Learning of bimodally distributed quantities

Saiwing Yeung (saiwing.yeung@gmail.com)
Institute of Education, Beijing Institute of Technology, China

Andrew Whalen† (aczw@st-andrews.ac.uk)
School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH UK

Abstract

Previous research has shown that people are able to use distributional information about the environment to make inferences. However, how people learn these probability distributions is less well understood, especially for those that are not normal or unimodal. In this paper we focus on how people learn probability distributions that are bimodal. We examined on how the distance between the two peaks of a bimodal distribution and the numbers of observations influence how participants learn each distribution, using two types of stimuli with different degrees of perceptual noise. Overall, participants were able to learn the various distributions quickly and accurately. However, their performance is moderated by stimuli type—whether participants were learning a distribution over numbers (low noise) or over sizes of circles (high noise). This work suggests that although people are able to quickly learn a variety of distributions, many factors may influence their performance.

Keywords: probability distributions; learning; subjective belief; intuitive statistics.

Introduction

Previous research has found that individuals are sensitive to distributional information in the environment, and are able to use that knowledge to make good judgments and predictions, even when the distributions are non-normal. In many cases, this is a particularly challenging task because non-normal distributions are encountered less often than normal ones. Yet individuals still perform well, across different tasks that have different underlying distributions. For example, individuals were able to make approximately optimal predictions for the total box office of a movie based on its current take, a quantity roughly in a power-law distribution (Griffiths & Tenenbaum, 2006; Lewandowsky, Griffiths, & Kalish, 2009). The ease at which individuals are able to make these and distribution-based estimates (e.g., Sanborn, Griffiths, & Shiffrin, 2010; Maye, Werker, & Gerken, 2002) opens up the question of how individuals learn the probability distributions that underlay reality (Peterson & Beach, 1967; Posner & Keele, 1968).

Because of the importance of distribution learning in solving many tasks, it is perhaps not surprising that recent research has found that people are quite adept in learning a range of distributions. Goldstein and Rothschild (2014) taught participants various unimodal distributions using samples in the form of numbers, and then evaluated how well they learned by requiring them to graphically reconstruct the distribution. They found that individuals were able to build distributions with statistical properties that matched well with the original stimuli. Similar methods have been successfully applied in eliciting people's implicit statistical knowledge (Haran, Moore, & Morewedge, 2010), suggesting that individuals are capable at learning and retaining a range of unimodal distributions.

Although many real world distributions are unimodal, other types of distributions, like bimodal distributions, are also necessary for every day and scientific tasks because they help us discover latent categorical differences. For example, male gorillas are on average twice the size of their female counterparts. Therefore, learning the bimodality of one observable attribute (size) enable us to infer another (sex) that is harder to diagnose. Moreover, examining how people learn bimodal distributions provides a foundation on which we can study broader questions of how individuals learn complex, non-normal distributions, and how they use the knowledge to make inferences and predictions about the world.

Prior studies have used a variety of stimuli to examine learning of distribution, including numbers (Goldstein & Rothschild, 2014), sizes (Xu & Griffiths, 2010), sensorimotor noise (Körding & Wolpert, 2004). A brief review of these studies suggests that learning bimodal distributions can take a substantial amount of training. Moreover, how well and how long it takes individuals to learn a distribution may depend on the kind of stimuli being learned. However, the methodological differences between studies confound the question of stimuli choice and learning rates, making it hard to draw any conclusions based on these studies alone. Therefore, here we also aim to understand the influence of stimuli type in the context of distribution learning.

In this paper, we first review the existing literature on learning of bimodal distributions. We then present a series of experiments that examine how individuals learn bimodal distributions over numbers, and analyze how the learning outcome varies as a function of the overall shape of the distribution and the number of training examples. We then replicate these experiments in a slightly more naturalistic domain, by looking at how individuals learn distributions over different sized circles. This procedure allowed us to examine how factors external to the distribution, such as perceptual noise, interact with distributional properties to influence people's learning.

Learning Bimodal Distributions

A large body of previous work has found that individuals can learn and make inferences based on bimodal distributions. For example, infants and adults were found to be sensitive to the distributional properties of sounds in learning phonetic categories (Maye et al., 2002). Infants who were exposed to sounds sampled from a bimodal distribution were found to later distinguish between sounds from the endpoints of

† Corresponding author

the continuum, whereas infants who were exposed to sounds from a unimodal distribution did not. In the domain of social cognition, it has been found that people's estimates concerning frequency of social behaviors that exhibit bimodality partially track the actual behaviors (Nisbett & Kunda, 1985).

