© 2018 American Psychological Association 0096-1523/18/\$12.00

2018, Vol. 44, No. 9, 1472–1481 http://dx.doi.org/10.1037/xhp0000544

Intrinsic Whole Number Bias in Humans

Santiago Alonso-Díaz, Steven T. Piantadosi, Benjamin Y. Hayden, and Jessica F. Cantlon University of Rochester

Humans have great difficulty comparing quotients including fractions, proportions, and probabilities and often erroneously isolate the whole numbers of the numerators and denominators to compare them. Some have argued that the whole number bias is a compensatory strategy to deal with difficult comparisons. We examined adult humans' preferences for gambles that differed only in numerosity, and not in factors that influence their expected value (probabilities and stakes). Subjects consistently preferred gambles with more winning balls to ones with fewer, even though the probabilities were mathematically identical, replicating prior results. In a second experiment, we found that subjects accurately represented the relative probabilities of the choice options during rapid nonverbal probability judgments but nonetheless showed biases based on whole numbers. We mathematically formalized and quantitatively evaluated cognitive rules based on existing hypotheses that attempt to explain subjects' whole number biases during quotient comparisons rather than a compensatory strategy.

Public Significance Statement

When people learn or decide about fractions, they reveal a whole-number bias: The value of the numerator or denominator guides the assessments and causes erroneous choices and learning outcomes. A plausible explanation is that individuals fail to compute ratios, perhaps due to bad educational practices. However, this study demonstrates that the whole-number bias can also occur in the presence of good ratio estimates. Part of the difficulties children and adults exhibit may stem from a more generic feature of cognition that places importance on numerical magnitudes.

Keywords: numerosity, fraction, probability, whole number bias, Bayesian model

Supplemental materials: http://dx.doi.org/10.1037/xhp0000544.supp

Human children have significant difficulty learning to represent and compare quotients, and even college-educated adults struggle to order the values of fractions, ratios, and probabilities (Bonato, Fabbri, Umiltà, & Zorzi, 2007; DeWolf & Vosniadou, 2015; R. Gelman, 1991; Reyna & Brainerd, 2008; Siegler, Fazio, Bailey, & Zhou, 2013). For instance, when trying to pick the largest value between 27/42 and 43/77, adults and children are wrongly inclined to pick 43/77, because its numerator, 43, is larger than 27 (Ni & Zhou, 2005). This is known as the "whole number bias" in quotient comparison.

Educational research has described the whole number bias as an explicit fall-back strategy that students use to compensate for unsuccessful reasoning about quotients (Fazio, DeWolf, & Siegler, 2016). Adults and children are thought to use whole numbers to compare fractions when they fail to generate an integrative, holistic representation of the quotient (Reyna & Brainerd, 2008). Some have argued that humans do not have an intuitive sense of fractional magnitudes and mostly rely on whole number comparisons (Bonato et al., 2007). This explanation predicts that the type of strategy used to compare quotients, whether integrative or whole-number, will depend on the specific fractions being compared. People will compare fractions based on whole numbers when they are too difficult to compare with an integrative representation.

However, recent research has shown that under some conditions people can accurately represent the magnitudes of ratios (Fabbri, Caviola, Tang, Zorzi, & Butterworth, 2012; Jacob, Vallentin, & Nieder, 2012; Schneider & Siegler, 2010). When subjects are asked to choose the larger value from two fractions, for example, their accuracy is reliable and is modulated by the quantitative difference between the fractions, a "distance effect" (Meert, Grégoire, & Noël, 2009). Distance effects in fraction estimation show

This article was published Online First June 25, 2018.

Santiago Alonso-Díaz and Steven T. Piantadosi, Department of Brain and Cognitive Sciences, University of Rochester; Benjamin Y. Hayden, Department of Brain and Cognitive Sciences and Center for Visual Science, University of Rochester; Jessica F. Cantlon, Department of Brain and Cognitive Sciences, University of Rochester.

Benjamin Y. Hayden is now at the Department of Neuroscience, University of Minnesota.

Funding was provided by the James S. McDonnell Foundation (220020300) and the National Science Foundation (Education Core Research Grant DRL-1459625) to Jessica F. Cantlon. Data and codes are available through the Open Science Framework https://osf.io/8v2fj/.

Correspondence concerning this article should be addressed to Santiago Alonso-Díaz, who is now at the Department of Economics, Universidad Javeriana, Carrera 7 No 40 - 62, Bogotá, Colombia, 110311. E-mail: santiagoalonsodiaz@gmail.com

that adults approximate the holistic magnitude of the fractions when they compare them. There is even research showing that human infants accurately track the relative values of ratios nonverbally (Denison & Xu, 2010; McCrink & Wynn, 2007). Thus, there is tension between research showing that ratio representation is accurate, versus research showing that it is learned slowly, often incorrectly, and is riddled with whole number biases.

The possible origins of the whole number bias in cognitive processing of quotients have been discussed in several studies (Bonato et al., 2007; Denes-Raj, Epstein, & Cole, 1995; Fazio et al., 2016; Jacob et al., 2012; Reyna & Brainerd, 2008; Schneider & Siegler, 2010; Siegler et al., 2013). One possibility is that, when faced with a proportional comparison, humans strategically rely on whole numbers when denominators are close in value and only proceed to compute proportions when denominators are not comparable (Schneider & Siegler, 2010). For example, in 2/10 versus 5/10 the denominators are identical and comparing numerators leads to the correct identification of the larger quotient. But in 2/10 versus 5/100 the denominators are too far apart and the whole number strategy breaks. In this case, subjects will attempt to compute the proportion associated with each fraction and compare them. Thus, under this "strategic bias," subjects compare values based on either ratios or whole numbers, depending on the distance between denominators.

