Why are People Bad at Detecting Randomness? Because it is Hard.

Joseph Jay Williams (joseph_williams@berkeley.edu)
Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California Berkeley
Berkeley, CA 94720 USA

Abstract

People often detect structure and patterns in data that is random. This difficulty in accurately evaluating randomness manifests itself in mistaken beliefs that a fair coin has a bias towards heads or tails, detection of causal relationships between variables that randomly co-occur, or observation of illusory correlations between continuous variables. A computational analysis of an optimal reasoner's performance on these three tasks suggests that this difficulty does not arise simply because people have an irrational disposition to see meaning in randomness, but because the underlying inference problem is intrinsically hard– for both statistical inference and human intuition. This analysis suggests that randomly produced data is inherently ambiguous, because data which is randomly produced can often also be produced by a systematic process. An experiment is reported which provides evidence that inferences about randomness are inherently difficult.

Keywords: randomness; illusory correlation; judgment biases; rational analysis; Bayesian inference;

Does winning three out of four games mean the Red Sox will usually win against the Yankees, or does it just reflect the ups and downs of the sport? Does taking vitamins reduce the chances you will get sick, or have you noticed that health doesn't seem to depend on supplements? Is there a relationship between how healthy people are and how much they exercise? People have a remarkable capacity to detect patterns, and both children and adults are often preoccupied with finding the statistical regularity, causal structure, and predictive relationships that exist in the world. The alternative to finding structure in the environment is to realize that it is unsystematic: certain events and patterns occur in the absence of systematic forces and the processes that generate them are random. Discriminating between events that occur at random and observations that provide evidence for underlying structure is important in learning about the actual structure of the world.

Despite the value of this capacity, there is a range of empirical evidence pointing at the errors that people make in discerning the presence of random rather than systematic processes. Given random samples of equally likely events, people erroneously detect statistical regularity and infer that one event is more likely than another, which can lead to maladaptive decision-making. People often infer causal relationships between randomly co-occurring variables, which leads to errors in clinical diagnosis (Chapman & Chapman, 1967) and the perpetuation of stereotypes (Hamilton, 1981). The tendency to detect correlations in the random variation of continuous variables can result in persistent false beliefs, such as an association between arthritis pain and the weather (Redelmeier & Tversky, 1996).

Why do people have such difficulty in correctly assessing whether data is random, or reflects structure? One proposal (Kahneman & Tversky, 1972) is that people's judgments are not guided by detailed knowledge of probability, but by a heuristic to judge data as random if is intuitively *representative* of a random sample. This proposal (and the accompanying empirical research) is often interpreted as evidence that people's ability to reason about chance is inherently biased and inaccurate.

In this paper we argue that errors in detecting randomness are not necessarily evidence that people simply reason poorly about chance. We present a rational analysis of the inference problem people face in judging whether data provides evidence for a random or systematic process, and show why detecting randomness might be intrinsically difficult. We then extend the analysis to demonstrate that this difficulty is faced in identifying randomness in coin flips, contingency data, and the joint values of continuous variables. Across these inference tasks, the data that is likely under (and produced by) a random process is often data that is also likely under (and could have been produced by) a systematic process. This means that randomly produced data is inherently ambiguous as to its source, and only provides weak evidence for a random process. We empirically test this formal analysis through an experiment on detecting randomness and regularity in sequences of coin flips, which supports the conclusion that judging randomness is in fact inherently difficult in the way our analysis predicts.

A formal analysis of judging randomness

The problem of detecting randomness can be analyzed in terms of using observed data d to evaluate two hypotheses: h_0, the hypothesis that the data was produced by a random process, and h_1, the hypothesis that the data was produced by some systematic process. A rational solution to this problem is to use Bayes' rule to combine prior beliefs about these hypotheses with the evidence the data provides for each. Bayes' rule states that the logarithm of the *posterior odds* in favour of h_1 after seeing d is

$$\log \frac{P(h_1|d)}{P(h_0|d)} = \log \frac{P(d|h_1)}{P(d|h_0)} + \log \frac{P(h_1)}{P(h_0)}$$

being the sum of the logarithms of the *likelihood ratio* and the *prior odds*. The log likelihood ratio (LLR) is what we take as a measure of the evidence that d provides in favour of h_1 (or against h_0), as its value determines the way in which beliefs are changed by the data.

