UNIVERSITY OF CALIFORNIA, SAN DIEGO

Optimal Experimental Design as a Theory of Perceptual and Cognitive Information Acquisition

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Cognitive Science

by

Jonathan David Nelson

Committee in charge:

Martin I. Sereno, Chair
Javier R. Movellan, Co-chair
Garrison W. Cottrell
Gedeon O. Deak
Craig R. M. McKenzie
Jochen Triesch

2005

The Dissertation of Jonathan David Nelson is approved, and it is acceptable in quality and form for publication on microfilm:

Co-chair

Chair

University of California, San Diego

2005

TABLE OF CONTENTS

(use the links below to access the full articles)

LIST OF FIGURES AND TABLES

(note that these page numbers do not correspond to the original articles)

Chapter 6

ACKNOWLEDGEMENTS

I am grateful to the many people who have shaped and encouraged my intellectual development:

to my parents, for unfailing support;

to Flavia, for encouragement, ideas, and the sharpest of helpful critique;

to my brothers Andrew and Daniel, my sister Janelle, and my brother-in-law Alex;

to Marty Sereno, Javier Movellan, and Gary Cottrell, for helpful mentorship, advice, critique, and ideas;

to my professors at Wheaton College, especially Gary Larson, Bob Vautin, Thomas Kay, and Jim Mann;

to Marta Kutas, for being a tough, honest critic, and making the eye tracker in her lab available for my research;

to the regular attendees of the Cognitive Brownbag group, of GURU, and of MPLAB, for great ideas and helpful critique;

to Laura Kemmer and Alan Robinson, for always having time to read a manuscript or listen to a draft talk;

to Jon Baron, Nick Chater, Michael Lee, Craig McKenzie, Javier Movellan, Mike Oaksford, and Keith Rayner, for their thorough yet kind critique of my written work;

to Josh Tenenbaum, Bob Rehder, and Aaron Hoffman, for making the collaborative enterprise of science so exciting;

to Becky Burrola and Gris Arellano-Ramirez for providing encouragement and advice for navigating academic bureaucracy;

to Seanna Coulson, Aaron Cicourel, Gedeon Deak, Jochen Triesch, David Kirsh, Jim Hollan, and Jeff Elman, for support, advice, critique, and encouragement;

to Tim Marks and Eric Wiewiora, for for thinking deeply about science, always being able to talk, and for sharing their meticulous mathematical minds;

to my fellow graduate students who made graduate school both exciting and humane;

and to Johnny Weh, Ian Fasel, Luis Palacios, Dan Bauer, and Mark Wallen, for technical assistance that has made a great deal of my research possible.

Chapter 2 consists of a preprint of an article that has been accepted for publication, in which the dissertation author was sole author and investigator:

Nelson, J. D. (in press). Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain. Psychological Review, 112(4), to appear October, 2005.

Chapter 3, in full, is a reprint of the following publication, in which the dissertation author was first author and primary investigator:

Nelson, J. D., Tenenbaum, J. B., Movellan, J. R. (2001). Active inference in concept learning. In J. D. Moore & K. Stenning (Eds.), Proceedings of the 23rd Conference of the Cognitive Science Society, 692-697. Mahwah, NJ: Erlbaum.

Chapter 4, in full, is a preprint of an article that has been accepted for publication, in which the dissertation author was first author and primary investigator:

Nelson, J. D.; Cottrell, G. W. (conditionally accepted). A probabilistic model of eye movements in concept formation. Neurocomputing.

Chapter 5, in full, is a reprint of a published article in which the dissertation author was second author and co-investigator:

Movellan, J. R., Nelson, J. D. (2001). Probabilistic functionalism: a unifying paradigm for cognitive science. Behavioral and Brain Sciences, 24(4), 690-692.

Chapter 6, in full, is a reprint of a published article in which the dissertation author was second author and co-investigator:

McKenzie, C. R. M., Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. Psychonomic Bulletin and Review, 10(3), 596-602.

CURRICULUM VITAE

Education

Ph.D., Cognitive Science, University of California at San Diego, 2005

M.S., Cognitive Science, University of California at San Diego, 2002

B. A., Cognitive Science and Statistics (Magna Cum Laude), Wheaton College, 1998

A. A. S., Whatcom Community College, 1995

Reviewed papers

Nelson, JD; Cottrell, GW (conditionally accepted). A probabilistic model of eye movements in concept formation. Neurocomputing.

Nelson, JD (2005). Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain. Psychological Review, 112(4), to appear October, 2005.

McKenzie, CRM; Nelson, JD (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. Psychonomic Bulletin and Review, 10(3), 596-602.

Movellan, JR; Nelson, JD (2001). Probabilistic functionalism: a unifying paradigm for cognitive science. Behavioral and Brain Sciences, 24(4), 690-692.

Nelson, JD; Tenenbaum, JB; Movellan, JR (2001). Active inference in concept learning. In J. D. Moore & K. Stenning (Eds.), Proceedings of the 23rd Conference of the Cognitive Science Society, 692-697. Mahwah, NJ: Erlbaum.

Nelson, JD; Movellan, JR (2001) Active inference in concept learning. Advances in Neural Information Processing Systems, 13, 45-51.

