically isolate sensitivity (see Galvin, Podd, Drga, & Whitmore, 2003; Geurten et al., 2018; Nelson, 1984; Vo et al., 2014). However, these statistical measures to varying degrees conflate *bias* (the tendency to misuse the confidence scale by selecting one option more frequently than is warranted) and *sensitivity* (the underlying ability to differentiate levels of confidence; see Fleming & Lau, 2014) and become less reliable as overconfidence grows (Maniscalco & Lau, 2014; Nelson, 1984). In contrast, Baer et al. (2018) experimentally isolated confidence sensitivity by asking children to select the *most* certain of two answers (Barthelmé & Mamassian, 2009). One possibility, then, is that domain-general similarities in sensitivity were masked by domain-specific differences in bias (notably, Vo et al., 2014, reported that children were overconfident in numerical judgments but underconfident in emotion judgments).

These discrepancies raise two possible explanations for confidence representation in childhood. First, there may be a true developmental difference, whereby confidence is represented domain-specifically early in development and over time becomes a common currency (potentially as certain regions of the PFC develop or with changes in broader neural connectivity; see Fleming & Dolan, 2012). Under this view, the reported correlations between perceptual domains in Baer et al. (2018) could be induced by third variables like intelligence or motivation. Alternatively, confidence could be domain-general throughout development, but the tasks used to measure children's confidence in previous work may be strongly affected by children's confidence biases or lack sufficient power to detect shared individual differences.

Here, we provide a stricter test between these accounts. To alleviate third-variable explanations and low power, both accentuated in the correlational analyses of all three previous studies, we used a within-subject experimental design where each child serves as their own control. To alleviate the challenges of children's confidence biases present in reporting or betting measures, we tested whether children could directly compare confidence across perceptual domains, deciding whether they were more certain in one perceptual decision (e.g., number) or one from another domain (e.g., emotion; see de Gardelle & Mamassian, 2014). And, to equate across previous studies, we asked children to judge their confidence retrospectively, rather than prospectively. If confidence were represented domain-specifically in childhood, then children should be unable to compare confidence across distinct perceptual domains or would at least be substantially worse at across- than within-domain comparisons. However, if confidence uses a domain-general common currency even in childhood, then children should compare within and across perceptual domains with equivalent accuracy.

Method

Participants

Forty-eight children 6 and 7 years of age participated in the study (M = 6 years 11 months, range = 6 years 0 months to 7 years 11 months; 22 girls), an age group younger than the reported domain-specificity at age 8 by Vo et al. (2014) and Geurten et al. (2018) but where children are known to make relative confidence judgments (see Baer et al., 2018; Baer & Odic, 2019). Sample size (preregistered at https://aspredicted.org/ei5ux.pdf) was set to be similar to that in related studies (e.g., Baer et al., 2018; Vo et al., 2014) but anticipated to yield higher power because of the within-subject experimental design and the use of Bayesian statistical analyses.[1] Per our preregistration plan, one child was excluded from analyses for not completing the entire task. All participants were tested individually at local schools in Vancouver, British Columbia, where the most common ethnicities are White and East or Southeast Asian. All children spoke English, and most children came from middle-class families. Ethical approval for the study was granted by the University of British Columbia Behavioral Research Ethics Board under the project name *Kid Quantity and Language* (H14-01984).

Materials and Procedure

All stimuli are available online at https://osf.io/74dcv/ and were presented on a laptop. From these stimuli, we measured *perceptual accuracy* or *sensitivity* in number, area, and emotion decisions and *within- and across-domain confidence accuracy* or *sensitivity* over these same decisions. On each trial, children saw two perceptual discrimination decisions, one easy and one hard, presented in different spatial locations. After answering both questions, children provided a retrospective relative confidence judgment indicating which of the two answers they were more confident of (see Figure 1). To assess whether their confidence decisions were warranted, we examined whether the accuracy of *chosen* trials (i.e., those children were more confident of) was higher than *discarded* trials, as would be expected if their confidence judgments were tracking meaningful information about their chances of success (see Baer & Odic, 2019). On two thirds of the trials, the two perceptual decisions were drawn from different domains (e.g., Number and Emotion, across-domain condition), allowing us to test whether children could compare their confidence across per-

[1] We did not perform a power analysis a priori because there was no effect size estimate then of the accuracy difference from low to high confidence judgments in this type of task. One publication (Baer & Odic, 2019) has since reported an effect size of d = 0.74 in 3- to 7-year-old children on this task for number decisions, suggesting a sample size of 22 participants given α = .05 and power = .90.

