Ongoing Discussion "Thought Piece"

Prepared by

Gipsie Ranney

gbranney@comcast.net

January 2012

Ongoing Discussion Host: Bill Bellows

(william.bellows@pwr.utc.com)

for

Pratt & Whitney Rocketdyne's

InThinking Network

Blind Spots

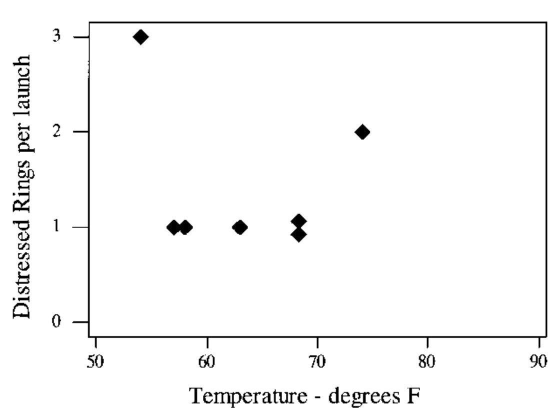
Gipsie B. Ranney

We all face an onslaught of information daily. We use some of that information to learn and make inferences. There are potential blind spots that it helps to know about and avoid in the thinking that goes on to learn and make inferences. In the following, I will attempt to point out some of these blind spots. I will frequently use examples taken from the reports and analyses of the Challenger and Columbia disasters. This is not because I wish to criticize NASA, but because there is publicly available information about those two events. Private organizations that have the same kinds of problems do not make them public.

Traps in "Learning from Failure"

The April, 2011 issue of the Harvard Business Review was devoted to "learning from failure." Little was said about the potential traps imbedded in studying failures and drawing conclusions about their causes. In 1998, Ian Bradbury and I published a paper titled "Improving Problem Solving."[1] In the paper, we discussed two examples that illustrate the inferential traps involved in studying only failures or defects. The first example came from the Challenger disaster. There was a teleconference on the evening prior to the launch. Thiokol solid rocket booster engineers argued that the launch should not proceed due to the low temperature. The argument made by Thiokol personnel was considered too weak to support a decision to delay the launch. The data considered in the discussion included the occasions on which there had been problems with O-rings on the solid rocket boosters and the temperatures at which these problems had occurred. A plot of number of distressed rings per launch versus booster joint temperature is shown below. Actually, the data were not displayed graphically and only two of the points on the plot below received much attention during the teleconference: the point at 55 °F and the one at 75 °F.

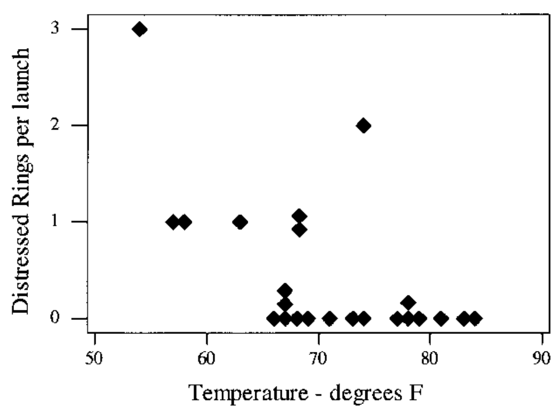
Challenger O-ring Problem Data



[1] Bradbury, I. and Ranney, G., "Improving Problem Solving," Report No. 167, Center for Quality and Productivity Improvement, University of Wisconsin, June, 1998.

There are no points displayed above or discussed during the teleconference for which the number of distressed O-rings (called failures for purposes of this discussion) was zero. That is, the relationship considered was confined to failures. Subsequent to the disaster, a plot was made that included the launches that had no distressed O-rings.

Challenger O-ring Data



Adding the information about launches in which there were no distressed O-rings changes one's view of a possible relationship, particularly considering that the projected launch temperature was near freezing. Since the accident, authors have argued that the second graph might have carried the day in the teleconference and led to a postponement of the launch.[2] Tufte argues that there was a clear proximate cause of the accident: "an inability to assess the link between cool temperature and O-ring damage on earlier flights. Such a pre-launch analysis would have revealed that this flight was at considerable risk."[3] Tufte discusses the 13 charts that were used during the teleconference and concludes, "In the 13 charts prepared for making the decision to launch, there is a scandalous discrepancy between the intellectual tasks at hand and the images created to serve those tasks. As analytical graphics, the displays failed to reveal a risk that was in fact present. As presentation graphics, the displays failed to persuade government officials that a cold-weather launch might be dangerous… *there are right ways and wrong ways to show data; there are displays that reveal the truth and displays that do not*. And, if the matter is an important one, then getting the displays of evidence right or wrong can possibly have momentous consequences."[4] So actually, there are two problems identified here. First, the omission of the prior launches that had not had distressed O-rings from the discussions, and second, the inadequacy of the displays of evidence.

