Age-Related Differences in the Influence of Category Expectations on Episodic Memory in Early Childhood

Kimele Persaud
kimele.persaud@rutgers.edu
Psychology, Rutgers University-Newark

Carla Macias
cm1172@scarletmail.rutgers.edu
Psychology, Rutgers University-Newark

Pernille Hemmer
pernille.hemmer@psych.rutgers.edu
Psychology, Rutgers University-New Brunswick

Elizabeth Bonawitz
lbaraff@gmail.com
Psychology, Rutgers University-Newark

Abstract

Previous research evaluating the influence of category knowledge on memory found that children, like adults, rely on category information to facilitate recall (Duffy, Huttenlocher, & Crawford, 2006). A model that combines category and target information (Integrative) provides a superior fit to preschoolers recall data compared to a category only (Prototype) and target only (Target) model (Macias, Persaud, Hemmer, & Bonawitz, in revision). Utilizing data and computational approaches from Macias et al., (in revision), we explore whether individual and age-related differences persist in the model fits. Results revealed that a greater proportion of preschoolers recall was best fit by the Prototype model and trials where children displayed individuating behaviors, such as spontaneously labeling, were also best fit by the Prototype model. Furthermore, the best fitting model varied by age. This work demonstrates a rich complexity and variation in recall between developmental groups that can be illuminated by computationally evaluating individual differences.

Keywords: Episodic Memory; Children; Computational Models; Category Knowledge; Color

Introduction

Reconstructing events from memory is an important facet of cognition, given that it informs how we perceive, interact with, and reason about the world around us. As with all computational processes, human memory is limited in its capacity and resolution, raising questions of how the mind handles the reconstruction of events from memory. That is, how do we strategically encode information that supports later use, while minimizing effort, error, and large demands on storage? This question is doubly interesting for young children whose memory systems are still developing. Relative to adults, children have comparatively limited cognitive resources (Davinson, Amso, Anderson, & Diamond, 2006; Diamond, 2006; Keresztes, Ngo, Lindenberger, & Newcombe, 2018), and their ability to maintain information in memory becomes compromised when faced with increased cognitive load (e.g., increased inhibition demands). Thus, an important question of development is what cognitive strategies might young learners employ to reduce uncertainty (i.e., noise or error) when retrieving information from memory?

To tackle strategic reconstruction of episodic events, research in adult cognition suggests that adults use prior knowledge and expectations to facilitate retrieval of information from memory. Adults develop prior knowledge and expectations that are well-calibrated to the statistical regularities of the environment (e.g., Griffiths & Tenenbaum, 2006), and use this knowledge to optimally perform on a broad range of cognitive tasks including: categorization (Huttenlocher, Hedges, & Vevea, 2000), reasoning (Oaksford & Chater, 1994), and generalization (Tenenbaum & Griffiths, 2001). In memory, well-calibrated knowledge and expectations for a stimulus category can improve average recall (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher et al., 2000). For example, Huttenlocher et al. (2000) found that people quickly develop expectations for the underlying categorical distribution of stimulus features, and use this knowledge to fill in noisy and incomplete memories. They demonstrated that responses regressed toward the mean of the overall category, thereby improving average recall.

This relationship between prior knowledge and episodic memory can be captured within a simple Bayesian framework which assumes that prior knowledge and expectations for the environment are optimally combined with noisy episodic content to produce recall of episodic experiences (Hemmer & Steyvers, 2009; Huttenlocher et al., 2000; Persaud & Hemmer, 2014; Steyvers & Dennis, 2006). Bayes rule provides a principled account of how to combine noisy memory representations with prior expectations to calculate the posterior probability for recall.

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

The posterior probability $p(\theta|y)$ describes how likely a recalled feature θ is, given prior expectations for the recalled feature $p(\theta)$ and noisy memory traces y. In this way, the Bayesian framework makes specific predictions about patterns that are explicitly borne out of the data, namely a regression to the category mean effect. It predicts that recall of stimulus features (e.g., different shades of red) is either over or under-estimated toward the mean of the category.

