The Rational Basis of Representativeness

Joshua B. Tenenbaum & Thomas L. Griffiths
Department of Psychology
Stanford University
Stanford, CA 94305-2130 USA
{jbt,gruffydd}@psych.stanford.edu

Abstract

Representativeness is a central explanatory construct in cognitive science but suffers from the lack of a principled theoretical account. Here we present a formal definition of one sense of representativeness – what it means to be a good example of a process or category in the context of Bayesian inference. This analysis clarifies the relation between representativeness as an intuitive statistical heuristic and normative principles of inductive inference. It also leads to strong quantitative predictions about people's judgments, which compare favorably to alternative accounts based on likelihood or similarity when evaluated on data from two experiments.

Why do people think that Linda, the politically active, single, outspoken, and very bright 31-year-old, is more likely to be a feminist bankteller than to be a bankteller, even though this is logically impossible? Why do we think that the sequence HHTHT is more likely than the sequence HHHHH to be produced by flipping a fair coin, even though both are equally likely? The standard answer in cognitive psychology (Kahneman & Tversky, 1972) is that our brains are designed to judge "representativeness", not probability: Linda is more representative of feminist banktellers than of banktellers, and HHTHT is more representative of flipping a fair coin than is HHHHH, despite anything that probability theory tells us.

Not only errors in probabilistic reasoning, but numerous other phenomena of categorization, comparison, and inference have been attributed to the influence of representativeness (or prototypicality or "goodness of example"; Mervis & Rosch, 1981; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975). However, a principled account of representativeness has not been easy to come by. Its leading proponents (Kahneman & Tversky, 1996; Mervis & Rosch, 1981) have asserted that representativeness should be defined only operationally in terms of people's judgments; an a priori, analytic definition need not be given. Critics have countered that this concept is too vague to serve as an explanation of intuitive probability judgment (Gigerenzer, 1996).

This paper presents a framework for constructing rational models of representativeness, based on a Bayesian analysis of what makes an observation a good example of a category or process. The goal is to identify precisely one sense of representativeness and show that it has a rational basis in normative principles of inductive reasoning. We will first point out some shortcomings of previous accounts based on likelihood or similarity, and show how a Bayesian approach can overcome those problems. We will then compare the quantitative predictions of Bayesian, likelihood, and similarity models on two sets of representativeness judgments.

Previous approaches

Likelihood. In trying to relate intuitions about representativeness to rational statistical inferences, a natural starting point is the concept of likelihood. Let d denote some observed data, such as a sequence of coin tosses, and h denote some hypothesis about the source of d, such as flipping a fair coin. The probability of observing d given that h is true, $P(d|h)$, is called a likelihood. Let $R(d,h)$ denote representativeness – how representative the observation d is of the generative process in h.

Gigerenzer & Hoffrage (1995) have proposed that representativeness, to the extent that it can be defined rigorously, is equivalent to likelihood: $R(d,h) = P(d|h)$. This proposal is appealing in that, other factors aside, the more frequently h leads to observing d, the more representative d should be of h. It is also consistent with some classic errors in probability judgment, such as the conjunction fallacy: a person is almost certainly more likely to match Linda's description given that she is a bankteller and a feminist than given only that she is a bankteller.

While likelihood and representativeness seem related, however, they are not equivalent. Two observations with equal likelihood may differ in representativeness. Knowing that HHHHH and HHTHT are equally likely to be produced by a fair coin does not change our judgment that the latter is the more representative outcome. Tversky & Kahneman (1983) provide several examples of cases in which a more representative outcome is actually less likely. Any sequence of fair coin flips, such as THHHTHT, is less likely than one of its subseqences, such as H or HHH, but may easily be more representative. More colorfully, "being divorced four times" is more representative of Hollywood actresses than is "voting democratic", but the former is certainly less likely.

