

Effect of Belief Bias on the Development of Undergraduate Students' Reasoning about Inference

Jennifer J. Kaplan
Michigan State University

Journal of Statistics Education Volume 17, Number 1 (2009), www.amstat.org/publications/jse/v17n1/kaplan.html

Copyright © 2009 by Jennifer J. Kaplan, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Statistics education; Statistical reasoning; Fundamental computational bias; Heuristics and biases.

Abstract

Psychologists have discovered a phenomenon called "Belief Bias" in which subjects rate the strength of arguments based on the believability of the conclusions. This paper reports the results of a small qualitative pilot study of undergraduate students who had previously taken an algebra-based introduction to statistics class. The subjects in this study exhibited a form of Belief Bias when reasoning about statistical inference. In particular, the subjects in this study were more likely to question the experimental design of a study when they did not believe the conclusions reached by the study. While these results are based on a small sample, if replicated, the results have implications for the teaching of statistics. Specifically, when teaching hypothesis testing, statistics instructors should be mindful about the context of example problems used in class, make explicit links between inference to experimental design and actively engage students in discussions of both believability of conclusions and the types of arguments they find convincing.

1. Introduction

Researchers in the field of psychology have been conducting studies designed to test the decision-making skills of humans since Peter Wason's work in the early 1960's ([Evans and Newstead, 1995](#)). In the early 1970's Tversky and Kahneman began to publish what is now known as the "heuristics and biases" literature (for example, see [Kahneman & Tversky, 1982](#) or [Tversky & Kahneman, 1974](#)). In the decades since Wason and Kahneman and Tversky's original work, many studies designed to understand the process of human reasoning have been conducted. While researchers who have followed in the wake of Wason and Kahneman and Tversky have disputed the original findings ([Hertwig and Gigerenzer, 1999](#)), the fact remains that subjects are prone to making errors in judgment on the designed tasks ([Stanovich, 1999](#)). The statistics education community can benefit from the results obtained by psychologists in studies of human rationality. The results, studied through the lens of statistics education, will inform statistics teaching because the irrationalities discovered by psychologists may represent misconceptions held by statistics students. This paper explores whether the psychological construct called "Belief Bias" is exhibited by beginning statistics students. While this work is a small pilot study, the recent report *Using Statistics Effectively in Mathematics Education* ([American Statistical Association, 2007](#)) suggests that such research should be published to expand the research base in statistics education. The work presented here can best be classified in the "generate" category of studies described in the report. It is an idea, generated from the literature in psychology, which could be the basis for a larger study of statistical reasoning.

2. Background

[Wild & Pfannkuch \(1999\)](#) proposed a framework for statistical thinking in empirical enquiry that was developed through in-depth interviews with Ph.D. level statistics students and professional statisticians. One aspect of statistical thinking they list is the ability to integrate statistical information with contextual information. This is the process by which a statistician decides how to appropriately use prior knowledge of the context of the inquiry situation with statistical knowledge and understanding ([Wild and Pfannkuch, 1999](#)). As the inquiry progresses, the statistician will move between the spheres of contextual and statistical knowledge, choosing processes, revising hypotheses, and forming conclusions in a cycle of inquiry. Research in psychology has shown that subjects tend to have difficulty correctly integrating contextual information when performing reasoning tasks. In fact, psychologists have identified a general class of biases, called Fundamental Computation Bias (FCB), that result from inexpert incorporation of contextual data ([Sloman, 1996](#)). The general description of FCB is beyond the scope of this work. This work is based on one such bias, "Belief Bias," which is the predisposition of subjects to be influenced by beliefs that are not logically relevant to the given task ([Evans, Brooks, and Pollard, 1985](#)).

Belief Bias arises in argument validation studies and can be traced to subjects' over reliance on

contextual information and a lack of use of statistical information. [Evans, Barston and Pollard \(1983\)](#) asked subjects to decide whether a given conclusion could be logically deduced from the given premises. There were four argument types: 1) valid with believable conclusion, 2) valid with unbelievable conclusion, 3) invalid with believable conclusion, and 4) invalid with unbelievable conclusion. The study found main effects for validity and believability: subjects were more likely to endorse a conclusion when the reasoning was valid rather than invalid and when the conclusion was believable rather than unbelievable. This latter effect exemplifies Belief Bias. There was also a significant interaction between the two factors in that subjects were much less likely to endorse the unbelievable conclusion in the invalid reasoning case than in the valid reasoning case. [Thompson \(1996\)](#) also found that believability of premises affects subjects' rating of strength of argument.