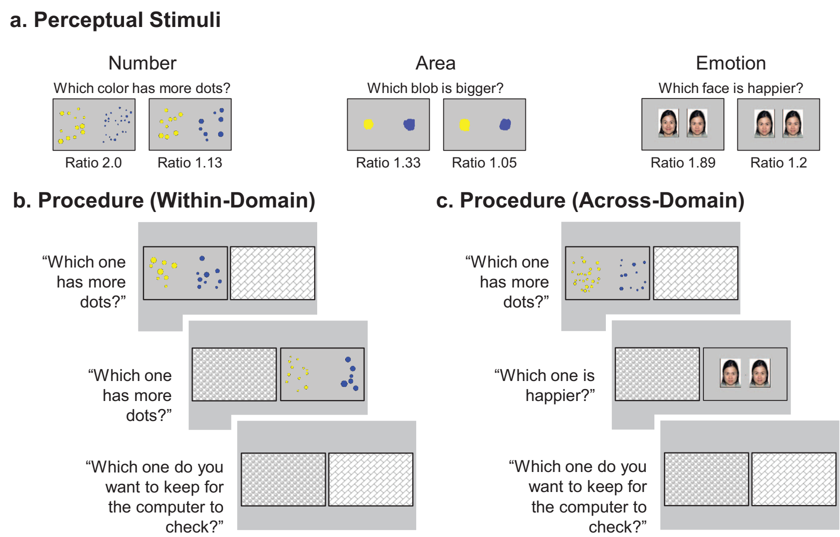

Figure 1. Example stimuli. Panel a: A high-confidence and low-confidence trial for the Number, Area, and Emotion tasks. Panel b: Procedure of the confidence task for a within-domain comparison. Panel c: Procedure of the confidence task for an across-domain comparison (full stimuli can be found at https://osf.io/74dcv/). Images of Model No. 18 are from "The NimStim Set of Facial Expressions: Judgments From Untrained Research Participants," by N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, & C. Nelson, 2009, *Psychiatry Research, 168*(3), p. 244. Copyright 2020 by Elsevier B.V. Reprinted with permission. See the online article for the color version of this figure.

ceptual boundaries. On the remaining trials, the two perceptual decisions were drawn from the same domain (e.g., Number and Number, within-domain condition), which served as a within-subject control.

Following Baer et al. (2018) and Vo et al. (2014), we used three dissociable perceptual domains: Number, Area, and Emotion (see Figure 1). On the Number discrimination trials, children saw groups of yellow and blue spatially separated dots and indicated "which side has more dots." On the Area discrimination trials, children saw one yellow and one blue amorphous blob and identified "which blob is bigger." On the Emotion discrimination trials, children saw two pictures of a female face displaying a mixture of happiness and anger and identified "which face is happier." Emotions were mixed by morphing images of happiness and anger in FantaMorph software (Version 4; Abrosoft, 2007) for each of four female models (two Asian and two White).

To assess confidence, we used a retrospective forced-choice confidence paradigm (e.g., Baer et al., 2018; Baer & Odic, 2019; de Gardelle et al., 2016; de Gardelle & Mamassian, 2014; see Figure 1), where children decided which of two trials they were more certain of answering correctly. To help children understand this task, the experimenter told children that they would keep only one answer for the computer to grade, and so they should keep the answer they were really sure was correct (see Baer & Odic, 2019; Hembacher & Ghetti, 2014, for similar instructions). Children did not receive feedback on the accuracy of their discrimination decisions or on their confidence choice but were given encouraging remarks periodically to keep them motivated (e.g., "You're going so fast!" "All right, let's do another one!"; see Baer & Odic, 2019).

