Journal of Experimental Psychology:
Learning, Memory, and Cognition
2013, Vol. 39, No. 3, 782–800

0278-7393/13/$12.00 DOI: 10.1037/a0029670

Naïve Point Estimation

Marcus Lindskog, Anders Winman, and Peter Juslin
Uppsala University

The capacity of short-term memory is a key constraint when people make online judgments requiring them to rely on samples retrieved from memory (e.g., Dougherty & Hunter, 2003). In this article, the authors compare 2 accounts of how people use knowledge of statistical distributions to make point estimates: either by retrieving precomputed large-sample representations or by retrieving small samples of similar observations post hoc at the time of judgment, as constrained by short-term memory capacity (the naïve sampling model: Juslin, Winman, & Hansson, 2007). Results from four experiments support the predictions by the naïve sampling model, including that participants sometimes guess values that they, when probed, demonstrably know have the lowest probability of occurring. Experiment 1 also demonstrated the operations of an unpredicted recognition-based inference. Computational modeling also incorporating this process demonstrated that the data from all 4 experiments were better predicted by assuming a post hoc sampling process constrained by short-term memory capacity than by assuming abstraction of large-sample representations of the distribution.

Keywords: point estimation, sampling model, intuitive statistics

At least since the works of Brunswik (1955), the human mind has been likened to an intuitive statistician (Gigerenzer & Murray, 1987; Peterson & Beach, 1967). It is often assumed that the adaptive problems that people meet in their environment are similar to those addressed in statistics and probability theory and, therefore, that the mind is likely to have adopted similar solutions to its adaptive problems (for more recent Bayesian incarnations of the intuitive statistician, see, e.g., Anderson, 1990; Griffiths & Tenenbaum, 2006; Oaksford & Chater, 2009).

Research on judgment and decision making has often—in very general terms—been motivated by the notion of bounded rationality, namely, the insight that people have limited time, knowledge, and computational ability (Simon, 1990). It is more rare that the implications of the specific nature of these constraints are carefully analyzed, such that people often, quite literally, have to rely on very small samples (but see Dougherty & Hunter, 2003; Gaissmaier, Schooler, & Rieskamp, 2006; Hansson, Rönnlund, Juslin, & Nilsson, 2008; Juslin, Winman, & Hansson, 2007; Kareev, Amon, & Horwitz-Zeliger, 2002; Stewart, Chater, & Brown, 2006) or, because of capacity constraints, have to rely on linear additive integration (Juslin, Nilsson, & Winman, 2009; Juslin, Nilsson, Winman, & Lindskog, 2011; Nilsson, Winman, Juslin, & Hansson, 2010).

In this article, we extend the previous research on man as an intuitive statistician by exploring how short-term memory constraints affect how people are able to use their knowledge of statistical distributions to make point estimates. More precisely, on the basis of four experiments and computational modeling, we evaluate if the accuracy of point estimates is constrained by the overall sample size stored in long-term memory (LTM) or by short-term memory (STM) constraints.

Point Estimates and Knowledge of Distributions

How do people make point estimates? Brown and Siegler (1993) emphasized the importance of knowledge of both the metric properties of the quantity, like its mean, variance, and distribution, and the mapping properties of the quantity, which involve knowledge about the ordinal relations based on domain-specific knowledge and heuristics (see also von Helversen & Rieskamp, 2008). Although mapping knowledge has been emphasized in research on heuristics (e.g., Gilovich, Griffin, & Kahneman, 2002), multiple-cue judgment (Juslin, Karlsson, & Olsson, 2008; von Helversen & Rieskamp, 2008), and function learning (DeLosh, Busemeyer, & McDaniel, 1997; Kalish, Lewandowsky, & Kruschke, 2004), the effects of metric knowledge on point estimates have been largely ignored (but see Pitz, Leung, Hamilos, & Terpening, 1976).

In the literature on forecasting (e.g., Goodwin, 1996) and in economic theory (e.g., Engelberg, Manski, & Williams, 2009), people are often assumed to use the central tendency of a subjective probability distribution as their point prediction, where this distribution manifests both their metric and their mapping knowledge. Intriguingly, however, there appears to be no single normative principle that can guide point estimates (Lehmann & Casella, 1998), and it is unclear from previous research how metric knowledge influences point estimates.

When people estimate unbiased descriptive statistics, like the arithmetic mean, the estimates generally coincide decently with the predictions from statistical theory (Peterson & Beach, 1967; Spencer, 1961, 1963), with estimates being uninfluenced by distribution

This article was published Online First August 20, 2012.

Marcus Lindskog, Anders Winman, and Peter Juslin, Department of Psychology, Uppsala University, Uppsala, Sweden.

This research was sponsored by the Swedish Research Council and the Swedish Tercentary Bank Foundation.

Correspondence concerning this article should be addressed to Marcus Lindskog, Department of Psychology, Uppsala University, Box 1225, SE-751 42 Uppsala, Sweden. E-mail: marcus.lindskog@psyk.uu.se

shape (skewed, normal, or bimodal) and variance (Malmi & Samson, 1983). People are responsive to variance (Evans & Pollard, 1985) but have problems making accurate estimates (Beach & Scopp, 1968; Kareev et al., 2002; Pollard, 1984; Slovic, Fishchoff, & Lichtenstein, 1977).

Research on knowledge of other statistical properties has mainly addressed whether people can estimate the shape of distributions they have encountered in their everyday lives (Fox & Thornton, 1993; Griffiths & Tenenbaum, 2006; Linville, Salovey, & Fischer, 1989; Nisbett & Kunda, 1985), such as the frequency of death causes (Hertwig, Pachur, & Kurzenhauser, 2005; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978), the grade point average of fellow students (Nisbett & Kunda, 1985), or the baking time of pastries (Griffiths & Tenenbaum, 2006). The results are mixed. In some cases, people's knowledge of the distributions appears biased by the external information in the environment (e.g., media exposure and death causes; Lichtenstein et al., 1978); in other cases, it is remarkably accurate (e.g., Griffiths & Tenenbaum, 2006; however, see Mozer, Pashler, & Homaei, 2008). People often put more weight on observations close to where they find themselves in the distribution (Fiedler, 2000; Nisbett & Kunda, 1985).

Several accounts of human judgment and decision making assume that people sample information from memory prior to making a judgment or decision (e.g., Busemeyer & Townsend, 1993; Fiedler, 2000; Kahneman & Miller, 1986; Stewart et al., 2006; Tversky & Koehler, 1994). In decision by sampling theory (Stewart et al., 2006), for example, the value of a target is determined by its relative rank in a small sample retrieved from memory, which explains, among other things, asymmetries in subjective value of gains and losses. Thus, it seems that people are prone to base judgments on small samples from memory (see also Dougherty & Hunter, 2003; Gaissmaier et al., 2006; Hansson et al., 2008; Kareev et al., 2002). Here we extend this research by exploring how point estimates are influenced by the reliance on small samples.

Two Models of Naïve Point Estimation

People often have to draw on their knowledge to make point estimates of unknown quantities. Assume, for example, that you have previously made 1,000 observations of yearly revenues of companies operating in a market and the distribution of these numbers is illustrated in Figure 1A (i.e., a unimodal distribution). Now a new and unknown company is randomly sampled from this market and your task is to predict or make a best guess about the revenue of the new company. What would be your best guess?

People can use knowledge of distributions to make point estimates in at least two different ways: by retrieving precomputed large-sample representations of distribution properties or by retrieving samples of observation post hoc at the time of judgment and use the sample properties to estimate the population properties. The first stages are similar with both processes: In the environment, there exist variables, natural or artificial, described by *objective environmental distributions* (OEDs). A biased or a representative subset of the observations in the OED is experienced by a person and becomes encoded in LTM in the form of a *subjective environmental distribution* (SED; Juslin et al., 2007). For example, a person may have stored in LTM the revenues of some companies operating on a market (the SED), which is a subset of all the companies operating on this market (the OED).

Constrained by LTM: Capitalizing on Experience

One possibility is that the point estimates derive from retrieval of large-sample representations. Precomputed large-sample-based representation may either derive from explicit attempts to abstract statistical properties during exposure to the distribution, as when a person keeps and updates a running mean of a variable as new observations are made, or, hypothetically, arise from corresponding preconscious and automatic computations (Zacks & Hasher, 2002). The hallmark of large-sample representations is that if a person benefits from an increasing number of observations in the SED with more experience, one expects the knowledge to become more consistent and accurate with more experience. For example, given random sampling from an OED with standard deviation σ, the law of large numbers implies a standard error of the mean computed from N observations in the SED of approximately $\sigma/\sqrt{N}$, which converges to zero as N increases. If people use large-sample estimates of the population mean for point estimation, with increasing experience, the point estimates should converge on the population mean.

Constrained by Short-Term Memory: The Naïve Sampling Model (NSM)

A second possibility is that observations are stored in LTM but, prior to the judgment, no abstraction of distribution properties occurs. Although, in principle, it could be possible that a person retrieves the entire SED at the time of judgment, we follow research suggesting that such online judgments are constrained by short-term memory capacity (e.g., Dougherty & Hunter, 2003; Gaissmaier et al., 2006; Hansson et al., 2008; Kareev et al., 2002; Stewart et al., 2006). At the time of judgment, a sample of n observations is retrieved, temporarily becoming active in short-term memory, and a property of this sample is used as a direct proxy for the population property (Juslin et al., 2007). The sample of n active observations in short-term memory, typically estimated to approximate 4 ± 2 observations (Cowan, 2001), is referred to as the *subjective sample distribution* (SSD). A person may, for example, not know the average salary of people working at his or her work place but may retrieve a number of known salaries and estimate the average salary, or "point estimate" the salary of a specific individual.

The qualification *naïve* in the NSM refers to the presumption of people that a sample property can be taken directly to describe population properties. For unbiased statistical properties, direct use of sample properties produces reasonable judgments, but direct (uncorrected) use of inherently biased properties of small short-term memory constrained (STMC) samples leads to bias. Sample proportion is an unbiased estimator, on average reproducing the population proportion, and direct use of sample proportion is likely to yield accurate assessments. By contrast, sample variability is inherently biased, which is why it needs to be corrected by $n/(n - 1)$ to be an unbiased estimate of the population variance. Uncorrected use of sample variance thus leads to bias. The implication is that the judgment is constrained and sometimes systematically biased by being generated from a small sample. For example, given random sampling from an