Figure 1 illustrates a simple version of the dissociation between representativeness and likelihood. Each panel shows a sample of three points from a Gaussian distribution. With independent sampling, the total likelihood of a sample equals the product of the likelihoods for

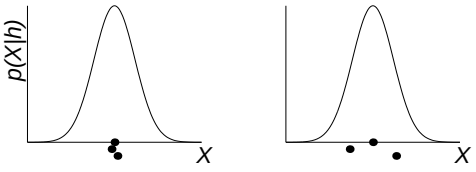

Figure 1: Given a normal distribution, the left sample has greater likelihood but the right is more representative.

each item in the sample. Thus the left sample has much greater likelihood, because each point is much closer to the peak of the distribution than in the right sample. Yet the more spread-out sample on the right seems more representative. We tested this intuition in a survey of 138 Stanford undergraduates. They were first shown a normally distributed set of thirty "widgets" produced by a factory. The widgets were simple drawings resembling nuts or bolts, varying only in their sizes. They were then shown three different samples, each with three widgets, and asked to rate on a scale of 1-10 how representative each sample was of the widgets produced by this factory. Each sample contained a point at the mean of the original distribution, and points at $z = \pm 2.85$ ("broad sample"), $z = \pm 1$ ("intermediate sample"), or $z = \pm 0.05$ ("narrow sample"). The intermediate sample, with a standard deviation similar to the population, received a significantly higher rating than did the much more likely narrow sample (7.1 vs. 5.2, $p < .05$). The broad sample, with lowest likelihood of all, also received a lower rating (6.9) than the intermediate sample, but not by a significant margin.

We also tested whether intermediate-range samples are more representative for natural categories, using as stimuli black-and-white pictures of birds. In a design parallel to the widget study, 135 different Stanford undergraduates saw three samples of birds, each containing three members, and rated how representative they were of birds in general. The samples consisted of either three robins ("narrow"); a robin, an eagle, and a seagull ("intermediate"); or a robin, an ostrich, and a penguin ("broad"). Although the robins were individually rated as more representative than the other birds (by a separate group of 100 subjects), the set of three robins was considered the least representative of the three samples. As with the widgets, the intermediate sample was rated more representative (6.3) than either the narrow (5.1) or broad (5.3) samples ($p < .05$ for both differences).

For natural categories as well as for the artificial widgets, a set of representative examples turns out not to be the most representative set of examples. Sample likelihood, because it is merely the product of each example's individual likelihood, cannot capture this phenomenon. At best, then, likelihood may be only one factor contributing to the computation of representativeness.

Similarity. Most attempts to explicate the mechanisms of representativeness, including that of Kahneman & Tversky (1972), rely not on likelihood but on some sense of similarity. That is, an observation d is representative of a category or process h to the extent that it is similar to the set of observations h typically generates.

Similarity seems to avoid some of the problems that likelihood encounters. `HHTHT` may be more representative of a fair coin than `HHHHH` because it is more similar on average to other coin flip sequences, based on such features as the number of heads or the number of alternations. Likewise, someone who has been divorced four times may be more similar to the prototypical Hollywood actress than someone who votes democratic, if marital status is weighted more heavily than political affiliation in computing similarity to Hollywood actresses.

However, the explanatory power of a similarity-based account hinges on being able to specify what makes two stimuli more or less similar, what the relevant features are and how are they weighted. Similarity unconstrained is liable to lead to circular explanations: having had multiple divorces is more representative of Hollywood actresses because marital status is more highly weighted in computing similarity to Hollywood actresses, but why is marital status so highly weighted, if not because having multiple divorces is so typical of Hollywood actresses?

Equating representativeness with similarity also runs into a problem when evaluating the representativeness of a set of objects, as in Figure 1. Similarity is usually defined as a relation between pairs of stimuli, but here we require a judgment of similarity between two sets of stimuli, the sample and the population. It is not immediately obvious how best to extend similarity from a pairwise to a setwise measure. The individual elements of the left sample are certainly more similar to the average member of the population than are the elements of the right sample. The left sample also comes closer to minimizing the average distance between elements of the population and elements of the sample. If similarity between sets is defined according to one of these measures, it will fail to match up with representativeness.

