Routledge
Taylor & Francis Group

Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: https://www.tandfonline.com/loi/pecp21

# Calculate or wait: Is man an eager or a lazy intuitive statistician?

Marcus Lindskog, Anders Winman & Peter Juslin

**To cite this article:** Marcus Lindskog, Anders Winman & Peter Juslin (2013) Calculate or wait: Is man an eager or a lazy intuitive statistician?, Journal of Cognitive Psychology, 25:8, 994-1014, DOI: 10.1080/20445911.2013.841170

**To link to this article:** https://doi.org/10.1080/20445911.2013.841170

*Journal of Cognitive Psychology,* 2013
Vol. 25, No. 8, 994–1014, http://dx.doi.org/10.1080/20445911.2013.841170

Routledge
Taylor & Francis Group

# Calculate or wait: Is man an eager or a lazy intuitive statistician?

**Marcus Lindskog, Anders Winman, and Peter Juslin**

Department of Psychology, Uppsala University, Box 1225, SE-751 42 Uppsala, Sweden

Research on people's ability to act as intuitive statisticians has mainly focused on the accuracy of estimates of central tendency and variability. In this paper, we investigate two hypothesised cognitive processes by which people make judgements of *distribution shape*. The first claims that people spontaneously induce abstract representations of distribution properties from experience, including about distribution shape. The second process claims that people construct beliefs about distribution properties post hoc by retrieval from long-term memory of small samples from the distribution, implying format dependence with accuracy that differs depending on judgement format. Results from two experiments confirm the predicted format dependence, suggesting that people are often constrained by the post hoc assessment of distribution properties by sampling from long-term memory. The results, however, also suggest that, although post hoc sampling from memory seems to be the default process, under certain predictable circumstances people do induce abstract representations of distribution shape.

***Keywords:*** Intuitive statistics; Numerical cognition; Sampling model.

We inevitably experience numerical variables in our everyday lives. We learn about the prices of groceries in our local supermarket, read about baseball players' batting averages or investigate revenues of companies in a foreign market. Often, we also make decisions based on the properties of such variables. For example, we might choose which supermarket to buy food in based on an estimate of the average food price. The question of how people represent knowledge of numerical distributions is further highlighted by the recent interest in 'rational' or Bayesian models of cognition (e.g. Oaksford & Chater, 2006; Tenenbaum, Kemp, Griffiths, & Goodman, 2011), which often presume that the cognitive processes are adaptations to distributions in the environment. How, in such cases, is the knowledge of numerical distributions represented in memory?

Since the early 1960s, psychologists have compared human judgement to statistical theory (e.g. Spencer, 1961, 1963) and at least since the works of Brunswik (1955), the human mind has been likened to an 'intuitive statistician' (Gigerenzer & Murray, 1987; Peterson & Beach, 1967). The conclusion has often been that whereas people can perform lower level arithmetical calculations resulting in unbiased estimates of central tendency, they are less sensitive to sophisticated properties like variance (Pollard,

Correspondence 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

This research was sponsored by the Swedish Research Council and the Bank of Sweden Tercentenary Foundation. We are indebted to Håkan Nilsson, Ebba Elwin and Maria Henriksson for reading and commenting on earlier drafts of the manuscript and to Anja Löfgren for help with the data collection.

1984; Slovic, Fishchoff, & Lichtenstein, 1977) and distribution shape (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978).

Although appealing, the notion of man as an intuitive statistician conflicts with a large body of research suggesting that judgements are the result of fallible heuristics and prone to biases (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). Part of the conflict has been addressed by arguing that people possess an ability to veridically record data in samples of experience but suffer from myopia with regard to the constraints that shape the samples (Fiedler, 2000). According to this notion, if man is a statistician, it seems to be a naive one (Fiedler & Juslin, 2006; Juslin, Winman, & Hansson, 2007).

Research in related areas, like categorisation learning (e.g., Ashby & Maddox, 2005; Nosofsky & Johansen, 2000), multiple-cue judgement (Juslin, Karlsson, & Olsson, 2008; von Helverson & Rieskamp, 2008) and function learning (DeLosh, Busemeyer, & McDaniel, 1997; Kalish, Lewandowsky, & Kruschke, 2004), have often been concerned with the cognitive basis of the judgements. However, research on statistical judgements has rarely addressed this issue (but see Brown & Siegler, 1993). In this paper, we therefore address people's knowledge of distribution shape from the perspective of a classic issue in cognitive psychology: Do people spontaneously induce abstract summary representations of the distribution shape, or do they primarily generate such judgements post hoc by the retrieval of concrete observations of the variable?

## Knowledge of distribution shape

Research on statistical judgements has been both extensive and enlightening but has mostly overlooked that variables contain information that descriptive parameters can only partially capture. The probability density functions of the two variables illustrated in Figure 1, for example, share the same central tendency but obviously have different properties. The most salient difference is the shape of the distribution (even though the variance also differs), which descriptive parameters can only partially capture without any assumptions. If, for example, you want to make a guess that is very likely to fall close to the value of a newly sampled observation from the distribution, in the unimodal distribution, guessing on the distribution mean is a good candidate. In the bimodal distribution, on the other hand, this is obviously not the case. An assumption of which class of distribution (e.g., normal or exponential) the variable belongs to would enable the distribution to be summarised by only a few parameters. For example, if people assumed that the variable in Figure 1A is normally distributed, the mean and variance would be sufficient to fully determine the distribution. However, as illustrated by the fact that the distributions in Figure 1 come from two beta-distributions—parameterised by the $\alpha$- and $\beta$-parameters—making an assumption of the distribution that generates a variable is not always straightforward.

Research concerning knowledge of distribution shape has mainly investigated if people can estimate the distribution shape of variables they have encountered in their everyday lives (Fox &

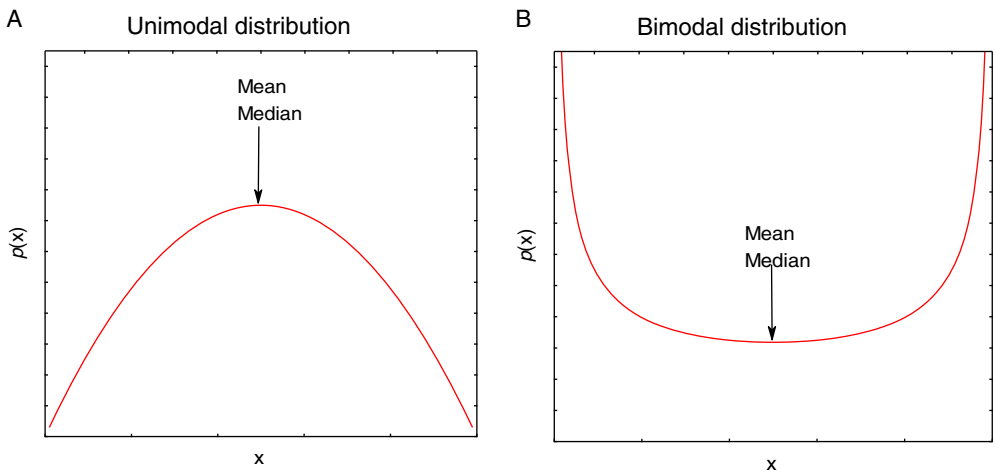


**Figure 1.** Illustration of a Unimodal (Panel A) and Bimodal (Panel B) distributions created with beta distributions.

Thornton, 1993; Griffiths & Tenenbaum, 2006; Jako & Murphy, 1990; Linville, Fischer, & Salovey, 1989; Nisbett, Krantz, Jepson, & Kunda, 1983; Nisbett & Kunda, 1985). Nisbett et al. (1983), for example, asked students to estimate the distribution of grade point averages among their peers, and Griffiths and Tenenbaum (2006) were concerned partly with distributions of baking times of pastries and movie runtimes. 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; but see Mozer, Pashler, & Homaei, 2008). It has also been noted that people often put more weight on observations close to where they find themselves in the distribution (Fiedler, 2000; Nisbett & Kunda, 1985). Here, we complement this research by studying the knowledge that people acquire about distribution shape in a controlled laboratory task.

Whereas the present study is concerned mainly with knowledge of distribution shape per se, this knowledge could potentially become manifest in and influence the quality of several judgement and decision-making tasks. Brown and Siegler (1993), for example, emphasised the importance of both metric (e.g., mean, variance and distribution) and mapping (ordinal relations within the domain) properties of a quantity for judgements about the quantity. Whereas research on heuristics (e.g., Gilovich et al., 2002), multiple-cue judgements (Juslin et al., 2008; von Helverson & Rieskamp, 2008) and function learning (DeLosh et al., 1997; Kalish et al., 2004) has been concerned with the influence of mapping knowledge on judgements, much less attention has been given to the influence of metric knowledge (but see, Pitz, Leung, Hamilos, & Terpening, 1976). Despite this, within several research areas it is generally assumed, often implicitly, that people are influenced by knowledge of distributional shape. For example, in both the literature on forecasting (e.g., Goodwin, 1996) and economic theory (e.g., Engelberg, Manski, & Williams, 2009), expert forecasters are expected to use the central tendency of a subjective probability distribution as point predictions for a variable under the assumption that the distribution is normal.