Another possibility is that subjects always compare both the whole numbers and the values of the quotients, regardless of difficulty. Under this "intrinsic bias" scenario, subjects would show evidence of whole number biases across all ratio comparisons. Both of these explanations predict more errors on "close" ratios than "far" ratios, but they differ in their predictions for when the whole number bias should emerge. Disentangling these explanations is important for understanding the sources of humans' difficulties learning fractions, and the degree to which the whole number bias is a fundamental and pervasive component of how humans conceive of ratios.

In order to measure adults' intuitions about ratios, and their spontaneous representations of whole numbers versus ratios, we tested them with a rapid two-alternative forced choice gamble task (2AFC). Subjects chose between hypothetical random draws of colored balls from urns to maximize their chances of selecting a "winner." The stimulus (and cognitive modeling) was mainly presented in nonsymbolic format, that is, visual dots presented on screen. Only one of the conditions used Arabic numerals. Nonsymbolic stimulus is relevant to the question of fraction cognition for two reasons: 1) It is more intuitive than symbolic fractions: It does not require formal education (Fontanari, Gonzalez, Vallortigara, & Girotto, 2014), and robust proportional reasoning with visual fractions has been reported across development and in many species (Denison & Xu, 2010; Fabbri et al., 2012; O'Grady, Griffiths, & Xu, 2016; Rakoczy et al., 2014; Xu & Garcia, 2008); and 2) whole number biases are not exclusively present in symbolic fractions; they also appear with visual proportions (Fabbri et al., 2012; O'Grady et al., 2016; Passerini, Macchi, & Bagassi, 2012). Thus, by using a nonsymbolic stimulus we made sure to capture proportional reasoning susceptible to undue numerosity influences.

It is important, however, to clarify that the source of symbolic and nonsymbolic whole-number biases may differ to some degree. For example, the time it takes to compare fractions is much longer in Arabic formats (Fabbri et al., 2012 vs. Schneider & Siegler, 2010), suggesting additional computations/strategies (Fazio et al., 2016). Beyond the potential differences, there is a link between symbolic and nonsymbolic systems (Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Halberda, Mazzocco, & Feigenson, 2008; Melnick, Harrison, Park, Bennetto, & Tadin, 2013; Park & Brannon, 2013). Perceptual systems are relevant for formal mathematical cognition because they seem to support generic intuitions in symbolic domains (Dehaene, 2009).

We found that subjects in two different experiments accurately represented the relative ratios of the choice options but showed biases to select the option with the larger numerator, even when the numerator was irrelevant or misleading. We mathematically formalized and compared influential hypotheses from the ratio processing literature to test the likely decision rule subjects used to make their selections. The results showed that subjects consistently factor both whole numbers and ratios into the mental comparison. We argue that whole number comparison is intrinsic to how adults reason about quotients.

Experiment 1: Whole Number Bias in a One-Shot Gamble

We first aimed to replicate and extend prior evidence of the whole number bias by testing subjects with speeded ratio comparisons in a gambling task, and assessing subjects' explicit estimates of the ratio values in the choice options. Each subject was given a one-shot gamble to pick the option most likely to yield a "winner." Based on prior research, we expect that participants will prefer gambling options that have a greater number of winning instances despite having equal ratios (e.g., 4/8 > 1/2; Denes-Raj et al., 1995). An open question is whether participants accurately represent the ratios of the choice options despite their whole number bias.

Method: Participants, Materials, and Procedures

All experimental procedures adhered to university standards, as approved by the Research Subjects Review Board. We approached individuals in a public location on the University of Rochester campus. We aimed to recruit around 90 participants per group for 3 conditions: nonsymbolic stimuli (dot size equal), nonsymbolic stimuli (cumulative surface area equal), and symbolic numeral stimuli (see Figure 1). Sample size was based on previous reports (Denes-Raj et al., 1995; Kirkpatrick & Epstein, 1992; Pacini & Epstein, 1999; Passerini et al., 2012). We confirmed the appropriateness of the sample size with a power calculation using the reported average effect in the original ratio bias study by Kirkpatrick and Epstein (1992). In their report, on average, 65% of participants picked the lottery with larger numerosity in the 0.1 win and 0.9 lose conditions of Experiment 2 (65% is an intermediate value; in a review made by Passerini et al., 2012 some ratio biases were as large as 80%). To obtain a power of 0.8 in a binomial test comparing chance level (50%) against 65%, the required sample is 81 subjects (power calculation with the binom package in R, function binom.power). The final sample was 284 subjects (145 female; mean age: 25.4 years, SD: 8.3).

