

**What do biased estimates tell us about cognitive processing?
Spatial judgments as proportion estimation**

Alexandra Zax¹, Katherine Williams¹, Andrea L. Patalano¹, Emily Slusser^{1,2}, Sara Cordes³,
and Hilary Barth¹

¹ Wesleyan University, ² San Jose State University, ³ Boston College

July 2019

In press, *Journal of Cognition & Development*

Keywords: spatial cognition, estimation, proportional reasoning, sequential reproduction

Word Count: 9065

Corresponding author:

Hilary Barth
Department of Psychology
Wesleyan University
207 High St
Middletown, CT 06459
hbarth@wesleyan.edu

Author Note: This work was supported in part by NSF DRL-0950252 to HB, by NSF DRL-1561214 to HB, ALP, and SC, and by a Wesleyan University Psychology Department Postdoctoral Fellowship to ES. We thank research assistants Ilona Bass, Ellen Lesser, Sheri Reichelson, Anna Schwab, and Jessica Taggart, and the participating students and families who made this work possible.

Abstract

Similar estimation biases appear in a wide range of quantitative judgments, across many tasks and domains. Often, these biases (those that occur, for example, when adults or children indicate remembered locations of objects in bounded spaces) are believed to provide evidence of Bayesian or rational cognitive processing, and are explained in terms of relatively complex Bayesian models (e.g., CAM; Huttenlocher, Hedges, & Vevea, 2000). Here, we suggest that some of these phenomena may be accounted for instead within a simpler alternative theoretical framework that has previously been found to explain bias in common numerical estimation tasks across development. We report data from university undergraduate students and 7- through 10-year-olds completing a speeded linear position reproduction task. Bias in both adults' and children's responses was effectively explained in terms of a relatively simple psychophysical model of proportion estimation. These data clearly show that the proportion estimation framework is a viable alternative to theories that explain biases as the result of a Bayesian cognitive adjustment process. We also discuss our view that these data are not easily reconciled with the requirements of the more complex Category Adjustment Model that assumes estimates should exhibit a central tendency bias.

What do biased estimates tell us about cognitive processing?
Intuitive proportional reasoning in spatial judgments

When we make quantitative judgments, they tend to be systematically biased, and the resulting patterns of bias are strikingly similar across very different kinds of tasks and cognitive domains. For example, related patterns of bias in quantitative judgments appear for tasks that require participants to judge the lightness of a grey square, the width of a schematic fish, a location on a 2D shape or a real-world object, the duration of a sound, the position of a number on a number line, the proportion of white dots in a mixed array, the frequency of the letter A in a mixed sequence, the proportions depicted by shapes in graphs, the probabilities involved in different gambles, and more (e.g. Barth, Lesser, Taggart, & Slusser, 2015; Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Duffy, Huttenlocher, & Crawford, 2006; Erlick, 1964; Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2000; Huttenlocher, Hedges, & Vevea, 2000; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Newcombe, & Sandberg, 1994; Landy, Guay, & Marghetis, 2017; Nakajima, 1987; Slusser & Barth, 2017; Slusser, Santiago, & Barth, 2013; Spence, 1990; Sullivan, Juhasz, Slattery, & Barth, 2011; Tversky & Kahneman, 1992; Varey, Mellers, & Birnbaum, 1990; Wu & Gonzalez, 1999; Zhang & Maloney, 2012). In general, judgments across these domains typically reveal a systematic overestimation of lower values and underestimation of higher values, resulting in a pattern of bias towards the mean value.

Despite the similarities of the biases that emerge when we make these different kinds of judgments, many different theories have been proposed to explain their origins. Some of these are relatively narrow, constructed to explain a small range of localized phenomena. Others are much broader, aiming to encompass a wide range of behaviors. In some cases, these theories lead to broad conclusions about the nature of cognition based on the observed patterns of bias. For example, a current prominent theme in cognitive science is the idea that perceptual and cognitive systems may operate in a manner that approximates Bayesian reasoning, in that responses may be improved through the combination of different information sources weighted by their reliability. Biases in quantitative judgments are often interpreted as providing direct evidence of this property of cognitive processing. The following sections will first discuss why these kinds of biases are often thought to provide evidence for Bayesian cognitive processes, and will then present a different perspective and accompanying experiments.

Building on early work in the spatial domain (Huttenlocher et al., 1991; Huttenlocher et al., 1994), J. Huttenlocher and her colleagues (Huttenlocher et al., 2000; see also Holden, Curby, Newcombe, & Shipley, 2010; Holden & Newcombe, 2013; Holden, Newcombe, & Shipley, 2013; Holden, Newcombe, Resnick, & Shipley, 2016; Huttenlocher & Lourenco, 2007; Sandberg, Huttenlocher, & Newcombe, 1996) developed a broad theory to explain how and why judgments are systematically biased for stimuli that vary along a continuum. The theory informed the development of the Category Adjustment Model or CAM, which was designed to account for judgments of these stimuli when they are made within a bounded space. An example of a judgment within an explicitly bounded space might be a location within a clearly defined geometric region, but the theory also applies to more implicitly bounded spaces (e.g., a judgment of the lightness of a shade of grey, with implicit boundaries of black and white). To give a brief description of the CAM and data from related literature, we use spatial judgments as an example, but we note that the theory is much broader, relating to the general role of categories in cognition.

One of the many examples of biases that arise in quantitative judgment appears when observers are asked to reproduce remembered locations within bounded linear regions by identifying a previously observed position in a rectangular space, or by attempting to retrieve a small object recently buried in a long thin rectangular sandbox (Huttenlocher et al., 1994). This work mapped out clear patterns of systematic bias in participants' spatial position estimates, along with characteristic developmental changes. Broadly speaking, younger children's estimates (assessed at the group level) are displaced toward the center of the entire rectangular space. Older children and adults, on the other hand, show a different pattern in which estimates for locations within each half of the rectangular space tend to be displaced toward the center of that half. The degree of displacement varies somewhat in different regions of the space.

The CAM explains the observed patterns of estimation bias in terms of a Bayesian combination process in which bias results from the hierarchical coding of information about a stimulus¹. That is, for a spatial location, the location is hierarchically coded at two levels: an inexact but unbiased fine-grained location, and a spatial category associated with the location. The category-level knowledge serves as prior information, combining in a weighted fashion with the remembered (unbiased, but imprecise) stimulus values. Bias is introduced because estimates are adjusted away from the remembered value toward the category prototype (for a spatial category, usually the center of mass of the category), resulting in a central tendency effect. The Category Adjustment Model predicts that estimates will generally be adjusted toward the prototypical center of a spatial category, with the degree and direction of adjustment varying with the number of categories into which the observer subdivides a space. For example, observers who treat a space as a single category generally adjust estimates toward the center of the entire space (the prototypical center of the spatial category). This corresponds to the pattern of bias observed in younger children's spatial position judgments in the task described above. Alternatively, for observers who subdivide the space into two halves, estimates within each half should generally be adjusted toward the center of that half. This corresponds to the pattern of bias observed for older children and adults.

The weighting of information sources is determined by their relative reliability, in this theory, with more variable sources weighted less. For example, when a memory of a specific stimulus value is less reliable (e.g., because it has been held in memory longer, leading to memory decay and greater variability), then category-level information should be weighted more heavily, leading to more adjustment toward the category prototype and therefore greater estimation bias (a larger central tendency effect). This process is considered rational (for discussion, see Bowers & Davis, 2012; Griffiths, Chater, Norris, & Pouget, 2012; Holden & Newcombe, 2013; Jones & Love, 2011) because, although it introduces bias, it also reduces variability and increases overall accuracy across trials (Duffy et al., 2006; Huttenlocher et al., 2000). Thus, this literature argues that children and adults use category information to improve estimation accuracy across many domains, in a manner that appears to approximate a Bayesian process. From this perspective, observations of bias in quantity judgments that turn up across many tasks should be interpreted as explicit evidence of this characteristic of cognition.

Studies exploring the CAM have covered a wide range of topics, from judgments involving geometric categories to those involving inductive categories. It has been applied across many domains and developmental periods. For example, adults' judgments of physical locations

¹ A different approach to explaining similar data is the dynamic field theory (e.g. Spencer, Austin, & Schutte, 2012; Schutte, Spencer, & Schoner, 2003).

in the real world (Holden, Newcombe, & Shipley, 2013) or locations within a photograph (Holden, Curby, Newcombe, & Shipley, 2010) and young children's reproductions of the sizes of simple fish-shaped stimuli (Duffy et al., 2006) have all been reported to lead to data broadly consistent with the Category Adjustment Model (Duffy et al., 2006). Recent work on category adjustment has further explored potential intriguing links to STEM knowledge, finding that expert geologists remember spatial locations in geologically-relevant scenes differently from experts in other fields (Holden et al., 2016). Not only does it have remarkable breadth, but the CAM was also ahead of its time when introduced: it anticipated a much later surge of interest in closely related Bayesian models of context dependence in quantitative judgments (e.g. Aagten-Murphy, Cappagli, & Burr, 2014; Cicchini, Arreggi, Cecchetti, Giusti, & Burr, 2012; Jazayeri & Shadlen, 2010; Sciutti, Burr, Saracco, Sandini, & Gori, 2015; see also Petzschnner, Glasauer, & Stephan, 2015; Shi et al., 2013). In these models, the context in which a stimulus appears plays a role similar to that played by the category in the CAM.