Finally, and most problematic for our purposes here, a definition in terms of similarity fails to elucidate the rational basis of representativeness, and thus brings us no closer to explaining when and why representativeness leads to reasonable statistical inferences. Hence we seem to be left with two less-than-perfect options for defining representativeness: the simple, rational, but clearly insufficient concept of likelihood, or the more flexible but notoriously slippery concept of similarity.

A Bayesian analysis

In this section we present a Bayesian analysis of representativeness that addresses some of the shortcomings of the likelihood and similarity proposals. As with likelihood, Bayesian representativeness takes the form of a simple probabilistic quantity, which in fact includes likelihood as one component. But like the similarity approach, it can account for dissociations of representativeness and likelihood, when a less probable feature of the stimuli is also more diagnostic of the process or category in question. Moreover, it applies just as well to evaluat-

ing the representativenes of a set of examples (e.g. Figure 1) as it does to individual examples.

Our notion of a "good example" is defined in the context of a Bayesian inductive inference task. As above, let d denote some observed data, and let $\mathcal{H} = \{h_1, \ldots h_n\}$ denote a set of n hypotheses (assumed to be mutually exclusive and exhaustive) that might explain the observed data. For each h_i, we require both the likelihood $P(d|h_i)$ and a prior probability, $P(h_i)$, which expresses the degree of belief in h_i before d is observed. Let $\bar{h}_i = \{h_j \in \mathcal{H} : j \neq i\}$ denote the negation of hypothesis h_i, the assertion that some hypothesis other than h_i is the true source of d. Then we define our measure of representativeness $R(d,h_i)$ to be the logarithm of the likelihood ratio

$$L(d|h_i) = \frac{P(d|h_i)}{P(d|\bar{h}_i)}. \quad (1)$$

This definition is motivated by Bayes' rule, which prescribes a degree of belief in hypothesis h_i after observing d given by the posterior probability

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)}. \quad (2)$$

Defining the posterior odds $O(h_i|d) = P(h_i|d)/P(\bar{h}_i|d) = P(h_i|d)/(1-P(h_i|d))$, and the prior odds $O(h_i) = P(h_i)/(1-P(h_i))$, we can write Bayes' rule in the form:

$$\log O(h_i|d) = \log L(d|h_i) + \log O(h_i). \quad (3)$$

Equation 3 shows why the log likelihood ratio, $\log L(d|h_i)$, provides a natural measure of how good an example d is of h_i: it indicates the extent to which observing d increases or decreases the posterior odds of h_i relative to the prior odds. Researchers in statistics (Good, 1950), artificial intelligence (Pearl, 1988), and philosophy of science (Fitelson, 2000) have previously considered $\log L(d|h_i)$ as the best measure for the weight of evidence that d provides for h_i, because it captures the unique contribution that d makes to our belief in h_i independently of all other knowledge that we have (reflected in $P(h_i)$).

To compute $R(d,h_i)$ in the presence of more than one alternative hypothesis, we express it in the form

$$R(d,h_i) = \log \frac{P(d|h_i)}{\sum_{h_j \in \mathcal{H}} P(d|h_j)P(h_j|\bar{h}_i)}. \quad (4)$$

$P(h_j|\bar{h}_i)$ is the prior probability of h_j given that h_i is not the true explanation of d: 0 when $i = j$ and $P(h_j)/(1-P(h_i))$ otherwise. Equation 4 shows that d is representative of h_i to the extent that its likelihood under h_i exceeds its average likelihood under alternative hypotheses.