Talks and posters

Nelson, JD; Cottrell, GW (2005). Intuitive experimental design: Toward a theory of questions' usefulness. ASIC, July, 2005.

Nelson, JD (2005, June). Ideal ideal observers. Annual Cognitive Neuroscience Retreat, Salk Institute.

Nelson, JD (2005). Late breaking results in intuitive experimental design. Psychology Dept., UCSD.

Nelson, JD; Cottrell, GW (2005). Eye movements for concept learning. COGS 200 seminar, April 29, 2005.

Filimon, F; Nelson, JD; Sereno, MI (2005). Parietal cortex involvement in visually guided, non-visually guided, observed, and imagined reaching, compared to saccades. Vision Science Society Conference, May, 2005.

Filimon, F; Nelson, JD; Sereno, MI (2005). Human parietal activations to visually guided and non-visually guided reaching versus saccades. Cognitive Neuroscience Society Conference, April, 2005.

Nelson, JD; Cottrell, GW; Movellan, JR (2004). Explaining eye movements during learning as an active sampling process. International Conference on Development and Learning, Oct, 2004.

Nelson, JD (2004) Statistical principles and intuitive experimental design. Experimental Philosophy Lab, UCSD, Oct, 2004.

Nelson, JD (2004). Finding useful questions in a natural environment. Cognitive Science Society Conference, Chicago, Aug, 2004.

McKenzie, CRM; Sher, S; Nelson, JD (2004). Framing effects and information leakage. Conference on individual decisions, Institute for Mathematical Behavioral Sciences, UC Irvine, May, 2004.

Nelson, JD; Cottrell, GW; Movellan, JR; Sereno, MI (2004). Yarbus lives: a foveated exploration of saccadic eye movement. Journal of Vision, 4(8), 741. Vision Science Society conference, May, 2004.

Nelson, JD. How should a test's usefulness be quantified-- with Bayesian diagnosticity, information gain, or probability gain? Implications for eye movement, medical decision-making, and cognitive psychological tasks.
-- Salk Institute for Neural Computation pre- and post-doc group, UCSD, Jan, 2004
-- Psychology Department, UCSD, Jan, 2004
-- Cognitive Science Department, UCSD, Jan, 2004
-- University of Auckland Psychology Department, New Zealand, Dec, 2003

Nelson, JD. When the ideal observer meets the brain: A new approach to modeling saccadic eye movement.
-- Cognitive Neuroscience Retreat, Del Mar, CA, May, 2003
-- 10th Joint Symposium on Neural Computation, UC Irvine, May, 2003
-- Psychology Department, UCSD, April, 2003
-- Kutas lab talk, UCSD Cognitive Science Department, April, 2003
-- Perceptual Expertise Network 6, Boulder, Colorado, Feb, 2003

McKenzie, CRM; Nelson, JD. What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. Annual Meeting of the Psychonomic Society, Orlando, FL, Nov, 2001.

Nelson, JD; Movellan, JR (2001). Inference by means of uncertainty. 8th Joint Symposium on Neural Computation, Salk Institute, May, 2001. (abstract)

Nelson, JD; Movellan, JR (2000). Concept induction in the presence of uncertainty. 7th Joint Symposium on Neural Computation, University of Southern California, May, 2000

Honors and Awards

NSF IGERT Predoctoral Fellowship, 2004-2005 (NSF grant DGE-0333451 to G. Cottrell)

NIH Predoctoral Fellowship, 2002-2004 (NIH grant 5T32MH020002-04 to T. Sejnowski)

Pew Graduate Fellowship, 1998-2000

PEMCO Fellowship, 1998-2003

Scholastic Honor Society, Wheaton College, 1998

Chair, Cognitive Science Society of Wheaton College, 1997-1998

National Merit Commended Scholar, 1995

Teaching and Professional Experience

IGERT Vision and Learning in Humans and Machines Bootcamp, 2004. Designed and mentored student project using gaze-contingent eye tracking experiment to study concept learning.

SCANS presents (COGS 91), 2003. Taught seminar course to introduce students to a variety of research agendas and applications of research in cognitive science.

Human Development (HDP I), 2000, 2001, 2002 (supervisors: Jeff Elman, Farrel Ackerman). Developed new content, with Jeff Williams of the UCSD Biomedical Library, to introduce MedLine and PsychInfo, and a new research component of the course. Led discussion sections for 70 students. Topics included neural and visual development, social development, and gender development.

Computing (COGS 3), 1999-2001 (supervisors: Mark Wallen, Mary Boyle). Designed new course components, including lectures and assignments, to introduce DHTML technologies, and assistive technologies for the disabled. Received teaching award.

Multimedia Design (COGS 187A), 1999 (supervisor: David Kirsh). Mentored student projects in Web design. Received teaching award.

Cognitive Ethnographer, Intel Architecture Labs, 1999 (supervisors: John Sherry, Brad Anders). Studied communication and workflow on the factory floor. Presented factory management with specific recommendations to improve factory efficiency and worker satisfaction.

Probability and Statistics, Wheaton College, 1998 (supervisor: Dr. Jim Mann). Led study sections, graded student work, and lectured in calculus-based mathematical statistics course.