Recent evidence suggests that children, like adults, adopt a similar process of integrating prior category knowledge with episodic traces to reconstruct events in memory. For example, Duffy et al. (2006) used assumptions of the Category Adjustment model (CAM) (Huttenlocher et al., 1991, 2000) to evaluate the contribution of category knowledge to memory for object sizes in children. CAM assumes that if category knowledge is integrated in memory, recall would exhibit regression to the mean effects. The model also assumes that the more noisy the episodic information, like memories in children, the stronger recall will regress to the mean. Duffy et al. (2006) found that like adults, children's recall regressed toward the mean of the underlying category distribution. This suggests that on an individual trial, a child might not have remembered the exact studied size, so they might use their learned category knowledge of the most frequently studied object sizes to help reconstruct the true size. They concluded that children use category knowledge to estimate stimulus features from memory.

Similarly, Macias and colleagues (in revision) used a simple episodic memory task, where children were shown shapes paired with different colors and were asked to recall the color-shape pairings. They found that children's recall regressed toward the mean of the seven color categories that were studied, indicating an influence of category knowledge on memory. To further assess episodic memory, they then evaluated the fits of three computational models of memory to explain the data: a Noisy Target model that assumes recall solely mirrors episodic information (i.e., the target color values plus random noise), a Noisy Prototype model that assumes that recall solely mirrors category information (plus noise), and an Integrative model that assumes that recall is an integration of episodic and category information. Quantitative model fits to the aggregate data favored the Integrative model.

These studies of memory in children, taken together, highlight an important role that category knowledge plays in episodic memory at early development (i.e., preschool age) and provide a watershed moment to explore the reconstructive nature of episodic memory at earlier stages. More specifically, this work facilitates the opportunity to perform a critical in-depth analysis of children's recall data to tease apart underlying individual and group-related differences in the reconstructive process. Exploring individual and age related differences is motivated by the Duffy et al. (2006) finding that not only do children rely on category knowledge, but also that memory in younger children exhibited steeper regression to the mean patterns, relative to older children. Recall based solely on category information could also result in steeper regression to the mean, and in turn, might be better fit by the Macias et al., Noisy Prototype ('category only') model. In other words, it could be the case that at the individual subject level, children might differ in the best fitting model, such that those with steeper regression might be better fit by the Noisy Prototype model, while less steep regression might be better captured by an Integrative model.

Furthermore, recall performance in children might not only differ at the individual subject level, but also at the individual trial level, especially if contextual strategies, such as spontaneously labeling study features, are employed to facilitate recall performance. For example, while running their study, Macias et al., observed that participants spontaneously labeled the colors, as they studied them and/or as they recalled them. For example, one older learner (age = 4.64years), stated, "Purple, purple, purple. I got this.", while studying a purple hue value. Counterintuitively, while labeling may boost the learner's ability to remember that an item was observed from a particular category, it may also lead to noisier storage of specific stimuli that deviate from category means, because the label provides a cheaper (albeit potentially less accurate) compression option than storing the details of the original. In this way, this individuating behavior of labeling might impact the reconstruction of events in memory at either the individual subject or trial level. Recent research suggests that labeling can influence recall of continuous color values, such that labeling results in information being lost gradually as opposed to suddenly (see, Donkin, Nosofsky, Gold, & Shiffrin, 2014 for discussion on the role of labeling, sudden death, and gradual decay in memory). To this end, there might be a difference in the best fitting models for children who spontaneously label colors or for specific trials where colors are labeled.

Therefore, the goal of this paper is to assess individual and age related differences in the reconstruction of events from memory in early development. More specifically, we sought to evaluate whether younger and older children employ different strategies to recall episodic events and whether the behavior of spontaneously labeling was better fit by a particular model. We hypothesized that young and older children would differ in their reconstructive processes, such that a different proportion of children from each group would be better fit by the three models. We expected that older children would be better explained by an Integrative model (i.e., combining noisy episodic traces with category knowledge), mirroring the behavior of adults, and younger children would be better explained by a Noisy Prototype model, given the degree of inexactness in their memory traces.

We also hypothesized that the individuating behavior of spontaneous labeling would impact memory reconstruction such that trials where labels were spontaneously provided would be better captured by the Noisy Prototype model. To test our hypotheses, we fit the Noisy Target, Noisy Prototype, and Integrative models to the experimental data from Macias et al., (in revision) at the individual subject level.