To illustrate the analysis concretely, consider the simple case of two coinflip sequences, HHHHH and HHTHT. Unlike the likelihood model, we cannot compute how representative an observation is of a hypothesis without specifying the alternative hypotheses that an observer might consider. In the interests of simplicity, we consider just three relevant hypotheses about the origins of HHHHH and HHTHT: a fair coin (h_F), a two-headed coin (h_T), and a weighted coin (h_W) that comes up heads with probability 3/5. The likelihoods of the two sequences under these hypotheses are, for the fair coin, $P(\texttt{HHHHH}|h_F) = P(\texttt{HHTHT}|h_F) = (1/2)^5 = 0.03125$; for the two-headed coin, $P(\texttt{HHHHH}|h_T) = 1$ while $P(\texttt{HHTHT}|h_T) = 0$; and for the weighted coin, $P(\texttt{HHHHH}|h_W) = (3/5)^5 = 0.0778$ while $P(\texttt{HHTHT}|h_W) = (3/5)^3(2/5)^2 = 0.0346$. For concreteness, we choose specific prior probabilities for these hypotheses: $P(h_F) = 0.9$, $P(h_T) = 0.05$, and $P(h_W) = 0.05$. Substituting these numbers into Equation 4, we have $R(\texttt{HHHHH}, h_F) = \log \frac{0.03125}{1\times 0.05/0.1 + 0.0778\times 0.05/0.1} = -2.85$, while $R(\texttt{HHTHT}, h_F) = \log \frac{0.03125}{0\times 0.05/0.1 + 0.0346\times 0.05/0.1} = 0.59$. This result, that HHTHT is more representative of a fair coin than HHHHH, accords with intuition and holds regardless of the prior probabilities we assign to the three alternative hypotheses. In a later section, we go beyond a qualitative reconstruction of intuitions to test a quantitative model of representativeness judgments for sequences of coin flips.

The Bayesian approach also accounts for cases where a sample with lower likelihood appears more representative. For instance, $P(\texttt{HHTHTHTTH}|h_F)$ is strictly lower than either $P(\texttt{HHTHT}|h_F)$ or $P(\texttt{HHHHH}|h_F)$, but HHTHTHTTH is no less representative than HHTHT. The Bayesian account also offers an intuitively compelling definition of representativeness for a set of examples, such as the widgets in Figure 1. We demonstrate by computing the representativeness for a sample X from a Gaussian population h_1. Let $\{x_1, \ldots, x_N\}$ be the N examples in X, m be the mean of X, and $S = \sum_i (x_i - m)^2$ the sum-of-squares. Let h_1 have mean μ and variance σ^2. We take the hypothesis space $\mathcal{H}$ to include all possible Gaussian distributions in one dimension – each a conceivable alternate explanation for the sample X. Because $\mathcal{H}$ is an uncountably infinite set, the sum in the denominator of Equation 4 becomes an integral. Assuming an uninformative Jeffreys prior on μ, σ (Equation 3 of Minka, 1998), our expression for Bayesian representativeness in Equation 4 then reduces to

$$R(X,h_1) = N \log S - \frac{1}{\sigma^2}\left[N(m-\mu)^2 + S\right], \quad (5)$$

plus a term that depends only on N and σ^2.

Equation 5 is maximized when $m = \mu$ and $S/N = \sigma^2$, that is, when the mean and variance of the sample X match the mean and variance of the population h_1. This result is intuitive, and it accounts for why people preferred intermediate samples of widgets or birds over broad or narrow samples in the surveys described above: the $N \log S$ term penalizes narrower samples and the $-S/\sigma^2$ penalizes broader samples. Yet this result is also not particularly surprising. More interestingly, Equation 5 gives a general metric for scoring the representativeness of any sample from a Gaussian distribution, which we will test quantitatively against people's judgments in the following section.

Quantitative modeling

In this section, we present quantitative models of representative judgments for two kinds of stimuli: sequences of coin flips and sets of animals. For each data set, we compare the predictions of Bayesian, likelihood-based, and similarity-based models.