Even so, bimodal distributions can be hard to teach in the laboratory. For example, while Körding and Wolpert (2004) were able to train people to adapt to bimodally distributed sensory noise, their experiment involved over 1,000 trials. McKinley and Nosofsky (1995) found that stimuli with bimodally distributed category boundaries were very difficult for participants to learn. A similar result was also noted by Xu and Griffiths (2010).

However, these three studies assessed participants' ability to make inferences based on the distributions, and did not test their explicit knowledge of what the distributions were. This leaves open the possibility that individuals may have correctly learned the shape of the distribution, but may have had difficulty applying that knowledge to these inferences. In order to more directly examine the learning of these distributions, we instead adopt a new experimental framework in which distribution information were explicitly elicited in the form of samples (Goldstein & Rothschild, 2014).

Experiments 1 and 2: Learning about numbers

In Experiment 1, we examined how individual's learning of distributions over numbers is influenced by the distance between the two modes of the distribution, and in Experiment 2, how it is influenced by the number of observations.

Participants and procedure

In Experiment 1, we recruited 200 participants from Amazon Mechanical Turk and in Experiment 2, 163 participants. Participants were required to be 18 years or older, reside in the U.S., and have a lifetime acceptance rate on MTurk of at least 95%. Each participant was compensated US$0.30. Participants who took Experiment 1 were ineligible to take Experiment 2.

Both experiments were web-based, administered using Qualtrics, and with animations of numbers displayed using JavaScript. In the learning phase, participants were told:

> Imagine we have large bag filled with ping pong balls. A number between 1 and 10 is written on each ball. On the next screen we will show you the numbers written on 48 of the balls. Please say the number as it is shown. Try to remember as many of the values as you can.

They were then shown 48 numbers, one at a time, in a random order. Each number was drawn in white on a black background and was displayed for 1000 milliseconds. After the display of each number, the stimuli area was cleared for 500 milliseconds before the next number was shown. After all the numbers were displayed they would move on to the elicitation phase, here they were asked:

> Now imagine we place all the balls back into the bag. We will now draw 100 balls from the bag. How many balls of each value (from 1 to 10) do you think we will draw?

Table 1: Stimuli distribution of the experiments. The distributions used in Experiments 1 and 3 are shown on top, and those for Experiments 2 and 4 in the bottom. Each column represents, for each value of a distribution, the number of samples shown during the learning phase. For example, in the $\text{diff}_\mu = 0$ condition, values of 5 and 6 were each shown 16 times.

	1	2	3	4	5	6	7	8	9	10
$\text{diff}_\mu = 0$	0	0	1	7	16	16	7	1	0	0
$\text{diff}_\mu = 2$	0	1	3	9	11	11	9	3	1	0
$\text{diff}_\mu = 4$	1	3	8	8	4	4	8	8	3	1
$\text{diff}_\mu = 6$	4	8	8	3	1	1	3	8	8	4
$\text{diff}_\mu = 8$	12	8	3	1	0	0	1	3	8	12
$n = 12$	0	1	2	2	1	1	2	2	1	0
$n = 24$	0	2	4	4	2	2	4	4	2	0
$n = 48$	0	4	8	8	4	4	8	8	4	0
$n = 96$	0	8	16	16	8	8	16	16	8	0

Participants then gave their expected distribution over the numbers 1–10 using a constant-sum scale that ensured the frequency of all numbers summed to 100. The order in which the response options were presented was counter-balanced: responses were either listed from 1 to 10, or 10 to 1. Finally, participants completed an optional demographics survey.

In Experiment 1, participants were randomly assigned to one of five between-subjects conditions, each representing a different degree of bimodality (Table 1). These conditions were created by combining two normal distributions with a variance of 1 and varying distances between the two modes, indexed by the condition label diff_μ. It can be seen that in condition 0 and 2, participants were shown a unimodal distribution, whereas in conditions 4, 6, and 8, participants were shown a bimodal distribution. In the learning phase, participants were shown numbers in frequencies that correspond to their conditions.

Experiment 2 followed the design of Experiment 1, except that, instead of always seeing 48 samples, participants saw either 12, 24, 48, or 96 samples. Because we wanted to create conditions with similar degrees of bimodality, the frequency of stimuli in each condition were multiples of the frequencies in the 12 samples condition. Note that the distributions here did not exactly match any condition from Experiment 1, but the relative frequencies are similar to the $\text{diff}_\mu = 4$ condition.