The difficulty inherent in evaluating randomness

Deciding whether data were generated by a random process is rendered inherently difficult because of the relationship between h_0 and h_1. The issue is that randomness is a special case of structure – namely, the case where no structure exists. An illustration of this relationship appears in Figure 1(a). Considering all possible data sets d in a particular domain, the hypothesis of randomness, h_0, assigns high probability only to the particular data sets with little or no structure. In contrast, the hypothesis of structure, h_1, assigns probability to the wide range of data sets containing any structure whatsoever.

This scenario can be illustrated in judging the randomness of sequences of coin flips. Sequences of coin flips with approximately equal numbers of heads and tails are likely to be generated by flipping a fair coin ($P(\text{heads}) = 0.5$), but they are also moderately likely to have come from a coin with a bias (if $P(\text{heads})$ is still close to 0.5, for example, 0.4 or 0.6). A fair coin is just a special case of a biased coin – a coin with zero bias. Formally, the hypothesis that a fair coin (h_0: $P(\text{heads}) = 0.5$) generated a sequence is *nested* within the hypothesis that a biased coin (h_1: $P(\text{heads}) \sim Uniform(0,1)$) generated the sequence.

What are the consequences of this nested relationship between the hypotheses of randomness and structure, for inferences about random processes? The inferences drawn from data depend on the evidence that randomly and systematically generated data sets provide for one hypothesis over another– in our analysis, this evidence is the log likelihood ratio of a particular data set d. The LLR ($\log \frac{P(d|h_1)}{P(d|h_0)}$) can be computed for each data set d, and the probability of d (for example, under a random process) influences how likely it is that that particular LLR will be observed in randomly generated data. The overall distribution of the LLRs of randomly generated data is derived by computing the LLR for every data set and taking into account how likely each data set is under a random process. This derivation gives the distribution of LLRs of randomly generated data (when h_0 is true), and the distribution of LLRs for systematically generated data (when h_1 is true), shown in Figure 1(c). Positive LLR values indicate that the data provides more evidence for h_1, and negative values show that the data provide more evidence for h_0. The greater the magnitude, the more evidence the data provides for one hypothesis over the other.

It is readily apparent that nested hypotheses produce asymmetric distributions of evidence, with most of the evidence for the nested hypothesis of randomness (h_0) being weak (negative but small LLRs), while the evidence for the alternative hypothesis of structure (h_1) is stronger (a wide range of positive LLR values). This follows from the nested nature of the inference: since h_1 assigns some probability to all the data sets for which $P(d|h_0)$ is large, there is a limit on the magnitude of the LLR in favour of h_0. In contrast, because h_1 assigns probability to datasets that are very improbable under h_0 it is possible to obtain a very large LLR in favour of h_1. Returning to judgments about coin flips, observing a sequence with a very uneven number of heads and tails provides very strong evidence for a biased coin, but observing a sequence with a slightly uneven number of heads and tails is more ambiguous, because it could be produced by a systematic process (e.g. one with $P(\text{heads}) = 0.6$). Even if this kind of sequence *was* randomly produced, it is not strongly diagnostic of a random process.