Previous work using the forced-choice confidence measure has found that children's and adults' ability to compare relative confidence is best when there is a large difference in confidence between the two discrimination trials and that confidence choices closely track accuracy (Baer et al., 2018; Baer & Odic, 2019; de Gardelle & Mamassian, 2014). Therefore, each pair of discrimination trials had one trial designed to elicit high confidence (children should answer correctly 90% of the time, on average) and one trial designed to elicit low confidence (children should answer correctly 60% of the time, on average). Because children's perceptual acuity in number, area, and emotion is different (see Baer et al., 2018), we relied on work mapping the developmental trajectory of perceptual acuity for each dimension to determine the key ratio that would produce 90% versus 60% expected accuracy, thereby roughly equating the confidence strength in each domain. High-confidence Number trials were set at a ratio of 2.0 (e.g., 12 yellow dots vs. six blue dots), whereas low-confidence trials were set at a ratio of 1.13 (e.g., nine yellow vs. eight blue dots); the high-confidence Area ratio was 1.33 (e.g., a 1,330-px^2 blob vs. a 1,000-px^2 blob) and the low at 1.05 (e.g., a 1,050-px^2 blob vs. a 1,000-px^2 blob); the high-confidence Emotion ratio was 1.89 (e.g., a 93.33% happy face vs. a 50% happy face) and the low 1.20 (e.g., a 70% happy face vs. a 56.67% happy face).

To evaluate whether children could compare confidence across domains, we had children see two comparison types: within-domain and across-domain. In within-domain comparisons, children first saw two perceptual discrimination questions from the same domain (e.g., Number and Number), one at the high-confidence ratio and one at the low, whereas in across-domain comparisons, children saw perceptual discrimination questions from two different domains (e.g., Number and Area), one at the high-confidence ratio and one at the low.

The experiment began with 12 practice trials (four from each dimension), where children were told how to complete the discrimination of each dimension. Afterward, each trial began with two gray occluders on the screen. When the child was ready, the experimenter pushed a button to reveal the discrimination trial on the left side and asked the Number, Area, or Emotion discrimination question. Children had unlimited time to answer, though most children did so within a second of viewing the trial. The experimenter pushed a button corresponding to the child's decision and revealed the discrimination trial on the right side while the left trial disappeared, again asking the Number, Area, or Emotion discrimination question. After the child answered the second trial, both trials were hidden and the child was asked to indicate which of those two trials they would like to "keep." In total, we ran 18 within-domain confidence trials (six per dimension), counterbalancing the left and right positions of the easier discrimination trial, and 36 across-domain confidence trials (12 per pair of dimensions), making sure that each dimension had an equal number of high- and low-confidence trials. Trials appeared in a random order and were counterbalanced to ensure that each domain appeared equally on the left and right sides of the screen. Thus, the experimenter did not know at the onset of any trial whether the child was going to view an across- or within-domain trial.

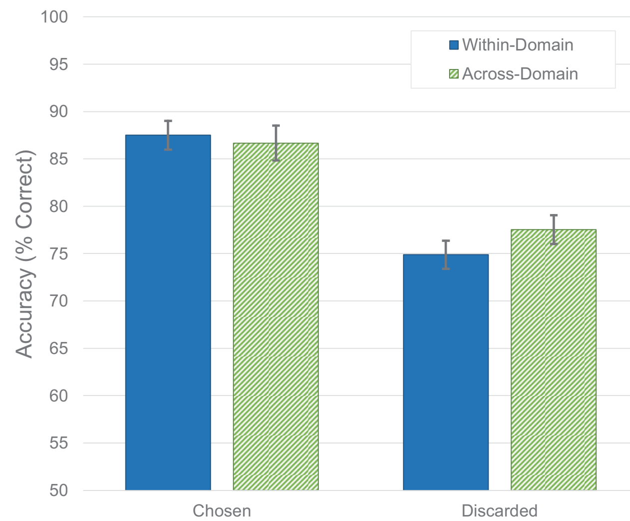

Figure 2. Children's accuracy (percentage of questions answered correctly) on perceptual decisions, grouped based on whether the child subsequently chose that question to keep (high confidence) or to discard (low confidence). Error bars represent 1 *SD*. See the online article for the color version of this figure.