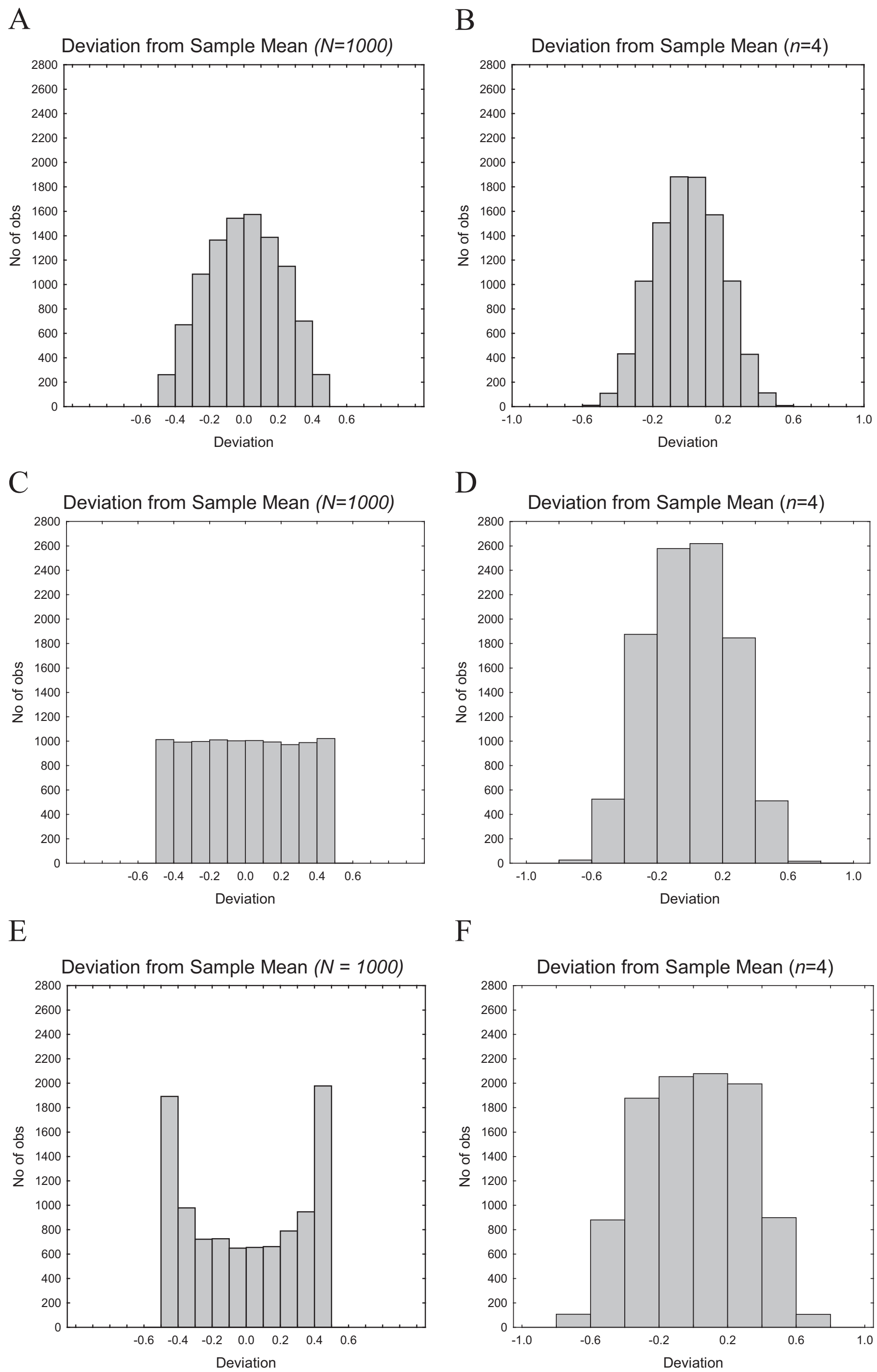

Figure 1. Average distribution shape at sample size n for a unimodal distribution ($n = 1{,}000$ in A, $n = 4$ in B), a uniform distribution ($n = 1{,}000$ in C, $n = 4$ in D), and a bimodal distribution ($n = 1{,}000$ in E, $n = 4$ in F). The *x*-axis is the observation minus the mean in the sample (e.g., 0 is an observation that coincides with the sample mean). The distributions were obtained by sampling repeatedly from beta distributions with parameters $\alpha = \beta = 3$ (unimodal distribution, A and B), $\alpha = \beta = 1$ (uniform distribution, C and D), and $\alpha = \beta = .3$ (bimodal distribution, E and F). No of obs = number of observations.

OED with standard deviation σ, the standard error of the estimated mean in the OED based on n observations in the SSD approximates $\sigma/\sqrt{n}$, regardless of experience (N in the SED). The counterintuitive prediction in this regard is that the point estimates should not converge on the population mean with more experience but show the same variability for people with little experience (small SED) and people with extensive experience (large SED).

An Illusion of Unimodality With the NSM

Figure 1 illustrates an intriguing property of small samples: Not only are they often insufficient as a basis for inferring the distribution shape, but, if anything, they are misleading about the distribution shape. One way to represent the impression of distribution shape that is conveyed in a sample is in terms of the deviations from the sample mean. A sample that mainly contains observations close to the sample mean and few observations far from the sample mean will naturally yield an impression of unimodal distribution shape (see Figure 1A). By contrast, a sample in which few observations are close to the sample mean but many observations have large deviations from the sample mean will signal a bimodal shape (see Figure 1E).

Figure 1 presents deviations from the sample mean at sample size n for a unimodal distribution ($n = 1{,}000$ in Figure 1A, $n = 4$ in Figure 1B), a uniform distribution ($n = 1{,}000$ in Figure 1C, $n = 4$ in Figure 1D), and a bimodal distribution ($n = 1{,}000$ in Figure 1E, $n = 4$ in Figure 1F). The score on the x-axis in these figures is the observation minus the mean in the sample (e.g., a score of 0 is an observation that coincides with the mean in the sample). With a unimodal distribution, the large sample in Figure 1A accurately portrays the underlying unimodal population distribution (a beta distribution with $\alpha = \beta = 3$). A sample size of 4 inherently makes drawing inferences about distribution shape difficult, but at least, on average, it will correctly suggest that most observations are close to the mean (see Figure 1B) and, on average, correctly convey the unimodal distribution shape.

However, although the large sample in Figure 1C from a uniform distribution (beta distribution with $\alpha = \beta = 1$) accurately conveys the distribution shape, the small samples in Figure 1D will, if anything, suggest a unimodal distribution, where most observations fall close to the sample mean. Moreover, although a large sample in Figure 1E accurately portrays the distinctly bimodal population distribution (beta distribution with $\alpha = \beta = .3$), the small sample in Figure 1F also suggests a unimodal distribution.[1] In short, perceiving the world through small samples not only makes it difficult to detect the underlying distribution shape, it will, if anything, convey an illusion of unimodality.[2] This illusion is, of course, confined to small samples ($n = [3, 8]$), but it is quite pervasive in the regions of sample size that are made relevant by the constraints on short-term memory capacity. Intuitively, in small samples, it is unlikely that both modes of a bimodal distribution are represented and the sample mean will chase the sampling error in the few observations in the sample. That people enter category learning tasks with what appears to be an implicit expectation of a normal distribution (Fried & Holyoak, 1984; Flannagan, Fried, & Holyoak, 1986) and exhibit a bias to assume normality when generating social distributions (Nisbett & Kunda, 1985) may be related to small samples looking unimodally distributed.

Summary of Predictions

We follow previous research suggesting that people can shift between strategies for making point estimates (Lawrence & O'Connor, 2005; Peterson & Miller, 1964; Weber, 1994; Winkler, 1970). This research shows that people have some ability to shift between different strategies for making point estimates as a function of strong and clear incentives and a sufficient amount of training. In the experiments reported in this article, however, our main interest lies not in the ultimate malleability of the strategies but in disclosing the strategies that people spontaneously engage in when making online point estimates in the absence of specific incentive structures during learning. We hypothesize that when a distribution is perceived to be unimodal, people are naturally inclined to make a point estimate close to the mean of the distribution, as suggested by the ambitions both to avoid large deviations (as formalized by least-squares minimization) and to maximize the chance of being correct. When a distribution is perceived to be bimodal, we propose that there is a conflict between the ambitions to avoid large deviations, suggesting a mean response, and to maximize the chance of being correct, suggesting an extreme value.

If people can accrue abstract large-sample representations, their knowledge of the distributions and their properties should improve

[1] Another way to represent the distribution shape is in terms of the interquartile distance (I) between the values of the 75th and the 25th percentiles, covering 50% of the distribution, relative to the distance R between minimum and maximum value (for finite dimensions). Define the following score, $U = .5 \times R - I$. In a uniform distribution, $U = 0$ (e.g., approximately true of the distribution in Figure 1C); in a unimodal distribution, $U > 0$ (e.g., in Figure 1A, $U = .17$); in a bimodal distribution, $U < 0$ (e.g., in Figure 1A, $U = -.18$). Defining the corresponding scores for samples of size 4 is inherently difficult, because the limits defining the corresponding intervals are less well defined (e.g., if A, B, C, and D are the values in the samples ordered in terms of their magnitude, any combination of cutoffs between A and B and between C and D will cover the central 50% of the distribution). However, even if we define the interval I as conservatively as possible, simply as the difference (C − D) between C and D, the strongly negative U of −.18 for the large sample bimodal distribution in Figure 1E turns into a slightly positive (unimodal) score of .03, when U is computed from samples of size 4 (see Figure 1F). This illustrates the same relative shift as in Figure 1 from the strong bimodality that is visible in a large sample (or the population) to the relatively much more frequent impression of unimodality conveyed in small samples.

[2] Although the phenomenon is limited to small samples, when we consider sampling error alone, it is worth remembering that similar effects will arise also from other sources of random error. In general, whenever a random error e is added to a score x to produce an observed score y (i.e., $y = x + e$) and regardless of the original distribution of the score x, the distribution of y very soon converges to a unimodal distribution. In Figure 1, the score is the deviation from the mean, and the error in this score arises from the uncertain estimation of the population mean, which is larger, the smaller the sample. Small samples thus convey the impression that most observations are close to the (sample) mean, as typical of unimodal distributions. Notably, the results in Figure 1 only address sampling error. To the extent that additional sources of random (neural) error arise in the processes of encoding, storage, or retrieval from LTM (e.g., Erev, Wallsten, & Budescu, 1994; Juslin, Olsson, & Björkman, 1997), this distortion toward a small-sample impression of a unimodal distribution will be further strengthened.

with experience. If people benefit from large-sample representations, with a unimodal distribution, we thus expect point estimates to converge on the population mean with training. With a bimodal distribution, we expect people to either (a) make point estimates that similarly converge on the population mean, if they are driven by the ambition to avoid large errors, or (b) increasingly make extreme responses as they learn more about the bimodal distribution shape, if they are driven by the ambition to maximize the probability of being right. The NSM suggests that the small samples that people have at their disposal are too small to reliably detect a bimodal distribution shape. Thus, after extensive experience and even if the payoff schedule explicitly encourages maximizing behavior (as in Experiments 3 and 4 below), there should be a tendency to rely on a mean strategy.

Note an intriguing prediction by the NSM. When people are probed for the proportions of the distribution that fall in predefined intervals, there should be a potential for accurate production of the distribution shape, because sample proportion is an unbiased estimator. Thus, despite making mean estimates for bimodal distributions, if people use sample proportion to assess the population proportion of the distribution falling in intervals, they should—if probed in this way—disclose accurate knowledge of the bimodal distribution shape and, accordingly, at the same time articulate the belief that an estimate is very unlikely to fall close to the mean.

Present Study

In Experiment 1, we tested the prediction that people are inclined to make point estimates according to a mean strategy, even with a distinctly bimodal distribution, while at the same time they disclose the ability to correctly reproduce the bimodal distribution shape when probed for the distribution with a proportion format. Experiment 2 verified that this tendency to make point estimates in the midinterval was truly a response to the distribution experienced. In Experiment 3, we tested if this effect persists with a loss function that explicitly and strongly rewards guesses that are close to the correct value and investigated if the effect can be manipulated by changing the format into one that invites the learning of proportions. In Experiment 4, we used the same loss function and investigated how judgments are affected by amount of training.

Because Experiment 1 documented the unpredicted operation of an additional mechanism for making the point estimates, we wait to introduce a more formal treatment of the models until after Experiment 1. After reporting Experiment 1, we introduce models of the processes that incorporate this unpredicted influence, allowing us to compare the assumptions that sample size is constrained by long-term or short-term memory.

Experiment 1: Guessing on the Unlikely

In a learning phase, participants observed the revenues of 60 fictional companies in a market with either a unimodal or a bimodal revenue distribution. Subsequently, they made point estimates of revenues both of companies seen in the learning phase and of companies not previously experienced but randomly drawn from the same distribution. Finally, they reproduced the overall population distribution by means of proportion estimates for predefined intervals. We predicted that the participants would often make point estimates close to the population mean for new companies, regardless of the distribution shape and despite their demonstrably accurate knowledge of the distribution shapes when probed with proportion estimates.

Method

Participants. Fifteen male and 15 female undergraduate students from Umeå University with an average age of 25 years each received 150 Swedish kronor (approximately $20) for their participation.