We then evaluated the log likelihood scores of the model fits to determine which account most often explained memory performance in younger and older children. In other words, we looked to see which model explained behavior for the greater proportion of children. After, we explored best fitting parameter values that would capture the amount of noise in the recall data for young and older children. A difference in the amount of noise in the data is one potential explanation for age related differences in the best fitting model. Finally we assessed whether labeling behavior affected the proportion of children fit by each of the models.

Three Models of Memory

Noisy Target Model The Noisy Target model assumes that information is stored in episodic memory as noisy traces of studied values (e.g., specific color values). In this way, reconstructed events are just inexact representations of true studied values (and not altered by category knowledge). If children are using the Noisy Target model, we should expect the noise (or error) in recall to be normally distributed around the true studied feature values, with no apparent bias toward a particular recall value. To evaluate this model relative to the data, we calculated the probability of responses given a Gaussian distribution centered on the target value, with noise in memory (we assume the same memory noise value learned from Macias et al.).

Noisy Prototype Model The Noisy Prototype model assumes that information is stored in episodic memory as categorical representations of studied features (e.g., the mean of the category to which the studied value belongs). In other words, under this model, the initial encoding of the representation is simply a pointer to the participant's prototype in that category. Other information about the studied value is not stored. Memory is simply a recall of the prototype – which we define as a sample drawn from this category, assuming a particular distribution, mean, and variance associated with it. To evaluate this model relative to the data, we calculated the probability of responses given a Gaussian distribution centered on the category prototype (i.e., mean) value given by participant ratings in Macias et al., (in revision), with noise on the category also calculated from noise given in a separate study [1].

Integrative Model The Integrative model amalgamates the assumptions of both the Noisy Target and Noisy Prototype models and assumes that recall is an integration of noisy episodic content and prior category knowledge. Under this model, prior category knowledge is used to fill in the gaps when episodic traces are noisy or incomplete. When the category representation is strong, and the memory trace is noisy, recall will resemble the category representation. The probability of responses under the Integrative model are relatively straightforward to calculate, because both the prior and likelihood distributions are Gaussian (which are self-conjugate). Furthermore, there are not specific weights assigned to the contributions of each model – this falls out naturally based on the degree of variance of each target and prototype models. We evaluate this model relative to the data, by calculating the probability of responses given the Gaussian that results from integrating these two Gaussian. Specifically, for the Integrative model, which integrates the Noisy Target and Prototype distributions, the standard solution for the mean and variance is given by,

$$\mu = \frac{1}{\frac{1}{\sigma_t^2} + \frac{n}{\sigma_p^2}}\left(\frac{t}{\sigma_t^2} + \frac{\mu_p}{\sigma_p^2}\right), \sigma = \frac{1}{\frac{1}{\sigma_t^2} + \frac{n}{\sigma_p^2}} \quad (1)$$

where σ_t refers to the memory noise on the target distribution, σ_p refers to the noise on the prototype distribution, t refers to the studied target value, μ_p refers to the mean of the prototype distribution to which the target value belong, and n=1.

Table 1: Frequency of Children Best Fit to Each Model

Model	Count(%)
Integrative	11 (33.33%)
Noisy Target	7 (21.21%)
Noisy Prototype	15 (45.45%)

In what follows, we first briefly explain the experimental methods employed by Macias et al., (in revision), to assess the role of category knowledge in episodic memory in children. We then discuss the results of the model fitting at the individual subject level in general, and age related differences, more specifically.

Experimental Methods and Results

Macias et al., (in revision) conducted two developmental experiments where they examined the relationship between prior color category knowledge and episodic memory in preschoolers (mean age: 54 mos.; range: 43 mos.-73 mos.). In the prior knowledge assessment, participants were presented with 9 color category labels (red, orange, yellow, green, blue, light blue, dark blue, purple, and pink) one at a time on a computer screen, along with a color wheel. The color wheel varied in hue only while luminance and saturation were held constant at 50 and 100 units respectively. Children were asked to point to a location on a color wheel to indicate the color that best represented the label.