Results of Experiment 1

The average age of the participants in Experiment 1 was 32.6, and 41.0% were female. The demographics of the remaining experiments are similar and will not be reported.

The aggregate responses of the five conditions are shown in Figure 1. The histograms represent the mean responses at each of the 10 possible values, with error bars showing ± 1 standard error. The blue lines represent the extrapolated *training distribution*—observed frequency of stimuli in the learning phase, scaled up linearly (from 48 to 100). Overall, the aggregated responses closely tracked the training distribution, suggesting that as a group, participants learned the challenging distributions very well.

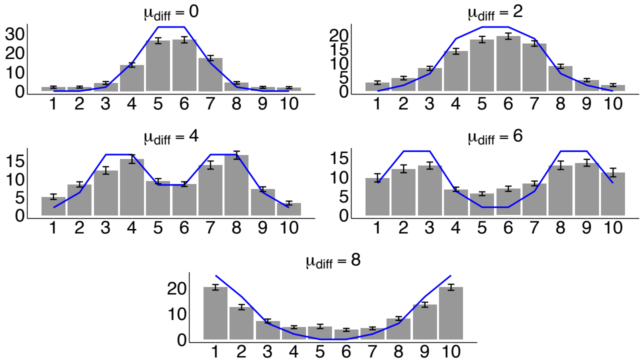

Figure 1: Aggregate results of Experiment 1. Each histogram represents a different condition, which is labeled on top. The histograms illustrate the mean human responses (with error bars of ± 1 *s.e.*) at each value. The blue lines correspond to the proportions of each value in the learning phase.

To quantitatively evaluate the responses we tracked two sets of statistics. We computed the Kolmogorov-Smirnov statistic D between each individual's responses and the training distribution to measure the raw accuracy of the responses. D gives a measure of the distance between two probability distributions, assigning a value of 0 to distributions that are identical, and 1 to distributions that are maximally different. D is computed as

$$D_n = \sup_x |F_n(x) - F(x)| \tag{1}$$

where $\sup_x$ is the supremum of the set of distances, F the cumulative distribution function, and F_n the empirical one.

While D gives insights to the fit of a distribution, it does little to tell us the degree of bimodality of a distribution. Therefore, we also computed the bimodality coefficient b (Freeman & Dale, 2013; *SAS/STAT 12.1 User's Guide*, 2012) and AIC_d (McLachlan & Peel, 2004). b is computed based on the assumption that a bimodal distribution should have very low kurtosis and/or high skewness:

$$b = \frac{g^2 + 1}{k + \frac{3(n-1)^2}{(n-2)(n-3)}} \tag{2}$$

where n is the number of samples, g the sample skewness and k the sample excess kurtosis. It ranges from 0 to 1, with values above 5/9 suggest bimodality. AIC_d is computed based on the difference in Akaike information criterion (AIC) of a model assuming a single normal distribution (unimodal) and a model assuming the combination of two normal distributions (bimodal). AIC_d does not have theoretical bounds, but higher values indicate better fits with bimodal distributions.

Figure 2 displays the means and standard errors of each of D, b, and AIC_d in each condition. Results of Experiment 1 are displayed by the graphs on the left. Overall, the D statistics were fairly low and stable across conditions, suggesting that participants were able to learn the distribution well, and their ability to learn the distribution was not largely influenced by the bimodality of the distribution. To test whether $diff_\mu$ indeed influences these statistics, we performed three regression analyses on each of D, b, and AIC_d, with $diff_\mu$ as the independent