ABSTRACT OF THE DISSERTATION

Optimal Experimental Design as a Theory

of Perceptual and Cognitive Information Acquisition

by

Jonathan David Nelson

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2005

Martin I. Sereno, Chair

Javier R. Movellan, Co-chair

Savage (1956) described how Bayesian statistics and decision theory could provide a theoretical foundation for identifying the most useful experiments to conduct. In the last quarter century, several researchers have suggested that this optimal theory might also describe human behavior. Research to date, though intriguing, has been limited by the simplistic, one-shot nature of most tasks studied, and the relatively arbitrary choices of normative models. The present dissertation seeks to clarify the theoretical and empirical bases of different optimal models of information acquisition (Nelson, in press), as well as to evaluate the possibility of extending these models to new types of tasks (Nelson, Tenenbaum, & Movellan, 2001; Nelson & Cottrell, accepted).

Several normative models for how people should identify useful experiments have been proposed, notably Bayesian diagnosticity, information gain (mutual information), Kullback-Liebler distance, probability gain (error minimization), and impact (absolute change). Existing results from human subjects do not discriminate between these norms as descriptive models of human behavior. Computational optimization found situations in which information gain, probability gain, and impact strongly contradict Bayesian diagnosticity, and in which diagnosticity's claims are clearly inferior. A new experiment strongly contradicts Bayesian diagnosticity. The other normative models behave similarly; each approximates human behavior reasonably well.

New results from modeling a number concept learning task suggest that people's queries can provide novel insights into their beliefs. We also introduce a principled probabilistic model that describes the development of beliefs on a classic (Shepard, Hovland, & Jenkins, 1961) concept formation task. We use that learning model, together with an optimal experimental design-inspired sampling function, to help explain eye movements on Rehder and Hoffman's (2003, 2005) eye movement concept formation task. The same rational sampling function can help predict eye movements early in learning, when uncertainty is high, as well as late in learning when the learner is nearly certain of the true category.

Together, the investigations reported in this dissertation help to clarify the theoretical foundation, and strengthen the empirical basis, of the theory that human evidence acquisition can be modeled as an optimal experimental design problem.

Chapter 1

Introduction

Two primary currents of ideas come together in this dissertation. The first is the study of "intuitive statistics" (reviews by Peterson and Beach, 1967; Edwards, 1968), in which several researchers sought to address whether statistical principles, such as use of Bayes' (1763) theorem, might provide insight into human subjects' thinking. The second is the statistical theory of optimal experiment selection, or optimal experimental design, as articulated by Savage (1954, chap. 6). Savage's work provided a compelling theoretical framework for deciding, in advance, which of several possible experiments to conduct, or which of several questions to ask. Each current of work, and how they come together in this dissertation, is introduced below.

Intuitive Statistics

One current of research that this dissertation draws on is the idea that optimal statistical principles could describe human perception of information, e.g. the "intuitive statistics" view, which was pursued extensively in the 1960s. Bayesian reasoning research, such as that described by Edwards's (1968) narration of a discussion among several scholars in that field, illustrates the types of tasks that were used with human subjects, and the types of statistical models that were applied to analyze the data. The idea of Bayesian reasoning research was to assess whether Bayes' (1763) theorem, which provides a statistically justified way to change beliefs based on new information, also describes how people's beliefs change as new information is obtained. ("Beliefs," in this context, are probabilities that describe the subjective plausibility of each possible category, or hypothesis, such as whether urn 1 or urn 2 was chosen in the example below.) A typical Bayesian reasoning task would concern two urns, with red and blue poker chips, in different, specified proportions. Suppose that urn 1 had 20% red chips and 80% blue chips, and that urn 2 had 80% red chips and 20% blue chips. An urn would be drawn at random, after which a sample of, say, 5 chips would be drawn from that urn. The sample would contain, say, 4 red chips and 1 blue chip, and the subject would be asked to state the odds that the one versus the other urn had been drawn. The subject would be asked which urn was more likely to have been drawn, and the odds for it, relative to the other urn. In this case, the sample clearly favors urn 2, with odds of about 64 to 1.

Subjects correctly identified which urn was more likely to have been drawn. The subject might state that the odds of urn 2 having been drawn were 4 to 1, relative to the odds of urn 1. Normatively (relative to the model of the task in the experimenter's mind, in which no possibility the data were presumed to be noise-free) the odds favoring urn 2 are more like 64 to 1. Thus, it was believed that human reasoning was qualitatively Bayesian, but "conservative," such that "it takes anywhere from two to five observations to do one observation's worth of work in inducing a subject to change his opinions" (Edwards, 1968, p. 18). Peterson and Beach's (1967) review of research on "Man as an intuitive statistician" describes a variety of tasks studied in this paradigm. The usual finding was that people's judgments' at least qualitatively conform to optimal statistical principles. In early research in this area, most tasks were simple and mundane, such as the urn-and-poker chip example mentioned above, as though the tasks were inspired by the exercises in introductory probability and statistics texts. Recent Bayesian theories of concept learning, such broaden accounts of cognition as intuitive statistics to include a variety of more interesting tasks. One example of a more interesting task is Shepard, Hovland and Jenkins's (1961) concept formation task, in which the subject must learn via experience which of several objects (e.g. circles and squares that are large or small and black or white) are consistent with an unspecified concept; Bayesian models of this task have been introduced by Anderson (1991), and by Nelson and Cottrell (in press). Another example is Tenenbaum's (1999, 2000, Tenenbaum & Griffiths, 2001) number concept game, in which subjects learn what number concept (such as "multiples of 7" or "odd numbers") is responsible for a set of numbers they are given.