Furthermore, in an empirical study of graduate students in science, [Koehler \(1993\)](#) found that his subjects tended to give more favorable ratings to research reports that made conclusions with which they agreed (called "belief-congruent" findings). The differences between the judgments were not based on harsh critiques of "belief-incongruent" results. This led Koehler to conclude that scientists are more likely to assume that a study with "belief-congruent" findings has been correctly conducted, but when study results are "belief-incongruent," scientists' skepticism is activated. A follow-up study of practicing research psychologists and scientists provided "further evidence for an agreement effect in...judgments of evidence quality" ([Koehler, 1993, pp. 46](#)) by experts.

In summary, the literature on Belief Bias in reasoning indicates that both novice and relative expert subjects' ratings of strength of arguments are affected both by the believability of conclusions and the believability of the premises. If, as Belief Bias predicts, the believability of the conclusion of an argument affects the reader's judgment of the argument, then Belief Bias could present a problem in the development of adaptive reasoning in statistics. For the purposes of this paper, adaptive reasoning in statistics is defined as knowing what can be inferred from data or statistical results and whether a conclusion is valid ([Kaplan, 2006](#)). This definition is based on delMas' description of statistical reasoning: "a person who can explain why results were produced or why a conclusion is justified demonstrates statistical reasoning" ([delMas, 2004, pg. 85](#)) and the National Science Education Standards (NSES) description of reasoning as being able to judge the quality of evidence and the validity of arguments and conclusions using communally understood standards for evidence ([NRC, 1996](#) and [NRC, 2003](#)). Adaptive reasoning in statistics, therefore, includes the ability to make statistically valid conclusions, critique statistical arguments, and discuss the scope of the conclusions generated. Reasoning is about knowing what can be inferred and what questions remain unanswered. The aim of this study is to illustrate some of the possible effects of belief bias on undergraduate learning of statistical reasoning.

3. The Study

3.1 Research Question

The primary research question addressed in this paper is: What role, if any, does Belief Bias play in the undergraduate students' reasoning about statistical inference? The study used open-ended interviews to assess differences in the development of reasoning about hypothesis testing by beginning statistics students.

3.2 Participants

Participants were selected from among all the students in the Introduction to Statistics courses offered by the department of mathematics at a large research institution in the U.S. Southwest. All of the participants had demonstrated procedural fluency and conceptual understanding of hypothesis testing on the final exam in the course. Procedural fluency, here, is the ability to produce a correct hypothesis test and conceptual understanding was tested through tasks such as telling the direction of change of a p-value based on changes in the sample size or the value associated with the null hypothesis ([Kaplan, 2006](#)).

Ten students participated in the study, 8 women and 2 men. There was one freshman in the sample and three students each who were sophomores, juniors and seniors, classified by credit hours. The students were all of a traditional age, ranging from 18 to 21 with a modal age of 19 (four subjects). Nine students reported having a G.P.A. above 3.7, and the tenth subject reported a G.P.A. of 2.89. Nine of the ten subjects reported earning a grade of A in the introduction to statistics course; the tenth subject had earned a C. The subject who earned a C and the subject with the low G.P.A. were not the same subject. Five of the subjects were in fields related to medicine: 2 in nursing, 2 in pre-pharmacy and 1 pre-med. There were two advertising students, and one each in public relations, psychology and Chinese.

3.3 Protocol

The results discussed in this paper were based on three tasks that were part of a larger protocol. The students were interviewed about three scenarios of varying degrees of believability. Each task described an experimental study, gave statistical conclusions including p-values and interpreted the results of the hypothesis test within the context of the study. Given the importance of pre-existing opinions to the enactment of belief bias, the students were questioned over the subject matter of the scenarios, prior to being presented the scenarios. This was done to assess the student's preexisting opinions within the domain of the tasks. The three scenarios were: 1. The Email from Dad, 2. The ESP study and 3. The Lyon Diet Heart Health Study.

In the Email from Dad task students responded, first in writing and later using interview questions to an email written by a father about a discussion with a doctor about possible medication that the grandmother might try for heart disease. The email included statistically significant results of a clinical trial of a new medication and asked the student to clarify the meaning of what the doctor said. In addition, the email provided anecdotal evidence against the use of the medication and asked the student to make a judgment on the best course of action for Grandma. The task was intended to have moderate believability because the study described is of the type that is actually done, but the number of subjects

included might appear low to students being interviewed. This task was modified from [Jordan \(2004\)](#).