Materials and apparatus. The computerized judgment task involved estimation of revenues of fictive companies. Two distributions of 60 revenues, a symmetrical unimodal distribution (beta distribution with $\alpha = \beta = 3.4$), and a symmetrical bimodal distribution (beta distribution with $\alpha = \beta = .33$) defined the two conditions. Both distributions were linearly transformed to a [1, 1,000] interval. For each participant, the revenues were randomly paired with one of 156 company names.

Design and procedure. Participants were randomly assigned to the unimodal or the bimodal conditions. During learning, participants were presented with a company on each trial and were asked to guess its revenue, after which they received feedback on the correct revenue. Learning continued until either a correct rate of 40% (24) of the 60 values had been achieved or a maximum of 400 trials was reached. A response was considered correct if it exactly matched the correct revenue, and a value was required to be correctly reported once. In a test phase, point estimates were obtained for each of the 60 old companies from the learning phase and for 60 new companies. Participants also assessed how many of the 60 companies that were encountered in the training phase fell into 10 predefined, equally wide intervals ([1, 100], [101, 200], . . . , [901, 1,000]).

Results and Discussion

Figures 2A and 2B present average assessed distribution shape based on assessing relative frequencies (proportions) that fall in 10 intervals on the continuum (the grey bars), compared with the objective distribution shapes (the white bars). The empirical distributions of point estimates for old exemplars seen in training are presented in Figures 2C and 2D, and the empirical distributions of point estimates for new exemplars first encountered in the test phase are presented in Figures 2E and 2F; in both cases, they are classified in the same 10 intervals on the continuum (grey bars). (The black bars in Figures 2C–2F for model predictions are discussed later.)

Figure 2 suggests three conclusions. First, with relative frequency estimates, the participants had fairly accurate knowledge of the underlying distributions, reproducing the unimodality in the unimodal condition and the bimodality in the bimodal condition (see the white bars in Figures 2A and 2B). However, it is also clear from Figure 2 that although the participants were able to reproduce the unimodal distribution almost perfectly, performance was somewhat poorer in the bimodal condition, and the mean absolute error from the correct distribution was larger in the bimodal condition (.048 vs. .023 in the unimodal condition), $t(27) = 3.77$, $p < .001$.

Second, although the point estimates for old exemplars (the grey bars in Figures 2C and 2D) reproduce the underlying distributions (the white bars in Figures 2A and 2B), in both conditions, there is

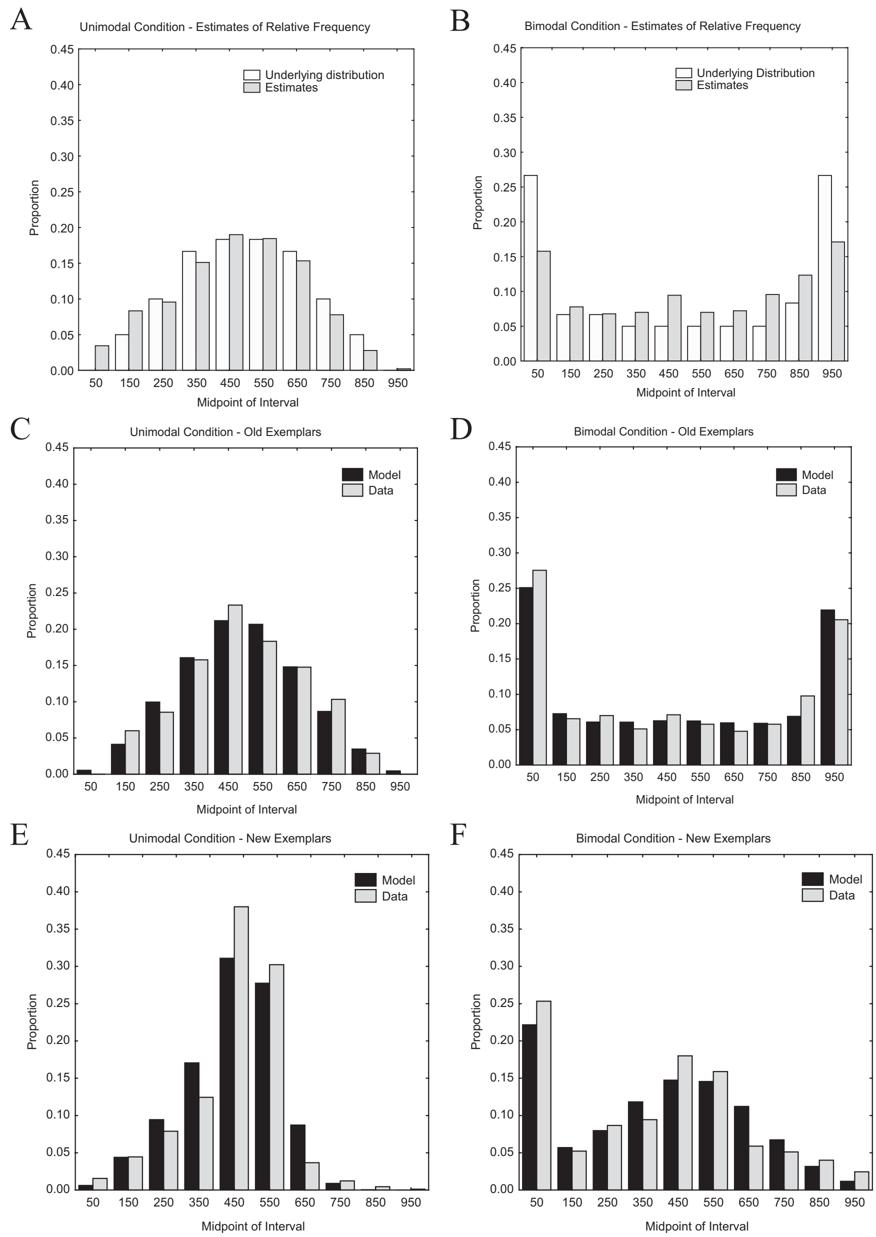

Figure 2. Average assessed distribution based on assessing relative frequency (A and B, grey bars) compared with the objective distribution shapes (A and B, white bars), and the empirical distribution of point estimates for old exemplars (C and D, grey bars) and the empirical distribution of point estimates for new exemplars (E and F, grey bars) compared with model predictions (black bars), shown for the unimodal (A, C, and E) and bimodal (B, D, and F) distributions separately.

a relative shift for the new exemplars toward more predictions in the intervals 450 and 550 close to the mean of the distribution (the grey bars in Figures 2E and 2F). In the unimodal condition, the proportion of point estimates in the 450 and 550 intervals increased from 26% on average for old exemplars to 57% on average for new exemplars, $t(15) = 4.20$, $p < .001$. In the bimodal condition, the proportion of point estimates in the 450 and 550 intervals increased from 7% on average for old exemplars to 27% on average for new exemplars, $t(15) = 2.62$, $p = .020$.

For each individual participant and each of the distributions presented in Figure 2 (assessed distribution shape, distribution of point estimates for old exemplars, and the distribution of point estimates for new exemplars), we fitted α and β parameters of a beta distribution to the distributions and classified them as unimodal if $\alpha > 1$ or $\beta > 1$ (or both) and as bimodal if both $\alpha < 1$ and $\beta < 1$. In the unimodal condition, 100% of the assessed distribution shapes, 100% of the distributions of point estimates for old exemplars, and 100% of the distributions of point estimates for new exemplars were classified as unimodal. In the bimodal condition, 87% of the assessed distribution shapes, 100% of the distributions of point estimates for old exemplars, but only 7% of the distributions of point estimates for new exemplars were classified as bimodal. In the bimodal condition, 87% of the participants predominantly made point estimates in the middle categories with the lowest probability of including the value, at the same time disclosing accurate knowledge about the bimodal distribution shape when probed in terms of proportions.

Third, there is a bias to make low point estimates that is especially pronounced for new exemplars. A two-way mixed analysis of variance (ANOVA) with condition (between subjects) and old versus new exemplar (within subjects) as independent variables and mean point estimate as the dependent variable shows a significant main effect of old versus new exemplar, $F(1, 28) = 10.16$, $p = .003$, $MSE = 7{,}560$, but no significant effect of condition, $F(1, 28) = 2.15$, $p = .153$, $MSE = 10{,}127$, and no significant interaction, $F(1, 28) = 1.20$, $p = .282$, $MSE = 7{,}560$. The mean point estimate for the old exemplars was 481.7 ($H_0 = 500$), $t(29) = 2.09$, $p = .045$, and the mean point estimate for the new exemplars was 410.1 ($H_0 = 500$), $t(29) = 3.91$, $p < .001$. The mean point estimate for new exemplars is significantly lower than 500 in both the unimodal condition ($M = 441.5$), $t(14) = 2.41$, $p = .030$, and the bimodal condition ($M = 378.7$), $t(15) = 3.18$, $p = .007$.

Notably, in the bimodal condition, there are two spikes in the distribution of point estimates for new exemplars, one across the two middle categories (450 and 550), as expected if the participants use a mean strategy, but also and more unexpectedly, one in the lowest category (50) that is not matched by a spike in the highest category (950). In principle, the spike of responses in the lowest category could represent an emergent insight about the bimodal distribution shape, along with a consequent shift from a mean strategy that implies guessing in a region that is very unlikely to include the value to a maximizing strategy in which participants instead attempt to maximize the probability of accuracy by guessing in a region that is more likely to includes the value.

This “shift-to-maximization” explanation of the bias is inconsistent with the results reported in this article on at least three grounds. First, the spike is only in the lowest category (50), not in the equally maximizing highest category (950). If the goal was to maximize, there are just as good reasons to make a guess in the highest category. Second, the maximizing account implies that the bias should only be observed for the bimodal distribution; in the unimodal distribution, the maximizing response is in the middle categories (450, 550), which does not imply any bias to make low point estimates. However, the bias to make low point estimates that is particularly strong with new exemplars exists for both unimodal and bimodal distributions. Third, and foreshadowing results from Experiment 4, the bias to make low point estimates, if anything, decreases rather than increases with additional knowledge of the distribution, which appears inconsistent with the effect arising from better insights about the distribution shape.

All of the above properties are consistent with a recognition-based inference. In the past decade, extensive research has indicated that recognition memory plays a large role in how people make judgments, as inspired by the *recognition heuristic* (RH), a term coined by Goldstein and Gigerenzer (1999). The journal *Judgment and Decision Making* recently dedicated three special issues to this construct. In an editorial note in the first of these issues, Marewski, Pohl, and Vitouch (2010) formulated the RH as follows: “*If there are N alternatives, then rank all n recognized alternatives higher on the criterion than the N − n unrecognized ones*” (p. 207, italics in original). The heuristic helps people to make inferences about an object’s criterion value when recognition memory is correlated with the criterion value, as is often the case (Marewski et al., 2010).

We surmise that the asymmetry in our results in terms of a disproportionately large amount of estimates in the lower ranges for new (unrecognized) items stems from a recognition-based strategy similar to the recognition heuristic, in that people infer that a recognition failure signals that the value of the object is low (“Small X if not recognized X”; see McCloy, Beaman, Frosch, & Goddard, 2010). Research has shown that people tend to rely more on the RH in environments in which it actually works and recognition validity is high (Gigerenzer & Goldstein, 2011). The cover story in the tasks addressed in this article concerned the revenue of companies, a criterion variable likely to be associated with high recognition validity in real environments. In other words, in general, the fact that you recognize a company suggests that it is a large company with large revenue rather than small company. A second and related possibility is that high numeric values may be intrinsically more memorable than low values, something that would offer recognition validity also directly in the laboratory setting. There are results in the following experiments that support this possibility (e.g., Experiment 2). Although the RH focuses comparisons between objects, we propose a mechanism that refers to inferences from recognition failures for single objects. Because of this discrepancy, we term this strategy a *recognition failure heuristic* (RFH), while we acknowledge that the concept is inspired by the RH.