Further, recent 'rational', or Bayesian, models of human cognition often assume that people update their beliefs in the light of new data based on knowledge of prior distributions, the shape of which presumably has to be specified somehow (e.g. Chater, Tenenbaum, & Yuille, 2006; Oaksford & Chater, 2006; Tenenbaum et al., 2011). This line of research has shown that people's responses are consistent with them having knowledge of the properties of an empirical prior distribution (Griffiths & Tenenbaum, 2006, 2011). However, this research has not relied on direct measures of participants' knowledge of the prior distribution and has generally refrained from explicit claims about how knowledge of the prior is represented (Griffiths & Tenenbaum, 2011, but see Vul, Goodman, Griffiths, & Tenenbaum, 2009). Other studies suggest that people store the information in a 'raw' form, much like non-parametric frequency counts in 'mental histograms' (Malmi & Samson, 1983).

In addition, research concerned with decisions on binary gambles has recently directed much attention to paradigms where probabilities and outcomes are learned inferentially from experience (e.g., Hertwig, Barron, Weber, & Erev, 2004) rather than explicitly stated and exactly known from the task (e.g., Kahneman & Tversky, 1979). In the former, estimates often need to be generated post hoc from long-term memory (LTM) at the time of judgement. Several accounts of human judgement and decision-making moreover assume sampling from memory prior to making a judgement or decision (see, e.g., Busemeyer & Townsend, 1993; Denrell, 2005; Fiedler, 2000; Kahneman & Miller, 1986; Stewart, Chater, & Brown, 2006; Tversky & Koehler, 1994).

We suggest that, as outlined in detail later, post hoc sampling from LTM is a generic process that people use to realise their knowledge of distribution properties (e.g. its shape). If this is the case, such judgements and decisions are likely to be influenced by constraints on the cognitive process, such that they have to be based on small samples of data retrieved from LTM that can be activated within working memory constraints, and that the information integration is constrained by the sequential real-time properties of a controlled judgement processes (Juslin et al., 2007). The present study complements previous research by investigating how knowledge of the statistical properties of a set of encountered numbers is influenced by a memory sampling process.

## Cognitive representation of the distribution shape

People could respond to the request for an assessment of the distribution shape of a numerical

variable in at least two principally different ways, either by actively abstracting summary representations during exposure or by post hoc assessment based on retrieved examples from the distribution. To appreciate the distinction, it may be useful to make a comparison with the corresponding issue in categorisation learning (e.g. Ashby & Maddox, 2005). Some models of categorisation learning assume that during training people actively induce abstract summary representations of the categories, like prototypes or classification rules. Other models assume that people store representations of the concrete category exemplars in memory, which are retrieved at the time of classification, and the similarity to these exemplars are used for the classification. To what extent is the knowledge of distribution shape guided by abstract representations?

### *Abstraction of explicit representations*

A first possibility is that people have the cognitive capacity to spontaneously abstract representations of the distribution properties from experience with the variable. This assumes that summary information is extracted online during exposure to the variable, much like a spontaneous calculation of the intuitive equivalents of running estimates of the mean and the variance, as each additional observation is presented.

Because abstracted parameters cannot by themselves support knowledge of distribution shape, this process is likely to entail a priori 'assumptions' about the distribution shape. To the extent that beliefs about distribution shape are explicit we do not expect them to be precise and quantitative, but to have a rough, qualitative character capturing prototypical distribution shapes, like uniform distribution (e.g., 'all values are equally likely'), unimodal distribution (e.g., 'most values are in the middle'), and bimodal distribution (e.g., 'most values are at the extremes'). Such a qualitative assumption together with summary statistics roughly specifies the distribution. There is evidence that such an assumption about the distribution shape is likely to involve a normal (or unimodal) distribution (Flannagan, Fried, & Holyoak, 1986; Fried & Holyoak, 1984). In this paper, we do not further address the specific nature of such putative abstract representations of distribution shapes, or the viability of different processes whereby people acquire such representations, but concentrate on the more fundamental question of whether, and under what circumstances, people induce abstract beliefs about distribution shape from experience with a variable.

### *Post hoc memory sampling*

A second possibility is that people do not spontaneously abstract explicit summary representations, but rather retrieve a sample of observations from memory post hoc at the time of judgement and compute the judgement, as suggested by the *Naive Sampling Model* (NSM: Juslin et al., 2007). At the time of judgement, a sample of 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). This process—similar to a 'lazy algorithm', as described in artificial intelligence (Aha, 1997) and cognitive science (Juslin & Persson, 2002)—is naturally limited by constraints on short-term memory (Dougherty & Hunter, 2003; Gaissmaier, Schooler, & Rieskamp, 2006; Hansson, Juslin, & Winman, 2008; Kareev, Arnon, & Horwitz-Zeliger, 2002; Stewart et al., 2006). The sample of active observations in short-term memory is typically estimated to approximate $4 \pm 2$ observations (Cowan, 2001). To illustrate, a person may not know the exact average salary of people working in his or her workplace, but may retrieve a number of known salaries and, on this basis, estimate the average salary.

The qualification 'naive' in the NSM refers to the presumption that sample properties can be taken directly to describe population properties. Whereas some sample properties, like mean and proportion, are unbiased, other sample properties, like variance and coverage, are biased. That is, whereas the expected value of the former coincides with the corresponding population property under repeated random sampling, the expected value of the latter systematically distorts population properties (which is why sample variance needs to be corrected by $n/(n-1)$ to be an unbiased estimate of the population variance). The naive presumption that a sample property can be used as a proxy for the population property thus affords accurate judgements with some sample properties (mean, proportion), but poor judgements with other sample properties (variance, coverage). The implication is that the judgement is constrained, and sometimes biased, by being naively projected in this way from a small sample.

The NSM suggests that because distribution shape is a global property that is inherently difficult to condense into a few observations, the

small samples people have at their disposal will in general be insufficient to detect the shape of the population distribution. Lindskog, Winman, and Juslin (2013) illustrated that 'perceiving' the world through small samples not only makes it difficult to detect distribution shape, but, if anything, it will convey an illusion of unimodality. That is, the sample distribution shape will often be misleading by suggesting a unimodal shape regardless of the shape of the population distribution.

## Predictions

The distinction between abstraction during training versus post hoc sampling is similar to the distinction between 'eager' and 'lazy' learning methods in artificial intelligence (Aha, 1997). The eager learning algorithms try to generalise the training data by performing computations before the time of a query (e.g. of regression slopes, means). Rather than pre-computing abstractions for every conceivable future demand, the lazy algorithms store data and postpone the computations to when the specific computations needed are known. It has been proposed (Juslin & Persson, 2002) that in a complex and unpredictable environment, such as the environment that confronts the human mind, lazy algorithms, such as exemplar models (Nosofsky & Johansen, 2000) and the naive sampling model (Juslin et al., 2007; Lindskog et al., 2013), afford greater efficiency and flexibility for future demands. Consistent with this notion, previous research indicates that people often access data after encoding rather than spontaneously extracting descriptive parameters during encoding (e.g. Malmi & Samson, 1983) and that estimates of descriptive parameters are constrained by short-term memory (e.g. Hendrick & Constantini, 1970). In addition, statistical judgements such as the production of confidence intervals (Juslin et al., 2007) and point predictions (Lindskog et al., 2013) seem to be calculated on small samples drawn from memory at the time of a judgement.

On the basis of these theoretical arguments and previous empirical findings, we hypothesised that in general people do not spontaneously induce abstract representations of distribution properties but rather construct them post hoc by sampling from memory. However, as in other learning tasks, we expected people to be able to generate abstract representations if they are encouraged to do so during training (e.g. if asked prior to training to use a sample of observations to determine if the distribution is unimodal or not). In the following, we derive specific predictions from this general hypothesis and suggest limiting conditions that affect whether the one or the other of the two processes is likely to determine performance.

### *Format dependence*

Defining properties of abstract representations are that they are independent of perceptual modality and response mode and often can be applied more flexibly to the problem. More specifically, the independence of an abstract representation with respect to response mode suggests that the same representation can inform judgements irrespective of response format. Accordingly, if the respondent masters two response formats and the judgements in both are derived from the same abstract representation, it is reasonable to expect similar performance with both. For example, if you have the abstract insight that body weight is normally distributed, you should find it equally difficult to verbally express this fact, to reproduce the general shape of the distribution, as well as to visually identify its shape (i.e. at least if you have basic acquaintance with graphical representations, as expected from the university students in the experiments reported later). This follows from the same abstract representation being used in all three formats.

By contrast, if representations about distribution shape are reproduced post hoc by retrieval of observations, there should be profound format dependence (Juslin et al., 2007). In the experiments reported later, we use two tasks with different response formats to assess participants' knowledge. In a *proportion-production task*, similar to methods used to elicit subjective probability distributions from experts (see, e.g., Hora, Hora, & Dodd, 1992; Ludke, Stauss, & Gustafson, 1977; Winkler, 1967), participants assess the proportion of values falling in pre-defined intervals of the target variable. In a *visual-identification task*, the participants select one of several graphical illustrations of possible distribution shapes. Reproducing the distribution by assessing proportions in predefined intervals implies fairly accurate knowledge (Lindskog et al., 2013) even if the proportions are based on small samples constrained by short-term memory, because sample proportion is an unbiased sample property that on average yields accurate representations in the long run. On the other hand, small samples generally provide a very

poor basis for identifying the population distribution shape from the distribution shape of a small, momentarily activated, sample in short-term memory, as when people are asked to identify the correct graph depicting the distribution. As noted, the small sample may even convey an illusion of unimodality regardless of the population distribution. With post hoc sampling, people should be able to express more accurate knowledge of the distribution shape when they make judgements of proportions for pre-defined intervals, which involves an unbiased sample property, than if they are confined to rely on an unreliable, biased sample property like sample distribution shape.