In the ESP study, students responded to a summary of a journal article that reported the results of a paranormal psychology experiment ([Bem and Honorton, 1994](#)) about the existence of Extra-sensory perception (ESP). The summary included a description of the experimental design of the ESP study and the results, which showed a statistically significant effect for ESP when subjects were focusing on dynamic pictures (short movie clips) and no effect for ESP when subjects were focusing on static pictures. This task was intended to have low believability because: 1. it is not a topic that people would think might be studied statistically and 2. people generally have an opinion about the existence or lack thereof of ESP.

In the Lyon Diet Heart Health study, students responded to the results of a study reported on the website of the American Heart Association ([American Heart Association, 2005](#)), in which subjects were randomly assigned to two diet groups. The subjects in first group were told to eat a healthy diet, but were not given specific menus or meals. The subjects in the second group followed a prescribed diet, described in the article as heart healthy and low in fat. The subjects on the prescribed diet were statistically significantly less likely to die of cancer or heart disease than those in the control group. This task was intended to have high believability because it is a study that students who were interviewed may have encountered. It has large sample sizes and reasonable conclusions; the diet that is found to be healthier has facets that should appeal to the students being interviewed.

While suggestions for interview questions were generated prior to conducting the interviews, the intent was to have guided, rather than scripted, interviews so that the interviewer had the leeway to explore further the thoughts and understandings of the subjects.

3.4 Procedure

All interviews were conducted in a conference room in the mathematics building on campus. The subjects completed the "Email from Dad" writing task first. The order of the other two tasks was determined at random. The interviews began with the preliminary questions designed to assess the prior beliefs associated with the Email from Dad task. After this, the researcher role-played being the father, calling to follow up on the email. For the other two tasks, the preliminary questions were asked before the subject read the article. After the subject finished the article, the protocol of questions generated prior to the enactment of the study was used. When the interview was over, the subject was debriefed. Subjects were paid \$20 for their time and the recordings were stopped after they left the room. The interviews lasted 50 – 70 minutes with most of them finishing in just over an hour.

3.5 Data Analysis

Using the audio recordings and transcripts, a summary was created for each subject. The summaries were used to find common themes across the interviews. The categories that had been identified through the open and axial coding of the Email from Dad writing samples that had been collected in a previous

study were used to identify relevant features of the interviews. In particular to this research each summary contained a description of the subject's discussion of experimental design. Further, particular care has been taken to include all types of evidence cited by the participant in each task.

4. Results

4.1 Reactions to unbelievable conclusions

Nearly every student in this study found the results of the ESP Study to be unbelievable, while none of the students questioned the results of the other two studies. There were two statistical results in the ESP study: a statistically significant effect for ESP when subjects were focusing on dynamic pictures (short movie clips) and no statistically significant effect for ESP when subjects were focusing on static pictures. The students who did not believe in ESP were surprised by the results for the dynamic pictures, which provided evidence to support the existence of ESP, and the students who believed in ESP were surprised by the results for the static pictures, which provided evidence to contradict the existence of ESP. In contrast, all of the students responded to the Email from Dad task by suggesting that Grandma stay on the new medication and indicating that the statistical results were strong and more useful for making decisions than the anecdotal evidence. For the students in this study, the context of the ESP task led to different responses than did tasks with contexts that led to believable conclusions. There were three main themes to the reactions subjects had when faced with an unbelievable conclusion: 1) critique the design of the experiment, 2) request more information, and 3) consider a rational explanation for the study.

There was a general propensity of subjects to look for problems in experimental design when the conclusions were unbelievable. This response to unbelievable conclusions was unexpected given that none of the subjects spontaneously discussed experimental design in the other two tasks. Hannah and Charlotte, both of whom believed in ESP, also believed that ESP could not be studied scientifically. They both indicated that the use of novices as subjects might have "diluted" the study results. According to Hannah and Charlotte's reasoning, novice receivers would have been guessing randomly. If there were enough novice receivers, their responses, which would have been guesses not based on true reception, might account for the non-significant results in the case of static targets.