Results

Our analysis plan was preregistered, and, unless otherwise noted, the results reported here follow that exact plan. Because we preregistered several secondary and exploratory analyses that are not central our hypothesis, we report the primary analyses here and the secondary in the online supplemental materials. No dependent variables interacted with age or gender (Fs < 2), so we collapsed across these variables in these analyses.[2] See https://osf.io/74dcv/ for data and annotated JASP (JASP Team, 2020) analyses.

Discrimination Decisions

Confirming that children understood the discrimination component of the three tasks, children were correct on 80% ($SD = 6$) of Area trials, above chance of 50%, $t(47) = 35.55$, $p < .001$, $d = 5.13$; 82% ($SD = 11$) of Emotion trials, $t(47) = 20.15$, $p < .001$, $d = 2.91$; and 84% ($SD = 7$) of Number trials, $t(47) = 32.60$, $p < .001$, $d = 4.71$. Reaction time (RT) data[3] revealed a slightly different pattern, with Area decisions reported fastest ($M = 2{,}780$ ms, $SD = 610$), followed by Number ($M = 2{,}930$ ms, $SD = 676$), and Emotion decisions slowest ($M = 3{,}094$ ms, $SD = 804$), suggesting that accuracy and RT were slightly dissociated in this sample.

Confidence Comparison

If children make relative confidence comparisons, their confidence choices should track their accuracy: *Chosen* trials should have higher accuracy than should *discarded* trials (see Baer & Odic, 2019; de Gardelle & Mamassian, 2014). Consistently, a 2 (confidence choice: chosen, discarded) × 2 (comparison type: within-domain, across-domain) repeated-measures analysis of variance (ANOVA) on discrimination accuracy found a significant main effect of confidence choice, $F(1, 47) = 30.22$, $p < .001$, $\eta_p^2 = .39$ (see Figure 2). Thus, consistent with the case in previous work, children's confidence choice reflected their accuracy: *Chosen* trials had higher accuracy than did *discarded* trials. Additionally, as shown in Figure 2, we found no main effect of comparison type, $F(1, 47) = 0.75$, $p = .392$, $\eta_p^2 = .02$, or Confidence Choice × Comparison Type interaction, $F(1, 47) = 1.69$, $p = .201$, $\eta_p^2 = .04$; the accuracy difference in confidence choice for within-domain trials ($M_{\text{Chosen}} = 88\%$, $SD = 10$, $M_{\text{Discarded}} = 75\%$, $SD = 13$) was not different from than for across-domain trials ($M_{\text{Chosen}} = 87\%$, $SD = 10$; $M_{\text{Discarded}} = 78\%$, $SD = 11$). In other words, children were equally sensitive in making their confidence choice across perceptual domains and within them.

Because the key finding here—that children's confidence is not affected by the manipulation of comparison type—is dependent on a null finding, we also preregistered a Bayesian repeated-measures ANOVA (using JASP with the default priors). As has been discussed at length elsewhere (Wagenmakers et al., 2018), a Bayes factor (BF) provides the relative weight of the evidence for the null versus the alternative hypothesis, therefore providing the graded strength or reliability for the null hypothesis. A BF_{10} of 1 indicates a lack of evidence for either hypothesis, whereas values that increase toward positive infinity indicate increasingly positive evidence for the alternative hypothesis and values that decrease

[2] Based on a reviewer's suggestion, we found a slight bias toward higher confidence in the right-hand answer (54%, $SD = 0.11$, always the second choice; see Sumner, DeAngelis, Hyatt, Goodman, & Kidd, 2019), $t(47) = 2.45$, $p = .018$, $d = 0.35$. It is important to note that this did not differ between the within- and across-domain conditions (53% within, $SD = 0.14$; 54% across, $SD = 0.11$), $t(47) = 0.72$, $p = .473$, $d = 0.10$, so we did not include this variable in further analyses.