Optimal Experimental Design

Another current of thought that is central to this dissertation is the idea of modeling experiment selection as a decision theory problem. Savage (1954, chap. 6) beautifully articulated how experiment selection could be modeled as a decision problem, in which the goal is to maximize subjective expected utility, relative to one's beliefs. Savage illustrated his theory with a simple example of a task of deciding how many pounds of grapes to buy. Savage noted that while subjective utility—the perceived value of the grapes to the person buying them—was likely to increase monotonically with the amount of grapes purchased, it was unlikely that three pounds of grapes would provide three times the subjective utility of one pound of grapes. Savage noted that by visually inspecting the grapes in the store, he could get some idea of their quality. Asking a fellow shopper their opinion of the grapes, however, could decrease the variance in his estimate of the grapes' quality, potentially allowing him to better determine how many grapes to buy, thus increasing his expected subjective utility. (In the example, the fellow customer was the lady standing next to him in the store, who, Savage noted, was likely to be a an excellent judge of the grapes' quality.) Savage noted that is typically a subjective cost to obtain new information (in this case, perhaps the embarrassment of letting an opposite-sex stranger know you lack expertise in judging grapes). If the estimated subjective value of increasing the precision in the estimate of the grapes' quality outweighs its subjective cost, then it makes sense to ask the question. Deciding whether or not to ask a question (or do a medical test, or conduct a scientific experiment) can then be modeled as a decision problem in which the goal is to maximize expected subjective utility. Statisticians such as Lindley (1956), Good (1950), Box and Hill (1967), and Fedorov (1972), each have proposed one or more means to quantify experiments (or questions') utility, and their ideas will figure prominently in this dissertation. However, the intricacies of particular utility functions for experiment selection should not obscure that these statisticians, like Savage, proposed quantifying an experiment's usefulness as the average (mean) utility of its possible outcomes.

Intuitive Experimental Design

Recent research has brought the above currents of ideas together, to see whether people's ideas of the value of information correspond to optimal statistical principles. Trope & Bassok's (1982, 1983; Bassok & Trope, 1983-1984) work on social hypothesis testing is an early example. A typical task included identifying what questions would be likely to help identify whether a person was an introvert or extravert. Examples of possible questions were whether the person is neat and tidy (a less useful question) and whether the person went out last Friday night (a more useful question). Subjects were asked to identify which questions were more useful. Baron (1985) gave a theoretical analysis of the value of information on several tasks, such as identifying which of several diseases a patient has, identifying which urn a ball was drawn from, and identifying the correct hypothesis on a simple model of Wason's (1960, 1966) 2-4-6 task. Skov and Sherman (1986) and Slowiaczek, Klayman, Sherman and Skov (1992) studied subjects' choices of questions on the Planet Vuma scenario. In that scenario, the task is to identify the true species of an invisible creature (Glom or Fizo) by inquiring as to whether it has particular features, such as breathing fire or wearing a hula-hoop. In each of the scenarios discussed above, the idea was that it should be possible to address whether human evidence acquisition (e.g. choices of questions to ask) corresponds to statistical principles for experimental design.

The present dissertation seeks to clarify some foundations of theories of experimental design and to broaden the range of tasks to which explicit value of information analyses are applied. To emphasize that this work draws on statistical theories of optimal experimental design, as well as empirical investigation of human subjects' intuitive competence on experimental design tasks, the area of research is termed "intuitive experimental design."

Rational models; heuristics and biases

In this dissertation, optimal models are used as descriptive models for human intuitions and behavior. The view underlying this approach could be caricaturized as the "innocent until proven guilty" view. This view is articulated in Brunswik's (1952) "molar analysis," Marr's (1982) computational analysis of vision, Anderson's (1990, 1991; Oaksford and Chater, 1999) rational approach, Movellan and Nelson's (2001) "probabilistic functionalism," and McKenzie's (2003) view that normative models should be treated as descriptive, rather than prescriptive, theories of cognition. Researchers taking this approach tend to belief that human cognition is well adapted, even optimal, with respect to the statistics of natural environments and the tasks that people face. The main contrasting viewpoint is the heuristics and biases view (Tversky & Kahneman, 1974; Gilovich, Griffin, & Kahneman, 2002; Baron, 2004). This view holds that human cognition fails when heuristic strategies that people use do not match the needs of a particular task. Researchers in the heuristics and biases tradition are more likely to discuss how psychologists might be able to help people improve their judgment on particular tasks, usually by explaining how a particular heuristic (mental shortcut) fails in one or another situation to help people achieve their own goals. As Baron (2004) put it, "if we can help people, then the failure to do so is a harm." Of course, no short summary can be entirely fair to anyone's perspective. The main purpose of the note above is to provide the reader with references to articles in which several researchers articulate their own perspectives, on what role theoretically optimal statistical models should have in the study of human cognition.