In the episodic memory task, 33 participants studied 15 shapes uniquely paired with 15[2] colors, one at a time on a computer screen. At test, participants were presented with a studied shape (filled in white with a black border), along with the color wheel used in the prior knowledge assessment. The task for the participants was to choose along the color wheel to indicate the color they recalled being paired with the presented shape. For complete experimental methodology, refer to the source publication (Macias, et al., in revision).

The results of the memory task revealed a regression to the category mean effect in a majority of the studied color categories such that studied hue values that were greater than the mean of the category were underestimated and studied hue values less than the mean of the category were overestimated. This regression to the mean effect is taken as evidence of an

[1] We also assessed a model in which we sample over variance, but best fit variance matched participant responses on Macias et al.'s prior knowledge task.

[2] One of the study trials was treated as a filler in order to counterbalance presentation order and was therefore removed from the data set prior to running any analyses.

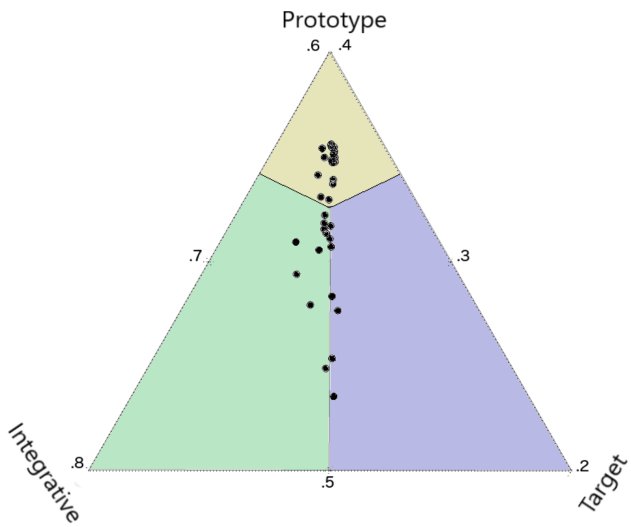

Figure 1: Ternary plot of the proportion of Log probabilities of the Integrative, Noisy Target, and Noisy Prototype models fit to each participant's data. Data points fall within the region of the model where it is best fit. Note that the figure has been zoomed in to the approximate center of the Ternary plot for better visualization of the data.

influence of category knowledge in episodic memory. Macias et al., (in revision) implemented three models and determined that the Integrative model provided the superior fit to the child data on aggregate. Here we fit the three models to individual subject data to assess for age related differences in the best fitting model.

Model Results

We sought to evaluate age and performance related differences between individual subjects and the fits of each model. Here we report the results of the model fits to children overall and then we evaluate the role of age.

Data Preparation for Evaluating Individual Differences

The data were prepared to perform four specific analyses: to evaluate individual differences in the best fitting model across the entire sample of children, to assess age related differences in the proportion of participants best described by each model, to assess additional group differences in the model fitting (e.g., the role of spontaneous labeling), and to evaluate age differences in best fitting model parameter values. We first fit the three models to each subjects' data. As with Macias et al., (in revision), the best fitting model was determined by the model with the largest log-likelihood value. To evaluate group differences, we performed a median split to classify children as younger and older learners (Table 2) and then compared the proportion of younger and older children described by each model. Of the 33 participants in the study, 16 were classified as young and 17 were classified as older. The median age of the total sample was 53 mos. (*sd*=6 mos.).

Table 2: Frequency of Model Fits by Age

Model	Count(%)	
	Young	Older
Integrative	6 (37.50%)	5 (29.41%)
Noisy Target	6 (37.50%)	1 (5.88%)
Noisy Prototype	4 (25.00%)	11 (64.71%)

The median ages for younger and older children were 49 mos. (*sd*=2 mos.) and 56 mos. (*sd*=5 mos.), respectively.

We also sought to evaluate group differences due to spontaneous labeling that was borne out of the experimental task. Of the 16 children classified as younger, 7 produced as least one label and of the 17 older children, 12 produced at least one label. This further suggests that labeling was a consistent strategy employed by children in this task. To evaluate best fitting models based on labeling, we first classified children into two groups: labelers and non-labelers. Labelers referred to learners who provided labels (at either study, test, or both) on more than 50% of trials (*n*=10/33) and non-labelers were all other children tested (*n*=23/33). We chose to use this classification because spontaneously labeling on more than 50% of trials suggests a consistent strategy of the individual to assist in recall.