Should we take bias in quantity judgments as evidence of Bayesian cognition? Is the CAM a good theoretical account of the cognitive processes that yield quantity judgments? Despite the large literature presenting patterns of bias that are broadly consistent with the CAM (cf. Friedman, Montello, & Burte, 2012; Sailor & Antoine, 2005; Duffy & Smith, 2017), there is also evidence that it may not be able to accommodate certain reliable patterns in the observed data. For example, the CAM is fundamentally a description of a central tendency mechanism, yet some quantitative judgments do not appear to exhibit the necessary kind of central tendency. In some tasks, patterns of spatial bias are directed *away* from category centers or prototypes in a manner that appears inconsistent with the category adjustment model. These findings may be most visible at the more rarely examined individual level, but in some cases this pattern is clear in group data as well (Barth et al., 2015; Crawford, Landy, & Presson, 2014; see also Crawford & Duffy, 2010; Crawford, Landy, & Salthouse, 2016; Sampaio & Wang, 2017).

In previous work, Barth et al. (2015) proposed that many tasks thought to provide evidence for the Category Adjustment Model could be productively conceptualized instead as proportion judgments (situations in which participants estimate the magnitude of some part within a bounded range) and that such judgments in themselves might be expected to produce the observed patterns of curvature (see also Hollands & Dyre, 2000). As in the CAM, the range in question need not be explicit and spatially delineated but might rather be more conceptually and/or implicitly bounded (as in the case of judgments of the lightness of a grey object within an implied range from black to white). One widely observed phenomenon is that estimates of proportions and frequencies tend to be biased such that smaller proportions are overestimated and larger ones are underestimated (see Hollands & Dyre, 2000; Zhang & Maloney, 2012, for reviews). Moreover, quantitative models of proportion estimation have been developed to describe exactly this kind of relative judgment (e.g., Spence, 1990; Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002). These relatively simple models have been shown to account for a wide range of situations that call for perceptual proportion judgments (Hollands & Dyre, 2000; Hollands et al., 2002). A converging body of work shows that the systematic patterns of estimation bias predicted by these models are also observed in tasks that require proportion judgments of abstract quantities, such as number line estimation (e.g. Barth & Paladino, 2011; Barth et al., 2016; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Dackermann, Kroemer, Nuerk, Moeller, & Huber, 2018; Link, Huber, Nuerk, & Moeller, 2014; Peeters et al., 2016; Peeters, Verschaffel, & Luwel, 2017; Rouder & Geary, 2014; Slusser et al., 2013; Slusser & Barth, 2017; Xing et al., under revision; Zax, Slusser, & Barth, 2019).

The patterns of bias predicted by models of proportion estimation are qualitatively very similar to those arising from tasks used to explore the Category Adjustment Model, and the relevant task structures are also highly compatible with these models. Therefore it may be that theoretical explanations of proportion estimation are able to describe and explain the data. It is also possible that these models will accommodate the reported lack of central tendency in some datasets that violate the predictions of the CAM, perhaps providing a more parsimonious and potentially unifying explanation of a wide range of phenomena. That is, previous estimation findings taken as evidence of Bayesian cue combination may be better explained by the simpler proportion judgment framework.

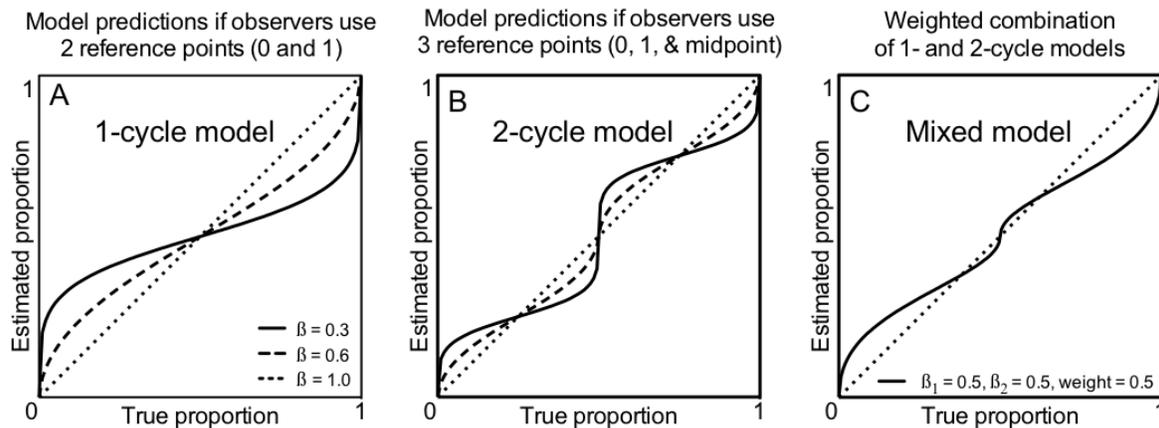


Figure 1. An overview of the proportion estimation framework. X -axes refer to true physical values, y -axes refer to estimates, and β is a parameter representing estimation bias ($\beta=1$ corresponds to no bias). **A.** Predicted estimation patterns if estimates are made within a bounded space relative to the two endpoints only (a 1-cycle model), shown for three possible values of β . **B.** Predicted estimation patterns when estimates are made within a bounded space relative to the two endpoints plus a middle reference point (a 2-cycle model), shown for three possible values of β . **C.** Predicted estimation patterns if a combination of strategies were used (e.g. if a middle reference point were used on some but not all trials, or by some but not all participants). The mixed model is a weighted combination of the pure 1- and 2-cycle models. Separate β parameters are associated with the 1-cycle and 2-cycle components of the mixed model, and a weighting parameter (ranging between 0 and 1) determines the relative contributions of each component such that values nearer to 1 reflect a larger contribution of the 1-cycle component. (These plots show example values of $\beta \leq 1$; an inverse pattern of over- and underestimation results if $\beta > 1$.)

Previous work (Barth et al., 2015) tested this hypothesis by asking whether a theoretical framework based on proportion estimation could explain performance in a simple spatial position reproduction task administered to adults and older school-age children (adapted from the work of Huttenlocher & colleagues). On each trial, a horizontal line appeared briefly on a screen, marked with a small vertical hatch to indicate a target position. After a short delay, an unmarked line appeared (at varying locations on the computer screen); the participant clicked on the line to reproduce the target location. Quantitative models of proportion estimation (Hollands & Dyre, 2000) were fitted to participants' position judgments. Figure 1 gives an overview of these models (see Hollands & Dyre, 2000, for a lengthier explanation; see also Slusser et al., 2013, for a fuller description of this theoretical framework and its application to a different family of tasks). Conceptually, the idea is that if there is bias associated with the absolute estimation of a magnitude, relative judgments should also incorporate the bias associated with the estimation of

each part². Estimates of single magnitudes in tasks that do not restrict estimates to fall within a bounded range (“unbounded” judgments) for many continua are often biased and well described by power functions. As a result, the estimation of each part magnitude in a relative, “bounded” judgment individually contributes to bias. The result is a proportional estimate with a characteristic S-shaped or reverse S-shaped curve arising from the combination of power functions (a “one-cycle” model, described by $y = x^\beta / (x^\beta + (1 - x)^\beta)$; Figure 1A; see Spence, 1990; Hollands & Dyre, 2000, for proof).

Hollands & Dyre (2000) generalized this idea to accommodate additional reference points (for example, using an imagined midpoint to partition the bounded space into two halves while making estimates). This strategic use of additional reference points leads to the repetition of the S-shaped or reverse S-shaped curve between each pair of reference points employed (two-cycle model; e.g. Figure 1B). Accuracy is better when more reference points are used (compare Figures 1A and 1B) and/or when there is less bias overall (when β is closer to 1). A mixed model was also introduced (Figure 1C); this is a weighted combination of the one-cycle and two-cycle pure models (Hollands & Dyre, 2000, see Eq. 9). This mixed model has three parameters: a bias (β) parameter for each component and a weighting parameter that determines the relative contribution of each component. The weighting parameter ranges between 0 and 1, with values closer to 1 reflecting a greater contribution of the one-cycle component. This mixed model reflects the likely possibility that multiple strategies might be used across participants or trials (i.e., some participants might consistently use middle reference points while others might consistently use none, or individuals might shift strategies within a session); it can account for asymmetries in estimation patterns that the simpler, symmetrical one- and two-cycle models cannot.

When these models were applied to older children’s and adults’ position judgments, two main findings emerged (Barth et al., 2015). First, the proportion judgment models – which require fewer free parameters than the CAM – explained the data well for both adults and children. The pure one-cycle model successfully captured the general pattern of bias with just one free parameter, and the three-parameter mixed model described above resulted in an even better explanation of the group data and of the majority of individual participants. Second, some individual data violated a key prediction of the CAM: that adjustment leads to a central tendency effect. Instead, for many individuals, numerous estimates were displaced away from the overall center of the line. In a few cases, nearly all estimates were displaced away from the center. This finding of bias *away* from category centers is not predicted by the Category Adjustment Model, whether a participant treats the space as a single category or subdivides it into two³. The proportional models, on the other hand, can take on either S-shapes or inverse S-shapes, modeling bias toward or away from category centers and accommodating the individual differences seen in the preliminary data through differing values of the bias parameter and strategic choice of reference points.

² This does not necessarily require formal proportional reasoning with exact values, but rather the intuitive cognitive processing of the ratio relationships of approximate quantities that is available even to infants (McCrink & Wynn, 2007) and young children prior to formal training (Barth, Baron, Carey, & Spelke, 2009; Boyer, Levine, & Huttenlocher, 2008; Jeong, Levine, & Huttenlocher, 2011; McCrink & Spelke, 2010; McCrink, Shafto, & Barth, 2017).