In short, in tasks where the judgement is based on post hoc sampling from memory we predict a profound format dependence, with better performance with the production format than with the identification format. This format dependence should be much smaller in situations where the participants can benefit from an abstract representation of the distribution shape.

### *Intentional learning*

If people, as suggested earlier, have an ability to generate abstract representations of distribution shape if they are actively encouraged to do so under training, this leads to distinct predictions when distribution shape is learned under incidental and intentional learning. In conditions of incidental learning, people should be confined to post hoc sampling from memory and therefore be victims of format dependence. In conditions with intentional learning, people should be able to induce abstract representations of distributions shape, improving the performance with the identification format, and thereby decreasing the format dependence.

### *Format order effects*

For similar reasons, under conditions of incidental learning, we predicted characteristic order effects, depending on the order in which the two formats are encountered. The participants encountering the identification format before the production format are fully exposed to the unreliability and bias in the small samples momentarily activated by sampling from long-term memory leading to poor performance with the identification format, which due to the format dependence is substantially improved with the later production format.

By contrast, if, as we predict, performance will be quite accurate with the production format also under incidental learning, participants that begin with the production format should perform much better in the identification task than those who begin with the identification format. If the production format yields a fairly accurate representation of the distribution shape, this in effect forces the participant to produce an abstract representation of the distribution shape that is informative in the later identification task. In other words, if you first make productions (by basic exemplar retrieval from long-term memory) that strongly imply, for example, a bimodal distribution, you gain insight from this format that makes you less likely to mistake the distribution shape for a unimodal one with the subsequent identification format.

In Experiment 1, we evaluated the participants' knowledge of distribution shape under conditions of intentional vs. unintentional learning, either through *visual identification* of the distribution shape or through the ability to *produce* the distribution shape by proportion judgements, after trial-by-trial exposure to a numerical variable. In Experiment 2, we replicated some of the key results from Experiment 1, whereas investigating the predictions of order effects more systematically.

## EXPERIMENT 1: DISTRIBUTION SHAPE AND INTENTIONAL LEARNING

In a learning phase, participants observed numbers presented as the quarterly revenue of companies. In a test phase, the participants' knowledge was elicited with the production and the identification formats described earlier. If people develop abstract knowledge of the distribution shape, we expect similar performance with the two formats, but if people rely on post hoc sampling from memory we expect a clearly superior performance with the production format that benefits from use of an unbiased sample property.

In Experiment 1, we investigate four main questions. First, the reliance on post hoc sampling will presumably make it difficult to detect the true distribution shape and introduce an impression of unimodality regardless of the true distribution shape (Lindskog et al., 2013). The revenues in Experiment 1 were from either a unimodal or a bimodal distribution and we predicted a response bias towards unimodality. Second, we wanted to test the prediction of a format-dependence effect, with substantially poorer performance with the identification format than with the production format. This should hold at least under incidental

learning. Third, we wanted to investigate if intentional learning influences this format dependence. Half of the participants were therefore told about the subsequent tests (intentional learning), the other half were not (incidental learning). If the participants induce abstract knowledge of the distribution shape when they are encouraged to do so, intentional learning should make the format dependence diminish. Finally, the within-subjects design naturally allowed us to test the prediction of specific order effects.

Previous research has shown that performance is sometimes impaired by intentionality when keeping track of frequencies (Zacks, Hasher, & Sanft, 1982; but see Greene, 1986). To investigate a salient potential predictor of individual differences in the ability to acquire knowledge of the properties of a variable's distribution, we also obtained a measure of basic mathematical skills and understanding of numbers, *Numeracy* (see, e.g., Reyna, Nelson, Han, & Dieckmann, 2009).

## Method

### *Participants*

Participants were 48 undergraduate (17 male and 31 female) students from Uppsala University ($M$ = 24.3 years, $SD$ = 4.8). They received a movie voucher or course credits as compensation for participating in the study.

### *Materials and procedure*

The computerised task consisted of two phases, a learning phase and a test phase, and was carried out on a PC. The objective for participants was to learn (remember) the revenues of fictitious companies. Two distributions of 60 values each, a symmetric bimodal distribution (Beta(.33, .33)) and a symmetric unimodal distribution (Beta(3.4, 3.4)), linearly transformed to the range [0, 1000], defined the unimodal and bimodal conditions, respectively (see Table 1). For each participant, the revenues were randomly paired with one of 156 company names. The learning phase consisted of six blocks, each company occurring once in each block. On each trial, participants saw the name of a company and had to predict/guess the value of the company's revenue. Predictions were followed by feedback and the presentation was self-paced. In the instructions to the learning phase, half of the participants received information about the upcoming production and identification formats (intentional condition) and half of the participants were withheld this information (incidental condition).

During the test phase, the participants were asked to produce and to identify the distribution of the target variable. With the production format participants assessed how many of the 60 companies fell into 10 equally wide intervals (frequency task) and the probability that a random company among those observed would have its revenue in that interval (probability task). The number of companies was required to sum to 60 and the probabilities were required to sum to 1.0. With the identification format, participants chose one of 11 graphs. An explanation of the graphs was given prior to their presentation. This explanation included several explicit exemplifications (e.g., 'A graph which is higher to the sides than in the middle indicates that most of the revenue values were either high or low.'). The graphs were provided without metric information, frequencies or probabilities, on the $y$-axis since we were only interested in non-metric, qualitative, knowledge of distribution shape. Participants indicated what graph that best described the distribution of observed values. All graphs were created from a beta distribution. One graph was uniform ($\alpha = \beta = 1$), three graphs were bimodal and symmetric ($\alpha = \beta = .2$, .33 and .8), three graphs were unimodal and symmetric ($\alpha = \beta = 2$, 3.4 and 7), two graphs were bimodal and skewed ($\alpha = .2$, $\beta = .8$ and $\alpha = .8$, $\beta = .2$) and two graphs were unimodal and skewed ($\alpha = 2$, $\beta = 8$ and $\alpha = 8$, $\beta = 2$). The participants performed identification once with graphs in the form of continuous density functions and once with graphs in the form of histograms with 10 equally spaced intervals. The order of these

**TABLE 1**
Characteristics of the sets of values used for the company quarterly revenue in Experiment 1, for the unimodal and bimodal conditions, respectively

| *Distribution* | *Mean* | *Median* | *Min* | *Max* | *SD* | *MAD* | α | β |
|---|---|---|---|---|---|---|---|---|
| Bimodal | 500 | 500 | 0 | 1000 | 389.2 | 355.2 | .33 | .33 |
| Unimodal | 500 | 500 | 127 | 873 | 182.9 | 150.9 | 3.4 | 3.4 |

variants was counterbalanced. Participants also made direct estimates of the median and mean absolute deviation (MAD) of the observed distribution, after a brief explanation and introduction to each of these concepts. Participants finally filled out a questionnaire consisting of 11 items covering numeracy. The questionnaire was a Swedish translation of the questionnaire used by Lipkus, Samsa, and Rimer (2001) (See also Lipkus & Peters, 2009; Peters et al., 2006).

## Design

Experiment 1 used a 2 × 2 factorial design with distribution (unimodal/bimodal) and intentionality (intentional/incidental) as independent between-groups variables. Participants were randomly assigned to the experimental conditions. The approximate length of the experiment was 120 minutes.

## Results

There were no significant differences between the production format with frequencies or probabilities, and between identifying histograms or density functions, and data in these conditions, respectively, were therefore collapsed.

### *Performance measures for knowledge of shape of distributions*

For each participant, we calculated a *mean absolute error* (MAE) for the production format ($MAE_P$) and the identification format ($MAE_{VI}$), respectively. $MAE_P$ was the mean absolute error between the rated and actual frequency of all intervals given by

$$MAE_P = \frac{\sum_{i=1}^{10} |r_i - a_i|}{10}, \quad (1)$$

where $r_i$ is the rated frequency of interval $i$ and $a_i$ is the actual frequency of interval $i$. Notice that $a_i$ in Equation 1 will differ in the unimodal and bimodal conditions.

For the identification data, we divided the range ([0, 1]) of each of the 11 underlying beta distributions into 10 equally wide intervals ([0, .1], [.11, .2]…[.91, 1]) and calculated the density ($d_i$) for each interval ($i$). Multiplying by 60 resulted in an expected frequency count ($f_i$) for each interval of each graph equivalent to the frequency count used for the production format. Using Equation 1 and inserting $f_i$ for $r_i$ and the $f_i$ of the Beta(.33, .33) or Beta(3.4, 3.4), depending on condition, for $a_i$ we calculated a MAE for each graph. The dependent measure of performance in the identification format ($MAE_{IV}$) was the MAE value of the chosen graph.