Coin flips

Methods. 278 Stanford undergraduates rated the representativeness of four different coin flip sequences for each of four hypothetical generative processes, under the cover story of helping a casino debug a new line of gambling machines. The sequences were d_1 = HHTHTTTH, d_2 = HTHTHTHT, d_3 = HHTHTHHH, and d_4 = HHHHHHHH. The generative processes were h_1 = "A fair coin", h_2 = "A coin that always alternates heads and tails", h_3 = "A coin that mostly comes up heads", and h_4 = "A coin that always comes up heads". The orders of both sequences and hypotheses were randomized across subjects. Representativeness judgments were made on a scale of 1-7.

Bayesian model. While people could construct an arbitrarily large hypothesis space for this task, we make the simplifying assumption that their hypothesis space can be approximated by just the four hypotheses that they are asked to make judgments about. We constructed simple probabilistic models for each hypothesis h_i to generate the necessary likelihoods $P(d_j|h_i)$. Priors for all hypotheses were assumed to be equal. To model h_1, "a fair coin", all likelihoods were set equal to their true values of $1/2^8$. To model h_3, "mostly heads", and h_4, "always heads", we used binomial distributions with $p = 0.85$ and $p = 0.99$, respectively. In some sense, these p values represent free parameters of the model, but their values are strongly constrained by the meaning of the words "mostly" and "always". Their exact values are not crucial to the model's performance, as long as "always" is taken to mean something like "almost but not quite always" (i.e. $p < 1.0$). To model h_4, "always alternates heads and tails", we used a binomial distribution over the seven possible state transitions in each sequence, again with "always" translated into probability as $p = 0.99$. All model predictions were then given by Equation 4.

Likelihood model. This model treats representativeness judgments simply as $P(d_j|h_i)$, as specified above.

Similarity model. We defined a simple similarity-based model in terms of two intuitively relevant features for comparing sequences: the number of heads in each sequence and the number of alternations in each sequence. Let α_j be the number of heads in sequence j, and β_j be the number of alternations. Then the similarity of sequences d_i and d_j is defined to be

$$sim(d_i, d_j) = \exp\left(-w_\alpha|\alpha_i - \alpha_j| - w_\beta|\beta_i - \beta_j|\right), \quad (6)$$

where w_α and w_β are the weights given to these two features. To compute similarity between a sequence and a generating hypothesis, we construct a prototype for each hypothesis based on the mean values of α and β over the whole distribution of sequences generated by that hypothesis. For example, for h_2, $\alpha = 4$ and $\beta = 7$; for h_3 (again assuming "mostly" means with probability 0.85), $\alpha \approx 6.8$ and $\beta \approx 1.8$. Lastly, we define the representativeness of sequence i for hypothesis j as $R(d_i, h_j) = sim(d_i, h_j)/\sum_k sim(d_i, h_k)$. The dimensional weights w_α and w_β are free parameters optimized to fit the data, giving $w_\alpha = 1$, $w_\beta = 0.4$.

Results. To compensate for nonlinear transformations that might affect the 1-7 rating scale used by subjects, the predictions of each model were first transformed according to a power function with a power γ chosen to optimize each model's fit, and then mapped onto the same interval spanned by the data. This gives both the likelihood model and the Bayesian model one free parameter plus two constrained parameters (corresponding to the meanings of "mostly" and "always"), while the similarity model has three free parameters (w_α, w_β, and γ) and the same two constrained parameters. All three models correlate highly with subjects' representativeness judgments, although the Bayesian model has a slight edge with $r = 0.94$, versus 0.87 for the likelihood model and 0.92 for the similarity model. Figure 2 presents an item-by-item analysis, showing that the Bayesian model captures virtually all of the salient patterns in the data.

Animals

Methods. We used data reported by Osherson, Smith et al. (1990; Tables 3 and 4) in a study of category-based induction. They asked one group of subjects to judge pairwise similarities for a set of 10 mammals, and a second group of subjects to judge the strengths of 45 arguments of the form $\{x_1$ has property P, x_2 has property P, x_3 has property P, therefore all mammals have property $P\}$, where x_1, x_2 and x_3 are three different kinds of mammals and P is a blank biological predicate. Such judgments of argument strength are not the same thing as judgments of representativeness, but for now we take them as a reasonable proxy for how representative the sample $X = \{x_1, x_2, x_3\}$ is of the set of all mammals.