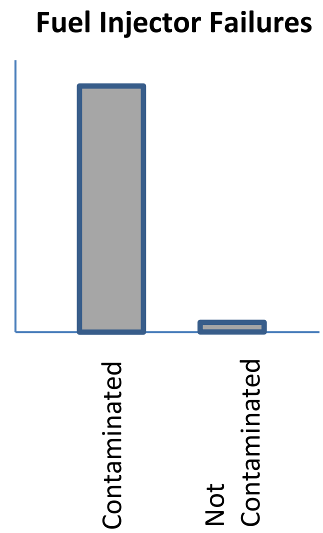
The second example comes from the automotive industry and was provided by our friend Mike Tveite:

---

[2] See, for example, Vaughan, Diane, *The Challenger Launch Decision*, University of Chicago Press, 1996.
[3] Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Graphics Press, 1996.
[4] It was not my intention here to discuss visual displays, but to advocate the use of all the data pertinent to the issue at hand in drawing conclusions. However, I recommend Tufte's books, including his landmark *Visual Display of Quantitative Information,* to all of us who analyze and present information.

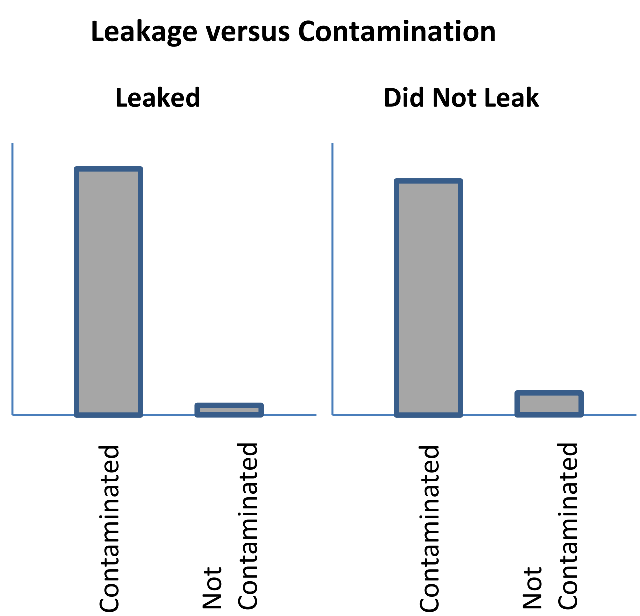
A manufacturer of fuel injectors had been having a problem with mild leakage. When the fuel injector should have been closed, small amounts of fuel were seeping past the fuel injector's sealing surface into the engine's inlet manifold. This often shows up first as a lack of smoothness of the engine at idle or reduced ease of starting.

Under the warranty agreement fuel injectors that were replaced at car dealerships for being leaky were returned to the manufacturer for problem solving analysis. Tear down of the fuel injectors and careful examination under the microscope revealed a relationship like that depicted below:



[The picture is intended to convey the idea that the vast majority of failed injectors examined contained contamination. It is not intended to be numerically accurate.] *A number of actions had thus been initiated to reduce the level of contamination. Possible actions that were being considered included redesign of the injector and fuel line filters, additional washing, flushing and inspection operations in the manufacturing process, increased air filtration and so on.*

The people working on the problem were asked whether they had examined any fuel injectors which were not leaky. They had not, but decided to do so. Doing so produced the following:



It thus became apparent that the contamination that they had been observing was present, independent of the occurrence of leakage. If "non-defective" injectors had not been examined, substantial time and capital could have been expended on reduction of the presence of contamination of the type observed without benefit to the problem at hand.

In the paper, we go on to note that it is important when a potential cause is of a binary nature; that is, it is either present or absent, to attempt to obtain data in **all** of the cells of a table such as the following:

	Problem or Failure	
Potential Cause	Present	Absent
Present		
Absent		

Tufte puts it more generally: "in reasoning about causality, *variations in the cause* must be explicitly and measurably linked to *variations in the effect*." He goes on to identify dual principles for reasoning about statistical evidence and for the design of statistical graphics: "(1) *documenting* the sources and characteristics of the data, (2) insistently enforcing appropriate *comparisons*, (3) demonstrating mechanisms of *cause and effect*, (4) expressing those mechanisms *quantitatively*, (5) recognizing the inherently *multivariate* nature of analytic problems, and (6) inspecting and evaluating *alternative explanations*."

I do realize that I have put forth two examples of failures in learning and inference to discuss traps in learning from failure. It is likely that there are cases in which these problems existed and the correct conclusions were reached nevertheless, but awareness of the potential for these failures is probably increased by providing examples of these traps.