In order to compare the two dependent measures, $MAE_P$ and $MAE_{IV}$, we used a measure, *mean absolute ratio* (MR), which scales the performance of participants ($MAE_S$) in each format against the difficulty ($MAE_R$) of the formats. MR is given by

$$MR = \frac{MAE_R - MAE_S}{MAE_R}, \quad (2)$$

where $MAE_R$ is a measure of the difficulty, given by the mean absolute error expected by random performance, and $MAE_S$ is the performance of the participant. MR is thus on a [ ∞, 1] scale where 1 represents perfect performance and 0 represents random performance. The procedure used to derive $MAE_R$ for the two formats is outlined in the Appendix.[1]

### *Producing the distribution*

Figure 2 presents the average assessed proportion of companies in each interval (grey bars) together with the proportions from the underlying

[1] Whereas the MR measure was created to standardise performance in the two tasks against random performance, it is possible that our model of random judgement will have affected the outcome of the analyses. To investigate the robustness of our results, we therefore reran all analyses with two separate changes in the assumptions. First, in the identification task we assume that there is an equal probability of any graph being chosen by a naive participant. However, a naive participant might choose an uninformative graph (i.e. the uniform graph) rather than any of the graphs with equal probability. We therefore changed $MAE_R$ in the identification task to equate the choice of the uniform distribution and reran all analyses of main effects and interactions in both experiments. These analyses revealed qualitatively equivalent results in both experiments. However, the format by order interaction in Experiment 1 was now only marginally significant. Thus, even when choosing a model of random judgement that suggests an uninformative rather than random response the results are similar. Second, and making an even stronger test of the robustness, we reran all analyses whereas assuming that both tasks were equally difficult under random performance. The results were similar to the original results with similar qualitative conclusions. However, in Experiment 2 the main effect of format did not reach significance. Thus, even under unrealistic assumptions, when equating the two tasks under random performance, we find comparable results suggesting that our results are fairly robust to the choice of dependent measure.

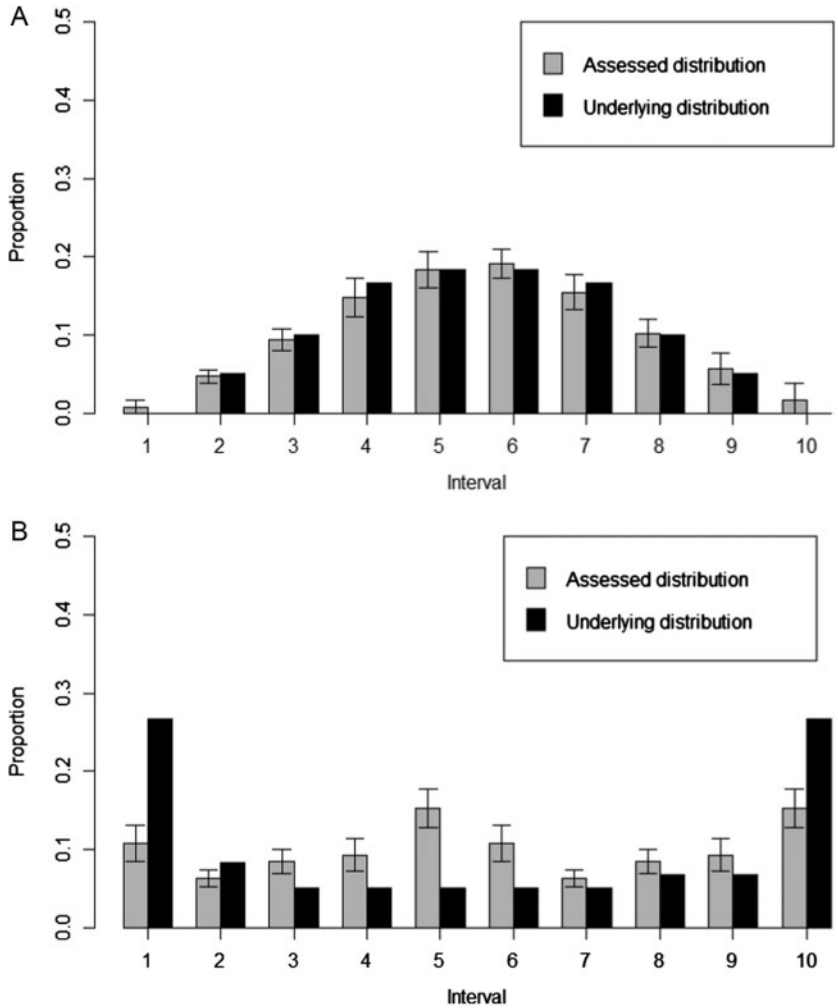


**Figure 2.** Assessed proportion of companies for each interval (grey bars) and the underlying distribution (black bars) for the unimodal (2A) and bimodal (2B) conditions separately in Experiment 1. Whiskers denote 95% confidence bars.

distribution (black bars) for the unimodal (Figure 2A) and bimodal (Figure 2B) conditions, respectively. The figure illustrates that participants in the unimodal condition are better at reproducing the underlying distribution than participants in the bimodal condition. Participants in the latter condition tend to underestimate the proportion of companies in the extreme intervals and overestimate the proportion in the middle intervals. The difference in performance in terms of MAE was significant ($t(46) = 3.4$, $p = .001$; unimodal: $M = 2.0$, $SD = 1.32$; Bimodal: $M = 3.3$, $SD = 1.3$).

### *Identifying the distribution*

Choices of the graphs were classified as having the same (congruent) or opposite (incongruent) shape as the underlying distribution. Most participants chose a graph that was congruent (83% and 75% in the unimodal and bimodal conditions, respectively). The interaction between congruent/incongruent choice and distribution was not significant ($\chi^2(1, N = 24) = 1.01$, $p = .31$). In terms of MAE, there was no significant difference between the two conditions ($t(46) = .75$, $p = .46$; unimodal: $M = 3.5$, $SD = 2.4$; bimodal: $M = 3.0$, $SD = 2.5$).

### *Performance on production vs. on identification*

As is evident in Figure 3, the variance in the two formats is not homogenous, preventing the formal analysis of interaction terms with a standard ANOVA. By standardising MR in the two formats, across each format separately, their variances (and means) will be equated, enabling the use of ANOVA. In the following, we use standardised variables to investigate interaction effects using ANOVA and non-standardised variables to

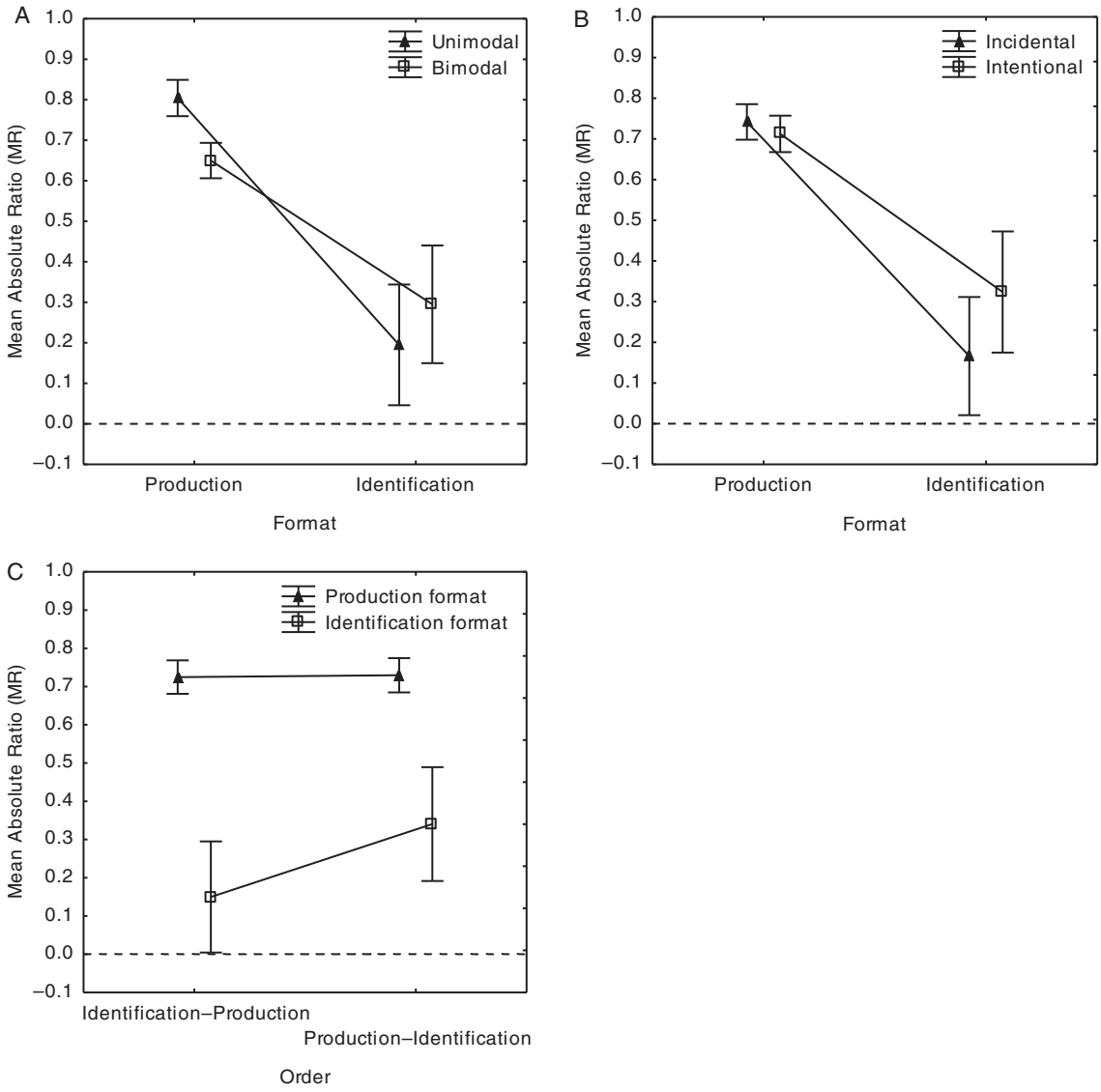


**Figure 3.** Mean absolute ratio for estimates in the bimodal and unimodal conditions (Panel A), incidental and intentional learning conditions (Panel B) and the production and identification formats (Panel C) as a function of task type (Panel A and B) or order (Panel C) in Experiment 1. Vertical bars denote 95% confidence intervals.

investigate main effects with non-parametric tests.[2] The figures illustrate the effects with means and 95% confidence intervals (CI) for unstandardised variables.