Bayesian model. We assume that people's hypothesis space includes the category of all mammals (h_M), as well as an infinite number of alternative hypotheses. For simplicity, we model all hypotheses as Gaussian distributions in a two-dimensional feature space obtained from a multidimensional scaling (MDS) analysis of the similarity judgments in Osherson et al. (1990). This allows us to apply essentially the same analysis used in the previous section to compute the representativeness of a sample from a Gaussian distribution (Equation 5), and also parallels the original approach to modeling category-based induction of Rips (1975). The MDS space for animals is shown in Figure 3. The large gray oval indicates the one-standard-deviation contour line of h_M, which we take to be the best fitting Gaussian distribution for the set of all ten mammals. We assume the set $\mathcal{H}$ of alternative hypotheses includes all Gaussians in this two-dimensions

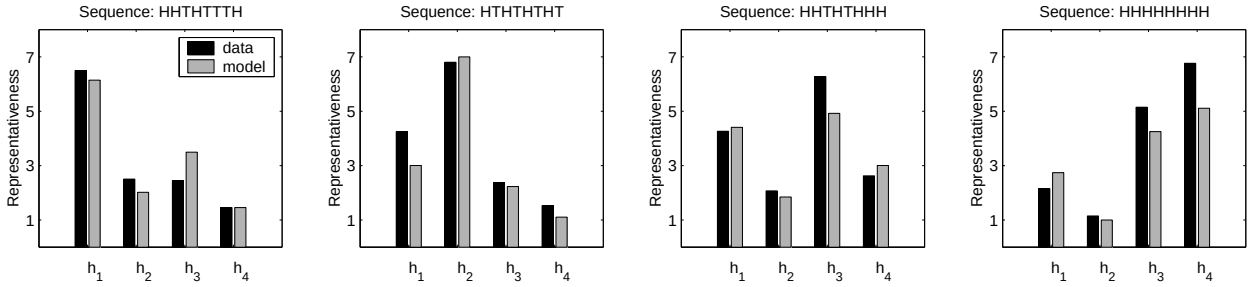

Figure 2: Representativeness judgments for coin flip sequences. Each panel shows subjects' mean judgments and the Bayesian model predictions for the representativeness of one sequence with respect to four different generating hypotheses: h_1 = "A fair coin", h_2 = "A coin that always alternates heads and tails", h_3 = "A coin that mostly comes up heads", and h_4 = "A coin that always comes up heads".

space, and we again use the uninformative Jeffreys' prior $P(h)$ (Minka, 1998; Equation 3). How representative a sample X (e.g. $\{\text{horse}, \text{cow}, \text{squirrel}\}$) is of all mammals can then be computed from a multidimensional version of Equation 5 (ignoring terms equal for all samples):

$$R(X, h_m) = N \log|\mathbf{S}| - N(\mathbf{m}-\mu)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{m}-\mu) - \text{trace}(\mathbf{S}\mathbf{V}^{-1}), \quad (7)$$

where m is the mean of X, $\mathbf{S} = \sum_i (\mathbf{x}_i - \mathbf{m})^{\mathrm{T}}(\mathbf{x}_i - \mathbf{m})$, $\mathbf{x}_i$ are the MDS coordinates of example i, N is the number of examples in X, and μ and $\mathbf{V}$ are the mean and covariance matrix of h_M (Minka, 1998). Equation 7 measures the representativeness of any sample X of N mammals in terms of the distance between the best fitting Gaussian for the sample (mean $\mathbf{m}$, covariance $\mathbf{S}$/N) and the best fitting Gaussian for the set of all mammals (mean μ, covariance $\mathbf{V}$). Figure 3 illustrates this graphically, by plotting one-standard-deviation contours for three samples that vary in how representative they are of the set of all mammals. Observe that the more representative the sample, the greater the overlap between its best-fitting Gaussian and the best-fitting Gaussian for the whole set.