I note here that the use of "The 5 Whys" to discover "The Root Cause" of a problem or a mistake or a failure is fraught with the danger of falling into the same kind of logical trap illustrated by the two preceding examples. In the paper, we go on to say, "If one only considers the 'problem' category of results from a system, one may either miss an important causal relationship (as in the Challenger Case) or erroneously infer presence of a causal relationship," as in the case of leaky injectors. Any analytical tool depends for the goodness of its outcome on the mind of the user.

Learning from Failures (or Successes) on a Large Scale

After the Challenger disaster, a commission chaired by former Secretary of State William Rogers was appointed to investigate. The U.S. House Committee on Science and Technology also conducted hearings and produced a report. Subsequently, Diane Vaughan did an intensive study of the context of the disaster and the practices and relationships that existed in the NASA system (including contractors), in the engineering profession, and in the government. After her study, Vaughan concluded that the Rogers Commission had failed to include important aspects of the context:

> *Both for easy public digestion and for NASA's survival, the myth of production-oriented, success-blinded middle managers was the best of all possible worlds. It removed from public scrutiny the contributions to the disaster made by top NASA officials, Congress and the White House, - it minimized awareness of the difficulty of diagnosing the risky technology. Locating blame in the actions of powerful elites was not in NASA's interest. And focusing attention on the fact that, after all this time, the technology still could defy understanding would destroy the NASA-cultivated image of routine, economical spaceflight and with it the Space Shuttle Program…*
>
> *Retrospection corrects history, altering the past to make it consistent with the present, implying that errors should have been anticipated. Understanding organizational failure depends on systematic research that avoids the retrospective fallacy…*

In hindsight, it is likely that an explanation can be found for any failure and an inference can be made that the parties involved should have known better.

It is just as dangerous to restrict study to successes. Peters and Waterman's book, *In Search of Excellence*, is an example. They identified some companies as "excellent." Then they studied each company to discover why they were excellent. They picked some practices that they saw as the reasons for excellence and discussed them in the book. My friend Bill Bellows pointed out that they didn't go to other companies that were not in the excellent category to see if they were also using those practices. Later, some of the companies identified as excellent were not so excellent.

Information from benchmarking exercises should be considered carefully from this perspective. Simply copying without careful thought about context and whether or not a practice will fit into an organization's system can be a dangerous practice.

The "Root Who"

The fundamental attribution error occurs when a behavior of an individual is attributed to the individual without considering situational factors that may have had an influence on the behavior. We have a tendency to attribute problems to the people who happened to be there when the problem occurred, rather than considering numerous other factors that contributed. (On the other hand, when we make a mistake ourselves, we tend to consider those other factors.)

In its discussion of the shuttle Columbia accident, the Columbia Accident Investigation Board stated:

> *Many accident investigations make the same mistake in defining causes. They identify the widget that broke or malfunctioned, then locate the person most closely connected with the technical failure: the engineer who miscalculated an analysis, the operator who missed signals or*

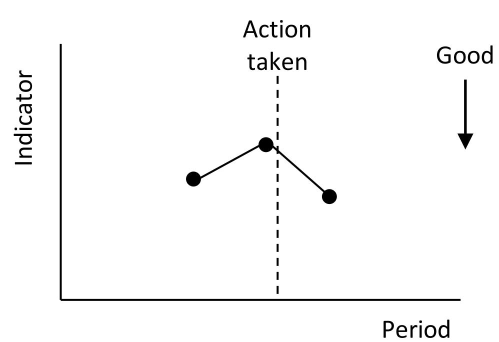
pulled the wrong switches, the supervisor who failed to listen, or the manager who made bad decisions. When causal chains are limited to technical flaws and individual failures, the ensuing responses aimed at preventing a similar event in the future are equally limited: they aim to fix the technical problem and replace or retrain the individual responsible. Such corrections lead to a misguided and potentially disastrous belief that the underlying problem has been solved. The Board did not want to make these errors.[5]

I recall hearing an executive in a large corporation refer to the mistake described above as "the ongoing search for the root who." Others refer to the kind of faulty investigation described by the Board as The Blame Game. If this is the kind of activity that goes on in an organization when a failure occurs, learning is suppressed. To avoid The Blame Game requires practice of the discipline of thinking about context; i.e., thinking systemically.

