
Active inference in concept induction

UCSD MPLAB TR 2000.03

Jonathan D. Nelson

Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0515
jnelson@cogsci.ucsd.edu

Javier R. Movellan

Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0515
movellan@inc.ucsd.edu

Abstract

People are active information gatherers, constantly experimenting and seeking information relevant to their goals. A reasonable approach to active information gathering is to ask questions and conduct experiments that maximize the expected information gain, given current beliefs (Lindley, 1956; Good, 1966; MacKay, 1992). In this paper we compare the behavior of human subjects with that of an optimal information-gathering agent (infomax) in a concept induction task (Tenenbaum, 1999, 2000). Results show high consistency between subjects in their choices of numbers to sample. However infomax generally fails to predict subjects' sampling behavior. It is unclear at this time whether the failure of infomax to predict human behavior is due to problems with Tenenbaum's concept induction model, or due to the fact that subjects use suboptimal heuristics (e.g., confirmatory sampling).

1 Introduction

In scientific inquiry and in everyday life, people seek out information relevant to perceptual and cognitive tasks. Scientists perform experiments to uncover causal relationships; people saccade to informative areas of visual scenes, turn their head towards unexpected (i.e., informative) sounds, and ask questions to understand the meaning of concepts.

Consider a person learning a foreign language, who notices that a particular word, "tikos," is used for baby moose, baby penguins, and baby cheetahs. Based on those examples, he or she may attempt to discover what tikos really means. Logically, there is an infinite number of possibilities: tikos could mean *baby animals*, or simply *animals*, or even *baby animals and antique telephones*. In practice a few examples are enough for human learners to reduce the space of possibilities and form strong intuitions about what meanings are most likely.

Suppose you can point to a baby duck, an adult duck, or an antique telephone, to inquire whether that object is "tikos". Your goal is to figure out what "tikos" means. Which question would you ask? Why? When the goal is to learn as much as possible about a set of concepts, a reasonable strategy is to choose those questions which maximize the expected information gain about those concepts, given one's current beliefs (Lindley 1956, Good 1966, MacKay 1992). The goal of this paper is to evaluate whether humans use such a strategy in a relatively unconstrained concept induction task.

1.1 Prior work: The card selection task

Prior work to understand human sampling from an information-theoretic perspective has focused on Wason's (1966, 1968) card selection task. The task was initially conceived as a deductive logic problem. Subjects are given a rule of the form *if p then q*, such as "*if there is a vowel on one side [p], then there is an even number on the other side [q]*." Then, the subject is given a choice of four cards, one each with a visible face showing a statement corresponding to *p*, *not-p*, *q*, and *not-q*. For example, the letter k might be written on a *not-p* card, the number 14 on a *q* card, etc. Each card with a *p* or *not-p* statement on one side has a *q* or *not-q* statement on the other side. The subject is asked to select those cards, and only those cards, that could falsify the rule "*if p then q*". According to deductive logic, the only way to falsify the rule *if p then q* is to select the *p* card and the *not-q* card. The most common subject responses, however, are selection of the *p* card and the *q* card (46 percent of subjects chose this combination of cards), or selection of the *p* card only (33 percent of subjects select only the *p* card). Only a small proportion of subjects, as few as 4 percent, choose the "correct" combination of cards, *p* and *not-q* (Johnson-Laird & Wason, 1970).

If subjects interpret the selection task to be a deductive task, then subjects' behavior on this task may demonstrate a systematic bias in human cognition. However, not all researchers have agreed that subjects interpret the selection task in a deductive manner. If subjects interpret the task in an inductive manner, norms of deductive logic do not apply. Oaksford and Chater (1994) proposed an inductive characterization of the task, under which subjects' behavior on this task may be considered optimal. They proposed that subjects see this task as one of choosing between two statistical hypotheses: (1) if *p* occurs, *q* invariably occurs; and (2) *p* and *q* are statistically independent of each other. Oaksford and Chater demonstrated that if *p* and *q* were both rare, the ordering of selection frequencies observed in studies with human subjects was monotonically related to the expected information gain about the two hypotheses. McKenzie and Mikkelsen (2000) demonstrated that in fact people do assume that both *p* and *q* are rare, when hearing statements of the form *if p then q*. Related work has shown that human sampling behavior is sensitive to variation in base rates of *p* and *q*, with subjects usually responding as predicted by information maximizing theory (Green, Over, & Pyne, 1997; Kirby, 1994).