Similarity-based models. Osherson et al. (1990) report pairwise similarity judgments for the animals, but to construct a similarity-based model of this representativeness task, we need to define a setwise measure of similarity between any sample of three animals and the set of all mammals. The similarity-coverage model proposed by Osherson et al. defines this quantity as the sum of each category instance's maximal similarity to the sample: $R(X, h_M) = \sum_j \max_i sim(i, j)$, where j ranges over all mammals and i ranges over just those in the sample X. A more traditional similarity-based model might replace the maximum with a sum: $R(X, h_M) = \sum_j \sum_i sim(i, j)$. Osherson et al. (1990) consider both max-similarity and sum-similarity models but favor the former as it is more consistent with their phenomena. However, there seems to be little a priori reason to prefer max-similarity, and indeed most similarity-based models of classification are closer to sum-similarity, so we consider both here.

Other models. We also compare the predictions of a simple likelihood model, which equates representativeness with $P(X|h_M)$, and Sloman's (1993) feature-based model. Heit (1998) also presented a Bayesian model of category-based induction tasks, but because his model depends heavily on the choice of priors, it does not make strong quantitative predictions that can be evaluated here.

Results. Figure 3 plots the argument strength judgments for 45 arguments versus the representativeness predictions of the probabilistic and similarity-based models. Both the Bayesian and max-similarity models predict the data reasonably well ($r = 0.80$ vs. $r = 0.88$), with no significant difference between them ($p > .2$). Neither of these models has any free numerical parameters. With one free parameter, the feature-based model performs slightly worse ($r = 0.71$). Interestingly, both the likelihood and sum-similarity models show a weak *negative* correlation with the data ($r = -.31$, $r = -.26$). This discrepancy directly embodies the insight of Figure 1: high likelihood can yield low representativeness when the sample is tightly clustered near the mean, as in the sample of $\{\text{horse}, \text{cow}, \text{rhino}\}$ (ellipse **C** in Figure 3). Sum-similarity performs as poorly as likelihood because it is essentially a nonparametric estimate of likelihood; likewise, max-similarity performs well because it correlates highly with Bayesian representativeness.

Discussion

Overall, the Bayesian models provide the most satisfying account of these two data sets. On the coinflip data, not only does Bayes obtain the highest correlation, but it does so with the minimal number of free parameters. On the animals data, Bayes obtains a correlation competitive with the best of the other models, max-similarity, even though it is based on less than half as much input data (20 MDS coordinates versus 45 raw similarity judgments) and may be hindered by information lost in the MDS preprocessing step. Most importantly, the Bayesian models are based on a rational analysis, which provides a single principled definition of representativeness applicable across the two quite different domains of coinflips and