To evaluate age related differences in the best fitting noise value, we implemented the Integrative model and for each participant, we searched over the space of possible noise values for the value that maximized the likelihood for each participant's data.

Model Fitting Results

Although the Integrative model is the best fitting model at the aggregate data level, it appears that at the individual level a greater proportion of children are better fit by the Noisy Prototype model (*n*=15/33), followed by the Integrative model (*n*=11/33), and then the Noisy Target model (*n*=7/33) (see Table 1). However, as can be seen in Figure 1, although a larger proportion of data points (each representing an individual child) fall towards the prototype apex, these points cluster towards the center (with near equal weight for the Target and Integrative models), suggesting that children who are classified as Prototype fits are nearly equally well fit by the other models. In contrast, for participants that are not best fit by the Prototype model, results skew significantly farther away from the center, suggesting that children who are better fit by other models are much more poorly fit by the Prototype. In light of this result, we next evaluated whether age plays a role in the proportion of children best fit by the models.

Age and Best Fitting Model To evaluate whether the proportion of children best fit by each of the three models was dependent upon age, we used the Freeman-Halton extension of the Fisher's Exact test to compute the (two-tailed) probability of obtaining a distribution of values in a 2(young vs older)x3(Integrative vs Noisy Target vs Noisy Prototype) contingency table, given the number of observations in each cell. The results revealed that the observed proportion of best fitting models was dependent on age (p=.031). In other words, there was a significant difference in the distribution of best fitting models between the age groups. Young children were evenly split in the number fit by the Integrative (n=6) and Noisy Target (n=6) models, followed closely by the Noisy Prototype model (n=4). Interestingly, however, older children had a different composition. An much larger proportion of older children were better fit by the Noisy Prototype model (n=11), followed by the Integrative model (n=5), and almost not at all described by the Noisy Target model (n=1) (see Table 2).

Age and Best Fitting Noise Parameter Macias et al., (in revision), demonstrated that for aggregated child data, the best fitting model was the Integrative model. To evaluate the fit of the Integrative model to young learners' data, they searched for the best fitting noise parameter value. Comparing this parameter to the best fit for adults revealed a significantly larger noise parameter for the children, suggesting that as children develop the fidelity of their memory gets sharper. Here we searched for the best fitting noise value at the individual subject level to test for age related differences within the preschool population. The goal was to assess whether a difference in the amount of noise between age groups could explain why young and older children were better fit by different models. In conflict with our prediction, there was a weak non-significant *negative* correlation between age and best fitting noise value (r-= -0.17, p=.35). This suggests that a difference in the best fitting model between age group was not a result of a difference in the amount of noise in the data[3]. We return to this point later.

Additional Group Differences and Best Fitting Model Similar to the evaluation of age, we then employed a Fisher's Exact test to evaluate whether the proportion of children best fit by the three models differed between *labelers* and *non-labelers*. To reiterate, we classified labelers as children who spontaneously provided a color label on more than 50% of trials. Figure 2 shows the composition of labelers and non-labelers fit by each model. A Fisher's Exact test yielded, p=.50 , suggesting no difference in the proportion of *labelers* and *non-labelers* best fit by the three models.

Although the difference between the proportion fits was not significant, there appeared to be a trend in which most *labelers* were described by the Noisy Prototype model (60%), while *non-labelers* were more diffused across the three models. Thus, to further evaluate the role of labeling, we separated participants' label trials from the non-labeled trials, creating two new datasets. We fit the three models to the

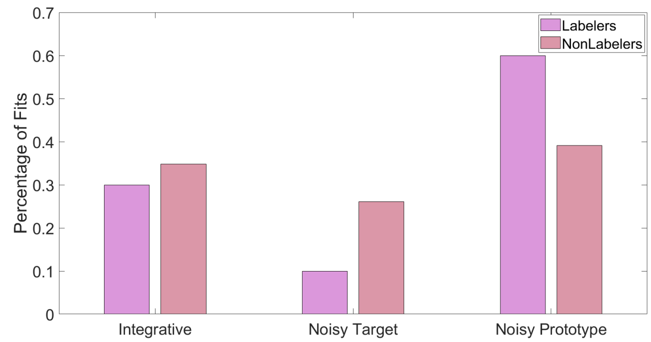

Figure 2: Proportion of Labelers and Non-labelers best fit by each model. Labelers were more likely to be best fit by the Noisy Prototype.

