Title
Developing Irrational Confidence? Metacognition in Probabilistic Decisions with Multiple Alternatives

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
MacColl, Cate
Sewell, David K
Redshaw, Jonathan

Publication Date
2024

Peer reviewed

Developing Irrational Confidence? Metacognition in Probabilistic Decisions with Multiple Alternatives

Cate MacColl (c.maccoll@uq.edu.au)
School of Psychology, The University of Queensland

David K. Sewell (d.sewell@uq.edu.au)
School of Psychology, The University of Queensland

Jonathan Redshaw (j.redshaw@uq.edu.au)
School of Psychology, The University of Queensland

Abstract

Prevailing theories propose that confidence in two-alternative forced-choice decisions is based on the probability that the selected option is correct. However, recent findings from three-alternative tasks suggest that adults' confidence might irrationally reflect the difference between the probabilities of the best and next-best options only, with other options disregarded. Using a novel probability task (in which participants guess the colour of a ball to be randomly selected from varying distributions) and a uniquely sensitive confidence measure, we investigated metacognition in multi-option decision making in children (N = 97, aged 6-9-years) and adults (N = 51). Contrary to previous findings, children's and adults' confidence was primarily explained by the probability of the best option. However, preliminary findings suggest that among older children and adults, additional irrelevant factors also accounted for unique variance in confidence. In some contexts, human confidence might be initially calibrated rationally but increasingly reflect irrational factors over development.

Keywords: cognitive development; decision making; development;

Introduction

Metacognition is studied in disciplines from basic perception science to applied educational and social psychology for the insights it can provide. From simple visual discrimination tasks to choosing who we vote for, metacognition informs the information we attend to, the emphasis we place on it, and the decisions we make (Denison et al., 2018; Mamassian, 2016). Among humans, metacognitive reflection on one's own thoughts is conventionally measured through explicit reports of confidence, which correlate strongly with the time taken to make the relevant decision (Yeung & Summerfield, 2012)and can predict decision accuracy, learning and subsequent actions (Flavell, 1979).

While decision-making is traditionally researched in the context of two-alternative forced-choice decisions, many real-world decisions do not involve simple binary choices. Even a basic decision of what to eat, for instance, will likely contain a myriad of options. It might be tempting to conceptualise such foraging activities as involving a series of binary choices based on the presence or absence of sustenance. However, each branching decision can contain multiple alternatives for prioritising food sources that are more nutrient dense or readily available, and many possible search locations that vary on dimensions like proximity and abundance at continuous scales (Abrams, 1991).

From this perspective, it is unclear how to apply insights from traditional two-alternative experiments to explain any decision-making behaviour, let alone complex human decisions. Perhaps it is the case that decisions such what we consume are made using a series of pairwise comparisons of two competing options, with the preferred option progressing to the next comparison until it is compared to a better alternative and ruled out. Alternatively, perhaps we routinely ignore less viable options based on implicit heuristics and compare only the two most favourable options. That is, despite many initial options, humans may be inclined to automatically reduce multiple-alternative decisions to binary choices. However, such competing hypotheses cannot be differentiated with the use of mere binary decision-making experimental tasks, as there are of course no additional options to be factored in or ignored. That is, information for the best option has a perfect inverse correlation with both the information for the second-best option and the information for all non-best options, since the second-best option is the only non-best option. Therefore, investigation of how humans differentiate the best option in complex decision-making environments is only possible through tasks involving decisions between more than two alternatives.

Paradigms using multiple alternatives can provide insight into the underlying mechanisms of metacognition, by measuring whether confidence reflects irrelevant as well as relevant evidence (Li & Ma, 2020). Conventional theories would imply that for a rational decision-maker, confidence should follow a Bayesian approach, merely reflecting the probability that the (chosen) best option is correct (Drugowitsch et al., 2014; Kepecs & Mainen, 2012), with the set of other options and their respective probabilities relevant only insofar as they intrinsically affect the probability of the best option. That is, at the point at which you enter a small raffle with a third of the tickets, the distribution of the other tickets (say, many participants each with a single ticket, or one raffle enthusiast with all of the remaining tickets) should not factor into your confidence in winning the raffle. Because most experimental paradigms have assessed confidence in

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

two-option contexts, however, it has not been possible to disentangle whether (i) non-best options are indeed appraised as a single set of alternative candidates that only indirectly influence confidence in the best option, or (ii) the discrete probabilities of non-best options in fact directly influence confidence in the best option.