³ Previous researchers (Crawford, Landy, & Presson, 2014) also noted that some prior findings of bias away from category centers are inconsistent with the Category Adjustment Model’s posited bias towards prototype values but offered a different explanation (see also dynamic field theory, e.g. Spencer et al., 2012; see also Sailor & Antoine, 2005

Some available data, therefore, do suggest that a large body of evidence used to argue for Bayesian combination in cognitive processes of estimation may be (1) explained well by a simpler model and (2) potentially incompatible with broad principles of the relevant Bayesian models. The present experiments address three main questions. First, can we replicate the finding that models of proportion estimation explain adults' spatial position estimates? To answer this question, we administered the same simple task to a new larger sample of adults (Experiment 1). Second, do these models explain children's spatial judgments earlier in development? Children aged 9-10 were tested previously (Barth et al., 2015), but this is a limited age range for tracking potential change over development, and indeed the age 9-10 group performed much like adults when given the same task presented at a slightly reduced speed. In the present study, we aimed to determine whether the proportion estimation framework can explain younger children's performance on the same computer-based spatial location reproduction task, and to begin to map out developmental change with respect to this theoretical framework (Experiment 2). Third, prior individual-level data were not necessarily compatible with the CAM: a central tendency effect did not always arise, and the data were sometimes even biased in the opposite direction. We asked whether this was the case in both Experiments 1 and 2 in order to evaluate whether children's and adults' estimates were compatible with the requirements of Bayesian central tendency mechanisms.

Experiment 1

Here we attempt to we replicate prior findings showing that models of proportion estimation can explain adults' spatial position judgments by administering a spatial reproduction task to a new sample of adults. We also address the question of whether estimates exhibit clear central tendency effects.

Method

Participants

Participants were 91 undergraduate students (23 male, 68 female; 18 – 22 years old) who received introductory psychology course credit for participation. Five participants (3 female, 2 male) were excluded for noncompletion, leaving 86 total contributing data.

Stimuli

Stimuli were created using MATLAB, and displayed in a different pseudorandom order for each participant. For each trial, a centered fixation rectangle (grey, 12.3 cm x 0.7 cm) was immediately followed by a stimulus screen and a response screen. Fixation rectangles and stimulus screens each lasted 500ms, and response screens were displayed for 1500ms. The stimulus screen displayed a centered 12.3 cm horizontal line enclosed by short vertical lines at each end, extending 0.3 cm above and below the horizontal line, and with a vertical hatch mark at a target position along the line. The response screen presented a non-centered pseudorandomly located line of identical size and orientation, with no target hatch mark. For coding purposes, the endpoints of the line were assigned values of 0 and 1000 for the purpose of identifying target locations along the line (these numbers were not presented or known to the participants). Target values to be estimated were sampled at intervals of approximately 50 units (a uniform distribution), with the presented values jittered such that, for example, positions corresponding to "47" and "51" were presented rather than two instances of "50". Target positions corresponded to the numbers 47, 51, 98, 102, 147, 153, 199, 202, 249, 252, 298, 302, 349, 351, 398, 403, 449,

453, 499, 502, 547, 552, 597, 601, 647, 652, 699, 703, 747, 753, 798, 802, 848, 853, 899, 901, 949, and 953, for a total of 38 distinct target positions.

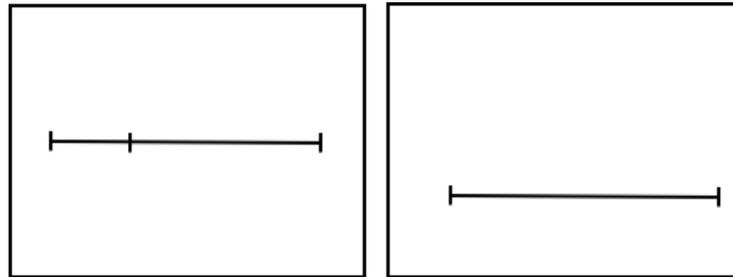


Figure 2. Schematic depiction of a stimulus screen (left) and a response screen (right).

Design

Participants completed the spatial estimation task and a number line estimation task (part of a separate study, not discussed further). Two number-line (N) trial blocks and two spatial estimation (S) trial blocks were completed in NSSN or SNNS order. There were 38 trials per block, with each target value represented once per block (twice total).

Procedure

Participants were seated in front of a computer, close enough to use a mouse, with blank paper covering the keyboard and top of the screen intended to obscure landmarks. Some participants used an HP ProBook 14" laptop with an external mouse to complete the study, and some used an HP Compaq desktop with a Dell 15" monitor. They were given written and spoken instructions for the first block. Two practice trials preceded the first block. Practice trials were identical to experimental trials, using target values drawn randomly from the list of experimental targets. Different practice trial values were presented to each participant and no feedback was given. For each trial, the task was to move the cursor and click the appropriate position (to match the previously presented target spatial location) on the blank horizontal line during the response screen. Mouseclicks were recorded as numbers from 0-1000, corresponding to locations along the response line. A 1000ms pause separated trials.

Results and Discussion

We compared participants' responses to the true locations of target positions. A participant's estimate for a target location was removed as an outlier if it differed from the group mean for that target location by > 2 SDs (4.79% of trials). We calculated percent absolute error (PAE) by dividing the absolute difference between the number corresponding to the participant's estimated position and the number corresponding to the presented position by the numerical range, then multiplying the quotient by 100 to yield a percentage. Adults were highly accurate on this task, with average PAE across individuals = 2.92%, similar to previous findings (Barth et al., 2015).

Adult participants' placements did exhibit systematic bias, despite their high accuracy. Figure 3 depicts group median estimates (left) and bias in group medians (right). Models of proportion estimation (see Figure 1) were fitted to the group medians and to individual data. We fitted one- and two-cycle versions of the model (Figure 1A, 1B) and also the mixed model (Figure 1C). The mixed model provided the best explanation of the adult group median data (see

Figure 3), yielding lower AICc scores ($\Delta \text{AICc} = 50.06$)⁴. This suggests that multiple strategies were used to perform the task, with variation either across individuals (perhaps with some individuals settling on a pure two-endpoint strategy and others on a pure endpoints-plus-midpoint strategy) or across trials (with each individual using multiple strategies). At the individual level, the mixed model was preferred for 67/86 participants (78%, similar to 79% reported in Barth et al., 2015). The pure one-cycle model was preferred for 5/86 (6%) of individual adults and the pure two-cycle for 14/86 (16%). (See Appendix A for individual fits).

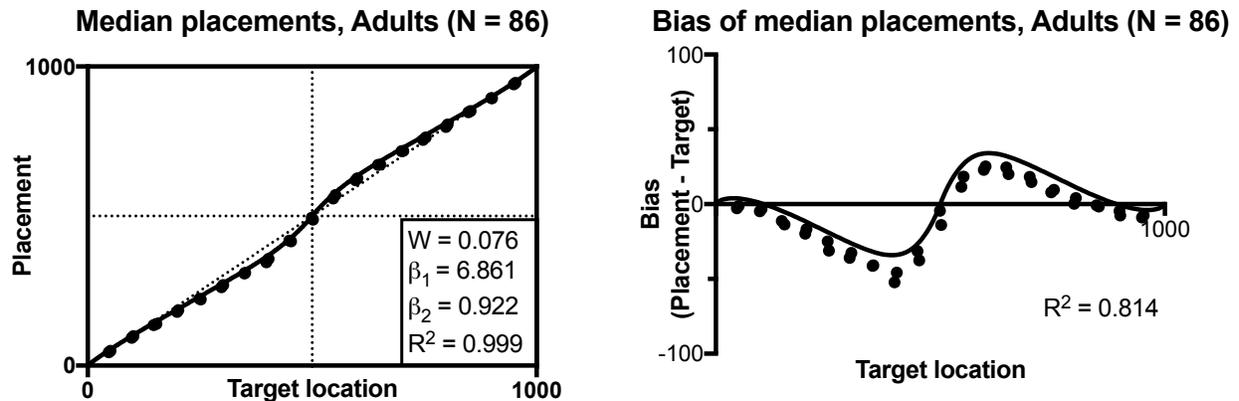


Figure 3. Findings from Experiment 1. Adults' median placements in the position reproduction task (left) and bias in adults' median placements (right), computed by subtracting actual locations from the corresponding placements. In the bias plot, placements too far to the left fall below $y = 0$, and placements too far to the right appear above $y = 0$. The best fit of the three-parameter mixed proportion estimation model is shown with both plots. Separate β parameters are associated with the one-cycle and two-cycle components of the mixed model, and a weighting parameter (ranging between 0 and 1) determines the relative contributions of each component such that values nearer to 1 reflect a larger contribution of the 1-cycle component.

These results show that the proportion estimation framework can indeed provide a good explanation of adults' spatial judgments in this task, consistent with prior findings (Barth et al., 2015). They further suggest that individual adults may have used a combination of reference point strategies, shifting from trial to trial. The fact that this mixed model, a combination of the pure one- and two-cycle models, fit the data so well suggests that participants did overwhelmingly rely on some combination of endpoints and midpoint in making their judgments. If participants had also incorporated quartile reference points, for example, this particular mixed model would not have been expected to accommodate the data as well.

In addition to determining whether proportion estimation models could explain adults' estimates, we aimed to test the possibility that behavior in this task, especially when examined at

⁴ AICc (Akaike Information Criterion, adjusted for small sample sizes), provides a measure of how well different models can explain data while taking varying numbers of parameters into account; Burnham, Anderson, & Huyvaert, 2011; Burnham & Anderson, 2002). ΔAICc refers to the difference in AICc values between a nonpreferred model and the preferred model, which always has the lowest score. Burnham and Anderson (2002) proposed the following interpretation guidelines: "As a rough rule of thumb, models having a Δ within 1-2 of the [preferred] model have substantial support and should receive considerations in making inferences. Models having Δ within about 4-7 of the [preferred] model have considerably less support, while models with $\Delta > 10$ have either essentially no support and might be omitted from further consideration or at least fail to explain some substantial structural variation in the data." (p. 446).

the level of individual trials, might violate the predictions of the Category Adjustment Model. Specifically, we asked whether estimates might exhibit bias away from the direction that would be expected if a central tendency mechanism were at work. To ask whether estimates were not consistent with central tendency bias, we first defined what should count as such a violation. We started with the assumption, consistent with both the CAM and our theoretical framework, that most participants would either treat the space as a single unpartitioned region, or partition the space into two halves – or else they would use a mixture of these two strategies. In the first case, treating the space as a single unpartitioned region, CAM predicts that estimates should generally be shifted toward the center of that single region. It does not necessarily predict that 100% of estimates should be shifted in this manner: CAM does allow for some bias in the opposite direction near the boundaries of a range (see Huttenlocher et al., 2000, for details). But the majority of estimates should certainly show a shift toward the overall center. In the second case, that of subdivision into two spatial categories, the CAM predicts that estimates would be biased toward the center of each half of the space, with values closest to the endpoints of the overall space biased towards the overall center and values closer to the middle biased away from the overall center. As a result, about half the total estimates would (coincidentally) end up showing displacement toward the *overall* center of the entire space, and about half would (coincidentally) show displacement away from the overall center of the entire space.