Table 3: Frequency of Model Fits based on Labeled and Non-Labeled Trials

Model	Count(%)	
	Label	Non-label
Integrative	3 (0.10%)	9 (27.27%)
Noisy Target	4 (0.12%)	7 (21.21%)
Noisy Prototype	26 (78.78%)	17 (51.51%)

aggregated label data and the aggregated non-label data. Unsurprisingly, the Integrative model provided the superior fit to both datasets, presumably because the model pays a lower cost for responses that, over the aggregate span between the observed target and category mean. Thus, to better understand the effects of labeling at the trial level, we then fit the three models at the individual subject level, again separating labelled trials from non-labelled trials. For the labelled trials, we found that 3 participants were best fit by the Integrative model, 4 by the Noisy Target model, but the majority of trials (26) were best fit by the Noisy Prototype model. In contrast, for the non-label trials, the distribution was less skewed, with 9 participants were best fit by the Integrative model, 7 by the Noisy Target model, and 17 by the Noisy Prototype model. A Fisher's Exact test revealed a marginally significant difference (p=.054) in the distribution of best fitting models between the labelled trials and non-labelled trials, such that most participants' label trials were best described by the Prototype model, while the non-label trials were slightly more dispersed.

Based on the finding of a difference in model fits between labeled and non-labeled trials, we re-examined the role of labeling on age. We had originally classified whole individuals as either labelers or non-labelers, and found no significant difference by age. Instead, we calculated the proportion of labeled trials provided by younger and older children, to test whether as a group, older children were more likely to provide labels during testing. A Fisher's Exact Probability Test revealed a significant difference in the proportions of la-

[3] An alternative explanation is that the sample sizes for young and older children split between each model was insufficient to detect a significant difference. However, the trending direction of the data ran counter to our developmental prediction, suggesting that even if greater power revealed differences, they would be in the unpredicted direction

beled and non-labeled trials contributed by each age group (p=.002). A larger proportion of labeled trials were generated by older (66%) compared to younger children (34%).

Discussion

Our goal was to evaluate whether age-related differences persist in the strategies young learners use to reconstruct events from memory. Recent work has found that young learners, like adults, adopt the strategy of integrating prior category expectations with noisy episodic traces to reconstruct events from memory (Macias, et al., in revision). This was evidenced by a model that assumes an integration of target and category information (i.e., Integrative model) providing a superior fit to the preschool data. Here we evaluate individual differences in the best fitting strategies. We first fit three models at the individual subject level and found that the larger proportion of children were better fit by the Noisy Prototype model compared to the other models.

In addition, there were marked differences in the proportion of young and older children best fit by each model. While young children were almost evenly split in fit across the three models, surprisingly, older children were most frequently fit by the Prototype model. This result might have been bolstered by the number of trials where older children spontaneously labeled. Recall that a significantly large proportion of labeled trials belonged to older children. In this way, spontaneously labeling during study and test might have induced older children to encode and/or retrieve the prototype of the category they verbally labeled. Thus, older children may have been more likely to adopt a general strategy (labeling) that instead led to less accurate recall of the specific observation. Future work might further explore the role of spontaneous labeling on children's recall performance. For example, it is unclear whether children were still using a labeling strategy on trials where they did not spontaneously label aloud. It is possible that they were silently labeling during the task. It is unlikely that this is the case, given that we found a significant difference in performance between labeled and non-labelled trials in terms of the model fitting. However, this is an empirical for future investigation. For instance, follow up studies could use verbal interference tasks to manipulate children's ability to provide verbal labels during encoding and retrieval to evaluate whether labeling alone encourages the use of the category prototype.

What might explain the finding that the Noisy Prototype model slightly outperformed the Integrative model in terms of best fit at the individual level? First, early memory development is marked by an up-prioritization of category information over nuanced episodic information (Keresztes et al., 2018). Such behavior would equate to encoding a red color value as a prototypical shade of red (e.g., the color of a red apple) as opposed to encoding the specific shade of red studied. Thus, during study, a majority of children may have encoded target information as a pointer to the category from which the target belongs, such as a category representative (i.e., the category mean) as opposed to encoding the exact color value studied.