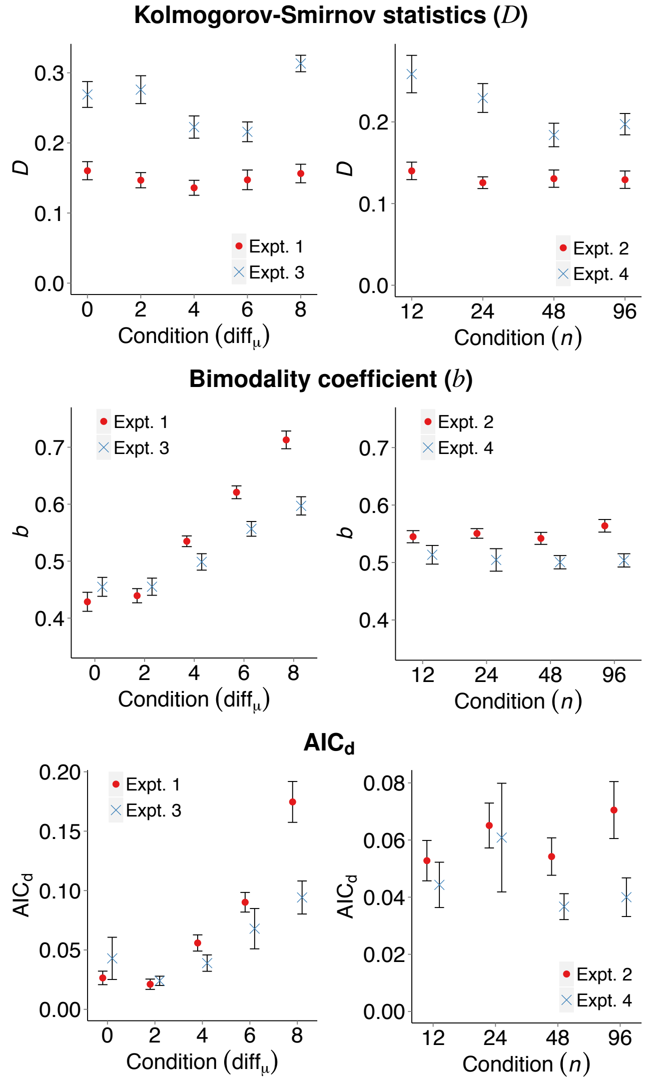

Figure 2: Comparison of the Kolmogorov-Smirnov statistics (D; top row), bimodality coefficient (b; middle row), and AIC_d (bottom row) between results of Experiments 1 and 3 (left), and of Experiments 2 and 4 (right). Error bars indicate ± 1 *s.e.*

variable. A regression showed that $diff_\mu$ did not influence D ($F(1,198) = 0.03$, $p = 0.86$), suggesting that individuals were able to learn the distributions equally well no matter the degree of bimodality.[1]

We next turn to b and AIC_d to evaluate the degree of bimodality at each $diff_\mu$. As expected, both b and AIC_d increase along with the distance between the two modes. A regression analysis showed that $diff_\mu$ indeed explained a significant proportion of the variance in the b and AIC_d scores (b: $F(1,198) = 300.35$, $p < 0.001$, $R^2 = .60$; AIC_d: $F(1,198) = 128.49$, $p < 0.001$, $R^2 = .39$), and suggested that bimodality in the stimuli influenced bimodality in the responses.

The increase in these values also closely tracked the values of the b and AIC_d for the training distribution (b^t and AIC_d^t,

[1]We used parametric statistical tests even though the data were not necessarily normally distributed. However, non-parametric tests gave similar results.

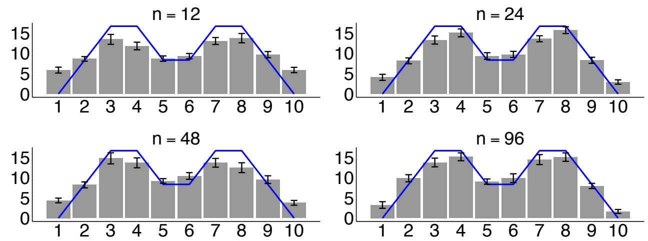

Figure 3: Aggregate results of Experiment 2, split according to conditions (shown on top). Please refer to Figure 1 for legend.

respectively). For example, in the $\text{diff}_\mu = 0$ condition, the b of the (unimodal) target distribution was 0.36, while the mean of participants' distribution was 0.43; whereas in the $\text{diff}_\mu = 8$ condition, the b of the target distribution was 0.81, compared to participants' response at 0.71. This suggests that although the participants were sensitive to the bimodality in the stimuli, at an individual level participant's responses were more noisy. However, the correlations between b and b^t ($r(198) = 0.79$, $p < 0.001$), and between AIC_d and AIC_d^t ($r(198) = 0.68$, $p < 0.001$) were both significant, suggesting higher bimodality in the stimuli corresponds to stronger inference of bimodality.

Overall the results suggest that participants were surprisingly good at inferring bimodality when the observations were indeed so, even with an observation of only 48 samples.