Subjects saw images of two urns, and their task was to choose one of them (Figure 1; stimuli were 21.6 by 27.9 cm laminated



Figure 1. Subjects were presented with a one-shot gamble in which they were to select one of two urns from which to pull a single ball. Their task goal was to pull a winning color (either orange or white). Subjects were tested in one of three conditions: Nonsymbolic with ball size equated between urns (top left, n = 88), Nonsymbolic with cumulative surface area covered by the balls equated between urns (top right, n = 83), and Symbolic numerals (bottom, n = 91). See the online article for the color version of this figure.

cards showing orange and white balls). They were told to imagine that if they pulled a white ball from the selected urn they would win \$100 dollars, and if they pulled an orange ball, \$0 dollars (the winning color was randomized and counterbalanced across subjects). We used a nonleading verbal instruction ("Please select the urn you prefer") to avoid inducing a response bias (Passerini et al., 2012). Subjects were instructed to report their choice as quickly as possible. The test stimuli were presented for 2 seconds. All subjects reported their choice within the 2-second display time. Subjects conveyed their answer by saying A or B, for left and right urn respectively (the side with more balls was counterbalanced randomly across subjects). All participants did one trial, and the numbers presented to them are shown in Figure 1 (i.e., all saw the same images).

We tested the accuracy of each subject's ratio estimations. In the nonsymbolic conditions subjects were asked to verbally report the proportion of white (or orange) balls in each urn. In the nonsymbolic and symbolic conditions subjects were asked, "If an urn has 15 green balls and 15 red balls, what is the probability of pulling a red ball?"

The effective sample was 262 subjects (92% of the total). Twenty-two subjects were excluded for lack of motivation or signs they did not understand the task or stimuli (e.g., wildly incorrect responses to explicit probability question). This was to ensure, conservatively, that results were not contaminated by confused subjects.

Results

Participants were attracted to the urn with more balls (see Figure 2). In the nonsymbolic condition, the ratios between orange and white balls were 1:1, or 50–50, for both choice options, yet the

majority of subjects chose the numerically larger option (63%, n = 108/171, binomial test: p < .001, 95% CI = 55%–70%). Subjects were biased toward the larger number regardless of whether they saw stimuli with dot size equal (Figure 2, NS; 63%, n = 55/88, binomial test: p = .02, CI = 51%-72%) or cumulative dot area equal (Figure 2, NSc; 63%, n = 53/83, binomial test: p = .01, CI = 52%-74%). The majority of subjects accurately reported the proportion of winning for each urn within 5% of the exact value (65% of subjects, n = 111/171; the mean smaller number urn estimate: M = 49.8%, SD = 0.01; larger number urn estimate: M = 47.5%, SD = 0.06). We analyzed the responses of subjects who had perfect estimates of the ratios of orange to white balls in



Figure 2. Ratio bias. When subjects faced two bags with equal probability they were biased to pick the one with more balls. This happened across experiments: Experiment 1: NS (nonsymbolic, i.e., dots), NSc (nonsymbolic cumulative area control), and S (symbolic, i.e., arabic numerals). Experiment 2: PPP (psychophysics of probability). Error bars are 95% confidence interval of a binomial test.

both choice urns, and they also showed a whole number bias for the numerically larger option (66%, n = 66/99, binomial test: p =.001, CI = 56%–75%). Finally, we conducted a logistic regression with the three experimental factors: number of winners, experimental condition (dot size equal, cumulative area equal), and explicit ratio estimate of the larger urn (accurate, inaccurate). The only significant effect was of the number of winners ($\beta = 1.0675$, std. error = 0.3243, p = .0009). Effects of experimental condition ($\beta = -0.1757$, std. error = 0.3208, p = .5839) and ratio estimate ($\beta = -0.2665$, std. error = 0.3419, p = .4357) were not significant. Conservatively, we tested an additional model with interactions, but it did not add additional explanatory power; if anything, it was worse (AIC simple model: 230.1; AIC interaction model: 234.68).

The whole number bias toward the numerically larger option also emerged in subjects from the symbolic condition. A significant majority of subjects chose the urn with the greater whole number of winning items (66%, n = 60/91, binomial test: p =.003, CI = 55%–75%). In this condition, *all* participants accurately reported that the urns had equal proportions of winning balls and an equal probability of yielding a win.

These results indicate that subjects' choices are biased by the whole numbers of items represented in the two prospects, despite their knowledge that the relative proportions of winning in the two urns were equal.

Experiment 2: Whole Number Bias in Rapid Nonverbal Probability Judgments

Experiment 1 replicated a well-known effect, the whole number bias, while confirming that participants have knowledge of the ratios involved. But we only tested urns with equal ratios, and the pre and post questionnaires are a qualitative confirmation that ratio estimates are available. Experiment 2 aimed to properly quantify the degree to which subjects' choices are influenced by relative ratio versus whole numbers during ratio comparison. For this we tested subjects with a rapid 2AFC ratio comparison task across a broad range of ratios and numerical values.