This limitation was addressed in a recent study by Li and Ma (2020), in which participants saw clusters of red, blue and green dots and a single target black dot. They were asked to take the perspective of a birds-eye-view scene of people wearing each coloured shirt and to assign the target person to one of the three colour groups. By varying the locations of each distribution of each colour, this task independently varied evidence for each option. Contrary to prevailing theories, results suggested that only the probabilities of the two best options influence adults' confidence judgements in three-alternative decision making, with the third-best option disregarded altogether. That is, rather than rationally treating the two non-best options as a single set of alternative candidates, adults seem to irrationally prioritise the second-best option as an alternative candidate even after a decision is made.

Developmental Perspectives

Studying the origins of metacognition and confidence in developmental contexts can provide critical insights into the foundational building blocks underlying complex human decision-making, including the propensity to account for rational and irrational factors. Evidence suggests that young children readily experience cognitive uncertainty and can reliably report on such uncertainty from between four and six years of age (Lapidow et al., 2022; Selmeczy et al., 2021). From a similar age, children can also report their confidence in simple two-alternative forced-choice decisions in paired association and recall tasks (García-Pérez & Alcalá-Quintana, 2013; Hofer & Pintrich, 1997; Roebers et al., 2019; Selmeczy et al., 2021).

Although such methods have facilitated important discoveries about children's confidence in simple decisions, they are nonetheless limited in at least three important ways. First, the measures are by nature restricted to assessing confidence in binary choices (e.g., was the target previously present or absent), precluding insights into the underlying mechanisms of children's decision-making and confidence as outlined above (cf. Li & Ma, 2020). Secondly, even in the context of two-alternative decisions, these paradigms can only provide a blunt assessment of children's confidence, given that judgements about the presence or absence of an item in a recall task are categorically true or false, rather than probabilistically accurate as a function of the evidence for and against competing options. Any variability in confidence beyond a two-point "high" or "low" judgement is attributable only to individual differences in factors beyond the scope of investigation. Finally, experimental measurements predicated on two alternatives necessitate that a rational agent should select an option that is more likely than unlikely (i.e., >50%), or at least equally likely and unlikely (i.e., 50%). The factors affecting confidence in such decisions may be distinct from the factors in cases where the rational option is overall unlikely (i.e., <50%), yet still more likely than any other available alternative. As a result, is it unclear how children experience and report on confidence in such contexts.

To our knowledge, the present study provides the first investigation of children's confidence judgements in multiple-alternative decisions (see Experiment 1). It also includes a comparison sample of adults (see Experiment 2), enabling inferences about the ontogenetic origins of and transformations in the processes underlying such judgements. Participants were presented with a computerised task which showed various proportions of coloured balls (red, blue, green, and yellow), such that the evidence for each of the four options could be varied independently. Across trials, participants were asked to predict the colour of a randomly selected ball from among 12 balls, and to rate their confidence in each prediction (see Figure 1). Confidence was measured with a 120-point continuous scale, enabling far more sensitive assessment of confidence than in previous developmental studies (typically using between 2- and 5-point judgements) or in Li and Ma's (2020) study with adults (using a 4-point scale). Overall, this study aims to assess how children and adults make judgements and metacognitive appraisals in the context of probabilistic uncertainty with multiple alternatives.

Experiment 1

Methods

Participants As preregistered[1], our final sample consisted of 6- to 9-year-olds (N = 97) recruited at a public museum. This sample size is consistent with previous research on children's confidence and allowed appropriate counterbalancing of conditions (see supplementary materials). Recruitment ceased when the intended sample size for each age in years (with a roughly balanced sex ratio) was reached. The final sample was comprised of 97 children (47 males, 50 females) aged between 6.01 and 9.96 years (M = 7.93, SD = 1.14), with 49 younger children (6 to 7 years) and 48 older children (8 to 9 years)[2].