Given the above predictions, we reasoned that in order to be broadly consistent with the CAM's predictions, about 50%-100% of participants' estimates should be displaced *toward* the overall center of the space. We also reasoned that the opposite pattern of about 50%-100% of estimates displaced *away from* the overall center should be considered incompatible with the CAM's predictions. We loosened this requirement further to accommodate the aforementioned allowable bias away from the center at the boundaries of the range, assuming that the percentage of estimates displaced away from the overall center might rise above 50% while still being considered permissible by the CAM. We therefore chose 60% (of individual estimates displaced away from the overall center) as a criterion to explore. (A schematic view of what one possible version of these patterns might look like is presented in Figure 4. We note that this is not the only possible way to characterize a lack of central tendency bias in placement patterns.)

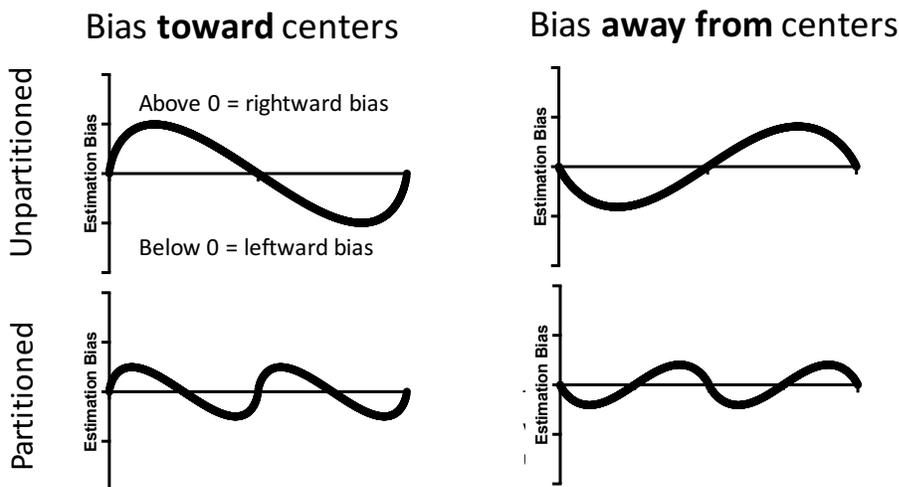
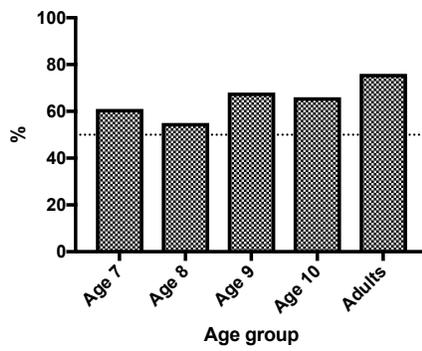


Figure 4. Schematic, approximate depiction of rough patterns of bias that would be compatible with the CAM's predictions (central tendency bias, on left) vs. patterns that would be incompatible with the CAM (bias away from central tendency, right). Top row depicts possible estimation patterns if observers treat the space as a single,

unpartitioned region; bottom row depicts possible estimation patterns if observers mentally partition the space into halves.

As would be expected based on the adult group median data shown in Figure 3, the majority of estimates at the group median level were displaced away from the overall center. A large percentage of adults' individual estimates were also biased away from the overall center: when we analyzed the data to determine how many participants met a criterion of 60% of individual estimates displaced away from the overall center, we found that the majority of adult participants did so (over 60%; see Figure 5). Over 40% of adults met a stricter criterion of 65% of individual estimates displaced away from the overall center, and close to 20% of adults even met a criterion of 75% displaced away from the overall center. In our view, this finding is difficult to reconcile with a cognitive process that is theoretically focused on central tendency bias.

Group medians: % estimates biased away from overall center



How many individual estimates were biased away from center?

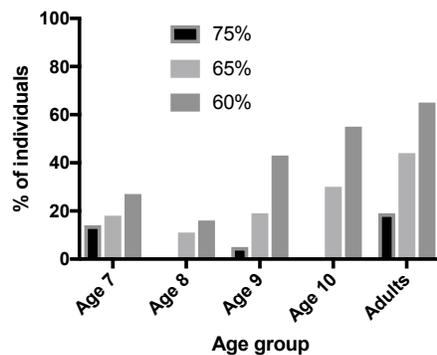


Figure 5. Results of tests for violations of central tendency. The top panel shows the percentage of group median estimates biased away from the overall center of the space for adults (Experiment 1) and children (Experiment 2). For every age group, over 50% of estimates were biased away from the overall center at the group median level. The bottom panel shows findings from individual estimates, giving the percentages of participants in each age group who met different thresholds for percentage of trials exhibiting bias away from the overall center. For each age group, the percentage of individuals is shown for whom 60%, 65%, and 75% of estimates were biased away from the overall center.

Overall, these results show that the proportion estimation framework can provide a good explanation of adults' judgments in this spatial position reproduction task. These results also suggest that adults' judgments are not necessarily compatible with the idea that they are biased

as a result of adjustment toward the center of spatial categories. Experiment 2 addresses the same questions in children.

Experiment 2

Experiment 2 was designed to determine whether the proportion estimation models applied in Experiment 1 to adults' spatial judgments can also explain younger children's performance on the same task, and to begin to map out developmental change with respect to this theoretical framework. We also aimed to test potential violations of the predictions of a central tendency mechanism in children's spatial judgments, using the same methods applied to adults' judgments in Experiment 1.

A broad developmental prediction consistent with both the CAM and the proportion judgment model is that children will not partition the space earlier in development, and later will begin to partition the space, treating it as two categories. The age at which this change typically occurs depends on the particular task. For example, Huttenlocher et al. (1994) reported evidence of subdivision in 10-year-olds but not 6-year-olds in their "sandbox" task, although a similar task using a rectangle drawn on paper yielded evidence of subdivision even in 4-year-olds. Therefore, it would be reasonable to predict a change from one-cycle to two-cycle patterns at some point in development, but not necessarily within in the age range tested here. Another reasonable prediction is that children's judgments will be more similar to adults' as they get older. The simplest interpretation of the mixed model, as initially introduced by Hollands & Dyre (2000), would not lead to a prediction that it should be the most sophisticated of the three models considered here (expected to appear in the oldest participants). Yet adults clearly made judgments best described by the mixed model, both at the group median level and for approximately 75% of individuals; the same is true for children aged 9-10 (Barth et al., 2015). We return to this issue in the General Discussion. In Experiment 2, we ask whether younger children differ from adults and older children, and if so how.

Method

Participants

Participants were 82 children aged 7 through 10 years, recruited from a database of local families. Testing took place in a quiet laboratory room. Three 7-year-olds were excluded from further analysis because their placements were uncorrelated with target locations. Participants contributing data included 19 7-year-olds (age range 7;0 – 7;11, mean age 7;5, 9 females, 10 males), 19 8-year-olds (age range 8;0 – 8;11, mean age 8;3, 10 females, 9 males), 20 9-year-olds (age range 9;0 – 9;11, mean age 9;4, 11 females, 9 males), and 20 10-year-olds (age range 10;0 – 10;10, mean age 10;5, 10 females, 10 males). Data from these 9- and 10-year-olds were previously reported as a single combined group in prior work (Barth et al., 2015) and are re-analyzed here separately by age, along with new data from younger children.

Stimuli

Stimuli were the same as those used for adults in Experiment 1, except that for children, fixation rectangles and stimulus screens each lasted 750ms and response screens were displayed for 2250ms.

Design

Experiment 2 used the same design as Experiment 1 except that children completed 19 trials per block (such that each child completed a total of one trial per unique target value), with blocks in NNSS or SSNN order to reduce the number of potentially confusing task switches.

Procedure

The procedure was the same as that for adults except for the following alterations. Along with their instructions, children also saw a picture of a blank response line on paper to demonstrate what the response screen would look like (no response was made and no feedback was given). Four practice trials preceded the first block, and a 1500ms pause separated trials. The experimenter also had the option to pause the game in case of distraction.

Results and Discussion

We compared children's responses to the true locations of target positions. As for the adults in Experiment 1, an estimate for a target location was removed as an outlier if it differed from the group mean for that target location by > 2 SDs (5.13% of trials for 7-year-olds, 5.24% for 8-year-olds, 5.10% for 9-year-olds, and 4.29% for 10-year-olds). General accuracy for each age group was again reported in the form of PAE, with an average PAE of 5.17% for 7-year-olds, 4.58% for 8-year-olds, 3.53% for 9-year-olds, and 3.58% for 10-year-olds.