### *Interactions*

To investigate the interaction terms, we ran a general linear mixed models analysis with format (production/identification) as independent within-subjects variable and distribution (unimodal/bimodal), intentionality (incidental/intentional) and order (production–identification/identification–production) as independent between-subjects variables and standardised MR scores as dependent variable. The analysis revealed three significant interaction effects (all other *p*s > .16).

*Format by distribution.* First, Figure 3A illustrates the significant format by distribution interaction ($F(1,40) = 6.37$, MSE = .55, $p = .016$), with larger format dependence for the unimodal than the bimodal distribution. Post-hoc analysis showed

[2] Whereas the standardisation is undertaken to allow use of ANOVAs by adherence to the homogeneity of variance assumption underlying the test, we verified that this standardisation procedure does not itself affect the conclusions of the analyses.

that only the differences in the production format between the unimodal and bimodal conditions were significant. The interaction was unexpected because we predicted an overall advantage in performance for the unimodal distribution regardless of task.

*Format by intentionality.* Second, Figure 3B illustrates a significant format by intentionality interaction ($F(1,40) = 6.38$, MSE = .55, $p = .016$). As predicted, if people induce abstract representations of the distribution when actively encouraged to do so by instruction (with intentional learning), the format dependence is smaller than if they do not (incidental learning), although the difference is fairly modest.

*Format by order.* Finally, Figure 3C shows the significant format by order interaction, ($F(1,40) = 4.17$, MSE = .55, $p = .047$). The figure illustrates that, as predicted, performance in the production format is unaffected by order. However, performance in the identification format improves when it is temporally preceded by the production format.

### *Main effects*

We investigated the main effects of distribution, format, and intentionality separately using non-parametric tests.

*Distribution.* We predicted an advantage in performance for the unimodal over the bimodal distribution. As is clear from Figure 3A, and in contrast to our prediction, there was no such overall advantage (Mann–Whitney: $U = 249$, $Z = .79$, $p = .43$).

*Format.* As illustrated in Figure 3A, the mean MR is substantially higher in the production format than in the identification format, with distinctly separated confidence intervals for the means for both the unimodal and the bimodal distributions. Collapsed across both distributions the mean MR was .71 ($SD = .15$) for the production format and .31 ($SD = .42$) for the identification format. This confirms the format dependence, with superior performance in the production format (Wilcoxon; $T = 78$, $Z = 5.23$, $p < .001$).

*Intentionality.* We predicted smaller format dependence with intentional learning than with incidental learning. It is possible that such a difference would also give a main effect of intentionality. Comparing performance with intentional versus incidental learning, however, revealed no significant difference (Mann–Whitney: $U = 256$, $Z = .64$, $p = .52$).

### *Estimates of descriptive measures*

The accuracy of the estimates of variability and central tendency was analysed with a 2 × 2 × 2 split-plot ANOVA, with distribution (unimodal/bimodal) and intentionality (intentional/incidental) as independent between-subjects variables, the two descriptive statistics (median/MAD) as independent within-subjects variable and the absolute deviation from the normative distribution parameter as dependent variable. The analysis revealed a significant main effect of distribution ($F(1, 45) = 10.9$, MSE = 11,785, $p = .002$) with the unimodal condition having lower absolute deviations ($M = 77.2$, $SD = 72.7$) than the bimodal condition ($M = 150.5$, $SD = 119.6$). None of the other effects were significant (all $p$s > .37).

To investigate possible over-/underestimation of the estimates (as opposed to absolute deviation), the signed deviation was entered as the dependent measure into an analogous ANOVA. This analysis revealed a significant main effect of type of statistic ($F(1, 45) = 15.5$, MSE = 15101.1, $p < .001$), with estimates of MAD underestimating the true variability ($M = 77.4$, $SD = 124.8$), whereas estimates of the median overestimated the central tendency ($M = 21.4$, $SD = 163.2$). The analysis further revealed a significant type of statistic by type of distribution two-way interaction ($F(1, 45) = 18.77$, MSE = 15101.1, $p < .001$). The participants in the bimodal condition overestimated the central tendency and underestimated the MAD, whereas the estimates in the unimodal condition show no such biases. No other effects in the analysis reached significance (all $p$s > .48). In sum, direct estimates of the distribution parameters were more accurate for the unimodal distribution.

### *Influence of numeracy*

To investigate the influence of numeracy on performance with the two formats, we calculated the correlation between numeracy and performance in the production and identification formats as well as with the measures of absolute deviation of the estimates of distribution statistics separately for the unimodal and the bimodal conditions. In the bimodal condition, neither $MAE_{IV}$ ($r(22) = -.01$, $p = .95$) nor $MAE_P$ ($r(22) = -.18$, $p = .41$) correlated significantly with numeracy. The same held for the numeracy–$MAE_{IV}$correlation ($r(22) = -.25$, $p = .23$) in the unimodal condition, but there was a significant numeracy–$MAE_P$ ($r(22) = -.44$, $p = .03$) correlation in the unimodal condition with a higher level of numeracy related to lower $MAE_p$.

There was no significant correlation between numeracy and accuracy of estimates of central tendency in either condition (unimodal: $r(22) = -.19$, $p = .38$, bimodal: $r(22) = -.15$, $p = .48$), and whereas the corresponding correlation for MAD did not reach significance in the unimodal condition ($r(22) = -.30$, $p = .15$), it did in the bimodal condition ($r(22) = .50$, $p = .01$), where higher numeracy was associated with more accurate MAD estimates. None of the differences in correlations between the unimodal and bimodal condition were significant (all $p$s $> .3$).

## Discussion

In Experiment 1, we investigated four main questions. First, we predicted better performance for the unimodal distribution than for the bimodal distribution with a response bias to produce and identify unimodal distributions. Second, we predicted format dependence with superior performance with the production format over the identification format. Third, we investigated how the performance was influenced by intentionality of learning. Fourth, the within-subjects design of the experiment allowed us to investigate order effects.

The hypothesis of a response bias favouring unimodal distributions was confirmed with the production format but not with the identification format, with the format by distribution interaction showing a significant difference between the two distributions in the production format but not in the identification format. This was unexpected, because we expected clearly better performance in the unimodal condition regardless of format. Further, the estimates of median and MAD were significantly more accurate in the unimodal distribution. There was strong format dependence with better performance with production than with identification. Further, this effect was stronger in the unimodal condition.

Intentionality improved accuracy only with the identification format, suggesting that performance with this format is related to the induction of abstract representations, whereas performance with the production format is not. In fact, with the production format performance was, if anything, impaired by intentional learning. That intentionality did not influence estimates of descriptive properties suggests that participants lack strategies to extract these properties online by keeping a running mean (or similar) that is updated on a trial-by-trial basis. However, it also indicates that people actually have quite good resources to function as intuitive statisticians using data stored in memory. Whereas performance with the production format was uninfluenced by order, performance with the identification format improved when it was preceded by the production format. This indicates that encountering the production format might effectively induce an abstract representation that participants can use to improve performance in the subsequent identification format.

Together these results provide evidence for at least two claims. First, the process spontaneously engaged by the participants seems to be a post hoc sampling from memory, which explains the large format dependence with extremely poor performance with the identification format that is especially strong under incidental learning. Second, when instructed to, or strongly invited by the format, people have the ability to induce abstract representations, which explains the improvement in identification with intentional learning and the order effects.

There was a significant correlation between numeracy and performance with the production format with higher numeracy associated with better performance. However, this was the case only in the unimodal condition. For the bimodal condition, numeracy instead correlated significantly with error in estimates of MAD, where higher numeracy was related to lower error. If people have a priori assumptions of unimodality, it may be that a high level of numeracy is required to 'override' such an assumption. Thus, a higher level of numeracy is especially beneficial when estimating descriptive properties when the underlying distribution is not unimodal.

# EXPERIMENT 2: ORDER EFFECTS AND RANGE SALIENCE

The purpose of Experiment 2 was to replicate the findings in Experiment 1 of a format-dependence effect and order effects, but also to investigate one additional factor that might affect the performance. A bimodal distribution is by definition associated with a large number of extreme observations, which should make the range better learned than in a unimodal condition. Indeed, learning the range of a variable would seem essential for encoding encountered numbers as of a low or high magnitude in a distribution. In Experiment 2, we thus investigated the possibility that the range end point salience might influence the knowledge of distributional shape.

## Method

### *Participants*

Participants were 48 undergraduate (32 male and 16 female) students from Uppsala University ($M$ = 25.6 years, $SD$ = 8.0). They received a movie voucher or course credits as compensation for participating in the study.