[3] Because the children did not push the buttons themselves, the reaction time (RT) measures are slightly inflated from the additional time it took the experimenter to push the button or due to issues of interpretation (e.g., an ambiguous point). Any RTs more than 2.5 *SD*s from a child's own mean RT were excluded, as preregistered.

toward 0 indicate increasingly positive evidence for the null hypothesis, providing a method for testing whether a null effect is meaningful.

Conducting an identical 2 (confidence choice: chosen, discarded) × 2 (comparison type: within-domain, across-domain) repeated-measure Bayesian ANOVA over discrimination accuracy, we computed the model comparison for the inclusion of the two variables and their interaction. With 79% probability given the data, the best fitting model included only confidence choice ($BF_{10} > 10^9$), providing overwhelming evidence that children's accuracy differed for *chosen* versus *discard*ed trials without any influence of comparison type. In comparison, the model including the interaction effect of Comparison Type × Confidence Choice had 6% probability and was 13.68 times less likely than the model including only confidence ($BF_{10} = .073$), providing strong evidence for the lack of an interaction, as predicted by a common currency account of confidence. Additional replications of this analysis can be found in the online supplemental materials (e.g., for Number trials alone).[4]

Confidence Processing Time

A second hypothesis of a strong domain-general account is that not only should confidence comparisons be effective across domains but they shouldn't incur any processing time cost. That is, if children's confidence is truly represented in a domain-general format and all confidence judgments use these representations, then there is no processing cost to "translate" from a domain-specific format to a domain-general format for comparisons (see de Gardelle et al., 2016). We therefore also examined whether the response time for making the confidence decision was equal on the within- and across-domain trials.

A preregistered paired *t* test of RT found that confidence judgments for within-domain comparisons ($M = 2{,}175$ ms, $SD = 749$) were faster than were across-domain comparisons ($M = 2{,}328$ ms, $SD = 904$), $t(47) = 2.89$, $p = .006$, $d = 0.42$. This was also confirmed with moderate evidence in a preregistered Bayesian analysis ($BF_{10} = 6.09$). Our data, therefore, suggest that children incurred about a 150-ms cost for making confidence decisions across domains.

Although a difference in RTs might signify a processing cost of converting a domain-specific signal into a domain-general one, it could also reflect the cost of switching tasks (de Gardelle et al., 2016). If so, one might expect to see a similar difference in RT for the across-domain discrimination decisions as well. Specifically, the second discrimination decision made on the across-domain trials should be slower than the second decision in the within-domain trials. To examine this possibility, we conducted two nonregistered exploratory analyses, finding a significant difference in RT between the second discrimination decision in the across-domain trials ($M = 2{,}825$ ms, $SD = 696$) compared to the within-domain trials ($M = 2{,}657$ ms, $SD = 557$), paired $t(47) = 2.78$, $p = .008$, $d = 0.40$. In addition, the discrimination RT cost (i.e., the difference in RT between within-domain second trials and across-domain second trials) correlated with the confidence RT cost, $r(46) = .44$, $p = .002$, strongly suggesting that task-switching costs explain the differences in confidence RTs.

Discussion

Here, we provide evidence in children as young as 6 years that confidence in distinct perceptual decisions is represented using a domain-general currency. Using a forced-choice confidence paradigm, we found that 6- to 7-year-olds compare their confidence equally well both across perceptual boundaries and within those boundaries. This entirely within-subject experimental method bypassed challenges faced by correlational designs, including subjectivity to third-variable explanations, allowing us to directly test for domain-general reasoning in children. Our results thus argue against the account of developmental change from domain-specific to domain-general confidence sensitivity in children between 8 and 10 years of age. Instead, they show that confidence representations are domain-general from at least age 6, as predicted by an account that confidence is domain-general throughout development. This in turn is consistent with theoretical accounts that treat confidence as a probability judgment (e.g., Pouget et al., 2016) or the result of an error signal (e.g., Boldt & Yeung, 2015), which could be available to humans throughout the life span.