Results of Experiment 2

In Experiment 2 we varied the number of observations and investigate how amount of training influences learning of distributions and inference of bimodality. We first visually inspected the aggregate results of Experiments 2 (Figure 3), and found that participants' responses track quite well with the observations across all conditions. The D, b, and AIC_d statistics for Experiment 2 are shown in the right column of Figure 2. Overall, participant accuracy on this task was high, with a relatively low D. We used a regression analysis to test whether the number of observations influenced the various statistics of learning and bimodality, using the logged number of observations as independent variable. We found that the number of observations did not influence D ($F(1,161) = 0.41$, $p = 0.52$). Similarly, there was no significant effect of number of observations on neither b nor AIC_d (ps > 0.05).[2] This shows that when learning numbers, participants' accuracy or inference of bimodality were not influenced by the number of samples shown, at least within the range of samples used here.

Discussion

We carried out two experiments to examine how the distances between the two modes and amount of training influence learning of bimodal distributions using numbers as stimuli. We found that the distance between the peaks of distributions influenced people's inference of bimodality. While these results might be expected, other findings were surprising. Participants appeared to be able to learn bimodal distributions based on a fairly small number of observations, even when the difference in the modes was small (e.g., $\text{diff}_\mu = 4$ condition in Experiment 1), or when receiving a limited number of observations (e.g., $n = 12$ condition in Experiment 2).

The results of Experiments 1 and 2 stand in contrast to previous experimental work on distribution learning in which participants needed dozens or hundreds of trials to learn the correct underlying distribution. One reason for this difference may be the types of stimuli that participants were presented with. For example, Körding and Wolpert (2004) found that participants could learn bimodally distributed uncertainties in a sensory feedback task. However, participants received 1,000 training trials and it is not clear whether the same performance could be obtained with less training. Similarly, in Xu and Griffiths (2010), participants were presented with graphical fish with widths sampled from a 2-mixture normal distribution, and they had to correct learn and categorize fish based on the width. Results showed that participants needed on average more than 60 samples to categorize the fish correctly.

In both Körding and Wolpert (2004) and Xu and Griffiths (2010), the task was a perceptual one where participants did not have access to external reference (i.e., estimated displacement and size of fish). This may lead to participants making noisy judgments of how far the image was displaced, or how large a fish was, making it harder to learn the underlying distribution. In contrast, Experiments 1 and 2 relied on the perception of numbers which are more familiar, easily distinguishable, and memorable, making it easier for the participant to perceive, encode, and recall the sample values. Furthermore, prior research has shown that continuous quantities are more noisily represented than discrete quantities (Feigenson, Dehaene, & Spelke, 2004), a fact that perhaps contributed to the difference in performance.

Experiments 3 and 4: Learning about circles

To explore how perceptual noise inherent to stimuli might influence the distribution learning task, in Experiments 3 and 4 we replicated Experiments 1 and 2 using a more perceptually challenging type of stimuli: circles of varying sizes. The results of these experiments should shed light on the role that stimuli type play in distribution learning.

Participants and procedures

Participants were recruited from Amazon Mechanical Turk, and were randomly assigned to a condition. The conditions and distributions in Experiments 3 and 4 were identical to those of Experiments 1 and 2, respectively (Table 1). We recruited 204 participants for Experiment 3 and 162 participants for Experiment 4. Participants who had taken a previous experiment were ineligible to take Experiments 3 or 4. Each participant was compensated $0.30 for their time.

Except for the following changes, Experiments 3 and 4 mimicked Experiments 1 and 2. Participants were told "Imagine we have a large bag filled with black ping pong balls of different sizes. On the next screen we will show you 48 of them. Please try to remember the sizes of the balls as they

[2]The results of the analyses here and in Experiment 4 were similar regardless of whether the number of observation is logged.