### *Materials and procedure*

The experiment was performed in the same way as Experiment 1 and used a $2 \times 2$ factorial design with distribution (unimodal/bimodal) and range salience (obvious/non-obvious) as between-subjects variables. Sixty values from a symmetric unimodal distribution (Beta(2.4, 2.4)), linearly transformed to the range [0, 1000], defined the revenue values in the unimodal/obvious condition. Adding 304 to each of these values generated the revenues for the unimodal/non-obvious condition. Revenue values for the bimodal/obvious condition were created by first calculating the interval frequency $Fbi_j = Fuf_j - (Fum_j - Fuf_j)$ for each of the 10 intervals [1, 100], [101, 200]…[901, 1000], where $Fuf_j$ is the frequency of the uniform distribution in interval $j$, $Fum_j$ is the frequency of the unimodal–obvious distribution in interval $j$ and $Fbi_j$ is the frequency of the bimodal–obvious distribution in interval $j$. For each interval, $Fbi_j$ numbers were then drawn uniformly on that interval to get the revenue values. This operation is the equivalent of a reflection of the unimodal–obvious distribution with respect to the uniform distribution. Adding 304 to each value of the bimodal/obvious distribution defined the revenue values for the bimodal/non-obvious condition. For each participant, the revenues were randomly paired with one of 156 company names. The values were presented in four blocks, in an individually randomised order, with each company occurring once in each block.

In the test phase, participants performed the production and identification formats described in Experiment 1. The reflection procedure ascertained that the mean absolute error at random performance for the production format was equal in the unimodal and bimodal conditions ($MAE_R$ = 8.3). We further choose the 11 graphs in the identification format to give the same level of chance performance in the unimodal and bimodal conditions ($MAE_R$ = 3.1). The approximate length of the experiment was 120 minutes.

## Results

We used the same performance measure (MR; see Equation 2) as in Experiment 1. As in Experiment 1, variance in the two formats proved not to be homogenous, and we therefore used the same standardisation procedure to investigate interaction terms with ANOVAs whereas main effects were analysed with non-parametric tests.

### *Producing the distribution*

Performance in the production task was similar to that seen in Experiment 1. Participants in the unimodal condition reproduced the underlying distribution significantly better than participants in the bimodal condition ($t(46) = 3.0$, $p = .004$; unimodal: $M = 1.8$, $SD = 1.2$; bimodal: $M = 2.7$, $SD = 1.0$). More specifically, participants in the bimodal condition underestimate proportions in the two extreme intervals.

### *Identifying the distribution*

Performance in the identification task mimicked those of Experiment 1 with most participants giving congruent choices (85% and 62% in the unimodal and bimodal conditions, respectively). The interaction between congruent/incongruent and distribution was significant ($\chi^2(1, N = 24) = 6.54$, $p = .01$). The difference in terms of MAE was not significant between the two conditions ($t(46) = 1.96$, $p = .06$; unimodal: $M = 1.6$, $SD = 1.6$; bimodal: $M = 2.7$, $SD = 2.3$).

### *Interactions*

To investigate the interaction terms, we ran a general linear mixed models analysis with format (production/identification) as independent within-subjects variable and distribution (unimodal/bimodal), range salience (obvious/non-obvious) and order (production–identification/identification–production) as independent between-subjects variables and standardised MR-scores as dependent variable. The analysis revealed two significant interactions (all other $p$s > .18).

*Distribution by order.* First, as is clear from Figure 4A and the significant distribution by order interaction ($F(1,40) = 4.71$, MSE = .96, $p = .032$,), the order effects were different for the unimodal and the bimodal distributions. With the unimodal distribution, the accuracy advantage of encountering the production format prior to the identification format is very modest with overlapping CIs,

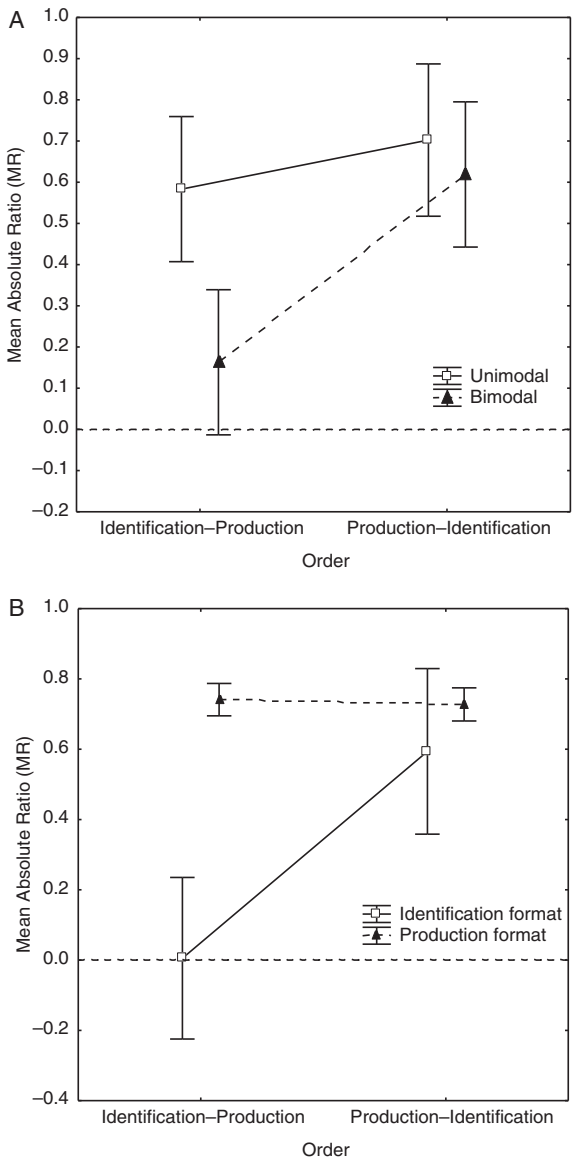


**Figure 4.** Mean absolute ratio in the unimodal and bimodal conditions (Panel A) and with the identification and production format (Panel B) as a function of order in Experiment 2. Vertical bars denote 95% confidence intervals.

but with the bimodal distribution, there is a clear advantage of performing the production format first. Encountering the identification format prior to the production format in the bimodal condition pulls the overall accuracy distinctly below the overall accuracy in the other three conditions in Figure 4A.

*Format by order.* Second, in Figure 4B, we see that the order effects observed in Experiment 1 were replicated in Experiment 2 with a larger difference between performance with the two formats if the identification format was encountered first ($F(1,40) = 13.5$, MSE = .57, $p < .001$). Performance with the production format was unaffected by the order, whereas performance with the identification format benefits substantially from experience with the production format. As in Experiment 1, this result suggests that the abstract representation needed to perform well with the identification format is not spontaneously available. Experience with the production format, however, seems to invite the participants to induce such an abstract representation.

### *Main effects*

We investigated the main effects of distribution, format and range salience separately using non-parametric tests.

*Distribution.* In Experiment 2, the overall predicted difference in performance between the unimodal ($M = .63$, $SD = .26$) and bimodal distribution ($M = .39$, $SD = .41$) approached significance (Mann–Whitney: $U = 195.5$, $Z = 1.89$, $p = .057$). Comparing performance between the two distributions for the two formats separately replicated the findings from Experiment 1 and revealed a significant difference in the production format ($t(46) = 3.04$, $p = .003$), whereas the difference in the identification format approached significance ($t(46) = 1.96$, $p = .056$). Thus in both formats, performance was better in the unimodal than in the bimodal condition.

*Format.* Experiment 2 replicates the format dependence observed in Experiment 1, with better overall performance with the production format ($M = .72$, $SD = .15$) than with the identification format ($M = .30$, $SD = .67$) (Wilcoxon; $T = 205$, $Z = 3.80$, $p < .001$).

*Range salience.* The ranges were created to be salient and non-salient. Participants estimated max and min values of the target variable. As a check of this manipulation, we calculated a composite measure of these two estimates as the mean of the absolute deviations from the normative min (0 and 304) and max values (1000 and 1304). An independent *t*-test revealed a significant difference ($t(45) = 3.88$, $p < .001$) in this measure between the two groups; participants in the obvious condition gave more correct range estimates ($M = 4.8$, $SD = 16.2$) than did participants in the non-obvious condition ($M = 32.9$, $SD = 30.9$), indicating that range end points in the non-obvious condition were as expected learned more poorly.

If learning the end points is essential for learning the distribution shape, we expected to find an effect of range salience. Comparing performance with obvious versus non-obvious range end points, however, revealed no significant difference (Mann–Whitney: $U = 211.5$, $Z = 1.57$, $p = .11$).

### *Estimates of descriptive measures*

To investigate the effect of the experimental variables on the estimates of variability and central tendency, we performed a 2 × 2 × 2 split-plot ANOVA with the two experimentally manipulated factors as independent between-subject variables, the two types of statistic (median/MAD) as independent within-subject variable and the absolute deviation from the normative value of the distribution parameter as dependent variable. The analysis revealed a significant main effect of distribution ($F(1, 44) = 18.37$, MSE = 6633, $p < .001$) with the unimodal condition being associated with lower absolute deviation ($M = 83.3$, $SD = 55.6$) than the bimodal condition ($M = 155.7$, $SD = 106.5$). None of the other effects were significant (all $p$s > .30). To investigate over-/underestimation, the signed deviation (as opposed to the absolute deviation) was entered as dependent measure into an analogous ANOVA. This analysis revealed a significant main effect of type of statistic ($F(1, 44) = 16.6$, MSE = 14482.7, $p < .001$) with estimates of MAD underestimating variance ($M = -106.5$, $SD = 117.5$) whereas estimates of median showed no bias ($M = -7.3$, $SD = 142.4$). There was a significant two-way interaction between type of statistic and type of distribution ($F(1, 44) = 5.14$, MSE = 14482.7, $p = .03$). As illustrated in Figure 5, this interaction is a result of participants in the bimodal condition underestimating MAD to a higher degree than those in the unimodal condition, whereas there is no apparent bias in either condition for estimates of central tendency. Further, there was a significant main effect of range salience ($F(1, 44) = 4.4$, MSE = 16745.3, $p = .04$) where the non-obvious condition was associated with a larger underestimation ($M = -84.6$, $SD = 135.3$) than in the obvious condition ($M = -28.1$, $SD = 138.5$). All other $p$s > .17.