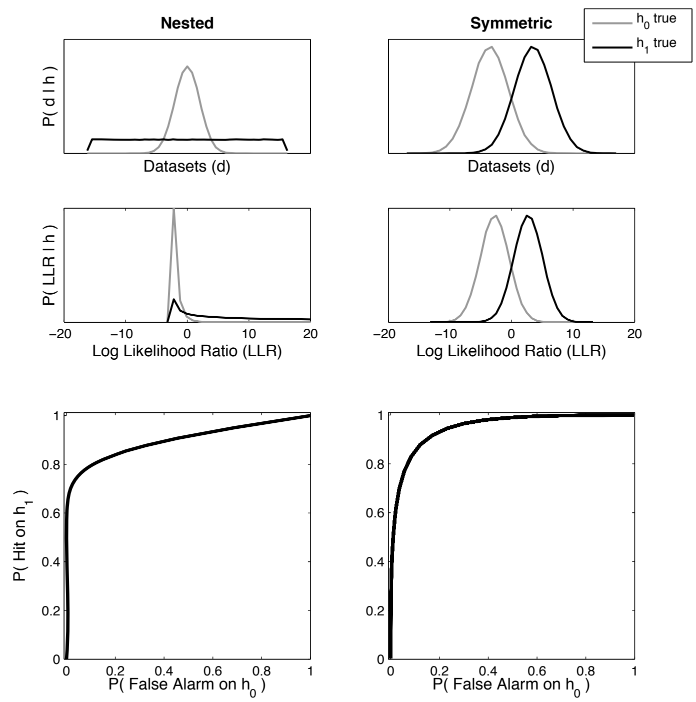

Figure 1: Comparison of nested and symmetric hypotheses: probability distributions over data (a,b), the distribution of the LLR (c,d), and ROC curves (e,f)

Making inferences about nested hypotheses is very different from the inference tasks typically considered in psychology (such as in signal detection theory). Figure 1(b) shows a more typical scenario, discriminating between two symmetric hypotheses, represented by normal distributions over the potential data sets, which partially overlap. As seen in Figure 1(d), symmetric hypotheses produce symmetric distributions of evidence: strong evidence can be obtained for either hypothesis and these distributions are reasonably distinct.

The relative difficulty of judgments about nested versus symmetric hypotheses is well-demonstrated in the receiver operating characteristic (ROC) curves in Figure 1(e) and (f). Taking the LLR as the "signal" for deciding between the two competing hypotheses, and using an exhaustive range of thresholds, the ROC curves plot the probability of correctly inferring h_1 when h_1 is true (detecting structure when a systematic process is present), $P(\text{Hit})$, against the probability of inferring h_1 when the data was generated under h_0 (detecting structure in randomly produced data), $P(\text{False Alarm})$. The curves demonstrate that a reasoner's sensitivity to the true source of the data is inherently lower for nested than for symmetric hypotheses, and producing a high hit rate requires producing a disproportionate number of false alarms.

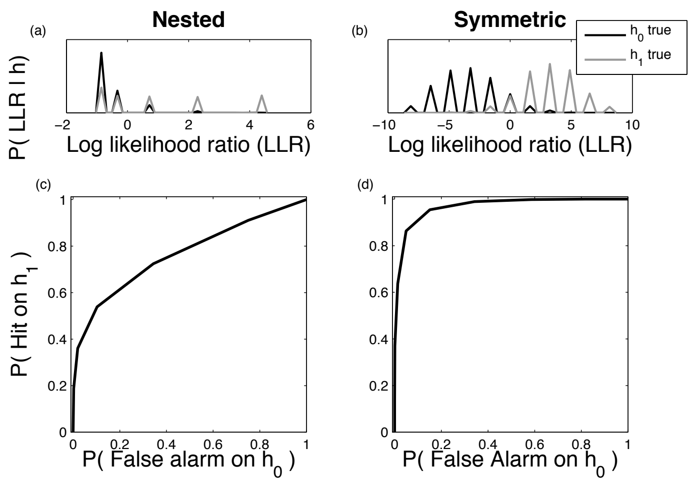

Figure 2: Distributions of the LLR (a,b) and ROC curves (c,d) for nested and symmetric hypotheses concerning coin flips.

False alarms are exactly the phenomenon that is interpreted as irrational– identifying structure in random data– but under this analysis false alarms are in fact an unavoidable cost imposed by the weak distribution of evidence for h_0.

This schematic analysis shows that the challenge of detecting randomness may be a direct consequence of its status as a nested hypothesis. In the following sections we show that the key points of this analysis hold in three settings where people erroneously detect structure in randomness: coin flips, contingency data, and two-dimensional continuous variables.