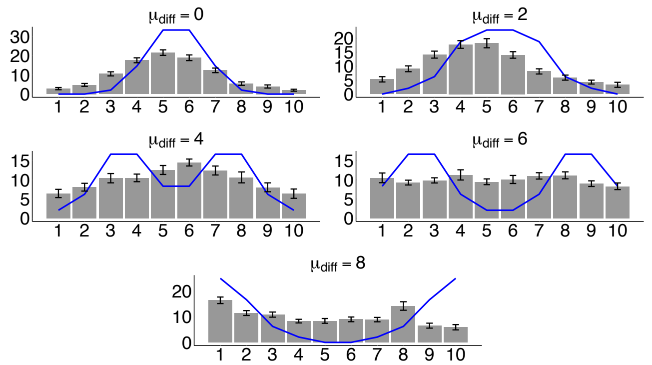

Figure 4: Aggregate results of Experiment 3.

are shown." The stimuli were solid black circles on a white background. The circles were of 10 different sizes, varying between 10 and 55 pixels wide in increments of 5 pixels. Circles of these 10 possible sizes are distributed the same way as numbers 1–10 in Experiments 1 and 2. Each circle was shown for 1000ms with a 500ms delay between presentations. There were additional, minor textual changes to address the fact that the stimuli were circles instead of numbers.

Results of Experiment 3

The mean responses for Experiment 3 are shown graphically in Figure 4. These means were not as close to the training distributions as in Experiment 1. This may be mainly due to perceptual noise and biases in reconstructing the distributions, resulting in variation in overall mean and placement of the modes of each participant's response. We see some hint of this in the $\text{diff}_\mu = 8$ condition where individuals place very little weight on the largest circles, but a large amount of weight on the third to largest circle. As a result, the distributions of these aggregate responses were quite different from the training distributions, especially in contrast to the results from the numbers experiments.

We next analyzed the distribution fit and bimodality statistics (Figure 2). Similar to Experiment 1, there was a significant increasing trend in b ($F(1,202) = 64.30$, $p < 0.001$, $R^2 = .24$) and AIC_d ($F(1,202) = 9.95$, $p < 0.01$, $R^2 = .05$), but not in D ($F(1,202) = 0.10$, $p = 0.75$), suggesting that participants reproduced increasingly bimodal distributions as diff_μ increased, but their learning performance was similar across conditions.

The correlations between b and b^t ($r(202) = 0.51$, $p < 0.001$), and between AIC_d and AIC_d^t ($r(202) = 0.25$, $p < 0.001$) were both significant. This demonstrates that, similar to Experiment 1, there was a close match between the stimuli and the learned distribution in terms of bimodality.

In order to assess the influence of perceptual noise on participants' inferences, we compared the distribution fit and bimodality statistics between Experiments 1 and 3. We first used a series of t-tests to compare D in corresponding conditions between the two experiments, e.g., $\text{diff}_\mu = 0$ conditions in both experiments. Results show that D in Experiment 3 is significantly higher in all five comparison (ps < 0.05), demonstrating that distributions presented as sizes of circles were harder for participants to learn.

As both b and AIC_d had an increasing trend, we used both stimuli type and diff_μ as independent variables, with b or AIC_d as the dependent variable in a regression analysis. Results showed that, for b, there was a significant effect of stimuli type and diff_μ (ps < 0.05). More interestingly, there was an interaction effect ($F(4,400) = 33.5$, $p < 0.001$), showing that the increasing trend is bigger when numbers were used as stimuli. Somewhat similar results were obtained in the regression using AIC_d, where the interaction effect was also significant ($F(4,400) = 17.5$, $p < 0.001$). Taken together, these results suggest that diff_μ had a bigger effect on b and AIC_d (steeper slope) when participants learned the distributions of numbers, compared to when they learned sizes of circles.

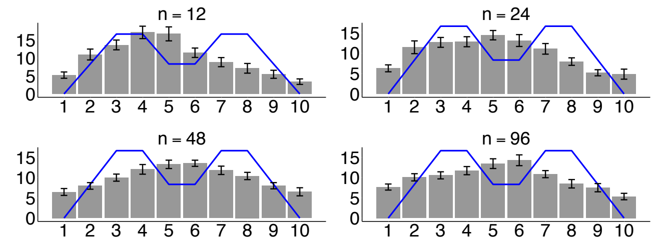

Figure 5: Aggregate results of Experiment 4.

Results of Experiment 4

As in Experiment 3, we found that the aggregate responses in Experiment 4 appeared to be substantially less bimodal than the corresponding numbers experiment (Figure 5). We next analyzed how well individual participants learned the distributions. We found that D decreased as (logged) number of observations increased ($F(1,160) = 8.48$, $p = 0.004$, $R^2 = .05$), showing that, overall, participants were learning the distributions better as they see more samples. In contrast, D between conditions in Experiment 2 were not significantly different. Bimodality coefficient b and AIC_d were not significantly different between conditions (b: $F(1,160) = 0.23$, $p = 0.63$; AIC_d: $F(1,160) = 0.54$, $p = 0.46$), showing that, as in Experiment 2, the distributions that individuals produced were not more bimodal with more observed samples.