Measures Our novel measure of explicit confidence was represented by a cup that participants were able to fill up by dragging across the screen. The measure was comprised of a sliding scale within the cup with 120 levels from empty (very unsure) to full (very sure). This measure allows a highly sensitive assessment of confidence in children in an intuitive and interactive format, without confounding metacognitive evaluations with emotion or affect (as for alternative child-oriented confidence scales that use faces). The measure was

[1] https://osf.io/tuwg9/?view_only=34b90cffbd8e41d4b96aedc8546416a9

[2] Effects for continuous age were followed up by examining simple effects among younger children and older children.

presented after the initial decision was made, such that children needed to reflect upon the decision they had just made when estimating their confidence instead of merely reporting the likelihood of their choice based on the visual display. Reaction time for the probability task was also recorded as an implicit measure of confidence. Reaction time is typically inversely correlated with explicit confidence ratings (Ackerman & Koriat, 2011; Roderer & Roebers, 2014), and so collecting these data enabled validation of the novel scale while also providing insight into implicit confidence processes like information processing and action execution (Ratcliff & McKoon, 2008).

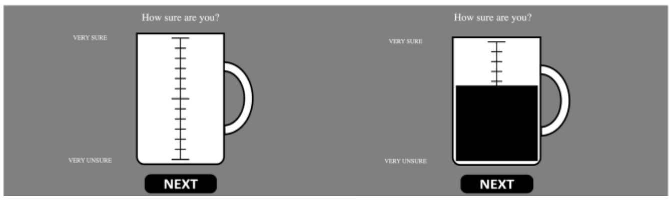

Figure 1: The novel confidence measure as seen by participants, at empty (0%, left) and partially filled (approximately 70%, right).

Procedure Participants were introduced to the Randomiser during a training phase in which they watched a video of two balls being unpredictably and quickly jumbled before one ball exited. Across four training trials, participants saw each possible combination of two colours and outcomes—black and black (black exits), black and white (white exits), black and white (black exits), white and white (white exits)—such that no outcome or initial starting position deviated from random chance levels. An initial check was then conducted to confirm children's understanding of the confidence measure (asking how sure they were of their own name, and the name of an unknown generic character) and the task itself (asking if two balls of the same colour were in the Randomiser, which colour would exit). Children that failed either of these measures were removed from the sample.

The main experimental task was comprised of 40 trials, each showing 12 boxes with red, blue, yellow and green balls inside the "Randomiser" (see Figure 2). On each trial, participants predicted which ball would exit the Randomiser by tapping on one of the four response buttons at the bottom of the screen. Probability was manipulated by varying the number of balls for the best option between 3 and 12 and the other colours were varied at every possible combination[3].

Results

Accuracy Generalised Linear Mixed Models with random intercepts were used to analyse children's accuracy as a function of age and the probability distribution of available information. Consistent with the broader literature on two-alternative forced-choice decisions, accuracy was analysed only on trials in which there was a single most likely option, and operationalised as selecting that option rather than any of the less likely options. Overall, children picked the most likely option from the array in 88% of trials (81% for younger and 96% for older children). Notably, this high performance reflects a sufficient understanding of probabilistic outcomes as a precursor for measuring confidence. As the probability of the best option increased (that is, as trials approached a guaranteed outcome) children became increasingly accurate $\chi^2(1, N = 97) = 39.56, p < .001$. For the present study, analyses of confidence (see below) were conducted using only accurate responses.

Alternative models of accuracy. To test whether children could have misunderstood the premise of the task, and tended to choose an option closest to the exit of the Randomiser, we also compared competing models for accuracy (i) based on selecting the most likely option (as above), and (ii) based on selecting a colour that matched one of the two balls closest to the exit (i.e., the two bottom middle balls). Comparisons suggested that participants' responses were much more likely to reflect the most likely option (88%) than the proximity of colours to the exit (57%), $t = 38.09$, p < .001.

Explicit and Implicit Confidence The explicit confidence model was built sequentially, with age and trial number entered at the first step. Similar to Li and Ma (2020), at Step 2 we compared separate models that contained fixed effects of competing factors: (i) the count for the best available option (Maximum model), (ii) the difference between the best two competing options (Difference model), and (iii) the