Method

In Experiment 2, new subjects performed a computerized version of Experiment 1 (Figure 3A) over several hundred trials (816). We recruited 21 participants from an undergraduate population (16 females; mean age: 19.7 years, SD: 1.48). Sample size was based on comparable multitrial perceptual tasks (Fabbri et al., 2012; Jarvstad, Hahn, Rushton, & Warren, 2013; O'Grady et al., 2016). To confirm the appropriateness of sample size we used the overall accuracy improvement found in congruent trials (larger ratio has larger numerosity) relative to incongruent trials (larger ratio has smaller numerosity). Improvements in previous reports were around 20%-25% (Fabbri et al., 2012; O'Grady et al., 2016). For sample size estimation, we conservatively used 20%: from 0.7 correct in incongruent trials to 0.9 in congruent trials (see Fabbri et al., 2012). The obtained sample size estimate for a Chi-Square test of proportions with 0.8 power was 23 subjects (pwr.p.test; pwr package in R). Our final sample of 21 subjects had a power of 0.76. This is slightly below the rule of thumb of 0.8, but we considered a power of 0.76 appropriate as whole number biases are routinely found in the literature (Fabbri et al., 2012; Ni & Zhou, 2005; O'Grady et al., 2016; Reyna & Brainerd, 2008). The added value of our work lies in testing hypotheses via a generative model (see below) rather than confirming a novel phenomenon.

Experimental tasks were implemented in Matlab's Psychtoolbox. A base payment of \$8 was provided. Instructions asked for a quick but accurate response. Feedback (a beep) signaled that the urn with higher probability was correctly selected. If both urns had equal probability a beep occurred regardless of choice. To ensure proper motivation, subjects were informed that at the end of the experiment one random trial was going to be selected and if in that



Figure 3. Probability task and main results of Experiment 2. (A) Subjects had to pick the option with the greater chance of selecting a winner by imagining that if they pulled an orange ball (lighter grey) they would win \$100 and if they pulled a green ball (darker grey) \$0. The panel has an example of a trial with a 0.25 probability distance (the numbers below the images are the actual probability values of the options). (B) Number intrusions: Accuracy improved when the larger probability also had more numbers of winning balls (the green trace (darker grey; the one higher in accuracy) is trials where the larger probability also had more winners; the red trace (lighter grey; the one lower in accuracy) is trials where the larger probability had fewer winners). There was also a clear distance effect. (C) Reaction time (RT) improved with probability distance between the urns. Shading is 2 s.e.m. See the online article for the color version of this figure.

trial the urn with the better ratio was chosen they would receive an additional \$2.

The probability of winning in each urn ranged from 0.125 to 0.94, and the probability distance between the two urns ranged from 0 to 0.38. There was an equal number of trials for each probability ratio, and the winning probability was balanced across sides (left, right).

The range of numerical values of the winning set ranged from 1 to 30 (the numerator), and the total number of balls ranged from 4 to 32 (the denominator). The number of winning balls was balanced in that on half of the trials the higher probability had the larger numerator and the remaining half, the smaller numerator. Under these conditions, a numerator strategy would result in chance performance. The sets had equal density, defined as the total number of dots divided by the spatial extent of the array (dashed circles surrounding the balls in Figure 3A; subjects saw the dashed circles). Half of the trials were equated for the cumulative area covered by the dots, and the remaining half were equated for dot size.

At the end of the probability task, subjects completed a nonsymbolic numerosity comparison task to measure number sensitivity. The stimuli were the same as those in Experiment 2, but participants reported the side with the larger number of dots. Subjects' sensitivity to numerical differences was quantified by measuring their numerical Weber's fraction using the procedure described in (Piantadosi, 2016). The Weber fraction represents the proportion difference needed between quantities to reliably discriminate them, with smaller Weber fractions reflecting better discrimination abilities. For numerosity discrimination, the mean Weber across subjects was 0.28 with a standard deviation of 0.08.

Results

Subjects performed above chance on the probability judgment task and accurately compared the ratios of orange to white balls between urns, t(20) = 12.91, p < .001. Choices were modulated both by the relative ratios of orange to white balls and by the whole numbers of winning items (Figure 3B). Subjects generally preferred urns with higher ratios of orange to white balls, and preferred them even more strongly when the whole number of orange balls (the numerator) was large compared to when it was small. A generalized linear model (family = Binomial, link = logit) on accuracy rates detected significant effects of probability distance $(\beta = 4.50, z = 11.70, p < .001)$, number of winning items ($\beta = -0.53$, z = -6.92, p < .001), and the interaction ($\beta = -2.02, z = -4.77, p < 0.001$) .001). A repeated measures ANOVA on response times (Figure 3C) also revealed modulation by probability distance (F(9, 180) =42.60, p < .001, $\eta_g^2 = 0.10$), number of winners (F(1, 20) = 32.41, $p < .001, \eta_g^2 = 0.01$), and the interaction (F(9, 180) = 5.85, p < 0.01) .001, $\eta_g^2 = 0.01$). The interaction reflects a greater influence of "number of winners" on accuracy at small probability distances than larger ones, because of a floor effect. Thus, subjects showed a whole number bias on top of the distance effect.

Source of the Whole Number Bias

The experimental results so far indicate that even during rapid nonverbal comparisons humans use representations of ratio that are skewed by their representations of whole numbers in the numerators (i.e., the winners). These findings are consistent with prior reports from the education and cognitive development literatures that adults and children have biases to erroneously factor whole number comparisons into judgments of quotients (Ni & Zhou, 2005; Siegler et al., 2013). There are several hypotheses about the sources of the whole number bias in cognitive processing of quotients, but these hypotheses have not been formalized. We formalize them under a single probabilistic framework that allows all to be quantitatively evaluated and compared (see Figure 4). The hypotheses tested are Denominator Neglect (Pacini & Epstein, 1999; Reyna & Brainerd, 2008), Whole Number Weighting (Bonato et al., 2007), Holistic Ratio Comparison (Jacob et al., 2012; Schneider & Siegler, 2010), Strategic Whole Number Bias (Fazio et al., 2016), and Intrinsic Whole Number Bias.