This account explains why there is a bias to make low estimates in the distribution, why it applies both to the unimodal and the bimodal distributions, and why its prevalence does not increase with additional training. It also implies that, although the bias should occur in both distribution conditions, it should be nominally larger in the bimodal condition than in the unimodal condition. Because of the distribution shapes, the examples of low values retrieved by a participant in the bimodal condition are likely to be lower than the examples of low values retrieved by a participant in the unimodal condition. In the next section, we incorporate this recognition-based inference within the NSM framework to contrast models assuming short-term memory versus LTM constraints on sample size.

A naïve point estimation (NPE) model. In this section, we outline a model for NPE. The formal model and the model-fitting procedure are presented in the Appendix. The model is intended to reproduce the four distributions presented in Figures 2C, 2D, 2E, and 2F, that is, the distributions of separate point estimates for old and for new exemplars classified in terms of the 10 intervals across both the unimodal and the bimodal conditions ($N = 40$, $df = 36$).

In regard to point estimates, the model illustrated in Figure 3 assumes there are three possible processes when an individual is presented with a probe *X* and asked to give a best estimate of the value of *X*. First, if it is possible to retrieve the value of *X* from LTM, this value is the point estimate (direct retrieval; left branch). Second, if the value cannot be retrieved from LTM, the response

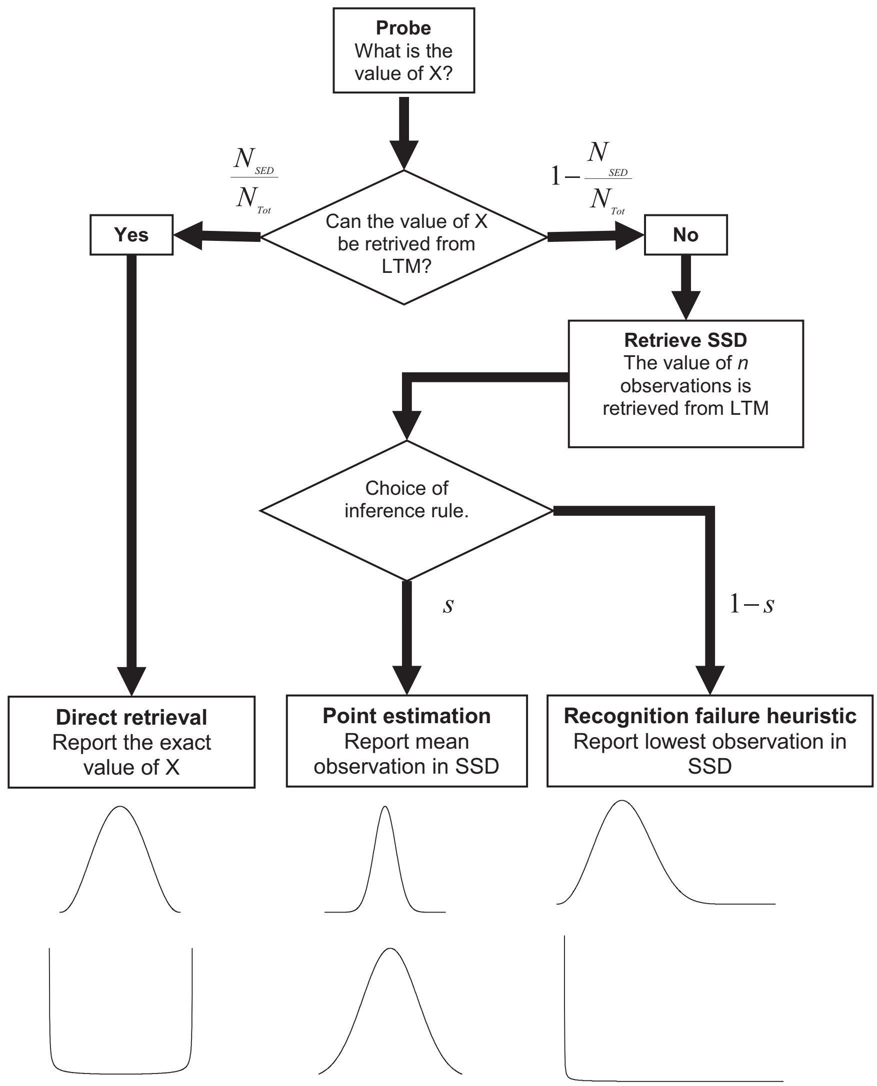

Figure 3. Schematic outline of the naïve point estimation model including three types of responses: direct retrieval, naïve point estimation, and recognition failure heuristic. The bottom two rows of probability density functions illustrates the predicted densities of the three respective processes when the underlying distribution is unimodal (top row) and bimodal (bottom row). The density function of the entire process will be a weighted average of these three densities. The weight each response type receives is determined by N_{SED} and *s*. LTM = long-term memory; SSD = subjective sample distribution; SED = subjective environmental distribution; Tot = total.

is derived from one of two inference rules. In both cases, a small sample (SSD) is activated in short-term memory. In the first case (NPE; middle branch), the mean of the SSD is reported as the point estimate (e.g., "the value of *X* is probably close to the sample mean, like most of the other values in the sample distribution"). In the second case (RFH; right branch), the lowest value in the SSD is reported (e.g., "Small *X* if not recognized *X*"). The distributions of responses, if all responses are given by a single strategy, are shown in the two bottom rows of Figure 3, when the underlying distribution is either unimodal (top row) or bimodal (bottom row). The probability that a response for an old probe is generated by retrieving a value from the SED in LTM is given by $N_{\mathrm{SED}}/N_{\mathrm{Tot}}$, where N_{SED} is the number of items seen in training that are stored in LTM and N_{Tot} is the total number of items in the training distribution. Further, the probability that a value, which cannot be retrieved from LTM, is point estimated by NPE is given by *s*. Thus, the resulting distribution of point estimates (DPE) will be a linear combination of the three types of responses (i.e., a weighted mixture of the direct retrieval, NPE, and RFH distributions) given by

$$DPE = \frac{N_{SED}}{N_{Tot}}DR + \left(1 - \frac{N_{SED}}{N_{Tot}}\right)(sNPE + (1 - s)RFH). \quad (1)$$

For old exemplars (grey bars in Figures 2C and 2D), the distribution is a mix of direct-retrieval responses (for old exemplars whose revenues can be retrieved) and NPE and RFH responses (for old exemplars whose revenues cannot be retrieved). For the new exemplars (grey bars in Figures 2E and 2F), no values can be retrieved and the distribution is a mix only of NPE and RFH responses.

Model fitting. We fitted two versions of the model described in Equation 1 to the participants' point estimates for new and old exemplars. In the STMC version of the model, the size of the SSD was set to 4 (i.e., $n_{\mathrm{SSD}} = 4$) and, in the long-term memory constrained (LTMC) version, the size of the SSD was set to be equal to the size of the SED (i.e., $n_{\mathrm{SSD}} = N_{\mathrm{SED}}$). In the STMC, n_{SSD} was fixed to 4 for three reasons: First, this makes the number of free parameters in the two models the same. Second, the capacity of short-term memory is typically estimated to approximately 4 ± 2 units (Cowan, 2001). Third, and a related point, previous applications of the NSM have made this assumption with good descriptive accounts of the data (see Hansson et al., 2008; Juslin et al., 2007).

These models had two free parameters (s and N_{SED}) that were fitted simultaneously to the 40 data points for the proportions of point estimates in the 10 equally wide intervals for new and old exemplars in the unimodal and bimodal conditions (for further details on the model fitting, see the Appendix). The STMC model achieved a good fit ($r = .97$, root-mean-squared deviation [$RMSD$] = .023, $s_e = .024$) with parameters indicating that the participants often used retrieval when making judgments for the old exemplars ($N_{\mathrm{SED}}/N_{\mathrm{Tot}} = .84$), while responses for new exemplars were predominantly generated by NPE and to a lesser extent by use of the recognition inference ($s = .72$). The model essentially saturates the data with an RMSD that is close to the standard errors (s_e) for the data points, suggesting that the model accounts for most the true variance in the data.

Figures 2 illustrate the ability of the STMC model to reproduce the qualitative pattern of responses in both the unimodal (see Figures 2C and 2E) and the bimodal (see Figures 2D and 2F) condition and for both new (see Figures 2E and 2F) and old (see Figures 2C and 2D) exemplars. Notice that the model is able to reproduce three important patterns in the data: (a) the different distributions for old exemplars in the two distributions; (b) the distributions of new responses that are a result of the RFH responses, with the distinctive spike of a relatively high proportion of responses in the lowest interval for point estimates of new exemplars in the bimodal condition (see Figure 2F) and the positive skew in the distribution of point estimates of new exemplars in the unimodal condition (see Figure 2E); and (c) the high proportion of responses, generated by use of NPE, in the midintervals for new exemplars in the bimodal condition. The model fits in Table 1 also indicate worse fit for the LTMC version ($r = .86$, $RMSD = .057$, $s_e = .024$) than for the STMC version, with RMSD that substantially exceeds the error variance in data. In addition, a sign test revealed a near significant difference in RMSD when the two versions of the model were fitted to individual data ($Z = 1.80$, $p = .07$; $Mdn_{\mathrm{STMC}} = .055$ vs. $Mdn_{\mathrm{LTMC}} = .068$).

In sum, participants have accurate knowledge of distribution shapes with relative frequency estimates but often prefer midinterval point estimates for the new exemplars in both conditions. In the bimodal condition, many participants guessed in the middle of the interval for new exemplars, despite manifest knowledge that the revenues are rarely found there. A model allowing all exemplars in the SED to be available as a basis for point estimates describes the data much more poorly than a model only allowing a small, STMC sample.

Table 1
Model Fit for Models Constrained by Short-Term Memory Capacity (STMC) and Long-Term Memory Capacity (LTMC) in the Four Experiments

Condition and model	s_e	*RMSD*	*r*	$N_{\mathrm{SED}}/N_{\mathrm{Tot}}$	*s*
Experiment 1					
STMC ($n = 4$)	.024	.023	.97	.84	.72
LTMC ($n = N_{\mathrm{SED}}$)	.024	.057	.86	.83	.70
Experiment 2					
Positive skew					
STMC ($n = 4$)	.025	.025	.94	.88	.66
LTMC ($n = N_{\mathrm{SED}}$)	.025	.077	.80	.87	.58
Negative skew					
STMC ($n = 4$)	.024	.040	.79	.83	.71
LTMC ($n = N_{\mathrm{SED}}$)	.024	.083	.37	.20	.73
Experiment 3					
Continuous					
STMC ($n = 4$)	.028	.027	.95	.84	.83
LTMC ($n = N_{\mathrm{SED}}$)	.028	.133	.74	.81	.61
Experiment 4					
Short					
STMC ($n = 4$)	.031	.044	.85	.74	.71
LTMC ($n = N_{\mathrm{SED}}$)	.031	.077	.68	.17	.67
Medium					
STMC ($n = 4$)	.029	.049	.81	.79	.74
LTMC ($n = N_{SED}$)	.029	.087	.54	.17	.69
Long					
STMC ($n = 4$)	.020	.036	.90	.77	.78
LTMC ($n = N_{\mathrm{SED}}$)	.020	.082	.70	.80	.70

Note. *RMSD* = root mean squared deviation; SED = subjective environmental distribution; Tot = total.

Experiment 2: A Response to the Distribution?

It could be that the results of Experiment 1 reflect a more superficial default response strategy to respond in the middle of the admissible response interval, when the participants have no knowledge at all of the company, rather than being based on properties of a sample retrieved from the distribution experienced in training. In Experiment 2, we used two skewed distributions to differentiate between a response strategy driven by the use of small samples and such a default midinterval response strategy. If the results reflect a default midinterval response strategy, participants having experienced distributions in the same range but of opposite skew will respond identically to new objects. We expected, however, that the responses would be sensitive to the central tendency of the respective distribution and thus differ between conditions.[3]

[3] Notice that a skewed distribution analogue to those on the right side of Figure 1 would be a unimodal distribution with the same mean as the original skewed distribution. Thus, using a positively skewed distribution and a negatively skewed distribution would also separate responses close to the mean of a unimodal sample distribution.