Children's placements were quite accurate, but like adults' they exhibited systematic bias. Figure 6 depicts group median estimates for each age group (left) and bias in group medians for each age group (right). Models of proportion estimation were fitted to the group median and individual data for all four age groups. We again fitted one- and two-cycle versions of the model (Figure 1A, 1B) and also the mixed model (Figure 1C). The mixed model provided a strongly preferred explanation of the group median data for 9- and 10-year-olds (see Figure 6). The mixed model was moderately preferred for 7-year-olds at the group median level, and the 8-year-olds' placements provided the same level of support for the mixed model and pure 2-cycle model (ΔAICc for mixed model vs. preferred pure model: 7-year-olds = 5.55, 8-year-olds = 0.13, 9-year-olds = 34.4, 10-year-olds = 68.4). (See Appendix B for individual fit information.) Figure 7 depicts the number of children in each age group with each preferred model. By age 10, 80%

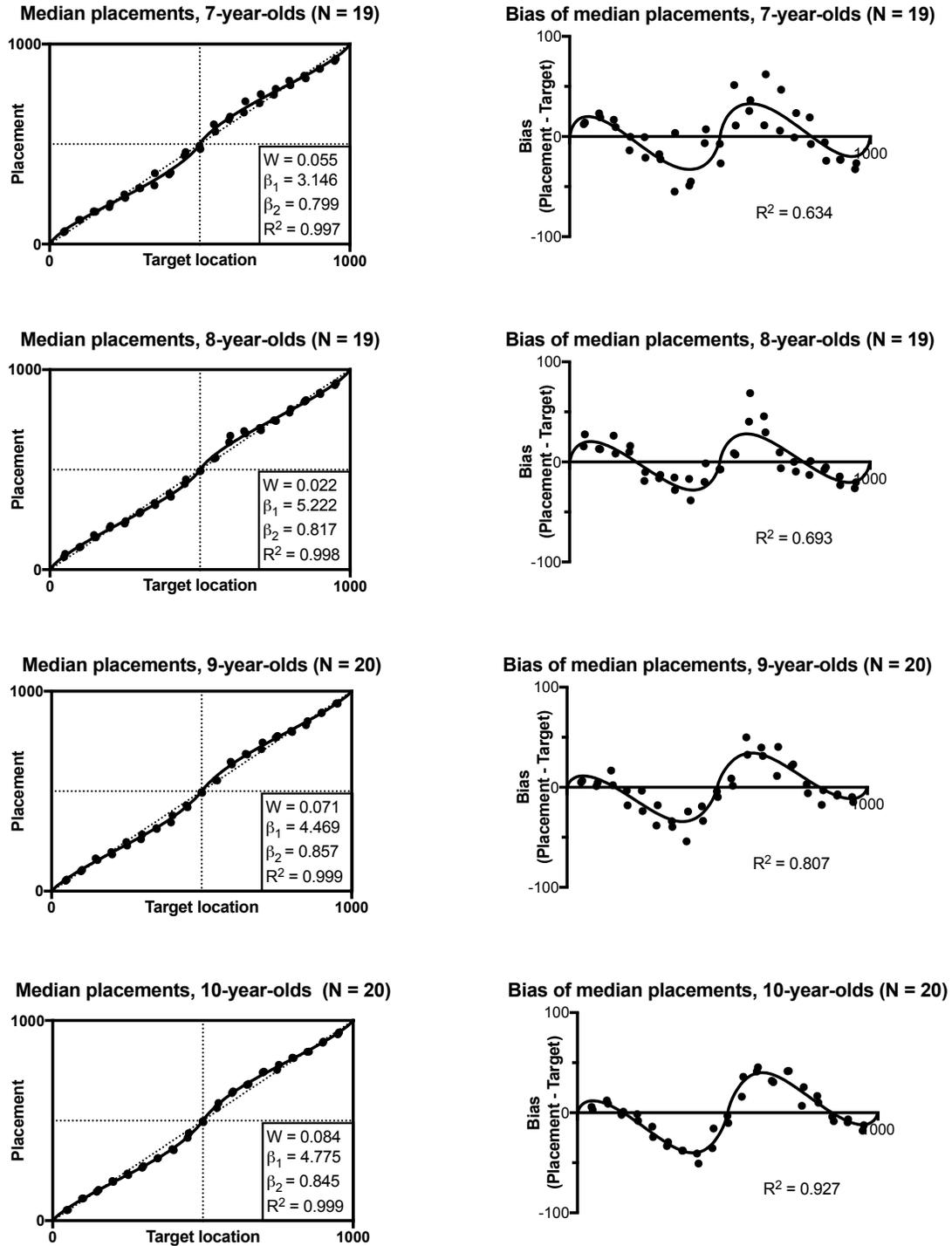


Figure 6. Findings from Experiment 2. Children’s median placements (left) and bias in their median placements (right), computed by subtracting actual locations from the corresponding placements. In the bias plot, placements too far to the left fall below $y = 0$, and placements too far to the right appear above $y = 0$. The best fit of the three-parameter mixed proportion estimation model is shown with both plots. Separate β parameters are associated with the one-cycle and two-cycle components of the mixed model, and a weighting parameter (ranging between 0 and 1) determines the relative contributions of each component such that values nearer to 1 reflect a larger contribution of the one-cycle component.

of individual children produced placements that were explained best by the mixed model, compared to 78% of adults in Experiment 1. We return to the question of what these developmental patterns might mean in the General Discussion.

These results show that the proportion estimation framework provides a good explanation of children's spatial judgments in this task, as it does for adults. They further suggest that older children may have used a combination of different reference point strategies, shifting from trial to trial, as adults appear to do, but that younger children may have been less likely to do so. We tested the possibility that children's estimates, like adults', might violate the predictions of the Category Adjustment model, exhibiting considerable bias away from the center. The majority of

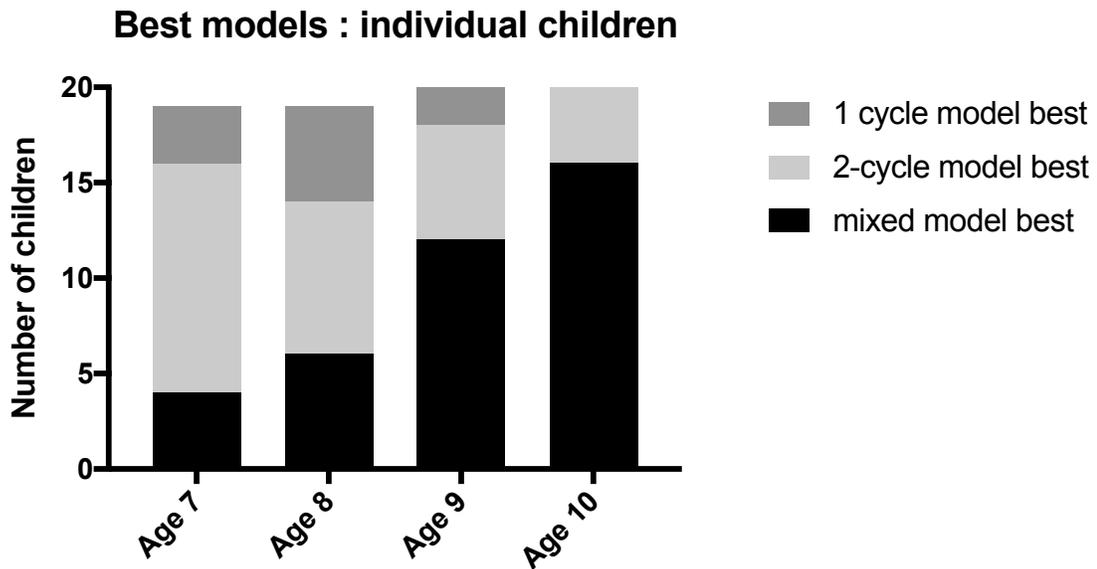


Figure 7. Experiment 2 preferred models for individual children in each age group.

estimates at the group median level for every age group were displaced away from the overall center. We analyzed individual data to determine how many participants met various criteria for the percentage of individual estimates displaced away from the overall center. A considerable number of children met a threshold of producing 60% or more of individual estimates that were biased away from the overall center, with a developmental trend toward producing more such estimates with age (see Figure 5 for breakdown by age). Again, we suggest that this result is difficult to reconcile with theories that describe a central tendency mechanism.

General discussion

Quantitative judgments exhibit systematic and remarkably similar biases across a wide range of tasks and domains. Despite this similarity, many different theories have been proposed to account for them. Often, biases in estimates are interpreted as providing evidence for broad theories about the nature of cognitive processing, such as the idea that perceptual and cognitive systems may carry out (or approximate) Bayesian processes in that they may improve or optimize responses by combining different information sources weighted by relative reliability. The Category Adjustment Model (CAM; e.g. Huttenlocher et al., 2000) is a prominent example of such a theoretical approach.

Some prior findings suggest that the large body of evidence used to argue for Bayesian combination in cognitive processes of estimation might instead be explained by a considerably simpler theoretical framework based on quantitative models of proportion estimation (e.g., Barth et al., 2015). Moreover, some existing data appear incompatible with key components of the posited Bayesian cognitive processes: namely, central tendency bias. The present work addressed three main aims in two experiments, focusing on the domain of simple spatial judgments as an example. First, we aimed in Experiment 1 to replicate the finding that models of proportion estimation can successfully explain adults' spatial position estimates, in a new sample of adults. Second, because previous work suggesting that these models can explain children's spatial judgments had only looked at 9- to 10-year-olds (who performed similarly to adults), we aimed in Experiment 2 to extend this investigation to a younger developmental range, and to identify potential patterns of developmental change. Third, in both experiments we aimed to test whether participants' estimates were compatible with the requirements of Bayesian central tendency mechanisms.

We applied the theoretical framework based on proportion estimation (Hollands & Dyre, 2000; Slusser et al., 2013), analyzing both group- and individual-level data with respect to the quantitative models described above (Figure 1). Specifically, we fitted the one-cycle and two-cycle proportion judgment models and the mixed model (a weighted combination of the two "pure" models) to group median and individual data. We found that these models could quantitatively explain the data very well for both children and adults. These findings show that, at minimum, this theoretical framework should be considered a viable alternative to theories that explain biases as the result of a Bayesian cognitive adjustment process, such as the Category Adjustment Model. The key take-home point is that the biases that arise in this family of tasks should not necessarily be interpreted as constituting evidence of Bayesian cue combination.