Taken together, these results suggest that research on the selection task has not demonstrated a failure of humans to understand simple principles of deductive logic, but rather a failure of experimenters to understand the inductive nature of Wason's task. Human sampling behavior on this task is consistent with the predictions of Oaksford and Chater's information-maximizing model. May we then extrapolate that humans are sensitive to maximizing information gain in a wide variety of contexts? Work on the selection task is promising, but the simplicity of the hypothesis space may limit its generality. In this paper we evaluate an information-maximizing approach to sampling behavior in a more complex hypothesis space, based on a concept induction task.

1.2 Tenenbaum's number concept space

Tenenbaum (1999, 2000) developed a Bayesian model of number concept induction. The model describes intuitive beliefs about simple number concepts, and how those beliefs change as new information is obtained. Earlier work on pairwise similarity judgments for the digits 0 through 9 (Shepard, Kilpatrick, & Cunningham 1975), suggested that intuitive number concepts could be divided into *mathematical* and *interval* concepts. Building on this work, Tenenbaum included both mathematical and interval concepts in his number concept space. Interval concepts were sets of numbers between *n* and *m*, where $1 \leq n \leq 100$, and $n \leq m \leq 100$, such as *numbers between 5 and 8*, and *numbers between 10 and 35*. Mathematical concepts included *odd numbers*, *even numbers*, *square numbers*, *cube numbers*, *prime numbers*, *multiples of n* ($3 \leq n \leq 12$), *powers of n* ($2 \leq n \leq 10$), and *numbers ending in n* ($1 \leq n \leq 9$).

Tenenbaum (2000) conducted a study, with 8 MIT students as subjects, to determine whether his model could describe human beliefs. He found a correlation of .99 between subjects' beliefs about what numbers belonged to the true concept, given some examples of that concept, and his model's predictions. To evaluate whether Tenenbaum's model might also describe the beliefs of our subjects, we replicated his study, with 81 undergraduate UCSD students participating. We obtained a correlation of .87 between model predictions and mean subject responses, with some systematic deviations between the model and the subjects' behaviors. In view of the high correlation between model predictions and mean responses, however, we hypothesized that the model provided a reasonable approximation to subjects' intuitions about number concepts. Therefore, we decided to use it as the probability measure attributed to subjects in a study of human sampling behavior.

1.3 Infomax sampling

Consider the following problem. An agent is given examples of numbers that are consistent with a particular concept, and is then allowed to pick a number between 1 and 100, to test whether it follows the same concept as the examples given. For example, the agent may be given the numbers 2, 6 and 4 as examples of the underlying concept. The agent may then choose to ask whether the number 8 is also a member of the concept. The goal of the agent is to guess the correct concept.

We formalize the problem using standard probabilistic notation: random vectors are represented with capital letters and specific values taken by those vectors are represented with small letters. Notation of the form “ $C=c$ ” is shorthand for the event that the random variable C takes the specific value c . We represent the examples given to the agent by the random vector X . The agent considers these examples and updates its beliefs about the probability of each possible concept, $P(C = c | X = x)$.

In the equations below, each Y_k is a random variable representing the outcome of testing a number “ k ”. For example, $Y_5 = 1$ represents the event that 5 is in the set corresponding to the true concept, and $Y_5 = 0$ the event that 5 is not in the true concept. The agent calculates the mutual information, or information gain, expected when asking each of the possible questions, given the example vector $X=x$:

$$I(C, Y_k | X = x) = H(C | X = x) - H(C | Y_k, X = x);$$

where $H(C | X = x)$ is the entropy of C given the event $X = x$,

$$H(C | X = x) = - \sum_{c \in C} P(C = c | X = x) \log_2 P(C = c | X = x),$$

and $H(C | Y_k, X = x)$ is the expected entropy remaining in C after the active sample,

$$H(C | Y_k, X = x) = - \sum_{c \in C} P(C = c | X = x) \sum_{v=0}^1 P(Y_k = v | C = c, X = x) \log_2 P(C = c | Y_k = v, X = x).$$

The infomax agent would ask the question with the highest information gain. Note that information gain is defined with respect to a subjective probability measure, which we approximate with Tenenbaum's model.

2 Human sampling in the number concept game

Twenty-nine undergraduate students, recruited from Cognitive Science Department classes at the University of California, San Diego, participated in the experiment. Subjects gave

informed consent, and received either partial course credit for required study participation, or extra course credit, for their participation.