Method

Participants. Nine male and 17 female undergraduate students from Uppsala University with an average age of 23.3 years participated in exchange for a movie ticket.

Materials and procedure. The experiment was performed in the same way as Experiment 1, except that the training phase involved 50 companies and the two conditions were defined by two bimodal distributions, one negatively skewed (beta distribution with $\alpha = .6$, $\beta = .4$) and one positively skewed ($\alpha = .4$, $\beta = .6$). Both distributions were linearly transformed to a [1, 1,000] interval. The test phase consisted of 50 old and 50 new exemplars. The learning phase stopped when the participants reached a total of 25 correct responses, using the same criterion as in Experiment 1, or had completed 400 trials.

Results and Discussion

As in Experiment 1, participants in Experiment 2 correctly reproduced the bimodal distribution shape for the bimodal distributions but, more surprisingly, found the distribution with positive skew more difficult to learn (.047 vs. .034 with negative skew), $t(22) = 3.7$, $p = .001$. Note that the finding that the distribution with negative skew is easier to learn is consistent with the assumption behind an ecological rational for the RFH responses; that is, larger values are more easily remembered. Point estimates in the interval [500, 700] in the negatively skewed condition and in the interval [300, 500] in the positively skewed condition were classified as mean responses (MRs; i.e., close to the mean). With negative skew, the rate of MR increased from 23% for old exemplars to 49% for new exemplars, $t(13) = 2.97$, $p = .012$. With positive skew, the rate of MR increased from 17% for old exemplars to 34% for new exemplars, $t(13) = 2.37$, $p = .036$. In statistical tests for the individual participants across both conditions, there was a significant shift toward more MR in 11 participants, a reverse shift in one participant, and a nonsignificant shift in 14 participants ($\alpha = .05$). Crucially, the mean point estimates for new exemplars was significantly higher in the condition with a negative ($M = 483$, $SD = 118$) than with a positive ($M = 345$, $SD = 125$) skew, $t(24) = 2.88$, $p = .008$, suggesting sensitivity to the sample central tendency.

We fitted the STMC and LTMC versions of the NPE model to the positive and negative skew conditions separately. The results, summarized in Table 1, indicate a better fit for the STMC version than the LMTC version in both the positive skew (for STMC, $r = .94$, $RMSD = .025$, $s_e = .025$; for LMTC, $r = .80$, $RMSD = .077$, $s_e = .025$) and negative skew condition (for STMC, $r = .79$, $RMSD = .040$, $s_e = .024$; for LMTC, $r = .37$, $RMSD = .083$, $s_e = .024$). It is also evident that STMC model provides somewhat better fit in the condition with positive skew. This is probably because the SED of the NSM fails to capture the memory advantage for large values and, thus, does not capture the boost in learning seen in the negative skew condition. Further, comparing the two models when fitted to individual data with a sign test revealed a significant difference in RMSD ($Z = 4.5$, $p < .001$; $Mdn_{\text{STMC}} = .061$ vs. $Mdn_{\text{LTMC}} = .096$).

Experiment 3: Inviting the Assessment of Proportions and Introduction of Explicit Incentives to Maximize

The results of Experiments 1 and 2 confirm the prediction that participants make point estimates close to the mean in the bimodal distribution even when they can correctly recreate the distributions shape with estimates of relative frequency. The NSM suggests that, at least in part, this is because different processes are elicited during retrieval, suggesting that it is possible to influence the perception of distributions by eliciting different cognitive processes during retrieval or encoding of the variable. In Experiment 3, we investigated this possibility by manipulating how people perceive a distribution by varying the cognitive processes during encoding.

In the first two experiments, the participants experienced a continuous variable, which probably required them to rely on small samples to make point estimates, inviting a perception of unimodality. However, partitioning the variable into ordinal categories using labels (*Small*, *Quite small*, *Medium*, *Quite large*, *Large*) should encourage the possibility of retrieving samples on the basis of these nominal categories and assessing the proportion in each category. This, in turn, should make it easier to perceive the distribution shape as bimodal and thus invite relatively more guessing in the most frequently occurring extreme categories of the interval.

In Experiment 3, half of the participants trained to make point estimates for a bimodal continuous variable; the other half trained to categorize values into one of five ordinal categories. In both cases, as in Experiments 1 and 2, the feedback training informed the participants of the correct value. We predicted that categorization training would give participants a better idea of the overall distribution shape and therefore reduce the rate of MRs. Note the alternative possibility: It could be that the MRs in the bimodal conditions are based on an accurate perception of the bimodal distribution shape at the time of judgment but that participants are nonetheless strongly guided by intuitions to make mean point estimates, for example, because they entertain some intuitive version of least-squares minimization. If this is true, training with categories, in addition to feedback about the continuous value, should have no effect on this basic normative intuition. However, if the MRs are mediated partially or wholly by reliance on small samples that fail to disclose the bimodal distribution shape, coding the observations in terms of category proportions should highlight the distribution shape.

A second purpose of Experiment 3 was to investigate the robustness of the mean strategy. It could, again, be argued that the participants in Experiments 1 and 2 had an accurate conception of the bimodal distribution shape at the moment of judgment, but they relied on some loss function that nonetheless reinforces a mean strategy. Although we find the reliance on such a loss function rather implausible already on psychological grounds, in Experiment 3, an explicit loss function reinforcing maximizing behavior in the form of a bonus for correct guesses (close to the distribution mode) was introduced to half of the participants. We expected the participants would already rely on such a loss function and thus to be unaffected by this manipulation.

Method

Participants. Eighteen female and 22 male undergraduate students from Uppsala University with an average age of 24.8 years were given a movie ticket or course credits in exchange for their participation.

Materials and procedure. The training phase involved 60 companies, and all conditions used a symmetric bimodal distribution (beta distribution with $\alpha = \beta = .33$) on the interval [0, 1,000]. The experiment was a $2 \times 2 \times 2$ factorial design with response format in the learning phase (continuous, category), response format in the test phase (continuous, category), and monetary incentive in the test phase (yes, no) as between-subjects variables. Participants were randomly assigned to one of the eight conditions. In the continuous format condition, participants responded with a number from 0 to 1,000, and in the category format condition they responded by choosing one of five text- and interval-labeled categories (*Small [0, 200]*, *Quite small [201, 400]*, *Medium [401, 600]*, *Quite large [601, 800]*, *Large [801, 1,000]*). After each guess, participants in both conditions were shown the correct value in the same continuous format. Learning continued until participants had made correct predictions on 50 of the 60 companies or 400 trials had been completed. A guess was considered correct if it was within a 200-unit interval around the actual value of the company's revenue. The test phase consisted of 60 old and 50 new exemplars. In the monetary incentive condition, participants were informed in the test phase that they would be paid a bonus proportional to the degree to which their answers were close to the actual values. They were instructed that for each point estimate that came within 100 units of the actual revenue, they would receive a bonus point and that a sufficient number of bonus points would earn them additional movie tickets. Thus, the instructions made it obvious to participants that an all-or-none incentive structure was used in the experiment.

Results and Discussion

There were no large or statistically significant effects of test condition or monetary incentive, and the data were collapsed into training with continuous or category predictions. Figure 4 presents the average assessed distribution by estimates of relative frequency, the distribution of point estimates for old exemplars, and the distribution of point estimates for new exemplars for the continuous (see Figure 4A) and category (see Figure 4B) conditions, respectively. Figure 4 suggests that the participants have accurate knowledge of the distributions when the knowledge is elicited with frequency estimates and that the shift of point estimates toward MRs occurs mainly in the continuous condition. The point estimates in the 500 ([401, 600]) interval and choices of the corresponding middle category were classified as MRs. The proportion of MRs was entered into a two-way ANOVA with response format in the learning phase (continuous/category, between subjects) and type of exemplar (new/old, within subject) as independent variables. There was a significant main effect of response format, $F(1, 38) = 4.46$, $p = .04$, $MSE = .03$, with more MR with the continuous ($M = .24$, $SD = .20$) than with the category ($M = .15$, $SD = .13$) format and a main effect of exemplar type, $F(1, 38) = 25.49$, $p < .001$, $MSE = .02$, with more MR for new ($M = .27$, $SD = .21$) than for old ($M = .12$, $SD = .07$) exemplars. The interaction was not significant, $F(1, 38) = 2.66$, $p = .11$, $MSE = .02$.

We fitted the STMC and LTMC models to data from the continuous training condition, where participants were training and memorizing point estimates as presumed by the models. The results, summarized in Table 1, show a better fit for the STMC version than the LTMC version (for STMC, $r = .95$, $RMSD = .027$, $s_e = .028$; for LMTC, $r = .74$, $RMSD = .133$, $s_e = .028$). Notice also that the fit of the STMC version is close to the standard error (s_e) of the data points. Further, a sign test revealed a significant difference in RMSD when the two versions of the model were fitted to individual data ($Z = 3.4$, $p < .001$; $Mdn_{\text{STMC}} = .081$ vs. $Mdn_{\text{LTMC}} = .145$). In sum, as predicted, the mean strategy was less prevalent after training with a category task that highlighted the bimodal distribution shape and the mean strategy effect proved resistant to monetary incentives.

Experiment 4: The Effects of Experience

Experiment 4 was designed to test the predictions related to experience (N_{SED}) and variation of point estimates. In Experiment 4, half of the participants trained with a unimodal distribution and the other half trained with a bimodal distribution. Further, half of the participants in each distribution condition had their training divided into three segments with tests occurring in between to capture the training effects, whereas half of the participants were tested only at the end of training. The NSM predicts that N_{SED} should increase with training but that the STMC version of NSM ($n = 4$) should provide better fit. If the tendency to place point estimates in the lowest category, seen in the previous experiments, is a shift to maximization based on accurate knowledge of the bimodality in the distribution, the proportion of these responses is expected to increase as learning progresses. If, however, these responses arise from a RFH, as assumed here, the proportion of responses should, if anything, decrease rather than increase with additional knowledge of the distribution. To replicate the findings of Experiment 3 of the robustness of the mean strategy even with an incentive to discourage participants from relying on it, we implemented the same all-or-nothing payoff structure.

Method

Participants. Thirty-two female and 16 male undergraduate students from Uppsala University with an average age of 24.7 years were given a movie ticket or course credits in exchange for their participation.

Materials. The learning phase used the same two distributions as in Experiment 1, that is, a symmetrical unimodal distribution (beta distribution with $\alpha = \beta = 3.4$) and a symmetrical bimodal distribution (beta distribution with $\alpha = \beta = .33$). Both distributions were linearly transformed to a [1, 1,000] interval. For each participant, the numbers were randomly paired with one of 156 company names.

Design and procedure. The experiment was a 2×2 factorial design with distribution (unimodal, bimodal) and exposure (full, segmented) as between-subjects variables. Participants were randomly assigned to one of the four conditions. The learning phase in the full condition was similar to that of Experiment 1. However, it consisted of five blocks in which each of the 60 items were shown once, in randomized order. At test, point estimates were obtained for a random selection of 15 of the presented (old) companies and 15 new companies. Participants also assessed the proportion (expressed as a percentage) of companies that fell into 10 predefined, equally wide intervals ([1, 100], [101, 200], . . . , [901, 1,000]). In the segmented condition, the learning phase was interrupted after the first, second, and fifth blocks with a test.