Stapler dissertation

This is a "stapler" dissertation, a compilation of articles that are relatively self-contained. This introductory chapter is designed to serve as a sort of glue that places the rest of the work in context, previews the other chapters, and discusses areas for future research. Each subsequent chapter of this dissertation is based on articles that have been published or accepted for publication. Pages have been shrunk and repaginated to fit formatting requirements, but the articles have not been altered in any meaningful way. There are two exceptions to this, in each case involving an article that has been accepted but not yet published. Chapter 2, "Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain," is based on the accepted text of the article that will appear in the October, 2005, issue of Psychological Review. The reader is advised to obtain and read the journal format of that article, rather than the version that appears in this dissertation, because (1) the formatting of the published article is more readable, and (2) additional references, to late-breaking work by other researchers, were added during the copyediting process. Chapter 4, “A probabilistic model of eye movements in concept formation,” is based on an article, coauthored with Gary Cottrell, that has been accepted at Neurocomputing. As of the writing of this dissertation, that article has not been through the copyediting process. Accordingly, the text of that article that appears in this dissertation is preliminary. The reader may consult http://www.jonathandnelson.com/ for electronic versions of the chapters in this dissertation. The content of each chapter of the dissertation is introduced below.

Chapters of the Dissertation

Chapter 2 reviews the history of work with human subjects using optimal experimental design principles to understand human evidence acquisition. This chapter is formed by the author’s article, “Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain,” that will appear in the October, 2005, issue of the journal Psychological Review. An introduction to that work appears below. It is followed by suggestions for future work in this area.

Several normative models (utility functions) for how people should identify useful experiments have been proposed, notably Bayesian diagnosticity, information gain (mutual information), Kullback-Liebler (Kullback & Liebler, 1951) distance, probability gain (error minimization), and impact (absolute change). Existing results from human subjects were shown to not discriminate between these norms as descriptive models of human behavior. Computational optimization found situations in which information gain, probability gain, and impact strongly contradict Bayesian diagnosticity, and in which diagnosticity's claims are clearly inferior. Results of a new experiment strongly contradicted Bayesian diagnosticity. The other normative models behave similarly; each approximates human behavior well. It is concluded that Bayesian diagnosticity serves no useful purpose as a model of evidence acquisition. The new results, it should be emphasized, support the main theoretically important finding from the earlier empirical research, namely that subjects' evidence acquisition strategies are reasonably well-aligned to optimal experimental design (normative) principles. However, other normative models provide a better explanation of empirical results to date than Bayesian diagnosticity, even in articles where the original authors used Bayesian diagnosticity. The research presented in Chapter 2 suggests that information gain (mutual information), Kullback-Liebler distance, probability gain, and impact are approximately equally adequate to explain existing findings in research on human evidence acquisition.

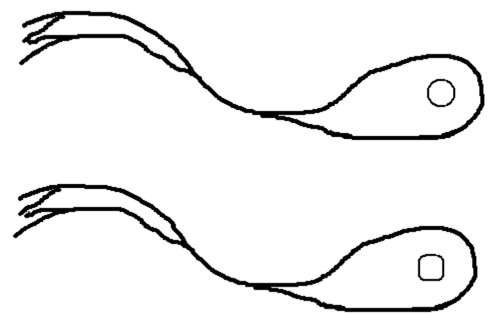

Figure 1. Illustrative plankton stimuli

The examples below differ according to the shape of the “eye” and whether the V-shaped feet are connected. Differences between forms of each feature are subtle, to facilitate a gaze-contingent eye movement task.

The research presented in Chapter 2 suggests a number of important issues to address in the future work. Many of these issues will require stimuli that are learned perceptually, rather with the words and numbers that have typically been used. Consider the “plankton” stimuli at right. There could be two or more species of plankton, and two or more features, such as an eye with different shapes (circle- or square-like), or different types (connected or not) of feet (Figure 1). This type of stimuli will be useful to facilitate comparison of eye movement-based means of information acquisition, in which the subject obtains information by looking at the location of a feature; with mouseclick- based means of information acquisition, in which the subject obtains information by clicking the mouse on a (not-yet-revealed) feature. Use of perceptually-learned stimuli will also enable addressing issues such as whether subjects are sensitive to class-conditional feature dependencies, such as bilateral symmetries. If enough data are collected from each subject, it may be possible to address several issues of interest on a subject-by-subject basis. Examples of these issues include which utility function best approximates each subject’s intuitions; how many steps into the future each subject plans; and how much, if any, softmax behavior is used when picking what questions to ask.

The Active Number Concept Game

Chapter 3 reports work to model an active number concept acquisition task. This work was originally reported as a NIPS paper (Nelson & Movellan, 2001). A further-developed state of the modeling (Nelson, Tenenbaum, & Movellan, 2001), a paper presented at the 2001 conference of the Cognitive Science Society, is reproduced in full as Chapter 3. The active number concept acquisition task is an extension of Tenenbaum's (1999, 2000) number concept game.

In the original task, subjects were given a randomly ordered collection of numbers, for instance 50, 80, 10, and 30, from an unspecified number concept, such as "square numbers" or "numbers between 60 and 90." Subjects were then asked to estimate the probabilities that a variety of numbers, besides those given as examples, were consistent with the true concept. Tenenbaum introduced a probabilistic model of subjects' beliefs on the original number concept task. This model explicitly specifies the prior probability of several number concepts, together with a likelihood function that describes the learner's putative belief that the example numbers were chosen at random from among those numbers that are consistent with a concept (the generative model). Inference, in Tenenbaum's model, is optimal Bayesian (Bayes, 1763) inference with respect to his generative model. An interesting feature of Tenenbaum's model is the large (about 5000) and diverse set of concepts it contains.