The experiment began with a brief description of the kinds of concepts in the hypothesis space:

Often it is possible to have a good idea about the state of the world, without completely knowing it. People often learn from examples, and this study explores how people do so.

In this experiment, you will be given examples of a hidden number rule. These examples will be randomly chosen from the numbers between 1 and 100 that follow the rule. The true rule will remain hidden, however. Then you will be able to test an additional number, to see if it follows that same hidden rule. Finally, you will be asked to give your best estimation of what the true hidden rule is, and the chances that you are right.

For instance, if the true hidden rule were multiples of 11, you might see the examples 22 and 66. If you thought the rule were multiples of 11, but also possibly even numbers, you could test a number of your choice, between 1-100, to see if it also follows the rule. Finally, based on that information, you could guess the true rule, and give your estimate of the chance it is the true rule.

Figure 1. Instructions for the study

In each question, one or more numbers were given as examples of a number concept, or “hidden number rule.” The task was to test the number that would provide the most information about the true concept:

Problem 1 of 15

The computer has just selected a new hidden rule.
The numbers **81, 98, 96, and 93** are example numbers that satisfy the rule.

Please test one number (1-100) to see if it follows the rule.

Number to test:

Answer: Yes, 90 does follow the rule.

Figure 2. Example of experimental stimuli

Subjects entered a single number, clicked “test,” and received feedback on whether the number they tested was consistent with the true concept, as shown above. Testing more than one number was not permitted, and subjects received an error message if they attempted to do so. In the second part of each question, which is not pictured here, subjects guessed what they believed the true concept to be, indicated their confidence in that choice, and were given the true concept.

The experiment included 13 questions, which were chosen to represent both mathematical and interval concepts, at various levels of overall uncertainty (see Table 1). The experiment was implemented via the Internet. Subjects logged on to the experiment at their convenience, from the location of their choice. Order of the questions was randomized, such that different subjects received different random orderings.

We implemented Tenenbaum’s theoretical number concept space, and simulated the behavior of infomax sampling agents in that space, on each of the questions in the behavioral study. We hypothesized that participants would sample numbers with high information value, as calculated relative to Tenenbaum’s probability model.

3 Results

Two findings stand out. First, the mean information value of subjects' sampling seldom neared that of infomax, when calculated relative to the formal probability measure in Tenenbaum's model. Second, in many cases there was high consistency between subjects in their choices of numbers to sample. The average information gain associated with subjects' sampling strategies, and each of three other sampling strategies, are given in Table 1, below. The confirmation strategy was calculated by considering the information value of the number that Tenenbaum's model considered most likely, but with probability less than 1.00, to be consistent with the true concept. The random strategy was calculated by averaging the information gain of each of the numbers between 1 and 100, weighting each number equally.

The questions are grouped into four types, according to the model's beliefs. On the *single example, high uncertainty* questions the model gives some probability to each of many concepts. On the *multiple example, low uncertainty* questions the model gives a single concept probability between .94 and 1.00. On the *interval* questions the model gives some probability to each of several interval concepts, but no probability to any of the mathematical concepts. On the *intermediate uncertainty* questions, the model gives both interval and mathematical concepts some probability, and no single concept dominates.

Question type	Single example, high uncertainty			Multiple example, low uncertainty			Interval			Intermediate uncertainty			
	16	60	37	16, 8, 2, 64	60, 80, 10, 30	81, 25, 4, 36	16, 23, 19, 20	60, 51, 57, 55	81, 98, 96, 93	21, 35, 41	6, 2	27, 7	14, 2, 6
Example numbers	16	60	37	16, 8, 2, 64	60, 80, 10, 30	81, 25, 4, 36	16, 23, 19, 20	60, 51, 57, 55	81, 98, 96, 93	21, 35, 41	6, 2	27, 7	14, 2, 6
Subjects	.70	.72	.73	.00	.06	0.00	.47	.37	.31	.57	.55	.27	.39
Infomax	.97	1.00	1.00	.01	.32	0.00	1.00	.99	1.00	1.00	1.00	.40	1.00
Confirmation	.97	1.00	1.00	.00	.00	0.00	.86	.81	.73	.62	.40	.32	1.00
Random	.35	.54	.52	.00	.04	0.00	.17	.20	.14	.49	.32	.22	.36

Table 1. Information gain of several sampling strategies on each question in the study, in bits, relative to Tenenbaum's probability model.