Participants were given a monetary reward contingent on the accuracy of their point estimates at test. They were instructed that for each point estimate that came within 100 units of the actual revenue, they would receive a bonus, which could be cashed in as additional

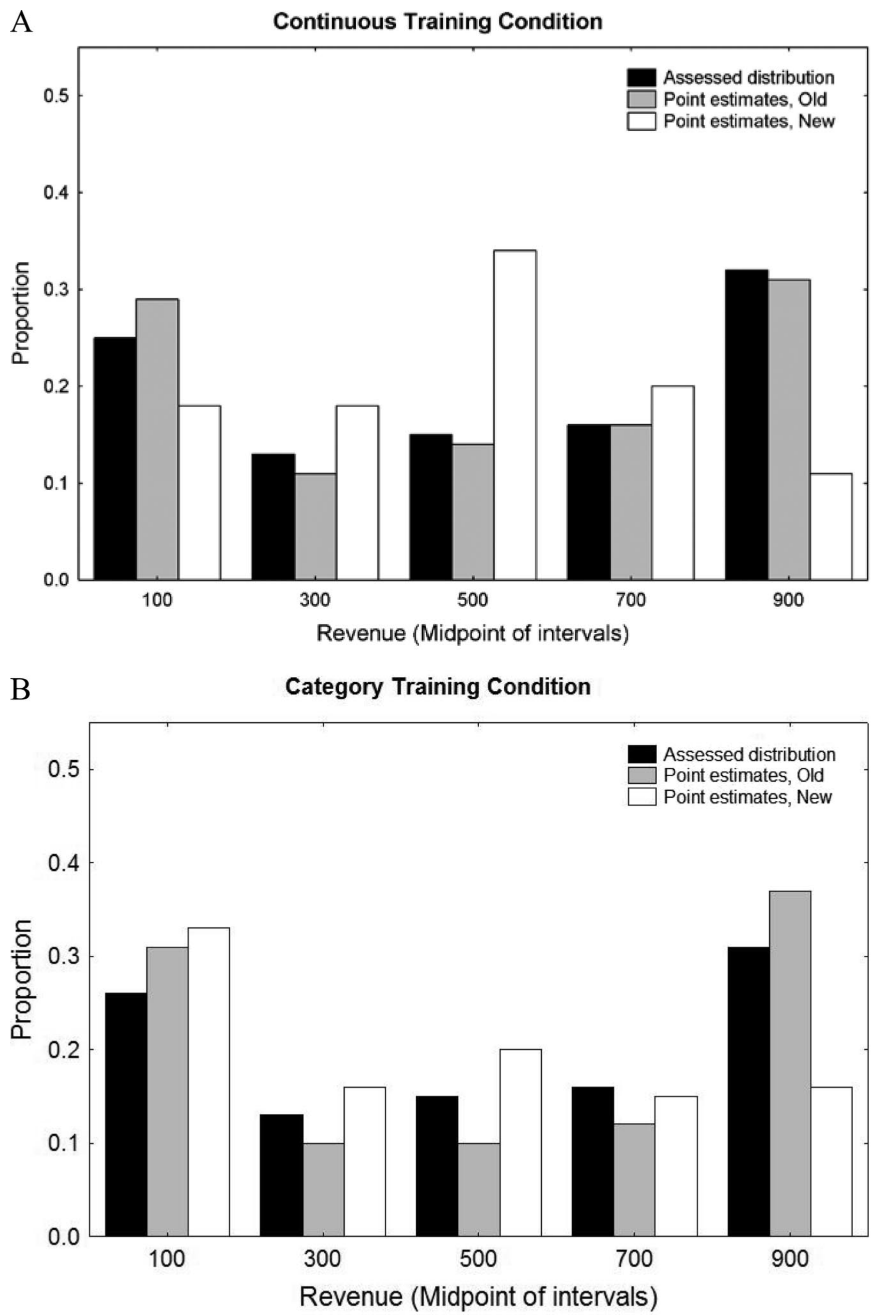

Figure 4. Average assessed distribution of frequency estimates, the distribution of point estimates for old exemplars, and the distribution of point estimates for new exemplars in Experiment 3, for the continuous training condition (A) and the category training condition (B).

movie tickets. Thus, as in Experiment 3, the instructions made it clear that an all-or-none incentive structure payoff was used.

Results and Discussion

Friedman ANOVAs revealed that the mean variance of the point estimates for new exemplars was unaffected by test round both in the bimodal condition, $\chi^2(2, N = 12) = 0.17$, $p = .92$, and in the unimodal condition, $\chi^2(2, N = 12) = 1.5$, $p = .47$. There was significantly lower variance (Mann–Whitney $U = 27.0$, $Z = 2.57$, $p = .01$) in the unimodal condition ($M = 188$, $SD = 161$) than in the bimodal condition ($M = 270$, $SD = 214$).[4] There was no difference

[4] The same results were obtained when analyzing point estimates on all exemplars in the segmented condition and both new and all exemplars in the full condition.

in the variance of point estimates in the full condition and the three tests in the segmented condition for either of the distributions.

The results were similar to those of Experiment 1, with estimates of relative frequency and point estimates for old exemplars mirroring the underlying distribution shape while a disproportional number of point estimates for the new exemplars in the bimodal condition come close to the distribution mean. Figure 5 shows the main findings, which highlight the proportion of point estimates for new exemplars that fall in the middle (400–600) and the extreme categories (low = 0–200, high = 800–1,000). Figure 5A shows the results in the full condition and the final test of the segmented condition. It can be seen that the responses in the middle category are the most numerous in both conditions and over both distributions, with a slightly stronger tendency for this in the segmented condition. There were no statistically significant differences between these conditions at the same amount of training. This shows that there is no effect of the repeated testing procedure of participants in the segmented condition.

The previous results showed behavior consistent with STMC NPE. If the results are a mixture of RFH and NPE responses, it is reasonable to assume that RFH responses will be more frequent initially when the task is unfamiliar and knowledge is scarce. If the point estimates in the lowest category, on the contrary, derive from an insight about the bimodal distribution shape along with an ambition to make estimates in a region where most values fall (maximizing), this tendency should increase with more training, as the participants get a better conception of the distribution shape. Figure 5B illustrates the proportion of the point estimates for new

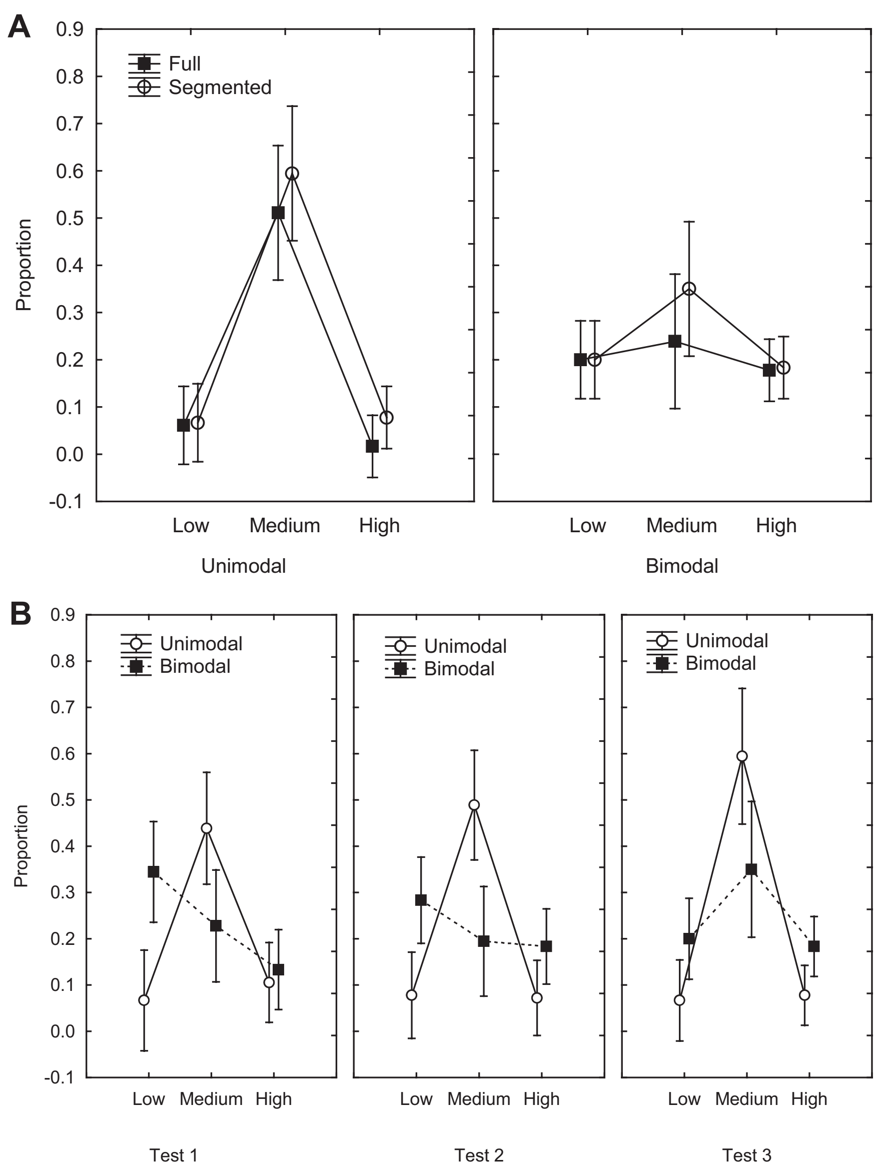

Figure 5. Proportion of point estimates on new exemplars in the low, medium, and high category in the unimodal and bimodal conditions for the full and segmented conditions, shown for the final test (A) and divided over three test rounds in the segmented condition (B).

exemplars as a function of learning in the segmented condition for both distributions. The proportion of low responses in the bimodal condition is initially high but decreases with training in parallel with an increase of mean responses. The proportion of mean responses also increases with training in the unimodal condition. This pattern is consistent with the idea that responses are a mixture of RFH and NPE responses, where the proportion of NPE responses relative to RFH responses increases with training, consistent with the STMC model with an increase in the value of *s* with training.

There were no significant differences between data from the full condition and from the third test round in the segmented condition. We therefore collapsed the data for these two tests and fitted the models to the data for short (segmented, Test 1), medium (segmented, Test 2), and long (segmented, Test 3 and full collapsed) training separately. For all three training lengths, this was done for both the STMC and the LTMC versions of the model. Two participants, who placed more than two thirds of point estimates on old exemplars in a single interval in one or more of the three tests and thus indicated poor learning, were removed from the fitting procedure. The results, summarized in Table 1, show a better fit for the STMC than the LTMC version in short (e.g., $r = .85$ vs. $r = .68$), medium ($r = .81$ vs. $r = .54$), and long ($r = .90$ vs. $r = .70$) training. The fit improves as training progresses (STMC $r = .85$, .81, and .90, respectively) and the relative use of NPE responses over RFH responses increases as training progresses (STMC $s = .71$, .74, and .78, respectively) as predicted. The proportion of responses directly retrieved from memory for point estimates of old exemplars increases from the first to the second test but then slightly decreases (STMC $N_{\text{SED}}/N_{\text{Tot}} = .74$, .79, and .77, respectively), indicating little additional learning from the second to the third test. Further, a sign test revealed a significant difference in RMSD when the two versions of the model were fitted to individual data for the long training ($Z = 4.7$, $p < .001$; $Mdn_{\text{STMC}} = .090$ vs. $Mdn_{\text{LTMC}} = .103$).

In sum: the variability of the point estimates was not significantly affected by the size of the SED (learning), supporting the idea that participants rely on a capacity-constrained sampling process, as implied by the NSM. Further, the tendency in the previous experiments to place a disproportionate proportion of point estimates for new exemplars close to the distribution mean was replicated. This tendency persisted with an explicit all-or-none monetary incentive structure. Participants in the bimodal condition placed 25% of their point estimates in an interval near the distribution mean ([400, 600]), when the distribution only contained 7% of its exemplars.

General Discussion

An extensive body of research suggests that people's judgments are constrained both by the reliance on small samples of information (e.g., Juslin et al., 2007; Stewart et al., 2006) and by limitations in short-term memory capacity (Dougherty & Hunter, 2003; Gaissmaier et al., 2006; Hansson et al., 2008; Kareev et al., 2002). In the present article, we extend this work, proposing that such limitations also influence how people make point estimates of unknown quantities. In four experiments, we compared the predictions by two possible accounts of how people use their knowledge of statistical distributions to make point estimates of unknown quantities.