The number concept game task contrasts nicely with the relatively simplistic scenarios of early research on the statistical person, and Bayesian reasoning, where a more typical task would be to infer which of two (rather than 5000) urns is responsible for a particular sample of poker chips.

Our active number concept acquisition task is an augmented version of the original number concept acquisition task. In the active task, subjects could pick a number to test, and receive feedback on whether or not that number was consistent with the true underlying concept. We sought to predict subjects' choices of numbers to test, by hypothesizing that subjects picked numbers to maximize the expected information gain (mutual information or Kullback-Liebler distance) with respect to their beliefs about the true concept, as approximated by Tenenbaum's model. An interesting feature of the active number concept task model is the large number of possible queries (each of the integers 1 through 100), compared with the handful of possible queries in most prior research (see Table 1 of Nelson, in press, reprinted here as Chapter 2).

The first attempt to model subjects' queries on this task using optimal experimental design utility functions failed (Nelson & Movellan, 2001). The behavior of the optimal experimental design model, and the human subjects, are described below. Suppose a subject was given the numbers 50, 80, 10, and 30 as an example of an unspecified number concept, and asked to test another number. Tenenbaum's original model, given those example numbers, believes that the most probable concept, given those random example numbers, is “multiples of 10,” and that “multiples of 5” and “even numbers” are the most plausible alternate possibilities. To maximize information gain (the utility function Nelson & Movellan, 2000, used) with respect to Tenenbaum’s model’s beliefs, one should test a number that is inconsistent with “multiples of 10” but consistent with either “even numbers” or “multiples of 5,” such as 25.

A typical subject, however, would test another multiple of 10, such as 20. This query does not differentiate between the working hypothesis and the alternate hypotheses, and, at least relative to Tenenbaum’s original model of belief development, appears to show evidence of “confirmation bias,” or a “positive test strategy” (Klayman, 1995, discusses these terms). In essence, results suggested that subjects consistently tested positive predictions of the model’s working theory, rather than predictions that would help differentiate between the working hypothesis, and alternate hypotheses.

It was not immediately clear whether the failure to predict subjects’ queries resulted from imperfections in the model that Tenenbaum (1999, 2000) used to describe the development of beliefs on the number concept game, or because subjects, in fact, were using a very inefficient sampling strategy on the task. A great deal of research (reviewed by Nelson, in press) suggests that optimal experimental design utility functions can help describe subjects’ intuitions on a variety of tasks. We therefore spent a great deal of time trying to differentiate between the two possibilities. It eventually became clear that although subjects were highly confident that the numbers explicitly presented to them were positive examples of the true concept (e.g. 50, 80, 10, and 30), subjects were not so certain that other numbers that were consistent with the working hypothesis ("multiples of 10"), but not explicitly presented as examples (such as 20), were in fact consistent with the true underlying concept.

We eventually built an augmented concept learning model, in collaboration with Josh Tenenbaum, that better described subjects' generalization behavior on the original (non-active) task than the original model. This model included a number of "foil" concepts, such as "multiples of 10 except the number 20." Inclusion of these concepts in the model enabled the augmented model be more faithful to subjects' beliefs, as measured by their generalization behavior. Relative to the augmented number concept learning model, testing a number such as 25, after the numbers 50, 80, 10, and 30 had been presented, made sense, when queries' usefulness was measured with information gain. Chapter 3 reprints this work (Nelson, Tenenbaum, & Movellan, 2001) in full.

Javier Movellan and I continued to explore this task following that point. That work has not yet been published, so highlights of it are given here. Let the random variable X represent a possible query, such as testing the number 25. X=1 if the number 25 is consistent with the true concept; X=0 otherwise. Let the random variable C represent the learner's beliefs about the true concept, represented as a probability distribution over all possible concepts. We had noticed an apparent relationship between the uncertainty in the outcome of a particular query, as measured with Shannon entropy (Shannon, 1948; Cover & Thomas, 1991), and the information gain of that query, with respect to reducing uncertainty about the true hypothesis. Numeric simulation of the model led us to conjecture that:

$$H(X)=H(C)-H(C|X).$$

In other words, it appeared that the learner's uncertainty about the outcome of their *question*, $H(X)$, was equal to the reduction in uncertainty about the true *concept* (the question's expected information gain), that asking the question was expected to provide. If the above conjecture were true, then it might not be necessary to explicitly model each concept that a subject might have in mind. Rather, it might be possible to ascertain the usefulness of potential questions according to the subject's beliefs about the questions' likely outcomes. Methodologically, it could be much easier to obtain reliable estimates of generalization probabilities for each possible number, and henceforth of the uncertainty in testing each possible query, than to ascertain what hypotheses subjects actually entertain on the task. In fact, our conjecture was correct. By the definition of mutual information (Cover & Thomas, 1991),

$$I(C,X)=H(C)-H(C|X),$$

the expected information gain of the query X. By the symmetry of mutual information, $I(C,X)=I(X,C)$, so

$$H(C)-H(C|X)=H(X)-H(X|C).$$

On the number concept task, if the true concept is known, there is no uncertainty about the answer to any question X, about whether a particular number is consistent with the true concept. In other words, on the number concept task, the conditional entropy in X given C:

$$H(X|C)=0.$$

So our conjecture is proven. On the number concept task, the information gain in a query X, with respect to the unknown true concept C, can be computed by calculating the uncertainty in the outcome of the query X.