The Denominator Neglect hypothesis argues that subjects exclusively rely on the numerator during quotient comparisons, and fail to factor the quotient denominator into the rule (Pacini & Epstein, 1999; Reyna & Brainerd, 2008). In the Whole Number Weighting rule, subjects represent the numerical value of the components and compare them (Bonato et al., 2007). Some research suggests that humans compare ratios holistically and would not predict a whole number bias in performance-so we tested this purely holistic rule for comparison with the others (Jacob et al., 2012; Schneider & Siegler, 2010). As described in the introduction, a dominant theory in the education and cognitive development literature is that people behave strategically during quotient comparisons and use the whole numbers of the numerators to compare quotients when denominators are close in value but compare quotients holistically when denominators are farther apart (Fazio et al., 2016). Finally, we tested a distinct possibility from these others, which is the hypothesis that people intrinsically compute and compare both the whole numbers of the numerators and the holistic quotient during probability comparison and always factor both outcomes into their decision about "which is greater."

The main idea behind the model is that when people compare visual ratios their choice on which one is larger is based on a weighted compromise among the (noisy) percepts of the actual values on-screen. The computational model infers these psychological percepts (white nodes in Figure 4) as well as the decision weights β_c for each of the values on-screen (winners, losers, and ratios). The inference is constrained by the data (gray nodes): the actual numbers presented to the participant and the observed choices.

In a Bayesian framework, we need to define our prior expectation on how the latent variables are distributed, as well as a likelihood for the choice data. These are written next to the graphical model in Figure 4. Number percepts follow traditional psychophysics of number in which, on average, people perceive the actual number with noise that scales with its magnitude (Figure 4; Whalen, Gallistel, & Gelman, 1999). Weber is the subjects' numerical acuity. It is a constant, determined by their accuracy in the numerosity comparison task of Experiment 2.

Ratio percepts follow a Beta distribution (Figure 4; Gallistel, Krishan, Liu, Miller, & Latham, 2014). This distribution is the result of a Bayesian agent that takes the number of winners (W) and losers (L) and infers the ratio with a binomial likelihood Binomial(W + L, Ratio) and a uniform prior Beta(1,1).

The likelihood of the choice data is a binomial distribution (see Figure 4). Total, is the total number of times a ratio comparison i



Figure 4. Cognitive framework (see main text for details). Conventions: W, L: winners and losers. Φ [W], Φ [L]: perception of W and L. Ratio: perception of ratios. Choice: choices favoring the bigger probability. β .: decision weights. The outer plate loops through i trials (n = 458; participants did 816 trials but some of them were redundant). The inner plate loops through both bags, one with bigger (B) and the other with smaller (S) ratio. Gray color is for data, white for latent psychological percepts. Squares and circles are discrete and continuous variables, respectively.

appeared across participants (i.e., a trial), and Choice_i is the number of times the bigger ratio B was selected in that comparison across participants. The probability of the binomial was determined with the Softmax expression pSM_i (Figure 4; Sutton & Barto, 1998). This expression takes values between 0 and 1, and as such it can be thought as a probability. f_{iB} and f_{iS} are functions weighting all the available percepts, winners, losers, and ratio, on the bag with big (B) or small (S) ratio in trial i. We used a flat prior for the decision weights β_c ($\sim U(-5,5)$). The decision weights for ratios were restricted to be greater than 0 under the assumption that humans follow positive expected values (i.e., negative weights would mean that subjects are repelled by large expected values).

This general framework can reveal whether participants followed any of the cognitive hypotheses mentioned above (except the strategic, which we will formalize in the next paragraph). If subjects use the Denominator Neglect rule, then the posterior probability density (i.e., after conditioning on human responses) for the decision weights for winners (numerator) should be positive and far from zero, while the remaining decision weights for ratios and losers should be negligible. Similarly, if the Number Weighting hypothesis is true, then the weight for winners and losers should be positive or negative and only the weights for ratio should be zero. If the Holistic Ratio hypothesis is valid then the weights for ratio are the only ones that should be relevant. But if the Intrinsic Number Bias is correct, then the posterior probability of the weights for numbers and ratios should all be distant from zero.

The final hypothesis not covered in the previous paragraph is that people are strategic, sometimes using winners or ratio, but never both. This assumes an additional psychological process not present in the general framework of Figure 4: a threshold on when the other percept is used. To formalize this, we create a second model, which we will compare with the basic one (see Figure 4) using a Bayesian information criterion (see below).

The strategic model main difference is that the cues used to decide depend on how close the denominators are. Thus, the model first compares perceived denominators ($\Phi[W] + \Phi[L]$), and if they are sufficiently close, the softmax probability pSM_i is solely based on the numerators (i.e., the weights for $\Phi[L]$ and Ratio are set to zero); or if on the contrary denominators are far apart, the softmax is solely based on ratio percepts (i.e., the weights for $\Phi[L]$ and $\Phi[W]$ are set to zero). Sufficiently close is implemented with a soft-threshold,

$$pUseRatio_i = \frac{1}{1 + e^{-k(Thr - rDen_i)}}.$$

pUseRatio_i is the probability of using ratio in comparison i. It is a sigmoid function employed in decision making literature as a soft threshold (Cisek, 2006). k and Thr are free parameters of the sigmoid, both with uniform priors: $k \sim U(0,20)$, Thr $\sim U(0, 1)$. rDen_i is the ratio between the smaller and larger denominator. In concrete terms, the sigmoid compares denominators to a threshold, and if they are too close, the probability that participants use ratio decreases.