Main Findings

The results comprise three main findings. First, across all four experiments, the results consistently favored the naïve sampling account of point estimation, suggesting that people base their point estimates on properties of a small sample constrained by short-term memory (the SSD) drawn from exemplars stored in memory (the SED) during training rather than having access to representations that benefit from most or all of the experience that is acquired during training.

Second, the results of the experiments documented the operation of an unexpected recognition-based inference, whereby the participants, in the absence of other information, guessed that an unrecognized company is more likely to have low rather than high revenue. Only after considering this additional mechanism could the models appropriately account for the data.

Third, the results also document limitations of the naïve sampling framework. In Experiment 1, it was clear that the participants found it much easier to learn a unimodal than a bimodal distribution. Although this finding, in general terms, is consistent with the claim that use of small samples may contribute to a belief or expectation that most distributions are normally distributed, there is no mechanism in the NSM that captures such an effect. Likewise, there is no mechanism that captures the better memory for large than small values in this specific task that drives the recognition-based inference and that makes the distribution with negative skew in Experiment 2 easier to learn than the distribution with positive skew.

Intriguingly, the NSM predicts point estimates close to the mean of the distribution regardless of the distribution shape, while at the same time it predicts an ability to reproduce the distribution when using unbiased estimators. Experiment 1 revealed that people do have a tendency to place point estimates of unknown quantities close to the distribution mean, both when the distribution is unimodal and when the distribution is bimodal. As illustrated in Figure 2B, even though the participants placed a disproportional proportion of point estimates for new exemplars in the midinterval of the distribution range, at the same time, they showed manifest knowledge of the shape of the bimodal distribution when reproducing it with estimates of relative frequency. Experiment 2 showed that participants were sensitive to the distribution mean and that the tendency to make point estimates close to the distribution mean was not the result of a superficial default response strategy. In Experiment 3, it was shown that a manipulation improving the participants' ability to detect the bimodal distribution shape reduced the rate of mean responses, as expected if they are at least in part based on an inability to detect the distribution shape from small samples.

Whether it is a normative error to place point estimates close to the mean of a distribution is, of course, dependent on the loss function assumed (Weber, 1994; Winkler, 1970). However, it is not self-evident why people should have developed an intuition for prediction that coincides with, specifically, minimization of least squares. One possibility is that people have a default loss function related to an a priori assumption of normality of the underlying distribution, because most real-world distributions are unimodal.

This would render people not only intuitive statisticians but intuitive parametric statisticians. Winkler (1970) showed that it is possible to change people's response strategies by implementing different loss functions on their judgments. In Experiment 3, we tested whether the tendency to place point estimates near the distribution mean persisted when a loss function that explicitly and strongly rewarded point estimates near the modes of the underlying bimodal distribution was introduced. We found that our participants continued to place point estimates close to the distribution mean even under the explicit all-or-none loss function, a finding that was replicated in Experiment 4. It should be noted that there is a crucial difference between the study by Winkler (1970) and ours. In Winkler's study, participants were explicitly told what loss function to use, whereas we introduced it with the aid of a monetary incentive. An important goal for future researchers is to understand what factors determine people's choice of strategy as a result of the implicit or explicit loss functions introduced.

In Experiment 4, we investigated how point estimates are influenced by learning. The NSM predicts that the variation in point estimates should be virtually unaffected by experience but reflect the variance in the underlying distribution. Further, people should continue to use a mean strategy even as they gain more experience, because small samples fail to signal the correct shape of the distribution even when the SED is large. As we have seen, this is the pattern of results in Experiment 4. The results from the modeling presented in Table 1 also suggest a continued use of small samples even when learning progresses. The results moreover suggest that the rate of low-category, recognition-based responses diminishes rather than increases with more training.

The finding in Experiment 4 that the proportion of point estimates close to the distribution mean increases with more experience suggests that the use of small samples may be contingent on the forming of the SED. That is, even though people are inclined to spontaneously sample from an SED when making point estimates, they will do so fully only when the SED is established with a sufficient number of exemplars in LTM. It remains for future researchers to investigate which strategies people use when knowledge is scarce (i.e., when the SED is not yet fully established) and how such strategies influence judgments. A further empirical question is at what point people switch to rely fully on NPE.

Limitations

Short-term memory capacity is a key limitation, but in none of the experiments do we manipulate this capacity, for example, by introducing cognitive load at encoding or at retrieval. Cognitive load at the encoding would probably only influence the formation the SED. Introducing cognitive load when the exemplars are retrieved from the SED to the SSD is theoretically more interesting, and, on the face of it, it seems plausible that this should produce a SSD with a smaller sample size. Yet, the net effect of such a manipulation is not entirely straightforward to predict, considering that decreased sample size may have different effects on the distributions generated by point estimation and by the recognition-based inference. Indeed, sometimes the effects cancel each other. For example, in a bimodal condition, a greater sample size contributes both to a reduced overall variance of the responses through its effect on the portion of responses from point estimation and to an increased overall variance of the responses through its effect on the portion of responses from recognition-based inferences. In addition, the cognitive load may affect the balance between these two processes or may even make the participants resort to strategies other than the ones suggested by NSM.

The results from Experiment 1 suggested that it is easier to learn a unimodal than a bimodal distribution. As noted, the NSM presently has no mechanism that explains such a difference. However, research indicates that people are inclined to expect normal (or at least unimodal) distributions in a variety of tasks (Flannagan et al., 1986; Fried & Holyoak, 1984; Lindskog, Winman, & Juslin, 2012). An interesting question for future researchers is to integrate such a priori expectations into models of how people learn distribution shapes.

In all four experiments, the NSM parameters were fitted to group data. It would be desirable, of course, to fit the models also to individual participant data. In the four experiments described above, the number of responses for each participant and each type of response (new and old), however, made this difficult. Nonetheless, comparing the RMSD for the two versions of the model over the four experiments when fitted to individual data with a sign test revealed significantly better fit for the STMC version over the LTMC version ($Z = 7.5$, $p < .001$; $Mdn_{\mathrm{STMC}} = .075$ vs. $Mdn_{\mathrm{LTMC}} = .096$). The same conclusion held for each experiment separately. The NSM includes three free parameters: s, n, and N_{SED}. In the current description of the model and in contrast to n and N_{SED}, s is theoretically underspecified and, at present, it is thus difficult to make strong predictions about s on the basis of theoretical or other grounds. It remains for future researchers to investigate the relationship between RFH and NPE responses to be able to formulate stronger predictions of the value of s and the change in value of s as knowledge goes from scarce to sufficient during learning.

The task in the experiments is to remember presented numbers and to learn to associate these with an appropriate label (a company name). It is probable that this undertaking is arduous and that partly misremembered numbers become associated with wrong labels or no labels at all. The N_{SED} parameter (i.e., the number of exemplars stored in LTM) is a simplification in that it disregards this distinction. The parameter has a dual function in the model: (a) It accounts for the number of correctly retrieved items for old items and (b) it provides the pool from which the numbers in the SSD are sampled for new or unrecognized items. In reality, the latter sampling of items in the model does not depend on whether the numbers are stored with an appropriate label; it merely samples stored numbers irrespective of associated label. This dual role of N_{SED} may explain why the estimated values of the N_{SED} in the experiments may appear surprisingly high given the stated learning criterion in the training phase.

For new items, a sufficiently large pool of numbers is needed in the SED, whether or not these numbers are stored with the appropriate labels. A more realistic but also more complex model could readily account for this by introducing a parameter denoting the probability that a number is stored with or without the correct label. We have refrained from introducing more parameters in favor of parsimony but conclude that this comes at the cost of an apparent overestimation of the proportion of observed items that are correctly remembered together with their labels.

This article contrasts two possible accounts of how people use knowledge of statistical distributions to make point estimates. It is, of course, also possible to formulate other accounts of how such estimates are made. For example, it might be the case that people are not limited to drawing just one sample prior to making a point estimate but, rather, that they draw several samples and integrate the information from them into a final judgment. Similarly, more sophisticated operations on the SSD, such as trimmed means or density estimation, could also be suggested. Such accounts, however, would place additional storage and computational demands on the human cognitive system, demands that lie beyond the capacity limitations documented in previous research (Cowan, 2001; Dougherty & Hunter, 2003; Gaissmaier et al., 2006; Hansson et al., 2008; Juslin et al., 2007; Juslin et al., 2011; Nilsson et al., 2009; Kareev et al., 2002; Stewart et al., 2006). In this article, we show that it is sufficient to assume that people retrieve a small sample (e.g., Juslin et al., 2007; Stewart et al., 2006), limited by short-term memory capacity (Cowan, 2001; Dougherty & Hunter, 2003; Gaissmaier et al., 2006; Hansson et al., 2008; Kareev et al., 2002; Stewart et al., 2006), from memory, assuming that they base their point estimate on the properties of that sample, to account for the data from our four experiments.

It might also be that a substantially different account could be equally successful in explaining the observed response patterns. One such account that recently has received a lot of attention is a Bayesian framework of human cognition (e.g., Chater, Tenenbaum, & Yuille, 2006; Oaksford, & Chater, 2009). In a Bayesian account, our participants would enter the learning phase with some prior assumption of how the revenue variable is distributed. In training, they would update their belief about the distribution using Bayes' theorem or some approximation of it. In the test phase, the participants would then express their posterior belief about the distribution.

To make point estimates, the participants could use the mean or mode from the posterior distribution or sample a value from the posterior distribution. In the latter case, values that are more probable will be sampled more often. Similarly, when asked to give the proportion of revenues in a given interval, a reasonable strategy would be to report the density of the posterior distribution in that interval. Thus, at least for estimated proportions and point estimates of new exemplars, the distribution of responses should mirror the posterior distribution.

The responses, however, suggest that if participants produce responses by a Bayesian account, they are using different posterior distributions in the two tasks. Although Figure 2B suggests that the posterior distribution in the bimodal condition is a symmetrical bimodal distribution, the posterior distribution suggested by Figure 2F is asymmetrical. In the proportion production task, the participants' responses mirror the underlying distribution. Considering that the data they encounter represent a symmetrical bimodal distribution, this would suggest that people enter the learning phase with a uniform prior for the revenue, which is updated to approximate the true distribution.

By contrast, Figure 2F suggests that participants in the bimodal condition have a prior with a lot of density in the lowest interval and the two middle intervals. It is unclear why such a prior should be chosen over a symmetrical prior or a positively skewed prior that is likely to map a real-world distribution of revenues. Further, comparing Figures 2E and 2F, the participants in the unimodal and bimodal conditions seem to enter the learning phase with different priors, which is not impossible but at least extremely unlikely. Thus, to adequately explain the data with such a Bayesian account, it would be necessary to assume (a) the priors are intricate (but only when the underlying distribution is bimodal), (b) that participants use different posteriors in different tasks, and (c) that participants' priors depend on what underlying distribution they are exposed to.

A final and general concern about the modeling and the experiments reported in this article may be that they involve a highly idealized and simplified task. The participants can only produce estimates in one of two ways: by retrieving the correct value from LTM (if possible) or by inferring a likely value on the basis of the premise that it has been sampled from the same population distribution as the exemplars (companies) previously encountered. By contrast, in most real-life point-estimation tasks, people benefit also from a variety of probe-specific cues and similarity relations that can be used as the input for much more elaborate reasoning.

Although ultimately, of course, the tenability of the conclusions presented here is an empirical question, we do note some considerations that mitigate this concern. The basic issues addressed in this article—how people represent knowledge of distributions in their environment and how this knowledge is translated into point estimates, for example, by reporting the estimated mean or mode as their best point estimate—apply also to more complex models that take additional cues and similarity into consideration. In the NSM, it would be relatively straightforward to make the probability that a previous observation is activated in short-term memory contingent on its similarity to the probe, in effect, transforming it into an instance- or exemplar-based model of the sort that has been considered in both judgment research (e.g., Dougherty, Gettys, & Ogden, 1999; Juslin & Persson, 2002) and categorization learning (e.g., Nosofsky & Johansen, 2000).