These findings open the possibility that confidence representations might be also domain-general in younger children. Preschool children, infants, and several nonhuman animals react strategically to uncertainty (Goupil, Romand-Monnier, & Kouider, 2016; Kepecs, Uchida, Zariwala, & Mainen, 2008; Lyons & Ghetti, 2013), just as children did here. It is important to note that these decisions cover many domains: memory decisions, perceptual access to a reward, and direct perceptual comparisons like those used here. At the same time, development in neural structures involved in metacognitive judgments (e.g., Filevich et al., 2020) could facilitate change in the structure of confidence representations before the age of 6 tested here. Given the diversity of decisions invoking the subjective experience of confidence in younger children, future research can test whether these decisions similarly use a common currency of confidence. This would provide even stronger evidence of the true nature of confidence representations and could test whether confidence reasoning is shared even among dramatically different systems like perception and memory (Rouault et al., 2018; Shea & Frith, 2019). We hope that this paradigm, when adapted for even younger children, can aid investigations of domain-generality in these populations.

Our findings also highlight an important distinction between sensitivity and bias in confidence reasoning (see Mazancieux et al., 2020; Winman, Juslin, Lindskog, Nilsson, & Kerimi, 2014). Whereas the current study and Baer et al. (2018) found evidence of domain-generality in a method that experimentally isolates sensi-

[4] A reviewer suggested that we conduct an exploratory analysis of whether children chose one domain more than the others as an additional test of whether children were making meaningful comparisons across domains, given that children's accuracy was better on Number than on Area or Emotion domain trials. Looking at the data from only the across-domain trials, children selected more Number trials (13.33 of 36 across-domain trials, $SD = 2.98$) relative to Emotion (10.83, $SD = 4.37$) and Area (11.83, $SD = 2.98$) trials, $F(1.53, 71.84) = 4.04$, $p = .032$, $\eta_p^2 = .08$ (Greenhouse-Geisser-corrected). Exploratory post hoc comparisons revealed a difference between Number and Area, $t(45) = 2.54$, $p = .043$, $d = 0.37$, and Number and Emotion, $t(45) = 2.50$, $p = .048$, $d = 0.36$ (Bonferroni-corrected). Therefore, children's confidence choices tracked the differences in accuracy between domains.

tivity, other studies saw domain-specificity when using calibration measures. Given evidence that bias in confidence reasoning appears domain-general in adulthood (see Mazancieux et al., 2020), bias may change from domain-specific to domain-general throughout development, whereas sensitivity remains domain-general throughout. Future studies using new statistical techniques like meta-d' (Maniscalco & Lau, 2012) or that combine measures could explore this possibility.

A domain-general confidence account provides a mechanism for integrating and comparing independent and modular perceptual representations. Several theories have argued that perceptual domains are translatable into a common format through language (Carruthers, 2002; Spelke, 2003). For example, young children do not appear to integrate information about a room's geometry with its visual features, such as color (Hermer & Spelke, 1996). Our findings, though, suggest that confidence representations are easily translatable across independent perceptual boundaries, allowing children to compare the reliability of perceptual information. If true, even prelinguistic children could compare their confidence across modular and otherwise unrelated domains, providing an account for how unified and centralized cognition could be informed by encapsulated perceptual analyzers (and see Shea & Frith, 2019, for a similar account about certainty in consciousness).

This work also signals the potential for domain-general transfer of confidence reasoning across perceptual tasks for children. For instance, giving young adults periodic feedback led to improved confidence sensitivity on an unrelated task (Carpenter et al., 2019). Our results suggest that similar training effects could also work in primary school, holding potentially powerful implications for educational practices given that metacognitive skills are considered important for learning (e.g., Lockl & Schneider, 2004).