Bias in sequences of coin flips

People often have difficulty in judging whether sequences of coin flips are random (heads and tails are equally likely) or whether they exhibit some bias and statistical regularity. The inference task people face can be characterized as using an observed sequence of coin flips, d, to decide between the nested hypotheses h_0, that the coin is fair and $d \sim \text{Binomial}(n = 10, 0.5)$, and h_1, that the coin is biased and $d \sim \text{Binomial}(n = 10, p)$ with $p \sim \text{Uniform}(0,1)$. This task can be compared to the (symmetric) judgment of deciding whether a coin is biased towards tails (h_0: $P(\text{heads}) = 0.3$) or biased towards heads (h_1: $P(\text{heads}) = 0.7$).

The distribution of LLRs for h_0 and h_1 for both the nested and symmetric cases are shown in Figure 2 (a) and (b). These distributions were obtained by generating 5000 sequences of 10 coin flips under h_0 and 5000 sequences of coin flips under h_1, computing the LLR for each sequence, and plotting these as a histogram. The results conform to the schematic analysis given above, with a skewed distribution of the LLR for the nested case. The ROC curves shown in Figure 2 (c) and (d) show that the nested case is more difficult, and has a very high false alarm rate.

Causal relationships in contingency data

A basic learning problem children and adults face is inferring the causal structure of the world, often in the form of judging whether one variable exerts a causal influence on another, raising the probability of its occurrence. Our formalization of the causal inference problem follows previous work on Bayesian models of causal induction (Griffiths & Tenenbaum, 2005). We assume that a person evaluates the relationship between two binary variables, a prospective cause C and an effect E, where C and E take on values of 1 and 0, indicating their presence and absence. If a causal relationship exists between C and E, then the probability that $E = 1$ is given by a "noisy-OR" distribution (stated below), while if there is no causal relationship the effect occurs with some base rate p_0. This gives us two hypotheses: h_0 is that there is no causal relationship, and $P(E = 1|C) = p_0$ with $p_0 \sim \text{Uniform}(0,1)$, while h_1 is that there is a causal relationship, with $P(E = 1|C)$ being $p_0 + (p_1 - p_0 p_1)C$, with $p_0 \sim \text{Uniform}(0,1), p_1 \sim \text{Uniform}(0,1)$. It is clear that h_0 is nested within h_1, as the absence of a causal relationship is equivalent to the cause producing the effect with a probability of 0, compared to a probability of p_1 (where $p_1 \sim Uniform(0,1)$) in the presence of a causal relationship.

The nested inference about whether there is a causal relationship can be contrasted to the symmetric judgment of whether a variable raises or lowers the probability of an effect: whether the variable is causally generative or causally preventative. The model for a preventative cause uses a noisy-AND-NOT parameterization (see Griffiths & Tenenbaum, 2005), giving us two hypotheses: h_0 is that there is a generative causal relationship, $P(E = 1|C) = p_0 + (p_1 - p_0 p_1)C$, with $p_0 \sim \text{Uniform}(0,1), p_1 \sim \text{Uniform}(0,1)$, and h_1 is that there is a preventative causal relationship: $P(E = 1|C) = p_0(1 - p_1)^C$, with $p_0 \sim \text{Uniform}(0,1), p_1 \sim \text{Uniform}(0,1)$.

Each sample drawn under a hypothesis was a 2×2 contingency table containing 10 individual cause-effect observations. For each sample, the values of p_0 and p_1 were drawn from a uniform distribution on $(0,1)$. Five observations were fixed to have the cause present, and five were fixed to have the cause absent, while the value of the effect was sampled from the probability distribution defined on E by the relevant hypothesis. A total of 5000 samples were drawn for each hypothesis. Figure 3 shows the distribution of the LLR and the ROC curves for the nested and symmetric hypotheses, bearing out the schematic analysis presented earlier.