3.1 Single example, high uncertainty questions.

The best agreement between infomax predictions and subjects' sampling behavior was observed on questions in which only a single example number was given. Infomax rated subjects' sampling as highly informative on each of these questions. In the case of the question with the single example number 16, the modal response was optimal. On this question eight subjects tested the number 4.

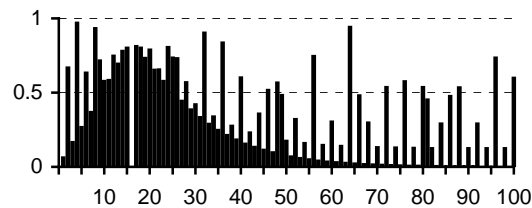


Figure 3. Information gain from sampling each number 1-100, in bits, given the example 16.

Results were similar on the questions with the example numbers 37 and 60. On those questions many subjects tested numbers with high information gain, but there was less consistency in the specific numbers tested.

3.2 Multiple example, low uncertainty questions.

On some questions, there was little or no information value in any sample. One might therefore conclude that sampling behavior on these problems is irrelevant to theory of human sampling. However, subjects' sampling was very consistent on these problems, appearing to follow a confirmatory strategy. In the question with the examples 16, 8, 2, and 64, the model gave probability 1.00 to the concept *powers of two*, and the most commonly tested numbers were 4 and 32 (each with eight subjects). In the question with the examples 81, 25, 4, and 36, the model gave probability 1.00 to the concept *square numbers*, and the most commonly tested numbers were 49 (eleven subjects) and 9 (four subjects).

On the question with the example numbers 60, 80, 10, and 30, the model gave probability .94 to the concept *multiples of 10*, and probability .06 to the concept *multiples of 5*. Here, information gain from testing any multiple of 10 was zero bits, although testing an odd multiple of 5 had information gain of 0.32 bits. Most subjects (21/29) followed the simple confirmatory strategy of testing numbers that were consistent with the most likely concept, *multiples of 10*, although five subjects tested an odd multiple of 5, such as 15, which was the information maximizing strategy on this question, returning 0.32 bits.

3.3 Interval questions

On these questions the model was certain that the true concept was of the interval form, although it was not certain of the precise endpoints. Numbers within the range of observed examples were given probability 1.00 of being consistent with the true concept and testing them was rated as uninformative.

Consider the case of the question with the example numbers 60, 51, 55, and 57 (see Figure 4). Numbers between 51 and 60 have probability 1.00 of being consistent with the true concept (right), and provide no information gain when sampled (left). On each of these questions, however, between 8 and 10 of the 29 subjects tested numbers within the observed range, but which were not explicitly given as examples. This behavior is uninformative, when information gain is calculated with respect Tenenbaum's formal probability model. A discrepancy between subjective probabilities and the formal probability model may be responsible for the apparent uninformativeness of subjects' behavior. In our earlier study of subjective probabilities, we found that subjects gave these numbers mean probability of .72 of being consistent with the true concept.¹ The information value of testing numbers just outside the range of observed examples, that have probability near .72, is high. In Figure 4, below, the numbers 50 and 61 each have probability .75 of being consistent with the true concept, and each provide .82 bits of information.

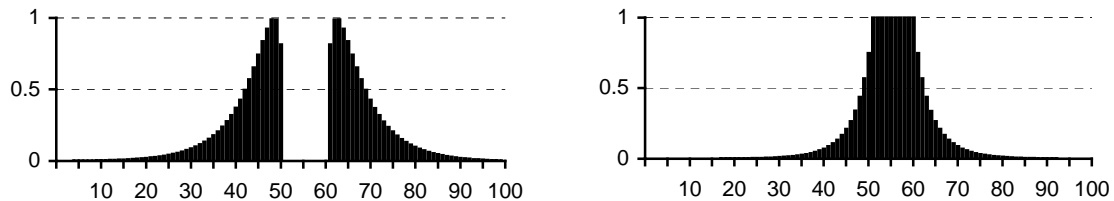


Figure 4. Information value of sampling each of the numbers between 1 and 100, in bits (left), and probability that each of the numbers between 1 and 100 is consistent with the true concept (right), given the example numbers 60, 51, 55, and 57.