The main critique of the study given by those who did not believe in ESP was that there were only four targets from which the receivers chose. Megan suggested that there should have been 30 targets. Megan's suggestion of 30 targets would lead to the following experimental design. First, subjects guessing randomly would be expected to be correct on roughly 3% of the trials. The minimum number of trials to meet the conditions for inference with 30 targets would be 450. In a study with 450 trials and 30 targets, the p-value reported in the paper would result if 6% (28 out of 450) of the trials were successful. Some subjects suggested that a higher percentage of successes would be necessary to convince them of the existence of ESP. The percentages they would have found convincing ranged from 45% to 95%. This suggests that it is unlikely that significant results from a study with 30 targets would be perceived as having more value in general. In addition, Sarah suggested that some of the films might

have been more eye catching and that would have biased the results. Finally, both Kyle and Bradford specifically mentioned that they could not find any flaws in the study.

The second type of response given by subjects in reaction to surprising results was a request for further information. The most common request for more information was to query about the existence of replication studies. In addition to that, Kyle, Natalie and Bradford asked clarification questions about the reported study. For example, regarding the ESP Study, Kyle asked how the researchers defined success and Natalie asked whether the senders and receivers knew each other. One unusual request came from Bradford who said that he would like to read critiques of the study. It is not clear whether he would have asked for this material if the results had been believable, but he seemed to have a very clear notion of the importance of counter-arguments in evaluating the evidence presented.

The third reaction to the unbelievable results was to search for a rational explanation for them. Bradford expressed concern about the ESP Study results because he could not think of a rational explanation for them. Four of the other subjects posited explanations that explain or negate the surprising result. Jewel said that she had heard that humans only use 18% of their brain so maybe ESP happens when a person can use more. Hannah also referenced brain function, suggesting that the difference in the results between static and dynamic targets might be the result of differences in the areas of the brain that process the two types of stimuli. Sarah noted that the static pictures should have been easier to choose correctly because they represent only one thing to be thinking of, while the movies are many images on which to be concentrating. Sarah also said that the results could just represent a good guessing day. While the students in this study did not, in general, question the findings of the other two studies, Kefira was shocked by the number of subjects in the Lyon Diet Study subjects who died during the study. Similar to the students who gave a rational explanation for the unexpected results in the ESP study, she explained the deaths in Lyon Diet Study by suggesting that the subjects were relatively old. In that case, she felt that the deaths would represent a more reasonable number.

4.2 Evidentiary Basis

In general, the subjects mentioned three types of evidence as being convincing: 1) statistical results, 2) a preponderance of evidence, and 3) a justification or rationalization. When the conclusion suggested by statistical evidence was incongruent to a prior belief, however, subjects tended to be less convinced by the statistics. In this case, subjects tended to search for a justification for the conclusion or rely on their pre-existing opinion. These opinions tended to be based on a preponderance of evidence. It should be noted that reliance on prior evidence is a reasonable reaction from a Bayesian perspective if the prior knowledge has an evidentiary basis.

5. Reflections and Discussion

One aspect of adaptive reasoning in statistics is that students should be able to discuss strengths and weaknesses of an experiment and comment on how the experimental design impacts the strength of conclusions that can be made ([Kaplan, 2006](#)). The results presented in the previous section suggest

mastery of this skill by beginning statistics students may be weak. In particular, they tend to spontaneously discuss experimental design features only when the conclusion presented is unexpected or counter to a previously held opinion. The difficulties exhibited by the subjects with regard to experimental design are disappointing. This is particularly true given [Alacaci's \(2004\)](#) claim that knowledge of experimental design "constitutes the backbone" of an expert knowledge of statistical techniques. Alacaci (2004), further, suggests that the teaching of inferential statistics should be linked explicitly to experimental design in order for novices to develop a knowledge base from which they will be better able to become experts ([Alacaci, 2004](#)).

The findings with regard to experimental design, while perhaps not surprising, present the following challenge for statistics educators: How can we encourage students to act with the same skepticism about experimental design whether the results of the study give evidence that either confirms or is in conflict with the opinion held by a student without instilling in them the sense that all statistical results are based on untruths or explicit attempts to bias the results? [Alacaci \(2004\)](#) gives four suggestions for improving the teaching of experimental design as part of the curriculum on inferential statistics: 1) make associations between statistical techniques and design aspects; 2) use think aloud techniques while completing examples of inference that include experimental design issues in the out-loud-thinking; 3) use examples of actual research in teaching inference; and 4) construct scenarios for teaching examples in which there are experimental design issues that will provide valuable class discussion.