Correlation between continuous variables

Detecting correlations between continuous variables is another important problem, whose solution allows people to act adaptively in the world. Despite this, there is ample empirical evidence that people detect positive and negative correlations between variables with no actual relationship (Redelmeier & Tversky, 1996) . For bivariate data, (x_1, x_2), the problem people have to solve can be formalized as deciding between h_0: there is no correlation and the variables are randomly related, and h_1: there is some correlation between the variables. Formally, we can capture the difference between these hypotheses by assuming that under h_0 the values of x_1 and x_2 are drawn independently from Gaussian dis-

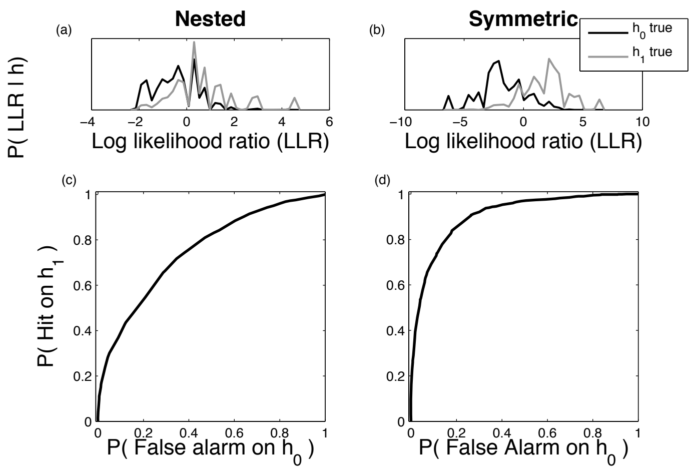

Figure 3: Distributions of the LLR (a,b) and ROC curves (c,d) for nested and symmetric hypotheses of causal relationship.

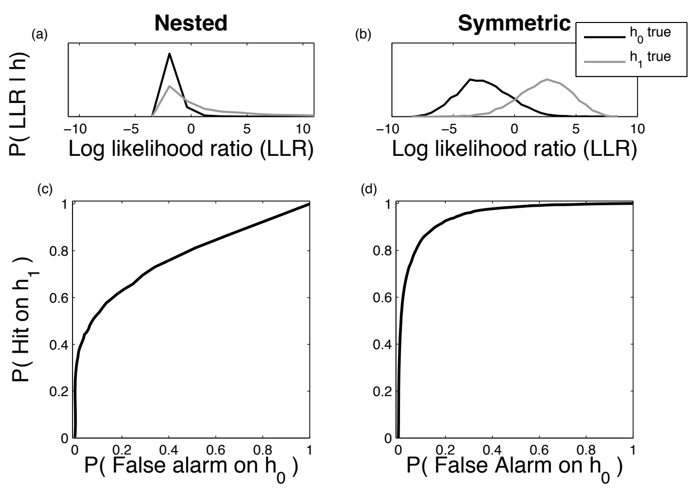

Figure 4: Distributions of the LLR (a,b) and ROC curves (c,d) for nested and symmetric hypotheses concerning correlations.

tributions with unknown means $\mu \sim N(0, \frac{\sigma^2}{k})$ and variances $\sigma^2 \sim \text{Inv-}\chi^2(k, \sigma^2)$, while under h_1 they are drawn from a multivariate Gaussian with mean $\mu \sim N(\mathbf{0}, \frac{\Sigma}{k})$ and covariance matrix sigma $\Sigma \sim \text{Inv-Wishart}(k\Sigma_0, k)$, where Σ_0 is the identity matrix, and $k = 3$ for h_0 and h_1. This allows x_1 and x_2 to be independent (have a correlation of 0) under h_0, but be correlated (have a correlation between -1 and 1) under h_1.

A symmetric analogue of this nested problem is judging whether there is a positive or negative correlation between x_1 and x_2. Here the hypotheses are h_0, that there is a negative correlation, and h_1, that there is a positive correlation. Both of these hypotheses are represented with the same distributions and parameter values as h_1 above, except that the parameters Σ_0 and k are adjusted to represent positively and negatively correlated data. Under h_0, $\Sigma_0 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$, while under h_1, $\Sigma_0 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. We take $k = 10$ in both cases, representing a reasonably strong belief in the appropriate correlation. For both the nested and symmetric cases, 5000 samples were generated under each hypothesis, with each sample consisting of 10 bivariate observations. The plots for the distribution of LLRs and the corresponding ROC curves are shown in Figure 4, and again bear out the original analysis.