## Discussion

Experiment 2 was designed to replicate the main findings in Experiment 1 and to investigate the possibility that the range end point salience of a target variable could influence knowledge of the distributional shape of that variable. Again, we found strong format dependence with better performance with the production format than with the identification format for both distributions. Further, we replicated the effect of distribution with better performance in the unimodal condition in the production format.

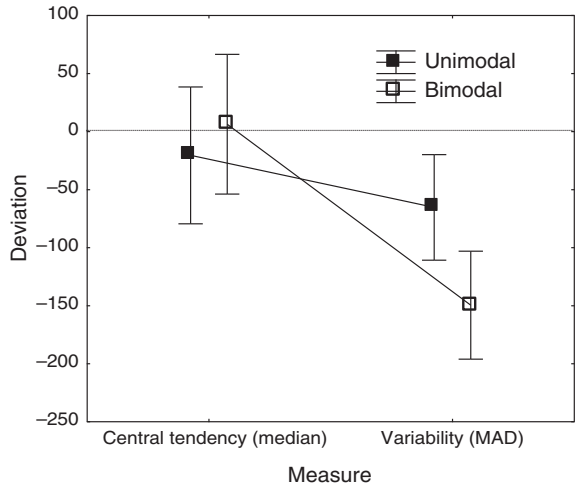


**Figure 5.** Deviation of estimates of Central tendency (median) and Variability (mean absolute deviation) for the unimodal and bimodal conditions in Experiment 2. Vertical bars denote 95% confidence intervals.

Even though participants in the obvious condition had better knowledge of the range end points than did those in the non-obvious condition, there were no main effects of this variable with the production and identification formats. This suggests that for learning the shape of the distribution knowledge of the range end points is of less importance. However, the effect of range salience on estimates of descriptive statistics (actual deviation) indicates that estimates become less biased when range end points are obvious, suggesting that people can utilise this information when making the estimates.

Finally, Experiment 2 confirmed that estimates of central tendency are more veridical than estimates of variance (see Pollard, 1984) and showed that the tendency to underestimate variance (Kareev et al., 2002) is stronger when the underlying distribution is unimodal rather than bimodal.

# GENERAL DISCUSSION

The ability of people to act as intuitive statisticians has been investigated by a large body of research. In this paper, we extend previous work by

investigating people's knowledge of higher order properties of numerical variables. More specifically, the present paper addresses the question of whether knowledge of a numerical variable is abstracted online during encoding, similar to a running mean or a posterior distribution, or if data are stored in a raw format during encoding and calculations made first at the time of judgement. Put differently, is man an eager intuitive statistician that generalises the training data before the time of a query or does he, like a lazy intuitive statistician, wait to perform calculations until after a query? The results from the two experiments suggest that people are constrained by the post hoc assessment of distribution properties by sampling from long-term memory (Juslin et al., 2007). However, although post hoc sampling from memory seems to be the default process, under certain predictable circumstances people do have a capacity to induce abstract representations of distribution shape.

In the two experiments, we tested three main predictions, along with limiting conditions, from the general hypothesis that people do not spontaneously induce abstract representations of distribution properties but rather construct them post hoc by sampling from memory. First, the post hoc memory sampling account predicted a format dependence effect with better performance when people could use an estimator which is unbiased than when they could not. In both experiments, we found a strong format dependence effect where participants performed better with the production format than in the identification format. It could be argued that such an effect might emerge because of general problems with interpreting graphs. Indeed, the use of graphical illustrations is not always straightforward (Friel, Curcio, & Bright, 2001; Galesic & Garcia-Retamero, 2011). However, most of our participants were university students with at least some basic training in statistics. This in conjunction with the fact that level of numeracy did not correlate with performance in the identification format in Experiment 1 makes a comprehension explanation even less probable.

Second, if people have an ability to generate abstract representations of distribution shape, when actively encouraged to do so, we predicted effects of intentional learning. When learning is incidental, people should be confined to post hoc sampling from memory and therefore be victims of format dependence whereas intentional learning should encourage people to induce abstract representations of distributions shape, improving the performance with the identification format, decreasing the format dependence. There was no main effect of intentionality in Experiment 1. However, as expected the format by intentionality interaction indicated a smaller format effect with intentional learning than with incidental learning. It might be that the instructions given in the intentional condition were not a sufficiently strong manipulation to induce intentional learning. A possibility that should be investigated in future research is to interrupt the learning phase with several reminders or tests. Our results, however, indicate that this might change the representation altogether, not as a result of intentionality but as result of eliciting knowledge.

Finally, we predicted characteristic order effects, depending on the order in which the two formats were performed. Whereas performance with the production format was expected to be uninfluenced by format order, performance with the identification format was expected to improve if it was preceded by the production format. This is because participants performing the identification format first are fully exposed to the unreliability of small samples whereas those performing it after the production format would be able to benefit from being forced to produce an abstract representation of the distribution shape that is informative in the later identification format. The characteristic order effects were present in Experiment 1 and replicated in Experiment 2 with significant format by order interactions. In both experiments, performance with the production format was unaffected by order, whereas performance with the identification format benefited substantially from being preceded by the production format.

Taken together these results suggest two major conclusions. First, the process spontaneously engaged by the participants seems to be a post hoc sampling from memory. Second, when instructed to, or strongly invited by the format, people have the ability to induce abstract representations. Whereas the first conclusion is supported by the strong format dependency effects in both experiments, an effect that was especially strong under incidental learning, the second finds support in the improvement in identification with intentional learning seen in Experiment 1 and the order effects seen in Experiments 1 and 2. It might be that the degree to which an abstraction is induced depends on how often exemplars in the underlying distribution are activated in memory. That is, a strong abstraction might be formed only after exemplars from the underlying distribution have

been activated repeatedly (see Kahneman & Miller, 1986, for a similar argument related to norms). It remains for future research to determine how often exemplars in a distribution need to be activated before a reliable abstraction is formed.

In the present study, we elicit explicit judgements of distributions using two specific tasks. Whereas we acknowledge that such explicit distribution judgements are seldom elicited in real life, our results have important theoretical implications beyond the specific tasks. Our results indicate a generic process that people will engage in whenever they are prompted for a judgement that requires the evaluation of statistical properties of an experienced variable. Thereby they also suggest how metric knowledge of this variable is realised at the time of judgements, regardless of the judgement task. In decision tasks where metric knowledge is important, for example in tasks using the decisions from experience paradigms (Hertwig et al., 2004), our findings could give an indication of how the distribution properties of the experienced binomial distribution are realised. Further, some Bayesian accounts of cognition have suggested that priors are realised by a sampling process from memory similar to a MCMC-sampling procedure (e.g. Vul et al., 2009). If this is the case, our results predict that (1) people will generally realise priors that are normally distributed and (2) the representation of a prior may change if it is repeatedly elicited. Similar consequences could be expected for several models that assume sampling prior to judgements (e.g., Denrell, 2005; Fiedler, 2000). Finally, whereas not very common in everyday life, explicit distribution judgements are often elicited from experts in various domains (O'Hagan et al., 2006). The findings from the present study may help inform and improve such judgements by suggesting how they are formed and what cognitive constraints shape them.

The present study tests the two alternative accounts of how people make judgements about distribution shape by evaluating a priori predictions derived from the hypothesis that people in general do not spontaneously induce abstract representations of distribution properties but rather construct them post hoc by sampling from memory. As such, the two accounts are evaluated in an indirect manner and our conclusions will depend on the validity of our predictions. Therefore, even though the two accounts are derived from results and theoretical accounts found in previous research (Juslin et al., 2007) and the results converge with previous findings (Lindskog et al., 2013), further research including more direct approaches is warranted before drawing strong conclusions. A more direct approach would be to formalise both accounts as computational models and compare predictions from the models with judgements by participants. Such an approach would be an interesting and promising venue for future research.

In addition to the three main predictions, we also investigated the prediction that people will have a general response bias towards unimodality. Indeed, previous research has indicated that people might expect unimodally distributed variables (Flannagan et al., 1986; Fried & Holyoak, 1984). There was no overall advantage in performance when the underlying distribution was unimodal but in both experiments we found an effect of the shape of the underlying distribution in the production format (the effect was marginally significant in the identification format in Experiment 2) with better performance when the distribution is unimodal as opposed to bimodal. The results also indicated that participants in the bimodal condition underestimated the target variable's variance to a larger extent than did those in the unimodal condition. These results taken together suggest that people might have a general inclination to view variables as unimodally distributed, regardless of the actual distribution shape. Whereas the results are expected if people utilise small samples to make judgements about distribution properties (Lindskog et al., 2013), they could also be accounted for if people have strong a priori assumptions of unimodality. Whereas the present study was not designed to distinguish between the two possibilities, it is an interesting question for future research to investigate if strong a priori assumptions exist or if the effect is due to post hoc sampling from memory.