3.4 Intermediate uncertainty questions

On these questions subjects often sampled numbers consistent with the most likely concepts, intermediate within the range of observed examples, to which the theoretical

¹ Subjects gave numbers explicitly given as examples probability 1.00, and seldom tested them.

probability model assigned probability of 1.00. Information value of testing numbers with probability 1.00, on these and all questions, was 0 bits. The question with the examples 6 and 2, is the most dramatic example of this. The modal response (ten subjects) was to test the number 4, which infomax rated as uninformative. But if the subjective probability of the number 4, given the examples 6 and 2, is less than 1.00, this behavior might in fact be informative. In the case of this question, and similar questions, it is not clear whether the theoretical probability model is failing to approximate subjective probabilities, or whether subjects are using an uninformative sampling strategy.

4 Discussion

Two findings stand out. First, subjects' sampling strategies seldom neared the information value of infomax, when evaluated according to the formal probability measure in Tenenbaum's model. Second, there was high consistency between subjects in their choices of numbers to sample. This supports the possibility of a descriptive theory of human sampling in the number concept space, irrespective of the optimality of human behavior.

This research contributes to the rational approach to cognition (Anderson, 1990; Chater & Oaksford, 2000), in which empirical findings of actual behavior are used to suggest appropriate properties of normative models. We are collaborating with Tenenbaum to improve the descriptive accuracy of the probability measure in his number concept space. As the descriptive accuracy of the formal probability measure improves, the calculation of the information value of sampling particular numbers will also become more reliable. At present, it is not clear whether our results show suboptimality in human sampling, or simply the need to modify aspects of the theoretical probability measure in the number concept space.

Future work in the number concept space will explore whether sampling strategies change when people are allowed to sample multiple times, as found by Ginzburg and Sejnowski (1996) on the selection task. Future work will also consider whether efficient heuristics may approximate the computationally intensive calculation of infomax sampling.

There is often a trade-off between studying domains that are simple enough to be formally understood, and complex enough to be relevant to interesting real world situations. Previous research in the rational study of human sampling behavior has considered a very simple space, in which the goal is to decide between two possible hypotheses. Therefore, a definitive description of subjects' beliefs, and sampling behavior, in the more complex number concept space, would bring new real-world relevance to theory of active inference. Active sampling in the number concept space does not itself approach the complexity of sampling in the real world, but it may be more easily understood. We hope that a good description of human sampling behavior in the number concept space will also help to illuminate behavior in more realistic situations, such as saccades in visual scenes.

Acknowledgments

We thank Josh Tenenbaum, Dan Bauer, Iris Ginzburg, Craig McKenzie, Terry Sejnowski, and Kent Wu, for their ideas and help in this research. Jonathan Nelson was partially supported by a Pew graduate fellowship during this research.

References

- Anderson, J. R. (1990). *The Adaptive Character of Thought*, LEA, Hillsdale, NJ
- Chater, N.; Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese*, 122, 93-131.
- Ginzburg, I & Sejnowski, T. J. (1996). Dynamics of rule induction by making queries: transition between strategies. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, La Jolla, California, 121-125.

- Good, I. J. (1966). A derivation of the probabilistic explication of information. *Journal of the Royal Statistical Society, series B*, 28, 578-581.
- Green, D. W.; Over, D. E.; Pyne, R. A. (1997). Probability and choice in the selection task. *Thinking and Reasoning*, 3, 209-236.
- Johnson-Laird, P. N.; Wason, P. C. (1970). Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, 22, 49-61.
- Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, 51, 1-28.
- Klayman, J.; Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211-228.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27, 986-1005.
- MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, 4, 590-604.
- McKenzie, C. R. M.; Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review*, 7, 360-366.
- Oaksford, M.; Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.
- Shepard, R. N.; Kilpatrick, D. W.; Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, 7, 82-138.
- Tenenbaum, J. B. (1999). *A Bayesian Framework for Concept Learning*. Ph.D. thesis, MIT.
- Tenenbaum, J. B. (2000). Rules and similarity in concept learning. In *Advances in Neural Information Processing Systems*, 12, Solla, S. A., Leen, T. K., Mueller, K.-R. (eds.), 59-65.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New Horizons in Psychology* (pp. 135-151). England: Penguin.
- Wason, P. C. (1968). On the failure to eliminate hypotheses -- A second look. In P. C. Wason & P. N. Johnson-Laird (Eds.), *Thinking and Reasoning* (pp. 165-174). England: Penguin