In other work on this task, we explored using a “conservative” likelihood function for belief updating. This work suggested that a conservative update rule approximates subjects’ belief development somewhat better than the original likelihood function. We used the type of conservatism discussed by Edwards (1968), in which an exponent somewhat less than one is placed over the likelihood, P(datum|hypothesis), the probability of obtaining a particular outcome given a particular hypothesis, or concept. In future work modeling the active number concept task it could prove useful to explore the behavior of the conservative likelihood functions used by Oaksford and Chater (1998) and Nelson and Cottrell (accepted), which are discussed in Chapter 4. Another issue we began to explore is whether heuristic sampling strategies, such as the positive test strategy, could approximate optimal strategies, if both strategies were constrained to use softmax rules to pick what questions to ask. In future work it will also be important to evaluate the behavior of other optimal experimental design utilities for evidence acquisition besides information gain, such as impact and probability gain, on this task.

Eye Movements as Experimental Design

Chapter 4 introduces a principled probabilistic model that describes the development of beliefs on a classic (Shepard, Hovland, & Jenkins, 1961) concept formation task. That learning model is used, together with an optimal experimental design-inspired sampling function, to help explain eye movements on Rehder and Hoffman's (2003, 2005) eye movement concept formation task. Results show that same rational sampling function can help predict eye movements early in learning, when uncertainty is high, as well as late in learning when the learner is nearly certain of the true category. Chapter 4 reproduces the complete text of the article, "A probabilistic model of eye movements in concept formation," coauthored with Gary Cottrell, that has been conditionally accepted at Neurocomputing. The main note to make presently is that this article illustrates how aspects of eye movement, itself a quintessentially perceptual task, may be modeled in the same way as other intuitive experimental design situations. The visual system's choice of where to direct the eyes' gaze may be fundamentally similar to a subject's choice of what question to ask on another task.

This article represents early stages of work in a new domain. There are several areas we hope to explore in the future. One of these is use of Boolean complexity as a source of prior probabilities. Another area is to explore the reasons why probability matching (or a similar, apparently suboptimal response function) is necessary for the model learner's error rate to qualitatively approximate human error rates. Finally, we intend to explore the possible use of theoretically optimal, rather than simply intuitively plausible, utility functions, perhaps derived with use of dynamic programming, in the eye movement model.

Probabilistic Functionalism

Chapter 5 reproduces, in full, a comment in Behavioral and Brain Sciences, in which Javier Movellan was primary author and I was coauthor (Movellan & Nelson, 2001). This comment was written in response to Tenenbaum and Griffiths' (2001) article. It does not discuss information acquisition, but rather addresses the theoretical bases of our approach to the study of cognition, which we termed "probabilistic functionalism." This chapter can be viewed as providing theoretical background on our rationale for using optimal statistical theories to describe human subjects' cognition and behavior.

What a Speaker's Choice of Frame Reveals

Chapter 6 reproduces, in full, an article in Psychonomic Bulletin and Review, in which Craig McKenzie was primary author and I was coauthor (McKenzie &

Nelson, 2003). This article, which is not directly tied to modeling information acquisition as optimal experimental design, concerns the circumstances under which speakers and listeners adopt particular "frames" in communication. The article provides experimental evidence that when deciding how to phrase a particular situation, speakers do not choose at random from available logically equivalent frames. The article also provides evidence that when perceiving communication, listeners have intuitions about when speakers choose particular frames. The article shows that listeners' intuitions about speakers' choices of utterances, and speakers' actual choices of utterances, are in good correspondence. For instance, the article results show that if a speaker describes a glass as "half empty," rather than as "half full," a listener could reasonably infer that the glass has lost water; and that listeners, indeed, tend to make this justified inference. Future research could explicitly model listeners' and speakers' sense of the usefulness of different frames. For instance, people could be explicitly asked whether the statement "the glass is half empty" is more or less useful than the statement "the glass is half full," with respect to inferring the previous state of the glass. Optimal experimental design utilities for evidence acquisition, such as impact, probability gain, and information gain, could then be compared to human judgments in this domain.

Conclusions

Chapter 2 provides evidence that several principled utility functions for evidence acquisition—information gain, Kullback-Liebler distance, impact, and probability gain—may serve to explain human intuitions about the usefulness of different pieces of evidence, on several tasks. Chapter 3 extends the optimal experimental design account of human evidence acquisition to an active number concept acquisition task. Chapter 4 presents a new probabilistic model of belief development on Shepard, Hovland, and Jenkins's (1961) concept formation task. Chapter 4 further shows how this explicit model of beliefs can serve to predict eye movements on Rehder and Hoffman's (2005) eye movement version of Shepard et al.'s task. Chapter 5 discusses meta-theoretical issues pertaining to probabilistic functionalism, or the rational analysis of cognition. Chapter 6 provides evidence supporting the belief that people's perceptions are well-calibrated to their environments, as well as areas that could in the future be studied explicitly within an optimal experimental design framework.