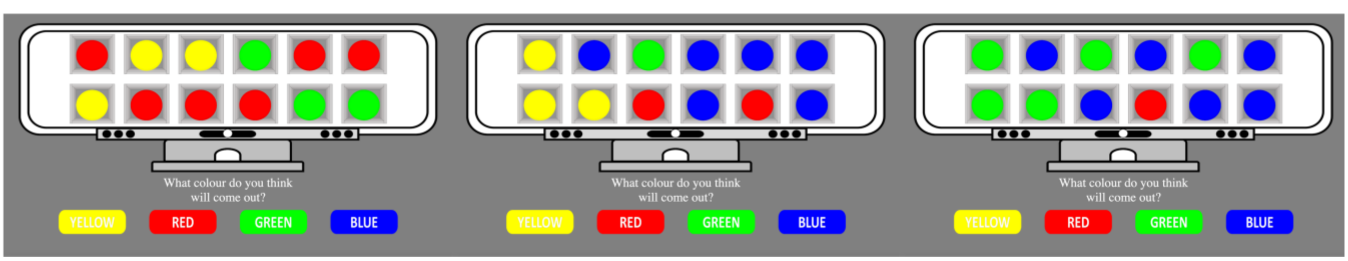

Figure 2: Sample trials (as seen by participants) in which the best option has six balls of that colour. However, the second-best option differs between 3 (left and right) and 5. Response buttons each have the colour name written in a box of that colour.

[3] For instance, when six boxes were of one colour, children saw one trial each of all possible distributions of the other six boxes—6-6, 6-5-1, 6-4-2, 6-4-1-1, 6-3-3, 6-3-2-1, and 6-2-2-2 distributions—randomised across colours.

overall entropy of all options at the trial level (i.e., the Shannon entropy value; Entropy model). In addition, we also compared a model that included (iv) the number of colours in the display (Colours model), as unlike in Li and Ma's (2020) paradigm where all colours were always present, this varied between one and four across our trials. We compared models using the Akaike Information Criterion (AIC; Akaike, 1974).

The best fitting Step 2 model (Maximum) included the total number of balls for the best available option ($F = 1162.10$, $p < .001$, Δ AIC = -992.90). This differs from Li and Ma (2020), who found that the best fitting model (Difference) included the difference between the amount of evidence for the best and next-best options. The models of Difference (Δ AIC = -935.56), Entropy (Δ AIC = -874.77), and Colours (Δ AIC = -588.63) had much lower explanatory power than the Maximum model.

To substantiate these findings, the same model comparisons were conducted for implicit confidence, with reaction time as the outcome variable. Findings indicated the same pattern, with the Maximum model producing a better fit ($F = 1421.27$, $p < .001$, AIC = -1176.44) in comparison to the Difference (Δ AIC = -1118.94), Entropy (Δ AIC = -772.58) and Colours (Δ AIC = -403.25) models. Therefore, contrary to previous findings in adults, children's explicit and implicit confidence is best predicted by the evidence for the best available option. See Figure 3 for the effect of the Maximum count on implicit and explicit confidence for each age group.

Interactions with age. Taking the Maximum model, an interaction between the Maximum effect and age was introduced at Step 3. The interaction showed a significant improvement in model fit ($F = 81.01$, $p <.001$, Δ AIC = -78.03). The results indicate that the effect of Maximum had a stronger effect on explicit confidence for older ($t = 31.91$, $p < .001$) than younger children ($t = 25.66$, $p < .001$). This interaction was also significant for response latency ($F = 97.51$, $p < .001$, Δ AIC = -94.07), again with a stronger effect for older children ($t = -33.24$, $p < .001$) than younger children ($t = -22.11$, $p < .001$) and. This indicates that with increasing age, the likelihood of the best available option has an increasing influence on both implicit and explicit confidence.

Exploratory Analysis Based on the discrepancy with Li and Ma (2020) and the significant developmental trajectories observed in the Step 3 model, an additional set of analyses were conducted to determine if other variables besides the Maximum effect could explain additional variance and improve the model fit. These Step 4 model comparisons