To sample from the posteriors defined by the model (see Figure 4), we used an MCMC sampler (Stan, 2016) with 8 chains, each one with 10,000 iterations (half as warm-up). Convergence was determined with the r hat measure. This measure was <1.1 for all latent parameters. Winners (21 unique winners) and losers (23 unique losers) percepts are estimated differently to account for

potential artifacts, say aversion to losers. Also, ratios with different numbers are estimated on their own (75 unique ratios). That is, we allowed for the possibility for ratios like 3/4 and 21/28 to be perceived differently. We reasoned that this complexity could help some of the hypotheses, like holistic ratio or number weighting, to better explain human performance (see the supplemental materials for the inferred percepts).

For model comparison we used two measures: WAIC, and R^2 . The first is a method to compare Bayesian models with penalties for complexity (Vehtari, Gelman, & Gabry, 2016). Lower values indicate better models (as a rule of thumb, differences greater than 10 are considered relevant). The second one (R^2) compares human data to the mean of the posterior predictive distribution p(Choice_{pred}|Choice).

Results

The first notable outcome is that the decision weights for winners, losers, and ratios percepts were not zero (see Figure 5). The credible intervals (95% intervals obtained from the posterior) were as follows: winners [0.045, 0.070, 0.099]; losers [-0.027, -0.009, 0.007]; and ratios [3.040, 3.584, 4.263]; the center value is the median of the posterior. These are unnormalized weights, which explains why for winners and losers the values are smaller: The units are different. Importantly, the mass of the posterior is clearly tilted away from zero (see Figure 5), consistent with a mental comparison that includes ratio and numerical values: an intrinsic whole number bias.

It is still possible that subjects were strategic, changing between a decision based on ratio to one based in whole number (Schneider & Siegler, 2010), and the model that weights all the cues is just blind to this possibility. Because of this we implemented a strategic model to directly compare both alternatives. Intrinsic number biases (Figure 6A) and strategic behavior (Figure 6B) reproduced the qualitative patterns of observed performance: probability distance effects and whole number biases. A visual inspection of Figure 6 also reveals that the intensity of whole number biases relies on probability distance, suggesting again that the interaction found in Experiment 2 is related to discernibility of the probabilities (floor effect) and how perceptual information is weighted. The obtained 95% credible intervals for the sigmoid parameters of the strategic model were as follows: k [5.224, 7.366, 10.183] and threshold [0.577, 0.638, 0.697]; the center value is the median.



Figure 5. 95% credible intervals for decision weights. For ratios and winners, the weights are highly unlikely to be zero, suggesting that participants computed ratios but continued to use numerosity: an intrinsic number bias. The blue dot in the center of the intervals is the median. See the online article for the color version of this figure.



Figure 6. Mean accuracy obtained by the models. Dotted lines are human data, solid lines mean of the posterior predictive. Larger R^2 and lower WAIC indicate better fits to the data. See the online article for the color version of this figure.

On both measures of model comparison, WAIC and R^2 , the "intrinsic number bias" rule is better at predicting the data (Figure 6A). The strategic account flips between ratio in some trials and whole number in others, which produces lower accuracy across probability distances (Figure 6B). To further confirm that strategic behavior was not driving participants, we tested a numerator and ratio strategy. Instead of using the closeness of denominator to determine ratio or number-based choice, we used the distance of numerators (if too close compare losers, else ratios: e.g. in 10/12 vs 10/15 the numerators are identical so the model picks, with high probability, the one with fewer losers, but in 10/12 vs 5/15 the numerators are apart and the model picks based on the largest ratio, namely 10/12) and ratios (if too close compare numerators, else ratios: e.g. in 10/20 vs 5/10 the ratios are identical so the model picks, with high probability, the larger numerator, while in 10/20 and 9/10 the ratios are apart and picks based on the larger one). The outcomes were subpar from the intrinsic number bias (numerator strategy: WAIC = 3727 (250); $R^2 = -0.45$; ratio strategy: WAIC = 3316(152); $R^2 = 0.36$).

To see how denominator neglect, whole number weighting, and holistic ratio would behave, we fixed the appropriate weights to zero and ran the sampler (e.g., in denominator neglect the weights for losers and ratio were set to zero, and only the percept of winners was inferred). In general, the implementation of these models fared worse than the intrinsic number bias (supplemental Figure 3).

Overall, the "intrinsic number bias" rule, in which people simultaneously compare the whole numbers of the numerators and the holistic probability, and factor both into their judgment, explains behavior the best. This is not an artifact of data aggregation. At an individual level the same pattern emerges (see Figure 7): People appear to factor whole numbers into their ratio comparisons across all problems, not just strategically selected ones.