Summary

For three domains in which people erroneously detect structure in randomness, a rational analysis identifies that the hypothesis of randomness is nested within the hypothesis of structure, causing the evidence for randomness (as measured by the distribution of LLRs) to be inherently weak. ROC curves constructed using the distribution of LLRs for randomly and systematically generated data (whether coin flips, contingency data, or 2-D data points) demonstrate that this weak evidence makes discriminating these processes inherently difficult– to obtain a high proportion of hits on h_1, a reasoner will incur a high proportion of false alarms on h_0 (detect structure in randomly produced data). The rational analysis shows that data from symmetric hypotheses have more diagnostic distributions of evidence, which facilitates more accurate judgment. We now present an experiment on people's judgments about the processes which generate sequences of coin flips. The experiment compares judgments about nested and symmetric hypotheses to test whether people's difficulty in detecting random processes does in fact arise from the relatively weak evidence provided by random data.

Experiment: Testing the model predictions

The experiment involved one of two judgment tasks about coin flips: deciding whether a coin was random or biased, and deciding whether a coin had a bias towards heads or tails. People's judgments were investigated for the sequences often observed in judging randomness versus bias (*nested* condition), and the sequences often observed in judging a bias towards heads versus tails (*symmetric* condition). In the critical *matched* condition, people judged whether a bias was to heads or tails, but for sequences whose distribution of evidence was chosen to match that of the *nested* condition.

Procedure and Design

For the *nested* condition, 50,000 sequences of 40 coin flips were generated from a fair coin, and 50,000 sequences of 40 coin flips were generated from a coin with a bias, following the sampling procedure for coin flips outlined earlier. All 100,000 samples were pooled and arranged in order of increasing LLR. Starting at the first percentile, sequences were selected at every 2nd percentile (1st, 3rd, 5th, ... , 99th percentile), for a total of 50 sequences which covered the range of LLR values. These 50 sequences of coin flips were presented for judgments about randomness versus bias in the *nested* condition. Another 50 sequences were selected similarly for the *symmetric* condition: by combining 50,000 sequences from a 'coin' with $P(\text{heads}) = 0.3$ and 50,000 from a coin with $P(\text{heads}) = 0.7$, ordering by increasing LLR, and selecting from every 2nd percentile. For the *matched* condition, 50 sequences were selected from the 100,000 in the *symmetric* pool, but such that their LLRs were as close as

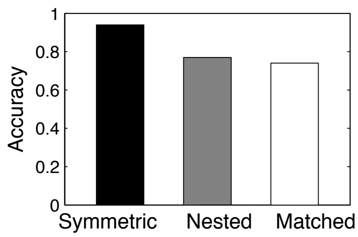

Figure 5: Judgment accuracy across conditions

possible to the LLRs used in the *nested* condition. When multiple sequences had the same LLR (as was often common), a sequence was randomly selected.

Participants were 120 undergraduate students: 40 in each of the three conditions. Participants in the *nested* condition were instructed that they would see sequences of 40 coin flips that each came from either: (1) a fair coin, or, (2) a coin biased to show heads with some probability other than 50%, and would have to decide which coin had produced the sequence. In the *symmetric* and *matched* condition they were instructed to discriminate between: (1) a coin that came up heads 30% of the time, and (2) a coin that came up heads 70% of the time. Participants were given 16 practice trials, followed by the actual experiment of 50 trials. Responses were made by pressing one of two buttons, with the button-response pairing randomly chosen for each participant.

Results

People's judgment accuracy in each of the three conditions is shown in Figure 5. An accuracy score was constructed for each participant as the proportion of correct inferences out of 50, with an inference scored as correct if the participant chose the hypothesis assigned higher posterior probability under the rational analysis (assuming equal priors on h_0 and h_1). Accuracy in the *symmetric* condition was significantly better than in the *nested* condition, replicating the finding that people are relatively inaccurate in their judgments of randomness, ($t(39) = 6.9, p < 0.001$). However, when the distribution of evidence (the LLRs of the sequences presented) was equated in both tasks (judging randomness and judging direction of bias), people's accuracy was equivalent. Accuracy in the *nested* and *matched* conditions did not differ significantly ($t(39) = 1.6, p = 0.12$), although accuracy in the *matched* condition was impaired relative to the *symmetric* ($t(39) = 8.6, p < 0.001$).