In addition to addressing whether the relatively simpler proportional reasoning framework would provide a good quantitative explanation of the data (in addition to its readily apparent qualitative fit), we sought to determine whether the data provided evidence *against* models such as the CAM. We did so by testing the degree to which individual estimates violated the predictions of the CAM in the sense that it is fundamentally a description of a mechanism that creates central tendency effects. Consistent with previous findings (Barth et al., 2015), participants' estimates were difficult to reconcile with the idea of a central tendency mechanism: numerous estimates were biased away from the predicted directions (see also Crawford, Landy, & Salthouse, 2016). This was true at the individual level – in some cases, nearly all estimates were displaced in the "wrong" direction) – and was also visible at the group median level.

Although we believe the direction of bias observed in our participants' estimates is not compatible with what would be expected if a CAM-style mechanism were at work, it would certainly be possible to accommodate data like these that show outward bias (rather than central tendency bias) with a theory preserving the spirit, if not the letter, of the CAM. For example, the category prototypes toward which estimates are adjusted in the CAM are usually understood to be the centers of mass of geometric regions, but this is not absolutely required. One could develop an explanation in which category prototypes are located at the outer edges of the space, or beyond the space, to accommodate outward bias (see also Crawford et al., 2016; Crawford et al., 2014). Such an explanation would become more compelling to the degree that it led to specific, testable *a priori* predictions about the cases in which category adjustment processes should lead to outward bias rather than the inward central tendency bias that has been the focus

of nearly all CAM-related literature to date. The appeal of this type of approach to researchers will likely vary depending on theoretical preference.

From our perspective, given the focus on explaining patterns of central tendency bias in CAM and related literatures (e.g. Cicchini et al., 2014; Sciutti et al., 2015), the existence of replicable data that do not conform to a central tendency pattern substantially decreases the appeal of this family of theories. Moreover, a strong focus on one family of explanations or one key targeted characteristic (such as Bayesian mechanisms and/or optimality) may lead researchers to underreport or underemphasize findings that don't fit (Rahnev & Denison, 2018; Barth, Cordes, & Patalano, 2018). There is a clear case to be made, therefore, for attending to and following up on what "doesn't fit" and for exploring explanations based in multiple theoretical frameworks.

The proportion estimation framework provides an attractive quantitative explanation of the data from this family of spatial tasks, coupled with a sensible conceptual match between the type of task it is designed to model (situating a magnitude within a bounded range) and the structure of the tasks that comprise the CAM literature. That is, the proportion judgment explanation yields good quantitative fits while making good sense. The CAM has the appeal of good sense also, but in our view the data do not conform to the CAM's requirements as they are usually presented in the published literature (where the CAM is generally understood to function as a central tendency mechanism). The proportional framework also has the appeal of relative simplicity when compared to the CAM, requiring fewer free parameters.

Like the CAM, the proportion estimation framework also has the advantage of generality. It has emerged as a powerful explanation of the bias that characterizes performance and developmental change in similarly structured abstract magnitude judgments such as bounded number line estimation (e.g., Barth & Paladino, 2011; Barth et al., 2016; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Dackermann et al., 2018; Link et al., 2014; Peeters et al., 2016; Peeters, et al., 2017; Rouder & Geary, 2014; Slusser et al., 2013; Slusser & Barth, 2017; Xing et al., under revision; Zax et al., 2019). This approach has led to novel predictions about the specific patterns of bias that should arise for number line judgments made under different task conditions and by different age groups over development, with these predictions supported by data (Zax et al., 2019). The appeal of this theoretical framework is increased by its demonstrated ability to account for a range of different quantity judgments.

Despite the proportion estimation framework's advantages, such as good quantitative fits to data, good conceptual fits with task structure, relative simplicity, and generality, there are many remaining questions to be answered. One question that specifically makes contact with the model fits in the present study concerns how we should interpret the mixed model. For both the present spatial reproduction task in children and adults (see also Barth et al., 2015) and the bounded number line estimation task in children (Zax et al., 2019), a mixed proportion estimation model (Figure 1C) provided a good quantitative explanation of performance. This mixed model, as described earlier, is a weighted combination of the pure models (a one-cycle model, Figure 1A, and a two-cycle model, Figure 1B). The mixed model can accommodate differences in strategy use across or within individuals, unlike the pure models that reflect single strategies: making estimates relative to two endpoint reference points (treating the space in question as a single region) or making estimates relative to two endpoints plus a midpoint (treating each half of the space as a region).

However, though the mixed model accounts well for the data in both tasks, the way in which it does so differs across tasks. The original theoretical motivation for the mixed model

(Hollands & Dyre, 2000) is the idea that we might well expect a mix of strategies to shape performance: some individuals might use one strategy consistently (for example, always mentally partitioning the space into two halves when making estimates) while others might use another (for example, never mentally partitioning the space). It is also possible that strategies might vary across trials within individuals. If the mixed model is simply the reflection of a combination of different partitioning strategies, then it makes sense that a two-parameter version of the mixed model should account for the data: a weighting parameter would reflect the relative contributions of each component, and a single beta parameter would describe the bias associated with both components⁵. This is exactly what is observed when children carry out number line estimation tasks, indicating the proper location of a given numeral on a line with two marked endpoints (Zax et al., 2019). A two-parameter mixed model accounted for number line data, and parameter values changed exactly as predicted by previous theoretical applications of the proportion judgment framework to number line tasks (e.g., Slusser et al., 2013). More specifically, task properties that would be predicted to support young children's partitioning of the number line (such as cues to the location of the midpoint) were associated with a larger contribution of the two-cycle component. A larger two-cycle contribution was also associated with increased age (Zax et al., 2019). For both components, resulting beta values were less than 1. Finally, the mixed model described group median but not individual level data: at the individual level, children were most commonly characterized by either the one-cycle or the two-cycle model, consistent with the idea that most children settled either on an endpoints-only strategy or an endpoints-plus-midpoint strategy, with a minority of children switching strategies across trials. Overall, in the context of children's bounded number line estimation, the functioning of the mixed model aligns extremely well with its original motivation as introduced by Hollands and Dyre (2000).

The situation is not so clear for the spatial tasks described in the present paper. Here, as in prior work with related tasks (Hollands & Dyre, 2000), a three-parameter mixed model best accommodates the data. The one-cycle and two-cycle components of the mixed model each require a different value of the beta parameter. To accommodate the placements of adults in Experiment 1, for example, the mixed model involves a very small contribution from a two-cycle component with a beta just below 1, and a very large contribution from a one-cycle component with an unusually large (far above 1) beta value. This is not readily consistent with the simplest interpretation of the mixed model, in our view. Hollands and Dyre (2000) reported similar patterns when re-analyzing data from previous researchers. Their admittedly ad-hoc explanation of this finding was that participants in these tasks might have been judging different perceptual dimensions at different times. For example, participants judging the proportion of dots of one of two colors might have been judging color saturation (consistent with a beta value above 1) or area (consistent with a beta value below 1). We think this idea does not extend easily to the present data, and we therefore don't favor it as an explanation of the role of the three-parameter mixed model for the current spatial task. Overall, although the mixed model fits the data

⁵ Such a pattern would not necessarily be predicted in every case. For example, if a sample included a wide age range, a three-parameter mixed model might be appropriate. This is because younger children typically exhibit a larger degree of bias in their number line placements and they are also less likely to use additional inferred reference points. As a result, when a mixed model is fitted to group data, bias parameters might be more extreme for the one-cycle component (corresponding to data generated by the younger children in the sample). This pattern would still be consistent with a straightforward interpretation of the mixed model.

extremely well, the resulting parameter values from the present spatial task are not clearly aligned with the simplest interpretation of the cognitive strategies underlying the mixed model – unlike the number line estimation task. Similarly, the developmental course of performance on the present spatial task is less straightforward than it is in the number line task. For the number line task, children exhibited the predicted developmental progression from less to more frequent use of a middle reference point (Zax et al., 2019), with relatively few individuals best described by the mixed model. But in the present spatial task, adults and older children’s placements are best described by mixed models, though a developmental progression toward the mixed model is not predicted with respect to the simplest interpretation of the model. This suggests that different kinds of cognitive strategies might underlie performance on the present spatial position reproduction task and the spatial component of the number line estimation task. We do not offer evidence for a different interpretation of the mixed model here; ongoing work will build on these differing observations across task types to deepen our understanding of the cognitive and developmental processes and strategies involved in quantity judgments like these.

The present work does not make explicit connections to neural data or computational simulations, but an interesting direction for future work will be to investigate ways in which the performance patterns seen in these tasks might be consistent with characteristics of neural coding of quantity (e.g., Prather, 2012, 2018). A relevant example linking behavioral results to neural coding characteristics comes from previous work on operational momentum effects. Operational momentum (McCrink, Dehaene, & Dehaene-Lambertz, 2007) refers to a phenomenon in which participants overestimate the outcomes of addition operations while underestimating the outcomes of subtraction operations. Prather (2012) showed through simulation that behavioral observations of operational momentum could (and indeed should) arise from the nature of the neural coding of number.

Another important question for future experimental work concerns the extent to which the proportion judgment framework might explain performance in other tasks that have been used to develop and explore the CAM in previous work. In addition, as briefly described above, the same general form is shared by a wide variety of quantitative estimation biases appearing across many domains. Studies reporting these findings span multiple literatures in cognitive and developmental psychology, theoretical and perceptual neuroscience, and judgment and decision-making. The paradigms that produce these biases are in many cases conceptually compatible with the proportion estimation framework (for an example from judgment and decision making, see Xing, Paul, Zax, Cordes, Barth, & Patalano, 2019). Because many of these phenomena can reasonably be thought of as implicit or explicit estimates of proportion, it may be that this theoretical framework could provide a useful unifying explanation of biases in a wide range of quantity judgments. Ongoing work will explore the possibility that theoretical explanations of proportion judgment could provide a unifying account of the psychology of quantity judgments across domains.