References

Abrosoft. (2007). Fantamorph 4 (Version 4) [Computer software]. Retrieved from http://www.fantamorph.com/

Baer, C., Gill, I. K., & Odic, D. (2018). A domain-general sense of confidence in children. *Open Mind: Discoveries in Cognitive Science, 2,* 86–96. http://dx.doi.org/10.1162/opmi_a_00020

Baer, C., & Odic, D. (2019). Certainty in numerical judgments develops independently of the approximate number system. *Cognitive Development, 52,* 100817. http://dx.doi.org/10.1016/j.cogdev.2019.100817

Barthelmé, S., & Mamassian, P. (2009). Evaluation of objective uncertainty in the visual system. *PLoS Computational Biology, 5,* e1000504. http://dx.doi.org/10.1371/journal.pcbi.1000504

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

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General, 148,* 51–64. http://dx.doi.org/10.1037/xge0000505

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25,* 657–674. http://dx.doi.org/10.1017/S0140525X02000122

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE, 11,* e0147901. http://dx.doi.org/10.1371/journal.pone.0147901

de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science, 25,* 1286–1288. http://dx.doi.org/10.1177/0956797614528956

Filevich, E., Forlim, C. G., Fehrman, C., Forster, C., Paulus, M., Shing, Y. L., & Kühn, S. (2020). I know that I know nothing: Cortical thickness and functional connectivity underlying meta-ignorance ability in preschoolers. *Developmental Cognitive Neuroscience, 41,* 100738. http://dx.doi.org/10.1016/j.dcn.2019.100738

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London: Series B*. Biological Sciences, *367,* 1338–1349. http://dx.doi.org/10.1098/rstb.2011.0417

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

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10,* 843–876. http://dx.doi.org/10.3758/BF03196546

Geurten, M., Meulemans, T., & Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cognitive Development, 48,* 62–81. http://dx.doi.org/10.1016/j.cogdev.2018.08.002

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. http://dx.doi.org/10.1073/pnas.1515129113

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. http://dx.doi.org/10.1177/0956797614542273

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61,* 195–232. http://dx.doi.org/10.1016/S0010-0277(96)00714-7

JASP Team. (2020). JASP (Version 0.13.1) [Computer software]. Retrieved from https://jasp-stats.org/

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008, August 10). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455,* 227–231. http://dx.doi.org/10.1038/nature07200

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1,* 153–169. http://dx.doi.org/10.1080/17405620444000085

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

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

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of Type 1 and Type 2 data: Meta-d′, response-specific Meta-d′, and the unequal variance SDT model. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 25–66). http://dx.doi.org/10.1007/978-3-642-45190-4_3

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General, 149,* 1788–1799. http://dx.doi.org/10.1037/xge0000746

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. http://dx.doi.org/10.1037/0033-2909.95.1.109

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19,* 366–374. http://dx.doi.org/10.1038/nn.4240

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: Insights from individual differences and neuroimaging. *Personality Neuroscience, 1,* e17. http://dx.doi.org/10.1017/pen.2018.16

Shea, N., & Frith, C. D. (2019). The global workspace needs metacognition. *Trends in Cognitive Sciences, 23,* 560–571. http://dx.doi.org/10.1016/j.tics.2019.04.007

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 277–311). Cambridge, MA: MIT Press.

Sumner, E., DeAngelis, E., Hyatt, M., Goodman, N., & Kidd, C. (2019). Cake or broccoli? Recency biases children's verbal responses. *PLoS ONE, 14,* e0217207. http://dx.doi.org/10.1371/journal.pone.0217207

Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills: Metacognition in the numerical domain. *Psychological Science, 25,* 1712–1721. http://dx.doi.org/10.1177/0956797614538458

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology: Part I. Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25,* 35–57. http://dx.doi.org/10.3758/s13423-017-1343-3

Winman, A., Juslin, P., Lindskog, M., Nilsson, H., & Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Frontiers in Psychology, 5,* 851. http://dx.doi.org/10.3389/fpsyg.2014.00851

Received August 6, 2019
Revision received June 17, 2020
Accepted June 24, 2020 ■