Alternatively, it could be the case that the use of category knowledge happens at retrieval. After the initial testing phase, the original studied information could have degraded over time and instead of reproducing the degraded information, children reproduced a value closer to the category representative to reduce error or uncertainty. Whether the influence of category knowledge occurs at encoding, retrieval, or both is a question for future research.

A third potential explanation for why a slightly great portion of children were best fit by the Noisy Prototype model might be due to the particular information studied. It should be noted that the study values for each category were selected such that they fell one standard deviation above and below the mean of the category (mean and standard deviations learned from the prior knowledge task). Given that children only studied colors that fell in close proximity of the prototypes, this might have propelled learners to rely on their category expectations, that is, adopting the Prototype strategy. Thus, the finding of a large portion of older children who are better fit by the Noisy Prototype model might be a consequence of the study values falling relatively close to the prototype. Future work might explore whether the model fitting results vary when children are presented with colors that substantially deviate from the prototype (i.e., more than 1 sd).

There were a number of limitations in this study that warrant caution in the interpretation of the results. First, the initial goal of Macias et al., (in revision), was to compare children's episodic memory performance to adults. For this purpose, a sample of 33 child participants was sufficient. However, to evaluate individual and age-related differences, a significantly larger sample of participants is needed to achieve strong statistical power for analysis. Second, the goal of this paper was to assess age related differences. Although a median split of children revealed some clear trends in a difference in model fitting by age, a more diverse age sample of children could provide further insight into differences in memory strategy by age. For instance, we anticipated that older children might rely less on the prototype to facilitate recall (although this might interact with the contrary strategy to label as children get older), but it is possible that the sample of children used here did not contain a wide enough age-range to observe this pattern. To this end, a natural future direction would be to collect more data for the purposes of evaluating age differences.

Despite these limitations, this paper demonstrates clear trends in age related differences in model fitting. Furthermore, we hope to have demonstrated that an approach that applies model fits at the individual level can provide insight into how different cognitive strategies (such as labeling) may color recall.

Acknowledgments

This work has received support from the National Science Foundation Graduate Research Fellowship under Grant Number NSF DGE 0937373 (KP), National Institutes of Health, IMSD Minority Biomedical Research Support Program under grant number 2R25GM096161-07 (CM), National Science Foundation CAREER Grant Number 1453276 (PH), NSF SES-1627971 (EB), and the Jacobs Foundation (EB).

References

Davinson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.

Diamond, A. (2006). The early development of executive functions. Lifespan Cognition: Mechanisms of Change. *Lifespan Cognition: Mechanisms of Change*, *210*, 70–95.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2014). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *21*, 2–11.

Duffy, S., Huttenlocher, J., & Crawford, E. L. (2006). Developmental Science. *Children use categories to maximize accuracy in estimations*, *9*, 597–603.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian Data Analysis*. Boca Raton, Florida: Chapman & Hall.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science*, *17*(9), 767–773.

Hemmer, P., & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics in Cognitive Science*, *1*, 189–202.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and Particulars: Prototype effects in establishing spatial location. *Psychological Review*, *98*, 352–376.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why Do Categories Affect Stimulus Judgment ? *Journal of Experimental Psychology*, *129*, 220–241.

Keresztes, A., Ngo, C. T., Lindenberger, W.-B. M., U, & Newcombe, N. S. (2018). Trends in Cognitive Sciences. *Hippocampal maturation drives memory from generalization to specificity*, *22*, 676–686.

Macias, C., Persaud, K., Hemmer, P., & Bonawitz, E. (in revision). Evaluating episodic memory error in preschoolers: Category expectations influence episodic memory for color.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.

Persaud, K., & Hemmer, P. (2014). The Influence of Knowledge and Expectations for Color on Episodic Memory Knowledge and Expectations for Color. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 1162–1167). Quebec City, Canada.

Persaud, K., & Hemmer, P. (2016). The Dynamics of Fidelity over the Time Course of Long-Term Memory. *Cognitive Psychology*, *88*, 1–21.

Steyvers, G.-T., Mark, & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*, 327–334.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–641.