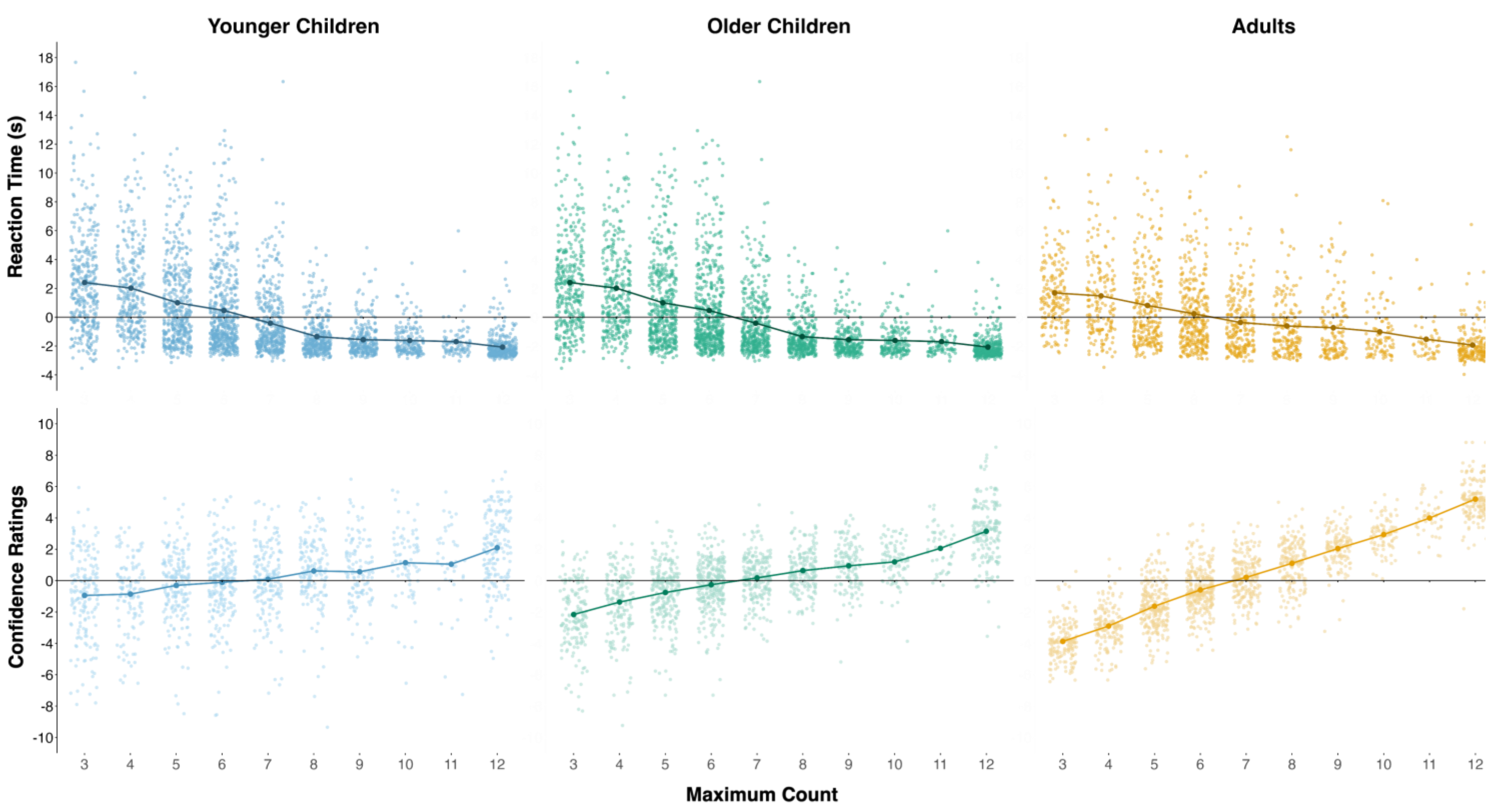

Figure 3: Graphs of the effect of Maximum Count on implicit confidence (mean-centered reaction time as denoted by darker colours, top row) and explicit confidence (mean-centered confidence reports as denoted by lighter colours, bottom row) for younger children, older children and adults (left to right respectively). The x-axis represents the number of balls for the best option in the display. Scores show raw data and trend lines represent the group mean at each level of Maximum Count.

indicated that neither Difference, Entropy, nor Colour explained additional unique variance in children's confidence over and above the Maximum model. Given the high degree of collinearity between Maximum and Difference (VIF = 10.19), Maximum and Entropy (VIF = 40.22), and Maximum and Colours (VIF = 17.28) factors, however, these findings need to be interpreted with caution.