A more general notion of a priori assumptions would be that people incorporate prior knowledge, not necessarily related to unimodality, into the task which may influence both their interpretation and representation of the data and its distribution shape. Thus, the way people represent and judge distribution shape could be dependent on situational factors. For example, the participants in the present study might have entered the task with the prior beliefs that most companies have low revenues and very few have high revenues. In turn, this belief, or knowledge, might have influenced their judgements of distribution shape. This could be

the case even though we explicitly instructed participants that the revenues were fictitious.

The idea of combining prior knowledge with experienced data could be summarised in a Bayesian model. Whereas some previous research has indicated that people use prior knowledge in similar tasks (Griffiths & Tenenbaum, 2006, 2011), other research has indicated that they do not (Lindskog et al., 2013). Because the present study does not elicit participants' prior knowledge before exposure to the data, it does not allow for the evaluation of a possible Bayesian process. Not knowing the prior would make the problem of evaluating a Bayesian process an ill-defined one. Further, the results of the present study indicate that eliciting a prior might actually change the representation of the distribution altogether. With respect to the influence of prior knowledge and possible Bayesian updating processes, it seems important for future research to address three interesting questions. First, to what extent does prior knowledge influence the representation and judgement of distribution shape. Second, are new data combined with prior knowledge by a Bayesian process or simply stored as new raw data points? Although this is still an empirical question, our results suggest the latter. Third, to what extent will the elicitation of prior knowledge before the exposure to new data influence the representation of distribution shape?

We replicated the finding from previous research that people are quite accurate at giving estimates of central tendency (Peterson & Beach 1967; Pollard, 1984) but that variance is often underestimated (Kareev et al., 2002). The latter was extended to hold for numerical (as compared to the more commonly used perceptual) variables and the degree of the bias was shown to be related to the distribution of the experienced variable. In fact, a tendency towards underestimation of variance could be indicative of an implicit unimodality assumption of the mind.

One limitation of the present study is that we did not include other elicitation formats than the production and the identification formats. It might be that these formats are not the most appropriate to elicit knowledge. For example, in the production task participants estimated the proportion of values in 10 different intervals. It might be that this procedure, requiring several individual decisions, tends to emphasise individual data points rather than the aggregate. A different sub-division of the range, for example with quartiles, might have allowed participants to put more emphasis on the shape of the distribution and thereby performed better. Although several methods used to elicit experts' subjective probability distributions include frequency (or probability) estimates for intervals (e.g. quartiles and percentiles) (Hora et al., 1992; Ludke et al., 1977; Winkler, 1967), little is known about how the representation of distribution shape is influenced by different sub-divisions of the range of the experienced variable. It will be an important question for future research to map out such limiting conditions on the knowledge of distribution shape.

## CONCLUSIONS

In several situations, people are expected to have a more or less accurate knowledge of how a numerical variable is distributed. Previous research has indicated that people often have quite an accurate knowledge of such statistical properties (e.g. Griffiths & Tenenbaum, 2006; Nisbett & Kunda, 1985). However, little research has explored how this knowledge is represented. In this paper, we show that the representation of how a numerical variable is distributed is contingent both on task demands and the properties of the experienced distribution. However, the default process adopted by people seems to be post hoc sampling from memory. The naive intuitive statistician thus seems to be lazy. Nevertheless, when instructed to do so, or when the task strongly invites it, people can induce abstract representations. However, it is questionable if people store posterior distributions or higher order assumptions of distribution shape. It seems rather to be the case that such knowledge is created online during sampling and judgement.

Original manuscript received December 2012
Revised manuscript received August 2013
Revised manuscript accepted August 2013
First published online October 2013

## REFERENCES

Aha, D. W. (Ed.). (1997). *Lazy learning*. Dordrecht: Kluwer Academic.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*(1), 149–178. doi:10.1146/annurev.psych.56.091103.070217

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*(1), 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

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*, 951–978. doi:10.1037/0033-295X.112.4.951

Dougherty, M. R., & 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 and Economic Statistics*, *27*, 30–41. doi:10.1198/jbes.2009.0003

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

Fiedler, K., & Juslin, P. (2006). *Information sampling and adaptive cognition*. New York, NY: Cambridge University Press.

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

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. doi:10.2307/749671

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

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*, 444–457. doi:10.1177/0272989X10373805

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

Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, NY: Cambridge 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

Greene, R. L. (1986). Effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 489–495. doi:10.1037/0278-7393.12.4.489

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

Griffiths, T. L., & Tenenbaum, J. B. (2011). Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*, *140*, 725–743. doi:10.1037/a0024899

Hansson, P., Juslin, P., & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1027–1042. doi:10.1037/a0012638

Hendrick, C., & Constantini, A. F. (1970). Number averaging behavior – Primacy effect. *Psychonomic Science*, *19*, 121–122.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534–539. doi:10.1111/j.0956-7976.2004.00715.x

Hora, S. C., Hora, J. A., & Dodd, N. G. (1992). Assessment of probability-distributions for continuous random-variables – A comparison of the bisection and fixed value methods. *Organizational Behavior and Human Decision Processes*, *51*, 133–155. doi:10.1016/0749-5978(92)90008-U

Jako, R. A., & Murphy, K. R. (1990). Distributional ratings, judgment decomposition, and their impact on interrater agreement and rating accuracy. *Journal of Applied Psychology*, *75*, 500–505. doi:10.1037/0021-9010.75.5.500

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., & 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 naive intuitive statistician: A naive 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*(2), 136–153. doi:10.1037/0033-295X.93.2.136

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. doi:10.2307/1914185

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511809477

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., Arnon, 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

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. (2013). Naïve point estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 782–800. doi:10.1037/a0029670

Linville, P. W., Fischer, G. W., & Salovey, P. (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*(2), 165–188. doi:10.1037/0022-3514.57.2.165

Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education & Behavior*, *36*, 1065–1081. doi:10.1177/1090198109341533

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. doi:10.1177/0272989X0102100105

Ludke, R. L., Stauss, F. F., & Gustafson, D. H. (1977). Comparison of five methods for estimating subjective-probability distributions. *Organizational Behavior and Human Decision Processes*, *19*, 162–179. doi:10.1016/0030-5073(77)90060-5

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

Mozer, M., 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

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339–363. doi:10.1037/0033-295X.90.4.339

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. (2006). *Bayesian rationality*. Oxford: Oxford University Press.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). *Uncertain judgements: Eliciting expert probabilities*. Chichester: Wiley.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*, 407–413. doi:10.1111/j.1467-9280.2006.01720.x

Peterson, C.R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46. doi:10.1037/h0024722

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), 1–18. doi:10.1016/0030-5073(76)90050-7

Pollard, P. (1984). Intuitive judgments of proportions, means, and variances: A review. *Current Psychology*, *3*(1), 5–18. doi:10.1007/BF02686528

Reyna, F. V., Nelson, L. W., Han, K. P., & Dieckmann, F. N. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*, 943–973. doi:10.1037/a0017327

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, *28*(1), 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), 1–26. doi:10.1016/j.cogpsych.2005.10.003

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. doi:10.1126/science.1192788

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty – Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124

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

Vul, W., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 66–72). Presented at the 31st annual conference of the cognitive science society.

Winkler, R. L. (1967). Assessment of prior distributions in bayesian analysis. *Journal of the American Statistical Association*, *62*(319), 776–800. doi:10.1080/01621459.1967.10500894

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic enconding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *8*(2), 106–116. doi:10.1037/0278-7393.8.2.106

## APPENDIX

In this appendix, we outline the details of the procedure used to derive the standardising constant ($MAE_R$) in the MR measure. The performance of each participant is standardised against $MAE_R$ because the two tasks are not equally difficult under random performance. That is, a naive participant choosing one of the graphs at random and distributing 60 objects over 10 intervals at random would not receive the same $MAE_S$ for both tasks. By standardising against $MAE_R$, we thereby make the two tasks comparable.

### Identification task

In the identification task, we considered a random judgement to be one in which the probability that that a participant would chose a specific graph is equal for all graphs. That is, all of the 11 graphs have equal probability of being chosen by a naive participant choosing at random. Each graph ($j$) is associated with a $MAE_j$ quantifying the deviance from the correct graph. $MAE_R$ was therefore calculated as

$$MAE_R = \frac{\sum_{j=1}^{11} MAE_j}{11}, \tag{A1}$$

that is, the mean MAE of the 11 presented graphs. In Experiment 1, $MAE_R$ was 4.9 and 4.8 in the unimodal and bimodal conditions, respectively. In Experiment 2, we constructed the graph so as to get an equal $MAE_R$ (3.1) in both conditions.

### Production task

In the production task, we considered a random judgement to be one where a participant would create a frequency distribution where all possible perturbations of sub-partitions of frequencies of 60 objects over 10 intervals would have the same probability of occurring. To estimate the expected performance of a participant giving random judgements, we created 10,000 random frequency distributions of 60 objects over 10 intervals. These distributions were created by random allocation of frequencies in the interval 0–60 over the 10 intervals with the constraint that these frequencies sum to $N$ (i.e. 60). $MAE_R$ was then calculated as

$$MAE_R = \frac{\sum_{k=1}^{10000} \frac{\sum_{i=1}^{10} |r_i - a_i|}{10}}{10000}, \tag{A2}$$

where $r_i$ is the rated frequency of interval $i$ for the random distribution and $a_i$ is the frequency of interval $i$ for the distributions presented to participants. In Experiment 1, $MAE_R$ was 8.1 and 8.5 in the unimodal and bimodal conditions, respectively. In Experiment 2, the presented distributions were constructed to give the same $MAE_R$ (8.3) in both distributions.