Despite the finding that people appear to simply be poorer at evaluating randomness than performing related judgments about structure, once the pattern of evidence is equated in these two tasks, people's accuracy is exactly the same. This provides strong evidence that discriminating the nested hypotheses of randomness and structure is only difficult because of the uninformative distribution of evidence that is produced– equating the evidence available in an intuitively 'easy' task leads to similarly impaired performance.

For each of the 50 sequences, Figure 6 shows the posterior probability that the rational analysis assigns to h_1, and the proportion of participants who selected h_1. People's mean judgments closely track the probabilities assigned by the rational analysis: the correlations for the *symmetric*, *nested*, and *matched* conditions are, respectively, 0.99, 0.94, and 0.92. These strong correlations indicate that people's judgments are very sensitive to the magnitude of evidence that a sequence provides for a judgment (as measured by the LLR), and that people's failure to choose the correct hypothesis corresponds to the inherent uncertainty an optimal reasoner faces.

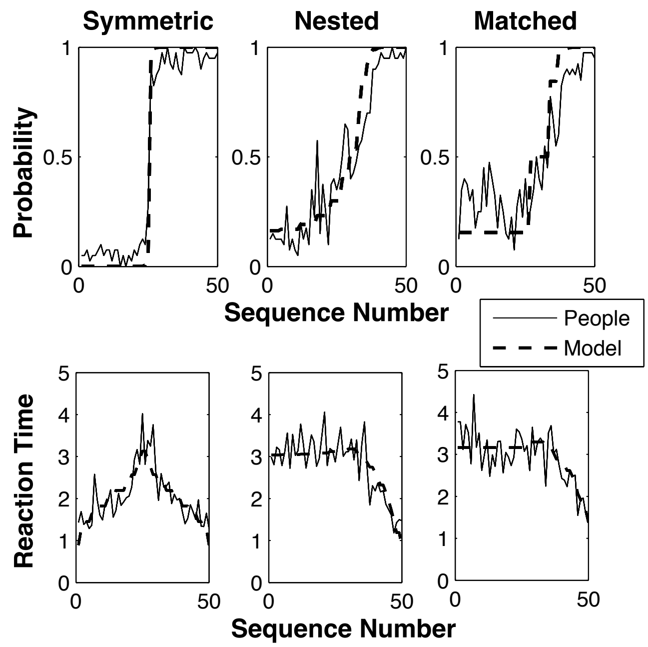

Figure 6: Human and model probability of choosing h_1. Mean reaction time from human data and model predictions.

A similar pattern of difficulty was found in the reaction time data: people took longer to make judgments in the *nested* ($t(39) = 2.5, p < 0.02$) and *matched* ($t(39) = 2.7, p < 0.01$) than the *symmetric* condition, although reaction times for these two conditions did not differ ($t(39) = 0.33, p = 0.74$). All reaction time analyses were carried out on data that was scaled for outliers (reaction times greater than 10 seconds were replaced by a value of 10 s). A particularly interesting feature of the reaction time data was the clear linear relationship between the time people needed to make a judgment about a sequence and the magnitude of the evidence that sequence provided, as represented by the LLR. The correlations between the time to make an inference from a sequence and the absolute value of the LLR of the sequence were -0.82 (symmetric), -0.84 (nested), and -0.75 (matched). The smaller the magnitude of the LLR, the longer the time to make a judgment, the larger the LLR, the quicker an inference was made. Figure 5 shows the mean reaction times. Reaction time was regressed onto the absolute value of the LLR using a linear function, so that Figure 5 shows the close match between people's actual reaction times (solid line) and the regression's prediction for how long people should take– a prediction based directly on the magnitude of the evidence a sequence provides (the LLR). The sequences which provide only weak evidence (as measured by the LLR in the proposed rational analysis) are the sequences that people find inherently difficult to evaluate.

