Supplemental Material

Table S1

Descriptive statistics, tests against chance, and average estimates of fit to the Weber model.

Dimension	*M* [95% CI]	*t*(40)	*p*	*d*	# Fit	*w* [95% CI]	Lapse rate [95% CI]
Non-Inverted Model Confidence Analyses							
Number	70.00 [63.04, 76.96]	5.81	<.001	0.92	31	.42 [.27, .56]	.10 [.04, .15]
Area	77.33 [69.98, 84.68]	7.52	<.001	1.19	32	.22 [.11, .32]	.13 [.07, .20]
Emotion	62.17 [56.47, 67.87]	4.32	<.001	0.68	29	.26 [.14, .38]	.19 [.11, .26]

Table S2

Correlations between tasks with and without controlling for age. Accuracy data is correlated using Pearson r, while model fit estimates are correlated using Spearman's ρ. Correlations controlling for age are shown in brackets.

	Accuracy (*r*)		*w* (ρ)	
	Area	Emotion	Area	Emotion
Non-Inverted Model Confidence Analysis				
Number	.72*** (.73***)	.71*** (.71***)	.06 (.06)	.60*** (.61**)
Area		.62*** (.62***)		.22 (.24)

^ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table S3

Factor loadings of measures by dimension.

	Non-Inverted Model Confidence Analysis
Measure	Component 1
Number	.916
Area	.876
Emotion	.872
Eigenvalue	2.37
Variance Explained	79%

Note: Lower w values indicate better precision.

Additional Confidence Analyses

As described in the main text, we found that while most children showed a clear pattern of selecting the easier of the two ratios on our relative Confidence task, 10 out of the 40 children showed the opposite pattern on at least one dimension, choosing the harder of the two ratios. While some of these children did communicate to us that they were doing this to challenge themselves, not all did. Nevertheless, given that these 10 children were consistently choosing the more difficult trials, we felt justified in using an inverted confidence model for their data, counting the harder trials as accurate and the easier trials as inaccurate. Here, we report on all of the analyses from the Confidence task without modelling these 10 children's data by using a reverse model, and instead treating their data as consistent with the typical model that states that children will seek to maximize performance by selecting the easier trials.

Using the same model for all 40 participants, children in the Confidence condition still performed above chance for all three dimensions (Table S1), choosing the easier of the two trials on 70% of trials (95% CI[63.90, 75.77], $t(39) = 6.76$, $p < .001$, $d = 1.07$). We again found that children were best on the Area trials and worst on the Emotion trials, $F(2, 78) = 16.04$, $p < .001$, $\eta_p^2 = .29$. Finally, we once again found no correlations between accuracy and age, all r's < .13, potentially due to our restricted range.

We also found the same metaratio effects: children were more likely to choose the easier question when the metaratio was higher in the Number task, ($F(2, 78) = 4.74$, $p = .011$, $\eta_p^2 = .11$), and the Emotion task ($F(2, 78) = 9.27$, $p < .001$, $\eta_p^2 = .19$), but not the Area task ($F(1.59, 61.90) = 1.91$ $p = .155$, $\eta_p^2 = .05$). This lack of effect, however, appears to be stemming from extremely good performance at all tested metaratios in the Area task (Figure S1b), suggesting that this model does not adequately fit children's data in this dimension.

We successfully fit all three dimensions for 22 children on the standard psychophysical model that assumes that children select the easier of the two trials with increasing metaratios. A total of 36 children could be fit on two of the three dimensions. Eight of these children were not fit by any model (as reported in the Main Text), while 10 were fit to the inverted model and thus excluded from the following analyses. Data from all fit dimensions is included where applicable, and accuracy data for all children is retained even if there is no model fit. The fit *w* data is presented in Table S1. Note that, as a result of this, many analyses using *w* estimates are comprised of a smaller sample (as low as 22 out of 40 children for the factor analysis), potentially leading to a loss of statistical power.

Finally, as shown in Table S2, we found correlations between Number, Area, and Emotion confidence discrimination for accuracy, replicating the domain-general effect even when all 40 children are assumed to have chosen the easier of the two trials. On the other hand, we found much weaker effects for *w*, possibly due to the ceiling effects in the Area task combined with the low sample size for the non-inverted *w* sample.

***Principal Component Analyses*.** Because we could not estimate all three *w*s for 18 children, our sample size for the PCA is very low. We therefore conducted a factor analysis over accuracy scores only, allowing us to use all 40 children in the sample. As shown by the factor loadings in Table S3, we replicated the effect in the Main Text and found a single factor underlies performance in all three Confidence tasks.

Comparison of PCA Between Conditions

An important question when evaluating the PCAs is whether the patterns of results in the Confidence and Discrimination conditions are meaningfully different. While we know of no way to directly compare the two given that the dependent variables are not identical, we adopted a bootstrapping approach to generate distributions of each PCA (see Babamoradi, van den Berg, & Rinnan, 2013)[1]. We conducted a non-parametric bootstrap with 10,000 samples (with replacement) on the PCAs in both conditions, testing whether the proportion of variance explained in the first component is different between the Discrimination and Confidence conditions. If the two PCAs generate a different number of factors, we should find that the 95% confidence intervals of the first component variance explained should not overlap. Consistent with this, we found that the 95% Bias Corrected and Accelerated interval in the Discrimination condition [31.34, 35.88] did not overlap with the interval in the Confidence [55.75, 85.71], consistent with the two PCAs generating different number of factors for the two conditions. A similar 95% Bias Corrected and Accelerated interval was found in the Confidence condition using the non-inverted data [66.52, 88.78].

References

Babamoradi, H., van den Berg, F., & Rinnan, Å. (2013). Bootstrap based confidence limits in principal component analysis—A case study. *Chemometrics and Intelligent Laboratory Systems*, *120*, 97-105.

[1] We thank an anonymous reviewer for suggesting this analysis.