We also compared the results between Experiments 2 and 4, by running a series of regression analyses using stimuli type and (logged) number of observations as independent variables, and D, b, or AIC_d as dependent variables. We focused on the interaction of the independent variables, because this test can assess whether stimuli type moderates the learning of bimodality with different amount of observations. We found that there was a significant interaction for D ($F(3,321) = 5.15$, $p = 0.02$), demonstrating that higher number of observations helped learning the distributions better when circles are being learned. In contrast, for both b and AIC_d, there was no interaction (both $p > 0.05$), suggesting that there was no difference between the two stimuli type in terms of whether more observations would lead to higher inference of bimodality.

This result suggests that in the numbers experiments, participants were able to learn the distribution so quickly that amount of observations did not influence learning. However, learning the distribution of sizes was more difficult that additional observations helped.

Discussion

We carried out two experiments to investigate how people learn distributions over the sizes of circles. Although in general, individuals were able to learn that these distributions were bimodal, there were substantial variations in where they located the modes of the distributions. Overall performance on a number of measures in Experiments 3 and 4 was lower than those in Experiments 1 and 2. Moreover, we found an interesting interaction. When the participants were learning numbers, their accuracy in learning the distributions remained stable for each number of training examples, but when they were learning sizes of circles their accuracy increased as they had access to more training examples. However, inference of bimodality with neither stimuli type increased with number of training. This suggests that amount of training might not be a major factor with respect to learning bimodality, at least within the range of training considered here.

General Discussion

In this paper we reported four experiments in which participants learned bimodal distributions. We found that when participants learned distributions over numbers, they were able to accurately learn bimodal distributions based on a small number of samples, including distributions with a relatively small distance between the two modes. However, when participants learned distributions of perceptually more challenging stimuli, sizes of circles, overall performance decreased, and participants were less likely to infer bimodality compared to when learning numbers. These results demonstrate that individuals are likely to be able to learn a wide range of distributions in a variety of settings, although not all distributions are learned equally—the difficulty of learning a distribution depends on what the distribution is of.

The current findings have potential implications for several different aspects of cognitive and decision research. First, our findings showed that people were able to learn bimodal distributions relatively quickly, even with few observations, especially with stimuli of lower perceptual noise. These results revealed a deeper correspondence between probabilities in the mind and those in the world than what was previously suggested, and complemented recent research on Bayesian decision theory in which people were provided with explicit priors (Acerbi, Vijayakumar, & Wolpert, 2014).

Second, one of the key parameters concerning learning of distributions seems to be stimuli type. Using experiments with the exact same parameterizations, we directly contrasted how people learned distributions based on stimuli that were different only in modes of presentation. Compared to learning distributions over numbers, learning distributions over sizes was far more challenging, albeit still possible, especially with more training. This difference may be driven by the relative difficulty in perceiving or encoding size of circles compared to numbers.

Third, in Experiments 2 and 4 people were sensitive to bimodality even in very small sample sizes. Results here are in agreement with prior works in which people were found to believe in the representativeness of small samples (Tversky & Kahneman, 1971), and show that the finding applies even in the learning of complicated continuous distributions.

Overall we found that individuals were able to quickly and accurately learn bimodal distributions, although their accuracy and inference depended on the interaction of many factors, including types of stimuli presented, distributional properties of the stimuli, and amount of evidence. While these results speak to people's ability to acquire statistical information from their environment, it also highlights the difficulty of the task. Understanding how we learn and use these statistical information will help explain how humans can carry out complex everyday tasks so efficiently.

Acknowledgments. This research was supported by John Templeton Foundation Grant #40128.

References

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLOS Computational Biology*, *10*(6), e1003661.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83–97.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*(1), 1–14.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773.

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*(7), 467–476.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cognitive Science*, *33*(6), 969–998.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 128–148.

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. Hoboken, NJ: John Wiley & Sons.

Nisbett, R. E., & Kunda, Z. (1985). Perception of social distributions. *Journal of Personality and Social Psychology*, *48*(2), 297–311.

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3-1), 353–363.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with markov chain monte carlo. *Cognitive Psychology*, *60*(2), 63–106.

SAS/STAT 12.1 User's Guide. (2012). Cary, NC: SAS Institute.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology*, *60*(2), 107–126.