People, however, seem to rely more strongly in numerosity in ambiguous trials, relative to the intrinsic whole number bias model. When both bags had the same ratio, the model and participants revealed a distance effect: The probability of picking the bag with more winning balls decreased when their values got closer (see Figure 8). Because the model also exhibits this behavior, a simple correlation is not sufficient to claim a pure



Figure 7. 95% credible intervals of decision weights (A–C) and model comparison by subject (D). For all subjects, decision weights were not concentrated at zero; the blue dot in the center of all the intervals is the median (A–C). In (D) each dot is the subject-level comparison of the WAIC for the strategic and nonstrategic models (lower values indicate better fit). The solid red lines above and below the dashed line are differences in WAIC that are considered substantial (>10). The dashed line is at zero. See the online article for the color version of this figure.

numerosity-based behavior. What is certain is that human subjects in these ambiguous trials seem to follow numerosity with more vigor (slopes and intercept in Figure 8). This human nuance is not fully captured by the framework presented here, at least not in terms of the effect size.

Discussion

When rapidly and nonverbally choosing between gamble probabilities, subjects preferred options with a greater numerosity of winners; this finding shows evidence of a whole number bias in humans' rapid nonverbal comparisons of ratios. However, contrary to some reports, this effect was not a consequence of weak ratio representations as revealed by subjects' explicit knowledge of relative ratios in Experiment 1 and by their overall accuracy and sensitivity to ratio distance in Experiment 2. Using a Bayesian computational framework, we tested 5 cognitive rules that could explain subjects' behavior. The cognitive model required holistic ratio and the whole numbers to successfully reproduce subjects' performance. Taken together, these results are consistent with the hypothesis that we rely on intuitive mental representations of ratio that are spontaneously affected by whole number comparisons.

People's intrinsic reliance on whole numbers during ratio comparison may be a naïve theory (S. A. Gelman & Noles, 2011). Naïve theories formed early in development can support (or hinder) more advanced notions and actions later in adulthood, for instance in mathematics and physics (Hespos & VanMarle, 2012). In our nonverbal probability judgment task, two core representations, number (Carey & Spelke, 1994) and ratios (Denison & Xu, 2010), were concurrently activated and influenced choice. The use of whole numbers during rapid nonverbal quotient comparisons (even when it is maladaptive) could suggest that adults' difficulties comparing quotients have a nonverbal, early developing origin. There is a rich set of results connecting perceptual systems for representing numerosity and mathematics achievement (Halberda et al., 2008; Park & Brannon, 2013), and it is possible that intuitions from perceptual systems that compute numerosity exert a persistent influence over concepts of symbolic quotients like fractions, probabilities, and ratios.

The second most plausible explanation of the whole number bias in our results is that participants were strategic, flipping between whole number and ratio. The computational implementation of the strategic hypothesis fell short in predicting our data. Although strategic behavior indisputably occurs during symbolic ratio processing (Faulkenberry & Pierce, 2011; Fazio et al., 2016), our results just suggest that strategic behavior does not fully capture whole number biases in rapid nonverbal probability judgments. These results raise the possibility that the emergence of the whole number strategy during quotient comparisons may have deeper roots in nonverbal perception.

An intriguing possibility is that adults preferring, for example, 4/8 to 1/2 may reflect an incorporation of sample size into proportional judgment that considers 4/8 as a better instance of a 0.5 proportion (assuming that subjects treat each image as a sample from an underlying population). Thus, our results are perhaps evidence for a broader mechanism at play in proportional reasoning: confidence in larger samples (Alonso-Díaz, 2017; Obrecht & Chesney, 2013). However, our generative model did have ratio representations modulated by sample size (i.e., Beta(Number of Winners + 1, Number of Losers + 1)). Only using such ratio representation to make decisions fell short in explaining human behavior. We think that sample size surely plays a role, particularly in tuning up ratio representations (Alonso-Díaz, 2017), but an intrinsic number bias seems to still be present even in the presence of sample-size effects.



Figure 8. Preference toward the more numerous bag, in trials when both bags had the same ratio, decreased when the difference in the numerator decreased. A linear regression (solid line) confirmed that both humans (Slope = -0.19, t = -3.61, p < .01) and the intrinsic whole number bias model (Slope = -0.07, t = -3.16, p < .01) had this distance effect in ambiguous trials. However, human subjects were much more affected by numerosity in uncertain trials (slopes and intercepts; note the difference in scale on the *y*-axis). See the online article for the color version of this figure.

A final consideration is that the intrinsic number bias may be part of a general phenomenon by which all available magnitudes, be that of numerosity, ratios, length, or other, are activated and affect choice. Even though our data cannot speak decisively on this issue, discrete fractions do seem to be especially hard for math learners. Young children can match two continuous proportions but struggle with discrete ones when numerical matches are possible (Boyer, Levine, & Huttenlocher, 2008). Thus, it is possible that the bias we reported comes exclusively from cognitive processors in charge of discrete, rather than continuous, magnitudes.