Together, the investigations reported in this dissertation help to clarify the theoretical foundation, and strengthen the empirical basis, of the theory that human evidence acquisition can be modeled as an optimal experimental design problem, in which people's intuitions are in good agreement with well-motivated statistical principles.

Sources Cited

Anderson, J. R. (1990). The Adaptive Character of Thought. Hillsdale NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. Psychological Review, 98(3), 409-429.

Baron, J. (1981). An analysis of confirmation bias. Paper presented at 1981 Psychonomic Society meeting.

Baron, J. (1985). Rationality and Intelligence. Cambridge: Cambridge University Press.

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making, pp. 19-36. London: Blackwell.

Bassok, M., & Trope, Y. (1983-1984). People's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? Social Cognition, 2(3), 199-216.

Box, G., & Hill, W. (1967). Discrimination among mechanistic models. Technometrics, 9, 57-71.

Brunswik, E. (1952). The conceptual framework of psychology. Chicago: University of Chicago Press

Chater, N.; Oaksford, M. (1999). Ten years of the rational analysis of cognition. Trends in Cognitive Sciences, 3, 56-65.

Cover, T. M., & Thomas, J. A. (1991). Elements of Information Theory. New York: Wiley.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal Representation of Human Judgment, (pp. 17-52). New York: John Wiley.

Fedorov, V. V. (1972). Theory of Optimal Experiments. New York: Academic Press.

Gilovich, T., Griffin, D. & Kahneman, D. (2002). Heuristics and Biases: the Psychology of Intuitive Judgment. Cambridge, UK: Cambridge University Press.

Good, I. J. (1950). Probability and the Weighing of Evidence. New York: Charles Griffin.

Klayman, J. (1995). Varieties of confirmation bias. Psychology of Learning and Motivation, 42, 385-418.

Kullback, S., & Liebler, R. A. (1951). Information and sufficiency. Annals of Mathematical Statistics, 22, 79-86.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. Annals of Mathematical Statistics, 27, 986-1005.

Marr, D. C. (1982). Vision: A Computation Investigation into the Human Representational System and Processing of Visual Information. San Francisco: Freeman.

McKenzie, C. R. M. (2003). Rational models as theories—not standards—of behavior. Trends in Cognitive Sciences, 7(9), 403-406.

McKenzie, C. R. M. & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. Psychonomic Bulletin and Review, 10(3), 596-602.

Movellan, J. R., & Nelson, J. D. (2001). Probabilistic functionalism: a unifying paradigm for the cognitive sciences. Behavioral and Brain Sciences, 24, 690-692.

Nelson, J. D. (in press). Finding useful questions: on Bayesian diagnosticity, probability, impact and information gain. Psychological Review.

Nelson, J. D., & Cottrell, G. W. (accepted). A probabilistic model of eye movements in concept formation. Neurocomputing.

Nelson, J. D., & Movellan, J. R. (2001). Active inference in concept learning. In T. K. Leen, T. G. Dietterich & V. Tresp (Eds.), Advances in Neural Information Processing Systems, 13, 45-51. Cambridge, MA: MIT Press.

Nelson, J. D., Tenenbaum, J. B., & Movellan, J. R. (2001). Active inference in concept learning. In J. D. Moore & K. Stenning (Eds.), Proceedings of the 23rd Conference of the Cognitive Science Society, 692-697.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631.

Oaksford, M., & Chater, N. (1998). A revised rational analysis of the selection task: exceptions and sequential sampling. In Oaksford, M., & Chater, N. (Eds.), Rational Models of Cognition (pp. 372-393). Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. Psychonomic Bulletin & Review, 10(2), 289-318.

Peterson, C. R., Beach, L. R. (1967). Man as an intuitive statistician. Psychological

Bulletin, 68(1), 29-46.

Popper, K. R. (1959). The Logic of Scientific Discovery. London: Hutchinson & Co.

Rehder, B., & Hoffman, A. B. (2003). Eyetracking and selective attention in category learning. In R. Alternman & D. Kirsh (Eds.), Proceedings of the 25th Annual Conference of the Cognitive Science Society. Boston, MA: Cognitive Science Society.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. Cognitive Psychology, 51, 1-41.

Savage, L. J. (1954). The Foundations of Statistics. New York: Wiley.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423, 623–656.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. Psychological Monographs: General and Applied, 75(13), 1-42.

Skov, R. B. & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. Journal of Experimental Social Psychology, 22(2), 93-121.

Slowiaczek, L. M., Klayman, J, Sherman, S. J. & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? Memory and Cognition, 20(4), 392-405.

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 S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), Advances in Neural Information Processing Systems, 12, 59–65. Cambridge, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. Behavioral and Brain Sciences, 24(4), 629-640.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. Journal of Personality and Social Psychology, 43, 22-34.

Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis testing. Journal of Experimental and Social Psychology, 19, 560-576.

Tversky, A. & Kahneman, D. (1974, September 27). Judgment under uncertainty: heuristics and biases. Science, 185(4157), 1124-1131.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 12, 129-140.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), New Horizons in Psychology (pp. 135-151). Harmondsworth, England: Penguin.

Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology. 20(3), 273-281.