Confirming that children across ages calibrated their confidence similarly, confidence did not vary significantly with age as a main effect (F = 0.87, p = .350). However, beyond the effect of the Maximum variable, reaction time was an additional significant predictor of explicit confidence (F = 20.22, p < .001, Δ AIC -18.15). The linear interaction of this effect with age was not significant (F = 3.39, p = 0.066). However, post-hoc analyses revealed that simple effects differed at the group level (F = 5.58, p = .018, Δ AIC -3.58), such that older children (t = -3.92, p < .001) but not younger children (t = -0.59, p = .554) showed a significant negative association between implicit and explicit confidence measures beyond variation in objective task parameters. Such a relationship provides preliminary evidence of a potentially non-linear developmental shift in confidence calibration between younger and older children. That is, in our cross-sectional sample at least, between the ages of 7 and 8 years[4], children's reaction time when making predictions became a unique predictor of their confidence in those predictions, such that faster reaction times were associated with higher confidence over and above the influence of the objective probability of the selected option.

Discussion

Experiment 1 demonstrates that children are highly sensitive to gradations in probability and can report on subtle nuances of confidence from as young as six years of age. Accordingly, the accuracy of children's predictions also increases with the probability of the most likely option. That is, children are less likely to select a lower-probability option as the proportion of the highest-probability option increases and approaches a guaranteed outcome. As expected, explicit and implicit confidence were strongly correlated in older children, such that these children reported being more confident as their reaction time decreased (and the probability of the best option increased).

These relationships provide evidence that even children can make primarily rational decisions informed by probabilistic information—with appropriately calibrated confidence in such decisions. Moreover, the apparently linear effects at the low end of the best-option probabilities (e.g., when the best option was more likely to *not* occur than to occur) indicates that such rationality extends beyond simple two-alternative choices as examined in traditional experimental paradigms. This is integral to a complete understanding of complex, real-world decisions.

[4] See Anderson (2002) for evidence of a consistent qualitative shift in executive function in children between the ages of 7 and 9

For a rational decision-maker, the count of the most likely option should be the sole factor influencing confidence, such that, in our task, it should not matter what distribution of colours make up the remaining options. Therefore, the selection of the Maximum Model is evidence that the children in our sample primarily factored in relevant information in their calibration of confidence. This finding stands in contrast to that from Li and Ma (2020), who found that an adult sample primarily based their confidence on irrelevant information (i.e., the probabilistic difference between the best and next-best options). However, the fact that older children also seemed to rate their confidence based on extraneous information such as reaction time indicates that, in some regards, children might become less rational in their calibration of confidence to objective probabilistic outcomes with age. Given the substantive differences between the current task and that of Li and Ma (2020), it was concluded that an additional adult sample was needed to contextualise these findings.

Experiment 2

Methods

Participants The adult study was added to the existing pre-registration before data collection. The sample was recruited from the UK pool of Prolific Academic. There were a total of 49 participants (to roughly match the size of each age group in Experiment 1; M = 39.76, SD = 15.00).

Measures The task was kept substantively the same, albeit with minor modifications to make it suitable for adults and for online delivery. This included minor changes in wording, and the removal of a manipulation check (designed to check children's understanding of the scale) that was not viable using the online platform.

Procedure Participants completed the task in their own time by accessing Prolific Academic using their own devices.

Results

Much like older children, adults were accurate in 97% of trials. Competing models were constructed using the same process as Experiment 1, with the exception that age was not entered in Step 1 (or as a factor in interactions) as no age-related effects were expected in the adult sample. The Maximum model was also the best fitting explicit confidence model for adult participants (F = 7451.84, p < .001, Δ AIC = -2953.79). Moreover, like older children in Experiment 1, adults showed a significant negative association between implicit and explicit confidence, above and beyond variation on maximum likelihood (F = 8.09, t = -2.98, p < .001, Δ AIC = -6.88).

Exploratory Analysis In a Step 2 model equivalent to the Step 3 models in Experiment 1, we added Difference, Entropy and Colours to separate models (also containing the Maximum variable) to explore the possibility of additional variance explained. Contrary to findings from Experiment 1, results indicate that Difference ($F = 11.97$, $p = .001$, Δ AIC = -9.93), Entropy ($F = 7.57$, $p = .006$, Δ AIC = -5.56) and Colours ($F = 10.00$, $p = .001$, Δ AIC = -7.98) separately improved the model fit over and above the Maximum effect. As above, however, these effects must be interpreted with caution due to the significant collinearity of variables in the model.