Discussion

The empirical finding that people are bad at detecting randomness in data (instead finding meaning and structure) is traditionally explained as a bias. A closer analysis of this inference problem from the perspective of Bayesian inference over hypotheses suggests that the bias does not emerge from some special relationship between cognition and randomness, but from the likelihood of observed data under the competing *nested* hypotheses of randomness and structure. Across all of the tasks analyzed, a core feature of the relationship between these hypotheses is that random data is often explained very well (assigned high probability) under the alternative hypothesis of structure, making the detection of randomness an inherently difficult problem. The reported experiment showed that this theoretical argument has significant psychological implications: people's accuracy and reaction time was directly related to a rational statistical measure of evidence, the LLR. Moreover, the inaccuracy and bias exhibited in randomness judgments did not seem to be due to some intuitive property of randomness, but to the low informativeness of the LLRs of random data: equating an intuitively easy task on this distribution made judgments about structure just as difficult.

The finding that people persistently detect structure and meaning in randomly generated data- across a range of tasks and domains- has frequently been taken as evidence that people's capacity to reason about chance and probability is fundamentally flawed. The rational analysis and supporting experiment presented in this paper suggest that errors in judging randomness shouldn't necessarily be taken as evidence for the broad theoretical conclusion that people are in general bad at reasoning about chance. A closer look at the computational problem people face suggests that judgments about randomness are inherently difficult, even with optimal statistical knowledge and extensive computational resources, and errors in judging randomness may reflect the difficulty of the problem rather than misleading intuitions about chance events.

This strategy of using a rational computational analysis to more precisely examine the inference problems people have to solve also sets up several useful directions for future research. The close empirical relationship between LLRs and people's judgments suggests that it may be valuable to think further on how analyzing LLRs can give insight into people's inferences and behaviour. Research on illusory correlation (Redelmeier & Tversky, 1996; Jennings, Amabile, & Ross, 1982) proposes that people erroneously detect structure by selectively attending to the subset of available data which provides evidence for structure and ignoring the data that provides evidence for randomness. We propose that if storage and processing limitations are imposed on a human reasoner, it might be adaptive to selectively attend to highly informative data with large LLRs (which we suggest is often evidence for structure) at the expense of weakly informative data with small LLRs (which may often be evidence for randomness). The magnitude of evidence that data provides about the state of the world (about the hypotheses under consideration) may be important in understanding biases in people's selective attention and memory.

In addition to demonstrating that judging randomness is difficult, the rational analysis presented here provides a context within which we can formally characterize the strategies people use to negotiate this difficulty. An analysis of the distribution of LLRs under h_1 demonstrated that a significant proportion of data produced by a systematic process in fact provided evidence for h_0 rather than h_1 (had negative LLRs). If a reasoner's first priority is discovering structure, then the optimal strategy is to set a generous threshold (e.g. a negative LLR) for inferences in favour of h_1. This strategy adaptively accepts the cost of a high false alarm rate to increase the detection of structure, but in a given individual situation will have the surface appearance of an irrational phenomenon.

The rational analysis in this paper aimed to demonstrate the extent to which people's judgments are guided by sophisticated statistical reasoning rather than biases, but it can still serve the function of identifying limitations in human cognition. The concept of nested hypotheses characterizes the inference tasks in which people face an inherent challenge in correctly diagnosing randomness, and will often erroneously detect structure. This paper demonstrates that this challenge is regularly faced in a broad range of ubiquitous and practically important judgment tasks: detecting statistical regularity, discovering causal structure, and identifying correlations.

Acknowledgments. This work was supported by a University of California Regents Intern Fellowship and a Trinidad and Tobago National Scholarship to JJW, and grant number FA9550-07-1-0351 from the Air Force Office of Scientific Research. We would like to thank four anonymous reviewers for their comments on an earlier version of this paper.

References

Chapman, L., & Chapman, J. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72*(3), 193–204.

Griffiths, T., & Tenenbaum, J. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Hamilton, D. (1981). Illusory correlation as a basis for stereotyping. In D. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 115–144). Hillsdale, NJ: Lawrence Erlbaum.

Jennings, D., Amabile, T., & Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211–230). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, *3*(3), 430–54.

Redelmeier, D., & Tversky, A. (1996). On the Belief that Arthritis Pain is Related to the Weather. *Proceedings of the National Academy of Sciences*, *93*(7), 2895–2896.