The relative lack of understanding of experimental design shown by the subjects may be partially attributed to the treatment of experimental design by the textbook the subjects used in their statistics course. Experimental design is covered in two chapters of the textbook. The problems in the textbook designed to assess understanding of experimental design are largely exercises in using the random digits table to select a sample or in drawing an outline of the experimental design. The chapters on inference, which might include the strength of conclusions that may be made based on the design features of a study or experiment, are not covered until much later in the course. In the discussions of conditions for inference, attention is paid to the selection of a simple random sample, the size of the sample and conditions about the distribution of the sample. Ideas about design issues, however, are not incorporated into the discussions about conclusions that can be made based on inference.

Current research in psychology on people's criteria for justifying causal claims predicts that people tend to value theoretical explanations over evidence ([Kuhn, 2001](#)). [Kuhn \(2001\)](#) states that the literature suggests that people "depend on explanations that allow their claims to make sense to themselves and others" (pg. 1). The tendency of the subjects to want an explanation for the unexpected outcomes is consistent with Kuhn's findings. Kuhn further suggests that recent research indicates that the "preference for explanation over evidence is dependent on context and on the strength of the evidence" (ibid) and that it diminishes developmentally, disappearing in highly capable undergraduate students. The data provided by the subjects in this study suggest that even high functioning undergraduates rely on explanation over evidence in the justification of claims when the claims are counter to a previous belief.

[Cobb \(1992\)](#) specified as a goal for introductory statistics classes the development of statistical thinking

and mentions, in particular, "recognizing the need to base personal decisions on evidence (data), and the dangers inherent in action on assumptions not supported by evidence" ([GAISE, pg. 3](#)). The comments made by the subjects when discussing unexpected findings indicate that these students have not met the goal of recognizing the value of evidence in this case. The challenge for statistics educators, therefore, is to help students develop an appreciation for statistical evidence even when it does not support a pre-existing belief.

This finding suggests some specific suggestions for instruction, including the writing of textbooks. If all classroom examples and homework problems in a statistics course are about contexts with which students are unfamiliar or lead to unexpected results, students may not accept the statistical process and reasoning as valid. Statistics instructors and textbook authors should consider the contexts of the examples and problems they choose to include in a course. Students should be familiar with the contexts of the problems. In addition, instruction should include some examples where the conclusion is entirely believable. Instructors should actively engage students in discussions of both believability of conclusions and the types of arguments they find convincing.

The findings of this study have led me to make some specific changes in the way I teach the algebra-based introductory statistics course. First, I begin the semester with the chapters on data collection and experimental design. This way, issues of experimental design and the differences between conclusions that can be made from surveys, observational studies and experiments are an overarching theme that recurs the entirety of the course. I have incorporated the ESP study as an example in my course for one-proportion hypothesis tests. While I teach a large class, I use personal response systems to elicit the students' prior knowledge about ESP and guide the discussion of the results of the study accordingly. The other area in which my teaching has been informed, is in the teaching of the idea that "association is not causation ([GAISE, 2006](#))." I try to ensure that I use examples in which a causal relationship between the two variables is plausible, such as years of schooling and income, and in which a causal relationship does not seem plausible, such as number of video game systems per capita and life expectancy. In this way, I attempt to help my students to focus on the design issues that allow us to conclude that a relationship may be causal rather than the plausibility of the causal relationship. I have not yet collected formal data about the effect of these practices on the understanding my students' develop, but these issues are ripe for future research.

The above conclusions are affected by certain limitations and delimitations of the research. The design of the study was intended to reduce sources of variation. By studying only students who had successfully mastered procedural fluency and conceptual understanding, the research would focus only on differences in other aspects of statistical proficiency. Unfortunately, this design limited the size of the sample. This is a delimitation of the study because it limits the generalizability of the conclusions. In this case, the findings from the interviews may not represent all of the types of understandings and misconceptions that are developed by beginning statistics students, even those students who are like the students in the sample. In addition, the transcripts of the interviews were only coded by one researcher, so there is no inter-rater reliability. The results, however, coupled with the findings in psychology and the ideas from statistics education about the goals for student learning, suggest that future research in this area is

necessary. In order to address the learning goal of developing students' statistical reasoning skills, the statistics education community should attend to factors, like Belief Bias, that affect how students make decisions, reason with data and react to statistical conclusions.

Acknowledgements

This research is based on a dissertation study by the author at The University of Texas at Austin, under the guidance of Dr. Uri Treisman. The author gratefully acknowledges his mentorship and support. In addition, the author would like to thank Yvette Nicole Johnson, Ed Corcoran, Bill Notz and the anonymous reviewer for their advice and comments regarding this paper.