Discussion

Similar to the findings from Experiment 1, these results are inconsistent with those from Li and Ma (2020). Rather than primarily reflecting the difference between the best and next-best available options, the confidence scores for adults in our study were instead best predicted by the value of the most likely option in isolation. That is, our participants were primarily rational in their approach. However, in line with findings from older children in the developmental sample, adults' reported confidence varies both as a function of their reaction time on the trial, seemingly in addition to information in the display beyond the maximum count. This indicates that adults and older children, perhaps more so than younger children, are factoring in irrelevant information when estimating their confidence in their initial judgement.

General Discussion

Across both samples, participants' confidence was explained predominantly by the Maximum model. This reflects a rational approach to decision confidence that, unlike Li and Ma's (2020) findings, is compatible with a traditional Bayesian framework. There are, of course, some notable differences between experiments that might explain the discrepancy.

Firstly, confidence may be a multifaceted process that functions differently in bottom-up perceptual contexts (as in Li and Ma's task) as compared to top-down prediction contexts (as in our task). This explanation may be considered unlikely, however, given that the stimuli of colour arrays, and the judgement of competing colours at varying levels of likelihood are quite similar between the tasks. Secondly, differences in the framing of the tasks may account for the differences in findings. While Li and Ma contextualised the three-alternative decision in a social context of groups of people wearing different colours, our task is founded in a physical probabilistic prediction about the outcome of a lottery-like selection. Thirdly, whereas Li and Ma's task had a highly sensitive manipulation of the multiple alternatives (i.e., hundreds of coloured dots) but a relatively blunt measure of confidence (a 4-point scale), our task had a relatively blunt manipulation of the multiple alternatives (i.e., 12 coloured balls) but a highly sensitive measure of confidence (a 120-point scale).

Another interesting finding was that, among older children and adults—but not younger children—irrelevant factors accounted for unique variance in confidence judgements over and above the primary factor of evidence for the best option. That is, younger children's confidence seemed to reflect rational factors only, whereas older children's and adults' confidence also reflected irrational factors. This ontogenetic pattern is observed in some other human cognitive traits and abilities, with foundational capacities developing early but later becoming more subject to metacognitive "overthinking" that, while utile in some contexts, can be deleterious in other contexts (Newport, 1990). One potential reason for this effect is that, as working memory develops with age (Alloway et al., 2006), there may be a corresponding increase in our capacity to account for a range of factors in making predictions and judgements, and the delayed presentation of the metacognitive measure may have also added working memory demands for younger children. Thus, because we can better process information in a display as we age, we may account for more and more information in our metacognitive judgements irrespective of its utility.

Regardless of the underlying explanation, the difference between our findings and those of Li and Ma (2020) suggests that distinct metacognitive decision-making processes may be deployed in different domains. Further research is needed to clarify the nature of information that is likely to result in the miscalibration of confidence beyond objective task parameters throughout development.

References

Abrams, P. A. (1991). Life history and the relationship between food availability and foraging effort. *Ecology*, *72*(4), 1242–1252.

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied*, *17*(4), Article 4.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71–82.

Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, *115*(43), 11090–11095.

Drugowitsch, J., Moreno-Bote, R., & Pouget, A. (2014). Relation between belief and performance in perceptual decision making. *PloS One*, *9*(5), e96511.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

García-Pérez, M. A., & Alcalá-Quintana, R. (2013). Shifts of the psychometric function: Distinguishing bias from perceptual effects. *The Quarterly Journal of Experimental Psychology*, *66*(2), Article 2.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*(1), Article 1.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), Article 1594.

Lapidow, E., Killeen, I., & Walker, C. M. (2022). Learning to recognize uncertainty vs. Recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Developmental Science*, *25*(2), e13178.

Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*(1), 2004. https://doi.org/10.1038/s41467-020-15581-6

Mamassian, P. (2016). Visual confidence. *Annual Review of Vision Science*, *2*(1), Article 1.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), Article 4.

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), Article 1.

Roebers, C. M., Kälin, S., & Aeschlimann, E. A. (2019). A comparison of non-verbal and verbal indicators of young children's metacognition. *Metacognition and Learning*, *15*(1), Article 1. https://doi.org/10.1007/s11409-019-09217-4

Selmeczy, D., Kazemi, A., & Ghetti, S. (2021). Developmental Differences in Subjective Recollection and Its Role in Decision Making. *Child Development*, *92*(6), Article 6. https://doi.org/10.1111/cdev.13611

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