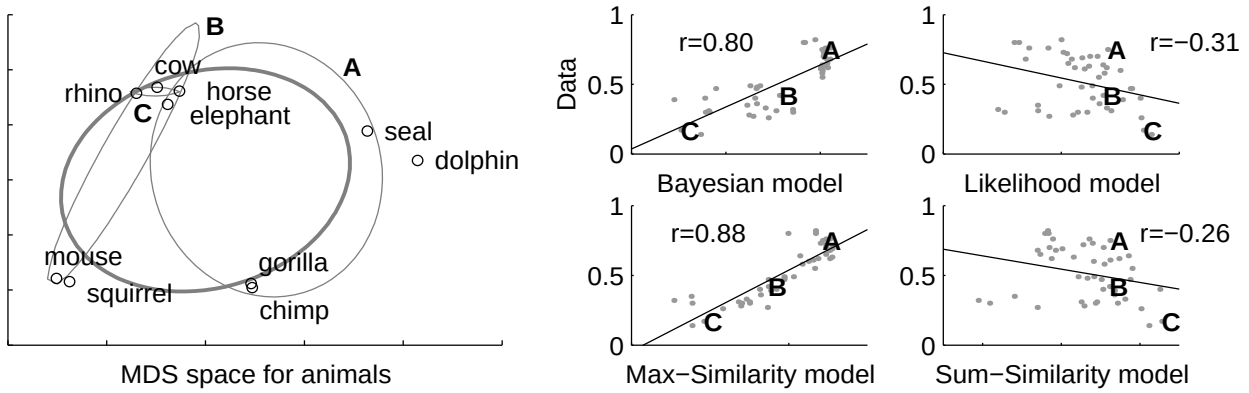

Figure 3: Modeling representativeness for sets of mammals. Ellipses in the MDS space of animals (left) mark one-standard-deviation contours for the set of all mammals (thick), a representative sample ($\{\text{horse}, \text{chimp}, \text{seal}\}$, **A**), a somewhat representative sample ($\{\text{horse}, \text{mouse}, \text{rhino}\}$, **B**), and a less representative sample ($\{\text{horse}, \text{cow}, \text{rhino}\}$, **C**). Scatter plots (right) compare strength judgments for 45 arguments with the predictions of four models (see text).

animals. In contrast, the similarity-based models have no rational grounding and take on very different forms in the two domains. They achieve high correlations, but only through the introduction of multiple free parameters, such as the feature weights on the coin flip data, or ad hoc assumptions, such as the choice of max-similarity over sum-similarity on the animal data. On the other hand, similarity-based models do have the advantage of requiring only simple computations. Thus both Bayesian and similarity-based models may have something to offer, but at different levels of analysis. Similarity may provide a reasonable way to describe the psychological mechanisms of representativeness, while a Bayesian analysis may provide the best explanation of why those mechanisms work the way they do: why different features of sequences are weighted as they are in the coinflip example, or why max-similarity provides a better model for inductive reasoning than does sum-similarity.

Conclusion

We have argued that representativeness is best understood as a Bayesian computation, rather than as a judgment of similarity or likelihood. Our analysis makes precise one core sense of representativeness – the extent to which something is a good example of a category or process – and exposes its underlying rational basis. Rational models have been successfully applied to a number of cognitive capacities (Shepard, 1987; Anderson, 1990; Oaksford & Chater, 1998) but not previously to analyzing representativeness, which is traditionally thought of as an alternative to normative probabilistic judgment. By clarifying the relation between our intuitive sense of representativeness and normative principles of statistical inference, our analysis may lead to a better understanding of those conditions under which human reasoning may actually be rational or close to rational, as well as those situations in which it truly deviates from a rational norm.

References

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

Fitelson, B. (2000). A Bayesian account of independent evidence with applications. Available at http://philosophy.wisc.edu/fitelson/psa2.pdf.

Gigerenzer, G. and Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological Review*, 102:684–704.

Good, I. J. (1950). *Probability and the weighing of evidence*. Charles Griffin & Co., London.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In Oaksford, M. and Chater, N., editors, *Rational models of cognition*, pages 248–274. Oxford University Press, Oxford.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cog. Psych.*, 3:430–454.

Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103:582–591.

Mervis, C. B. and Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, 32:89–115.

Minka, T. P. (1998). Inferring a gaussian distribution. http://www-white.media.mit.edu/ tpminka/papers/gaussian.html.

Oaksford, M. and Chater, N. (1998). *Rational models of cognition*. Oxford University Press, Oxford.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., and Shafir, E. (1990). Category-based induction. *Psychological Review*, 97:185–200.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Rips, L. J. (1975). Inductive judgments about natural categories. *J. Verbal Learning and Verbal Behav.*, 14:665–681.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Sloman, S. A. (1993). Feature-based induction. *Cog. Psych.*, 25:231–280.

Tversky, A. and Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90:293–315.

Acknowledgements. Supported by Mitsubishi Electric Research Labs and a Hackett studentship to TLG. N. Davidenko, M. Steyvers, and M. Strevens gave helpful comments.