References

- Alacaci, C. (2004). "Inferential statistics: Understanding expert knowledge and its implications for statistics education." *Journal of Statistics Education*, 12 (2), <http://www.amstat.org/publications/jse/v12n2/alacaci.html>
- American Heart Association (2005). "Lyon Diet Heart Study." [Online], (<http://www.americanheart.org>)
- American Statistical Association (2007). "Using Statistics Effectively in Mathematics Education Research." [Online], www.amstat.org/research_grants/pdfs/SMERReport.pdf
- Bem, D. and Honorton, C. (1994). "Does psi exist? Replicable evidence for an anomalous process of information transfer," *Psychological Bulletin*, 115 (1), 4 – 18.
- Cobb, G. (1992). "Teaching Statistics," In *Heeding the Call for Change: Suggestions for Curricular Action*, L.Steen (ed.). Washington, D.C: Mathematical Association of America, 3 – 43.
- delMas, R. (2004). "A comparison of mathematical and statistical reasoning," In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, Ben-Zvi, D. and Garfield, J. (eds.). The Netherlands: Kluwer Academic Publishers, 79 – 95.
- Evans, J. St. B. T. , Barston, J. L. & Pollard, P. (1983). "On the conflict between logic and belief in syllogistic reasoning," *Memory & Cognition*, 11, 295 – 306.
- Evans, J. St B. T. , Brooks, P.G. , and Pollard, P. (1985). "Prior beliefs and statistical inference," *British*

Journal of Psychology, 76, 469 – 477.

Evans, J. St. B. T. and Newstead, S. (1995). "Creating a psychology of reasoning: The contribution of Peter Wason," In *Perspectives On Thinking and Reasoning: Essays in Honour of Peter Wason*. Evans, J. St. B. T. and Newstead, S., (eds.). East Sussex, UK: Lawrence Erlbaum Associates, Ltd.

GAISE (2006). "College Report of the Guidelines for Assessment and Instruction in Statistics Education Project," [Online], <http://www.amstat.org/education/gaise/GAISECollege.htm>

Hertwig, R. and Gigerenzer, G. (1999). "The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors," *Journal of Behavioral Decision Making*, 12 (4), 275 – 305.

Jordan, J. (2004). "The use of writing as both a learning and an assessment tool," Presentation at ARTIST Roundtable Conference, [Online]. Lawrence University, Appleton, WI, <http://www.rossmanchance.com/artist/Proctoc.html>

Kahneman, D. and Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11, 123 – 141.

Kaplan, J.J. (2006). Factors in Statistics Learning: Developing a dispositional attribution model to describe differences in the development of statistical proficiency. Unpublished Doctoral Dissertation. [Online], <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/06.Kaplan.Dissertation.pdf>

Koehler, J. (1993). "The influence of prior beliefs on scientific judgments of evidence quality," *Organizational Behavior and Human Decision Processes*, 56, 28 – 55.

Kuhn, D. (2001). "How do people know?" *Psychological Science*, 12 (1), 1 – 8.

National Research Council, (2003). *Learning and Instruction: A SERP Research Agenda*, [Online] eds. Donovan, M. S. and Pellegrino, J. W., Washington, D.C.: National Academy Press, <http://www.nap.edu/openbook/0309090814/html/index.html>

National Research Council, (1996). *National Science Education Standards: National Committee on Science Education Standards and Assessment* [Online], Washington, D.C. : National Academy Press, <http://books.nap.edu/catalog/4962.html>

Ramsey, F. and Schafer, D. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis, second edition*, U.S.A. : Thomson Learning, Inc.

Sloman, S. (1996). "The empirical case for two systems of reasoning," *Psychological Bulletin*, 119 (1), 3 – 22.

Stanovich, K. (1999). *Who Is Rational: Studies of Individual Differences in Reasoning*, Mahwah, N. J. : Lawrence Erlbaum Associates, Inc.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185 (4157), 1124-1131.

Wild, C.J. and Pfannkuch, M. (1999). "Statistical thinking in empirical enquiry," *International Statistical Review*, 67(3), 223 – 265.

Jennifer J. Kaplan, Ph.D.
Michigan State University
Department of Statistics and Probability
A413 Wells Hall
East Lansing, MI 48824
kaplan@stt.msu.edu

[Volume 17 \(2009\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)