References

- Barth, H., Baron, A., Spelke, E., & Carey, S. (2009). Children's multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology, 103*, 441-454.
- Barth, H., Cordes, S., & Patalano, A.L. (2018). Suboptimality in perceptual decision making and beyond. *Behavioral & Brain Sciences, 41*, e225.
- Barth, H., Slusser, E., Kanjlia, S., Garcia, J., Taggart, J., & Chase, E. (2016). How feedback improves children's numerical estimation. *Psychonomic Bulletin & Review, 23*, 1198-1205.
- Barth, H., Lesser, E., Taggart, J., & Slusser, E.B. (2015). Spatial estimation: A non-Bayesian alternative. *Developmental Science, 18*, 853-862.
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 125-135.
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138*, 389-414.
- Boyer, T.W., Levine, S.C., & Huttenlocher, J. (2008). The development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*, 1478-1490.
- Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference. A practical information-theoretic approach (2nd ed). New York: Springer.
- Burnham, K. P., & Anderson, D. R., & Huyvaert, K.P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*, 23-35.
- Cicchini, G.M., Arregghi, R., Cecchetti, L., Giusti, M., & Burr, D. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience, 32*, 1056-1060.
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin and Review, 18*, 331-338.
- Cohen, D. J., & Sarnecka, B. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50*, 1640-1652.
- Crawford, L.E., & Duffy, S. (2012). Sequence effects in estimating spatial location. *Psychonomic Bulletin & Review, 17*, 725-730.
- Crawford, L.E., Landy, D., & Salthouse, T. A. (2016). Spatial working memory capacity predicts bias in estimates of location. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 42*, 1434-1447.
- Crawford, L.E., Landy, D., & Presson, A.N. (2014). Bias in spatial memory: Prototypes or relational categories? *Proceedings of the Cognitive Science Society*.
- Dackermann, T., Kroemer, L., Nuerk, H.-C., Moeller, K., & Huber, S. (2018). Influences of presentation format and task instruction on children's number line estimation. *Cognitive Development, 47*, 53-62.
- Duffy, S., Huttenlocher, J., & Levine, S. (2005). It's all relative: How young children encode extent. *Journal of Cognition and Development, 6*, 51-63.
- Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize estimation. *Developmental Science, 9*, 597-603.

- Duffy, S., & Smith, J. (2017). Category effects on stimulus estimation: Shifting and skewed frequency distributions—A reexamination. <https://doi.org/10.3758/s13423-017-1392-7>
- Erlick, D.E. (1964). Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology*, *67*, 475-482.
- Friedman, A., Montello, D.R. & Burte, H. (2012). Location memory for dots in polygons versus cities in regions: Evaluating the category adjustment model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1336-1352.
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*, 129-166.
- Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are). *Psychological Bulletin*, *138*, 415-422.
- Hollands, J.G., & Dyre, B.P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*, 500-524.
- Hollands, J.G., Tanaka, T., & Dyre, B.P. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 563-574.
- Holden, M. P., & Newcombe, N. S. (2013). The development of location coding: An adaptive combination account. In Waller, David (Ed); Nadel, Lynn (Ed). *Handbook of spatial cognition*, (pp. 191-209). Washington, DC, US: American Psychological Association, x, 309 pp.
- Holden, M. P., Newcombe, N. S., & Shipley, T. F. (2013). Location memory in the real world: Category adjustment effects in 3-dimensional space. *Cognition*, *128*, 45-55.
- Holden, M. P., Curby, K. M., Newcombe, N. S., & Shipley, T. F. (2010). A category adjustment approach to memory for spatial location in natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 590-604.
- Holden, M.P., Newcombe, N.S., Resnick, I., & Shipley, T.F. (2016). Seeing like a geologist: Bayesian use of expert categories in location memory. *Cognitive Science*, *40*, 440-454.
- Hund, A., & Plumert, J. (2003). Delay-induced bias in children's memory for location. *Child Development*, *73*, 829-840.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*, 352-368.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*, 220-241.
- Huttenlocher, J., & Lourenco, S.F. (2007). Using spatial categories to reason about location. In J. Plumert & J. Spencer (Eds.), *The Emerging Spatial Mind* (pp. 3-24). New York, NY: Oxford University Press.
- Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology*, *27*, 115-147.
- Jazayeri, M., & Shadlen, M.N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*, 1020-1026.
- Jeong, Y., Levine, S., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development*, *8*, 237-256.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral & Brain Sciences*, *34*, 188-231.

- Landy, D., Guay, B. & Marghetis, T. (2017). Bias and ignorance in demographic perception. *Psychonomic Bulletin & Review*, DOI 10.3758/s13423-017-1360-2
- Link, Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Unbounding the mental number line: New evidence on children's spatial representation of numbers. *Frontiers in Psychology*, 4, 1021.
- Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24, 2351-2360.
- McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics*, 69, 1324–1333.
- McCrink, K., Shafto, P., & Barth, H. (2017). The relationship between non-symbolic multiplication and division in childhood. *Quarterly Journal of Experimental Psychology*, 70, 686-702.
- McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, 116, 204-216.
- McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, 18, 740-745.
- Nakajima, Y. (1987). A model of empty duration perception. *Perception*, 16, 485-520.
- Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., Luwel, K. (2016). Children's use of number line estimation strategies. *European Journal of Psychology of Education*, 31, 117-134.
- Peeters, D., Verschaffel, L., Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology*, 2017, 1-19.
- Petzschner, F.H., Glasauer, S., & Stephan, K.E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 19, 285-293.
- Prather, R. W. (2012) Connecting neural coding to number cognition: A computational account. *Developmental Science*, 15, 589-600.
- Prather, R. (2018): Neural coding partially accounts for the relationship between children's number line estimation and number comparison performance. *Journal of Cognition and Development*, 19, 201-219.
- Rahnev, D., & Denison, S. (2018). Suboptimality in perceptual decision making. *Behavioral & Brain Sciences*, 41, e223.
- Rouder, J., & Geary, D. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science*, 17, 525-536.
- Sailor, K.M. & Antoine, M. (2005). Is memory for stimulus magnitude Bayesian? *Memory & Cognition*, 33, 840-851.
- Sampaio, C., & Wang, R.F. (2017). The cause of category-based distortions in spatial memory: A distribution analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1988-1992.
- Schutte, A. R., Spencer, J. P., & Schoner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development*, 74, 1393-1417.
- Sciutti, A., Burr, D., Saracco, A., Sandini, G., & Gori, M. (2015). Development of context-dependency in human space perception. *Experimental Brain Research*, 232, 3965-3976.
- Shuford, E. H. (1961). Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology*, 61, 430-436.

- Slusser, E., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Child Psychology*, *162*, 181-198.
- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*, 193-208.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 683-692.
- Spencer, J., Austin, A., & Schutte, A.R. (2012). Contributions of dynamic systems theory to cognitive development. *Cognitive Development*, *27*, 401-418.
- Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, *18*, 557-563.
- Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology*, *16*, 613-625.
- Xing, C., Paul, J., Zax, A., Cordes, S., Barth, H., & Patalano, A. L. (2019, in press). Probability-range effects on probability distortion in a gambling task. *Acta Psychologica*.
- Xing, C., Zax, A., Taggart, J., Bass, I., George, E., & Barth, H. (under revision). Numerical estimation strategies are correlated with math ability in school-age children.
- Zax, A., Slusser, E.B., & Barth, H. (2019). Partitioning and proportion estimation in children's numerical judgments. *Journal of Experimental Child Psychology*, *185*, 71-94.
- Zhang, H., & Maloney, L.T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, cognition, and action. *Frontiers in Neuroscience*, *6*, 1. doi:10.3389/fnins.2012.00001

Appendix A: Individual analyses of Experiment 1.

Adults

SN	Preferred Model	R²	β (if pure model)	β_1 (if mixed model)	β_2 (if mixed model)	Weight parameter
1	Pure 2-cycle Model	0.9914	0.9422			
2	Mixed Model	0.9949		1.131	0.759	0.3868
3	Mixed Model	0.9961		2.072	0.8231	0.1023
4	Mixed Model	0.9952		5.572	0.8356	0.1002
5	Mixed Model	0.9943		1.155	0.4818	0.7417
6	Mixed Model	0.9961		3.896	0.9818	0.05972
7	Mixed Model	0.9837		5.583	0.7959	0.08031
8	Mixed Model	0.9933		1.301	0.5366	0.6128
9	Mixed Model	0.9906		2.106	0.8737	0.127
10	Mixed Model	0.9894		~6	0.9184	0.1453
11	Mixed Model	0.9935		1.365	0.8548	0.2499
12	Mixed Model	0.9853		~6	0.9436	0.07004
13	Mixed Model	0.9917		1.066	0.2745	0.7986
14	Pure 2-cycle Model	0.9966	0.9057			
15	Mixed Model	0.9975		5.603	0.8294	0.03853
16	Mixed Model	0.9966		5.161	0.8044	0.1092
17	Pure 2-cycle Model	0.994	0.8699			
18	Mixed Model	0.9945		5.063	0.9551	0.1033
19	Pure 1-cycle Model	0.9943	1.054			
20	Mixed Model	0.9937		2.759	0.9324	0.09964
21	Mixed Model	0.9954		~6	0.8741	0.1459
22	Mixed Model	0.9937		2.297	0.8799	0.1883
23	Pure 2-cycle Model	0.9864	0.8037			
24	Mixed Model	0.9905		4.769	0.8706	0.05565
25	Mixed Model	0.9924		5.586	0.928	0.114
26	Pure 2-cycle Model	0.9919	0.8893			
27	Mixed Model	0.9913		1.1	0.03262	0.8644
28	Mixed Model	0.9948		1.494	0.7096	0.497
29	Mixed Model	0.9953		1.281	0.7125	0.3773
30	Mixed Model	0.9889		1.083	0.1599	0.8714