References

- Alonso-Díaz, S. (2017). *Number representation in perceptual decisions* (Doctoral dissertation). University of Rochester, Rochester, NY.
- Bonato, M., Fabbri, S., Umiltà, C., & Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1410– 1419. http://dx.doi.org/10.1037/0096-1523.33.6.1410
- Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, 44, 1478–1490. http://dx.doi.org/10.1037/a0013110
- Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/ CBO9780511752902.008
- Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *The Journal of Neuroscience*, 26, 9761–9770. http://dx.doi.org/10.1523/JNEUROSCI .5605-05.2006
- Dehaene, S. (2009). Origins of mathematical intuitions: The case of arithmetic. Annals of the New York Academy of Sciences, 1156, 232–259. http://dx.doi.org/10.1111/j.1749-6632.2009.04469.x
- Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 21, 1083–1092. http://dx.doi.org/10.1177/01461672952110009
- Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, 13, 798–803. http://dx.doi.org/10.1111/j.1467-7687.2009.00943.x
- DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction*, 37, 39–49. http://dx.doi.org/10.1016/j.learninstruc.2014.07 .002
- Fabbri, S., Caviola, S., Tang, J., Zorzi, M., & Butterworth, B. (2012). The role of numerosity in processing nonsymbolic proportions. *Quarterly Journal of Experimental Psychology*, 65, 2435–2446. http://dx.doi.org/ 10.1080/17470218.2012.694896
- Faulkenberry, T. J., & Pierce, B. H. (2011). Mental Representations in Fraction Comparison Holistic Versus Component-Based Strategies. *Experimental Psychology*, 58, 480–489. http://dx.doi.org/10.1027/1618-3169/a000116
- Fazio, L. K., DeWolf, M., & Siegler, R. S. (2016). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1–16. http:// dx.doi.org/10.1037/xlm0000153
- Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 17075–17080. http://dx.doi.org/10.1073/pnas.1410583111
- Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review*, 121, 96–123. http://dx.doi.org/10.1037/a0035232
- Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial

and transcendent constructions. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 293–322). Hillsdale, NJ: Erlbaum.

- Gelman, S. A., & Noles, N. S. (2011). Domains and naïve theories. WIREs: Cognitive Science, 2, 490–502. http://dx.doi.org/10.1002/wcs.124
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internetbased sample. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 11116–11120. http://dx.doi.org/10.1073/ pnas.1200196109
- Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455, 665–668. http://dx.doi.org/10.1038/nature07246
- Hespos, S. J., & vanMarle, K. (2012). Physics for infants: Characterizing the origins of knowledge about objects, substances, and number. WIREs: Cognitive Science, 3, 19–27. http://dx.doi.org/10.1002/wcs.157
- Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: The brain's code for proportions. *Trends in Cognitive Sciences*, 16, 157–166. http://dx.doi.org/10.1016/j.tics.2012.02.002
- Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuomotor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 16271–16276. http://dx.doi.org/10.1073/pnas .1300239110
- Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 63, 534–544. http://dx .doi.org/10.1037/0022-3514.63.4.534
- McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, 18, 740–745. http://dx.doi.org/10.1111/ j.1467-9280.2007.01969.x
- Meert, G., Grégoire, J., & Noël, M. P. (2009). Rational numbers: Componential versus holistic representation of fractions in a magnitude comparison task. *Quarterly Journal of Experimental Psychology*, 62, 1598– 1616. http://dx.doi.org/10.1080/17470210802511162
- Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L., & Tadin, D. (2013). A strong interactive link between sensory discriminations and intelligence. *Current Biology*, 23, 1013–1017. http://dx.doi.org/10.1016/j.cub .2013.04.053
- Ni, Y., & Zhou, Y. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, 40, 27–52. http://dx.doi.org/10.1207/s153269 85ep4001_3
- Obrecht, N. A., & Chesney, D. L. (2013). Sample representativeness affects whether judgments are influenced by base rate or sample size. *Acta Psychologica*, 142, 370–382. http://dx.doi.org/10.1016/j.actpsy .2013.01.012
- O'Grady, S., Griffiths, T., & Xu, F. (2016). Do simple probability judgments rely on integer approximation? In Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 1008–1013). Austin, TX: Cognitive Science Society.
- Pacini, R., & Epstein, S. (1999). The interaction of three facets of concrete thinking in a game of chance. *Thinking & Reasoning*, 5, 303–325. http://dx.doi.org/10.1080/135467899393959
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, 24, 2013– 2019. http://dx.doi.org/10.1177/0956797613482944
- Passerini, G., Macchi, L., & Bagassi, M. (2012). A methodological approach to ratio bias. *Judgment and Decision Making*, 7, 602–617.
- Piantadosi, S. T. (2016). Efficient estimation of Weber's W. Behavior Research Methods, 48, 42–52.
- Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition*, 131, 60–68. http://dx.doi.org/10.1016/j.cognition.2013.12.011

- Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18, 89–107. http://dx.doi.org/10.1016/j.lindif.2007 .03.011
- Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception* and Performance, 36, 1227–1238. http://dx.doi.org/10.1037/a0018170
- Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences*, 17, 13–19. http://dx.doi.org/10.1016/j.tics.2012.11 .004
- Stan, D. T. (2016). *RStan: The R interface to Stan.* Retrieved from http://mc-stan.org
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. Cambridge, MA: MIT Press.

- Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC (arXiv Preprint). Retrieved from http://arxiv.org/abs/1507.04544
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137. http://dx.doi.org/10.1111/1467-9280.00120
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. Proceedings of the National Academy of Sciences of the United States of America, 105, 5012–5015. http://dx.doi.org/10.1073/pnas.0704450105

Received July 24, 2017

Revision received February 12, 2018

Accepted February 14, 2018