We thus propose that the basic conclusions formulated in the context of the NSM are likely to apply also to the processes captured by more elaborate models. In other words, people are inclined to construct knowledge of distributions post hoc by retrieving samples, as constrained by short-term memory, and these samples are unlikely to disclose bimodal distribution shapes.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance, 3,* 109–123. doi:10.1016/0030-5073(68)90001-9

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real -world quantitative estimation. *Psychological Review, 100,* 511–534. doi:10.1037/0033-295X.100.3.511

Brunswik, E. (1955). Representative design and probabilistic theory in functional psychology. *Psychological Review, 62,* 193–217. doi:10.1037/h0047470

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic–cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459. doi:10.1037/0033-295X.100.3.432

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences, 10,* 287–291. doi:10.1016/j.tics.2006.05.007

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. doi:10.1017/S0140525X01003922

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 23, 968–986. doi:10.1037/0278-7393.23.4.968

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106,* 180–209. doi:10.1037/0033-295X.106.1.180

Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113,* 263–282. doi:10.1016/S0001-6918(03)00033-7

Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics, 27,* 30–41. doi:10.1198/jbes.2009.0003

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519–527. doi:10.1037/0033-295X.101.3.519

Evans, J. S. B. T., & Pollard, P. (1985). Intuitive statistical inferences about normally distributed data. *Acta Psychologica, 60,* 57–71. doi:10.1016/0001-6918(85)90013-7

Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review, 107,* 659–676. doi:10.1037/0033-295X.107.4.659

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 241–256. doi:10.1037/0278-7393.12.2.241

Fox, S., & Thornton, G. C. (1993). Implicit distribution theory: The influence of cognitive representation of differentiation on actual ratings. *Perceptual and Motor Skills, 76,* 259–276. doi:10.2466/pms.1993.76.1.259

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257. doi:10.1037/0278-7393.10.2.234

Gaissmaier, W., Schooler, L. J., & Rieskamp, J. (2006). Simple predictions fueled by capacity limitations: When are they successful? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 966–982. doi:10.1037/0278-7393.32.5.966

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6,* 100–121.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Inferences, heuristics, and biases: New directions in judgment under uncertainty.* New York, NY: Cambridge University Press.

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York, NY: Oxford University Press.

Goodwin, P. (1996). Statistical correction of judgmental point forecasts and decisions. *Omega, 24,* 551–559. doi:10.1016/0305-0483(96)00028-X

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L.-G. (2008). Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and Aging, 23,* 531–544. doi:10.1037/a0012782

Hertwig, R., Pachur, T., & Kurzenhauser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 621–642. doi:10.1037/0278-7393.31.4.621

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106,* 259–298. doi:10.1016/j.cognition.2007.02.003

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116,* 856–874. doi:10.1037/a0016979

Juslin, P., Nilsson, H., Winman, A., & Lindskog, M. (2011). Reducing cognitive biases in probabilistic reasoning by the use of logarithm formats. *Cognition, 120,* 248–267. doi:10.1016/j.cognition.2011.05.004

Juslin, P., Olsson, H., & Björkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making, 10,* 189–209. doi:10.1002/(SICI)1099-0771(199709)10:3<189::AID-BDM258>3.0.CO;2-4

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science, 26,* 563–607. doi:10.1207/s15516709cog2605_2

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114,* 678–703. doi:10.1037/0033-295X.114.3.678

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153. doi:10.1037/0033-295X.93.2.136

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review, 111,* 1072–1099. doi:10.1037/0033-295X.111.4.1072

Kareev, Y., Amon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 131,* 287–297. doi:10.1037/0096-3445.131.2.287

Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting, 21,* 3–14. doi:10.1016/j.ijforecast.2004.02.003

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation.* New York, NY: Springer.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 551–578. doi:10.1037/0278-7393.4.6.551

Lindskog, M., Winman, A., & Juslin, P. (2012). *Calculate or wait: Is man an eager or lazy intuitive statistician?* Manuscript submitted for publication.

Linville, P. W., Salovey, P., & Fischer, G. W. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical-evidence and a computer simulation. *Journal of Personality and Social Psychology, 57,* 165–188. doi:10.1037/0022-3514.57.2.165

Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning and Verbal Behavior, 22,* 547–559. doi:10.1016/S0022-5371(83)90337-7

Marewski, J. N., Pohl, R. F., & Vitouch, O. (2010). Recognition-based judgments and decisions: Introduction to the special issue (Vol. 1). *Judgment and Decision Making, 5,* 207–215.

McCloy, R., Beaman, C. P., Frosch, C. A., & Goddard, K. (2010). Fast and frugal framing effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1043–1052. doi:10.1037/a0019693

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science, 32,* 1133–1147. doi:10.1080/03640210802353016

Nilsson, H., Winman, A., Juslin, P., & Hansson, G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. *Journal of Experimental Psychology: General, 138,* 517–534. doi:10.1037/a0017351

Nisbett, R. E., & Kunda, Z. (1985). Perception of social distributions. *Journal of Personality and Social Psychology, 48,* 297–311. doi:10.1037/0022-3514.48.2.297

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7,* 375–402.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences, 32,* 69–84. doi:10.1017/S0140525X09000284

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46. doi:10.1037/h0024722

Peterson, C., & Miller, A. (1964). Mode, median, and mean as optimal strategies. *Journal of Experimental Psychology, 68,* 363–367. doi:10.1037/h0040387

Pitz, G. F., Leung, L. S., Hamilos, C., & Terpening, W. (1976). The use of probabilistic information in making predictions. *Organizational Behavior and Human Performance, 17,* 1–18. doi:10.1016/0030-5073(76)90050-7

Pollard, P. (1984). Intuitive judgments of proportions, means, and variances: A review. *Current Psychological Research & Reviews, 3,* 5–18. doi:10.1007/BF02686528

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1–20. doi:10.1146/annurev.ps.41.020190.000245

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39. doi:10.1146/annurev.ps.28.020177.000245

Spencer, J. (1961). Estimating averages. *Ergonomics, 4,* 317–328. doi:10.1080/00140136108930533

Spencer, J. (1963). A further study of estimating averages. *Ergonomics, 6,* 255–265. doi:10.1080/00140136308930705

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53,* 1–26. doi:10.1016/j.cogpsych.2005.10.003

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567. doi:10.1037/0033-295X.101.4.547

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General, 137,* 73–96. doi:10.1037/0096-3445.137.1.73

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 115,* 228–242. doi:10.1037/0033-2909.115.2.228

Winkler, R. L. (1970). Intuitive Bayesian point estimation. *Organizational Behavior and Human Performance, 5,* 417–429. doi:10.1016/0030-5073(70)90032-2

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *Etc. frequency processing and cognition* (pp. 21–36). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780198508632.003.0002

Appendix

A Naïve Point Estimation Model

Model Description

In this Appendix, we formally outline the naïve point estimation (NPE) model. The model is used to predict the proportion of responses (p_i) in a subinterval i, with starting point γ_i, end point $\gamma_i + l$ and length l, of the range of the underlying distribution when the entire range is divided into k nonoverlapping intervals. Notice that the model does not require l to be equal for all k intervals, only that the partition of the range be nonoverlapping and cover the entire range. However, the modeling in this study is done with $k = 10$ (Experiment 3, $k = 5$) and with equal l for all intervals. In this article, the underlying distribution is always a beta distribution with parameters α and β (Beta(α, β)). Notice that this distribution is defined on [0, 1]. The model has three parameters. N_{SED} is the number of exemplars stored in long-term memory (LTM). Thus, N_{SED}/N_{Tot} is the probability that a given probe can be retrieved from memory. The probability that participants will use the choice strategy NPE when they cannot retrieve the value of a probe from memory is given by s. Finally, n is the number of items contained in the subjective sample distribution (SSD) when sampling from memory and is thus equivalent to short-term memory capacity.

The model predicts that the proportion of responses in an interval is given by

$$DPE_i = \frac{N_{SED}}{N_{Tot}} DR_i + \left(1 - \frac{N_{SED}}{N_{Tot}}\right)(sNPE(n)_i + (1 - s)RFH(n)_i) \quad (1)$$

where

$$DR_i = \int_{\gamma_i}^{\gamma_i + l} Beta(\alpha, \beta) \quad (2)$$

and

$$NPE(n)_i = \int_{\gamma_i}^{\gamma_i + l} N\left(\frac{\alpha}{\alpha + \beta}, \sqrt{\frac{1}{n}\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}}\right) \quad (3)$$

where $N(\mu, \sigma)$ is a standard normal distribution with mean μ and standard deviation σ. Thus, NPE_i is the density of the sampling distribution of Beta(α, β) with sample size n on the interval $[\gamma_i, \gamma_i + l]$. Note that as n grows large, the central limit theorem indicates that the sampling distribution is approximately normal. However, as illustrated by Figure A1, the sampling distribution comes close to normal already at $n = 4$, motivating the use of $N(\mu, \sigma)$ for the NPE responses. Finally,

(*Appendix continues*)

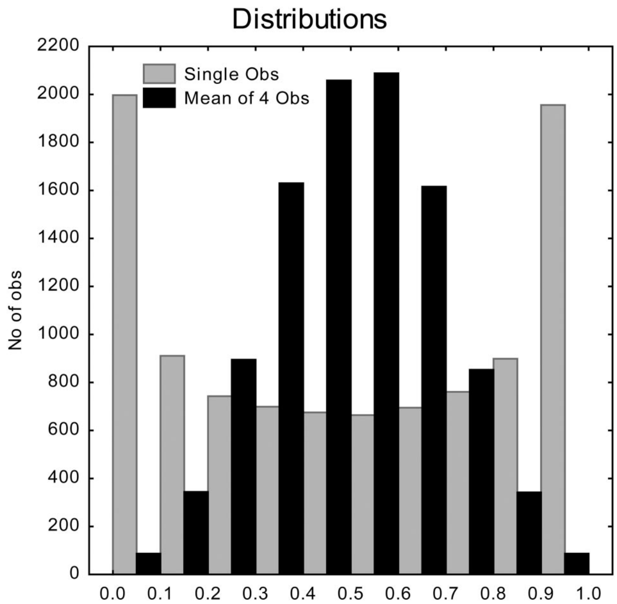

Figure A1. The distribution of means computed from four individually bimodal observations (black bars) and the distribution of single observations (grey bars) from a bimodal distribution. Obs = observations.

$$RFH(n)_i = 1 - \left(\int_{\gamma_i + I}^{1} Beta(\alpha, \beta)\right)^n - \sum_{t=0}^{i-1} RFH(n)_t. \qquad (4)$$

Thus, RFH_i is the probability that the lowest value in the SSD is found in interval i. This is a monotonically decreasing, negatively accelerated function, which, when n is small, resembles a step function (see Figure 3).

Model Fitting Procedure

The model was fitted to group data using least-squares nonlinear fitting implemented in a MATLAB program. It was fitted to data for the distribution of point estimates for new exemplars and point estimates for old exemplars. For point estimates for new exemplars, we used $N_{SED} = 0$. That is, when giving point estimates of new exemplars, participants are expected to use only NPE or RFH. For point estimates for old exemplars, N_{SED} was a free parameter fitted to data. We fitted the model in two versions, one where point estimates are constrained by short-term memory capacity (STMC) and one where point estimates are constrained by long-term memory capacity (LTMC). In the STMC version, we used $n = 4$, that is, the size of the SSD is constrained to four items. In the LTMC version, we set $n = N_{SED}$, that is, all items stored in LTM will be available in the SSD when making point estimates.

For all data sets, except the one in Experiment 3 in which we used five intervals, we divided point predictions into 10 equally wide intervals. Because this division of intervals is, in some way, arbitrary, we risk introducing rounding errors into the interval-partitioned distribution of point predictions. To avoid this, a small, normally distributed error ($\mu = 0$, $\sigma = .1$) was added to all point estimates before we partitioned the distribution into intervals and calculated the empirical interval proportions.

Received January 12, 2011
Revision received June 4, 2012
Accepted June 18, 2012 ■