31	Mixed Model	0.9931		~ 6	0.7755	0.07272
32	Pure 2-cycle Model	0.9951	0.9712			
33	Pure 1-cycle Model	0.9933	1.045			
34	Mixed Model	0.9948		3.175	0.8769	0.06355
35	Mixed Model	0.9943		~ 6	0.8838	0.08467
36	Mixed Model	0.9955		5.208	0.9449	0.17
37	Pure 1-cycle Model	0.9945	1.117			
38	Pure 2-cycle Model	0.9945	0.9102			
39	Mixed Model	0.9946		5.833	0.9002	0.1025
40	Mixed Model	0.9937		1.087	0.5729	0.6791
41	Mixed Model	0.9946		5.752	0.8681	0.08139
42	Mixed Model	0.9941		2.18	0.8362	0.1021
46	Mixed Model	0.9921		1.108	0.2429	0.7656
47	Mixed Model	0.9931		~ 6	0.8979	0.07553
50	Pure 2-cycle Model	0.985	0.9029			
51	Pure 2-cycle Model	0.9891	0.9615			
52	Mixed Model	0.9931		~6	0.8979	0.07553
53	Mixed Model	0.9924		2.902	0.8815	0.08339
54	Mixed Model	0.99		~ 6	0.8908	0.08241
55	Mixed Model	0.9964		1.164	0.006141	0.8947
56	Mixed Model	0.9916		3.432	0.9091	0.1497
57	Mixed Model	0.9935		~ 6	0.8157	0.1677
58	Mixed Model	0.9923		~ 6	0.9133	0.09143
59	Pure 1-cycle Model	0.9896	1.107			
60	Mixed Model	0.9945		1.449	0.7894	0.3288
61	Mixed Model	0.9925		5.978	0.9056	0.1134
62	Mixed Model	0.9944		2.218	0.8374	0.09908
63	Mixed Model	0.9853		1.144	0.273	0.7447
64	Mixed Model	0.9907		1.098	0.6536	0.6647
65	Pure 2-cycle Model	0.9933	0.8159			
66	Mixed Model	0.99		1.101	0.139	0.8975
67	Pure 2-cycle Model	0.9918	0.8382			
68	Mixed Model	0.9909		1.328	0.6899	0.3732

69	Mixed Model	0.9918		4.287	0.777	0.161
70	Pure 1-cycle Model	0.9902	1.103			
71	Mixed Model	0.9951		~ 6	0.9693	0.1173
72	Mixed Model	0.99		2.161	0.878	0.122
73	Mixed Model	0.9904		~ 6	0.8129	0.05608
74	Mixed Model	0.9932		1.137	0.5611	0.72
75	Mixed Model	0.9934		~6	0.8792	0.1664
76	Mixed Model	0.9962		3.582	0.8784	0.07015
77	Mixed Model	0.9961		2.815	0.9147	0.2253
78	Pure 2-cycle Model	0.9942	0.9806			
79	Pure 2-cycle Model	0.9909	0.8987			
80	Mixed Model	0.9895		~ 6	0.8938	0.08665
81	Mixed Model	0.9954		4.011	0.8919	0.1317
82	Mixed Model	0.9883		~ 6	0.9547	0.08718
83	Mixed Model	0.9868		~ 6	0.9077	0.09973
84	Mixed Model	0.9869		~ 6	0.7851	0.0757
85	Mixed Model	0.9918		~ 6	0.9323	0.04434
86	Mixed Model	0.9935		1.205	0.4918	0.6921
87	Mixed Model	0.9941		1.443	0.4272	0.6028
88	Pure 2-cycle Model	0.9909	0.8954			
89	Mixed Model	0.9953		1.201	0.1843	0.7568
90	Mixed Model	0.9971		5.245	0.7898	0.1124
91	Mixed Model	0.9923		3.417	0.838	0.0927

Appendix B: Individual analyses of Experiment 2 participants by age group.

Age 7

SN	Preferred Model	R ²	β (if pure model)	β_1 (if mixed model)	β_2 (if mixed model)	Weight parameter
4	Pure 1-cycle Model	0.9671	0.8541			
5	Pure 2-cycle Model	0.9679	0.7999			
7	Pure 2-cycle Model	0.9702	0.7231			
11	Pure 2-cycle Model	0.9159	0.6393			
12	Pure 1-cycle Model	0.953	1.172			
23	Mixed Model	0.9834		5.099	0.8076	0.3009
26	Pure 1-cycle Model	0.9708	0.8965			
34	Pure 2-cycle Model	0.954	0.8347			
35	Pure 2-cycle Model	0.98	0.8407			
36	Pure 2-cycle Model	0.9678	0.844			
37	Pure 2-cycle Model	0.9711	0.7457			
44	Pure 2-cycle Model	0.9309	0.6064			
45	Mixed Model	0.9862		0.9263	0.1871	0.8171
46	Pure 2-cycle Model	0.9815	0.8724			
47	Pure 2-cycle Model	0.9694	0.8913			
49	Mixed Model	0.9834		1.12	~ 1.198e-016	0.8441
51	Pure 2-cycle Model	0.9741	0.7607			
54	Mixed Model	0.9734		~ 6	0.8456	0.1137
55	Pure 2-cycle Model	0.9441	0.616			

Age 8

SN	Preferred Model	R ²	β (if pure model)	β_1 (if mixed model)	β_2 (if mixed model)	Weight parameter
3	Pure 2-cycle Model	0.9907	0.6755			
6	Mixed Model	0.9812		~ 6	0.7396	0.07806
8	Pure 2-cycle Model	0.9685	0.7598			
9	Mixed Model	0.9832		2.799	0.7934	0.1812
15	Pure 2-cycle Model	0.9783	0.8022			
18	Pure 1-cycle Model	0.9549	0.6236			
19	Mixed Model	0.9785		0.8341	0.5503	0.6477
20	Pure 2-cycle Model	0.9819	0.8637			
21	Mixed Model	0.9898		0.8133	0.7233	0.3722
5	Pure 2-cycle Model	0.9801	0.7616			
28	Pure 1-cycle Model	0.9716	0.898			
39	Pure 2-cycle Model	0.9744	0.6828			
43	Pure 1-cycle Model	0.9527	0.8036			
56	Pure 2-cycle Model	0.9898	0.8323			
58	Pure 2-cycle Model	0.9752	0.8201			
59	Mixed Model	0.9951		1.172	0.6843	0.4422
60	Mixed Model	0.9577		3.518	0.8002	0.2006
62	Pure 1-cycle Model	0.9604	1.266			
63	Pure 1-cycle Model	0.9869	0.9603			

Age 9

SN	Preferred Model	R ²	β (if pure model)	β_1 (if mixed model)	β_2 (if mixed model)	Weight parameter
3	Pure 2-cycle Model	0.9829	0.7747			
4	Pure 2-cycle Model	0.99	0.8436			
11	Mixed Model	0.9906		2.93	0.9319	0.208

13	Mixed Model	0.9855		3.418	0.7893	0.0882
14	Mixed Model	0.9848		~ 6	0.7808	0.08191
15	Mixed Model	0.9798		2.749	0.8802	0.1321
20	Pure 2-cycle Model	0.9911	0.8208			
23	Mixed Model	0.9904		4.672	0.8206	0.07354
27	Pure 1-cycle Model	0.9797	0.944			
30	Mixed Model	0.9923		2.248	0.9149	0.1114
31	Pure 2-cycle Model	0.9863	0.7388			
33	Mixed Model	0.9896		5.512	0.8262	0.1069
36	Mixed Model	0.9887		3.322	0.932	0.1534
39	Mixed Model	0.991		4.124	0.796	0.059
45	Pure 2-cycle Model	0.9927	0.8723			
46	Mixed Model	0.9905		~ 6	0.8901	0.06243
48	Mixed Model	0.9898		5.155	0.7677	0.1462
52	Pure 2-cycle Model	0.9895	0.8603			
57	Pure 1-cycle Model	0.9859	0.9504			
61	Mixed Model	0.9873		1.203	0.1353	0.7395

Age 10

SN	Preferred Model	R ²	β (if pure model)	$\beta 1$ (if mixed model)	$\beta 2$ (if mixed model)	Weight parameter
8	Mixed Model	0.995		1.07	0.1769	0.8518
10	Mixed Model	0.9886		4.236	0.6923	0.1373
16	Mixed Model	0.9924		~6	0.8092	0.1567
17	Mixed Model	0.9881		~ 6	0.7241	0.1235
18	Pure 2-cycle Model	0.9935	0.8123			
19	Mixed Model	0.9914		1.146	0.4696	0.7188
24	Mixed Model	0.9917		1.171	0.2962	0.7266
25	Mixed Model	0.9901		1.106	0.1735	0.8463
28	Mixed Model	0.9882		~ 6	0.7864	0.09669
34	Mixed Model	0.9882		~ 6	0.7761	0.05637
37	Mixed Model	0.9623		~ 6	0.8406	0.1302
41	Pure 2-cycle Model	0.9912			0.9128	
43	Mixed Model	0.993		1.092	3.337E-11	0.8035

44	Pure 2-cycle Model	0.9785	0.8404			
47	Mixed Model	0.9888		4.522	0.833	0.1082
49	Mixed Model	0.9874		5.864	0.871	0.1056
55	Mixed Model	0.9937		2.041	0.7843	0.168
56	Mixed Model	0.9923		2.35	0.8658	0.1026
62	Mixed Model	0.985		3.432	0.8093	0.07979
63	Pure 2-cycle Model	0.9822	0.8186			