Some Superstitious Learning

Performance indicators are watched at all levels in organizations. When they get worse from one period to the next, actions are often taken to correct the situation. If a given performance indicator then gets better the next period, the actor concludes that his or her action produced the improvement. I have attempted to illustrate the connection of the conclusion to the results in the following graph:

An Action Appears to Produce Improvement

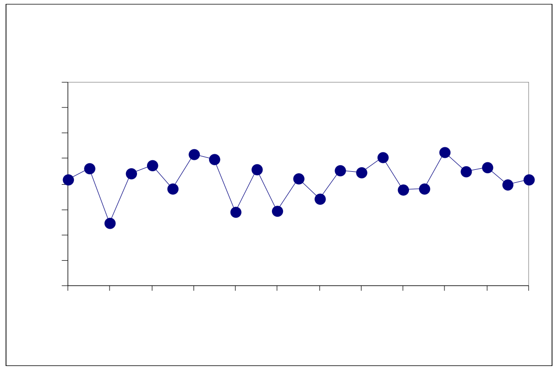


However, it is possible that the action had no immediate effect. Using a mathematical model for stable variation, Dr. Dennis Gilliland demonstrated that the likelihood that the third number pictured above would be lower than the second when the second was higher than the first was two thirds.[6] (The same likelihood would exist for the third number in a sequence high-low-high). This means that if the variation in a series of results is stable, the person who takes the action in this situation is more likely to have his or her belief confirmed than not when the action produces no effect at all. Plotting performance indicators over time would reduce the likelihood of this kind of superstitious learning. In

[5] Report of the Columbia Accident Investigation Board, August, 2003.
[6] Gilliland, Dennis C., *Experiences in Statistics*, Dubuque: Kendall/Hunt, 1990, p.54.

the following time ordered plot, note that the first three points in the series follow the pattern shown above, as do points four, five and six, as well as other sequences of three points in the series.

Performance Indicator Plotted Over Time



The Dangers of Induction

W. Edwards Deming stated, "No number of examples establishes a theory, yet a single unexplained failure of a theory requires modification or even abandonment of the theory."[7] Deming's statement implies that you can pile up empirical examples that appear to support a theory from here to the moon, but that does not constitute proof of the theory. One's degree of belief that the theory is correct may increase, but one has not proved that the theory will hold up in the future. Inductive proof can be done in mathematics, but not in the world of experience. Deming used the example of the mythical rooster Chanticleer to explain what he meant by the statement above:

> *The barnyard rooster Chanticleer had a theory. He crowed every morning, putting forth all his energy, flapped his wings. The sun came up. The connexion was clear: His crowing caused the sun to come up. There was no question about his importance. There came a snag. He forgot one morning to crow. The sun came up anyhow. Crestfallen, he saw his theory in need of revision. Without his theory, he would have had nothing to revise, nothing to learn.*

If Chanticleer had been a two year old rooster, he would have had nearly 500 observations that appeared to support his theory. However, those 500 observations did not prove the correctness of his theory. As another example, think of the many centuries during which our species collected empirical observations that were consistent with the theory that the sun revolved around the earth. Nevertheless, Milton Silviera, Chief Engineer, NASA headquarters wrote, "the first flight [of the shuttle] represented a proof of the design concept." I am not sure of the meaning of his statement, but I'm certainly unsure that the word proof is appropriate to describe one observation, successful though it might have been.

[7] Deming, W. Edwards, *The New Economics, 2nd Ed*., M.I.T., 1994.

In his important book, *The Structure of Scientific Revolutions*, Thomas Kuhn describes the doing of science as puzzle-solving within the current scientific paradigm until anomalies occur that force scientists to reexamine their basic theory and develop new theories that provide the ability to explain the anomalies.[8] Vaughan observes, "In the Kuhnian sense, a paradigm is a fundamental component of scientific culture. It is a world-view based on accepted scientific achievement, which embodies procedures for inquiring about the world, categories into which these observations are fitted, and a technology that includes beliefs about cause-effect relationships and standards of practice and behavior."

In the engineering disciplines, when it is extremely costly or virtually impossible to construct tests of ideas, it may be tempting to rely on a series of observations in practical application as "proof." In the case of the Challenger, this appears to have happened. Vaughan writes,

> *From integrated sets of assumptions, expectations, and experience, individuals construct a worldview, or frame of reference, that shapes their interpretations of objects and experiences. Everything is perceived, chosen, or rejected on the basis of this framework. The framework becomes self-confirming because, whenever they can, people tend to impose it on experiences and events, creating incidents and relationships that conform to it. And they tend to ignore, misperceive, or deny events that do not fit. As a consequence, this frame of reference generally leads people to what they expect to find. Worldview* [paradigm] *is not easily altered or dismantled because individuals tend ultimately to disavow knowledge that contradicts it. They ward off information in order to preserve the status quo, avoid a difficult choice, or avoid a threatening situation. They may puzzle over contradictory evidence but usually succeed in pushing it aside – until they come across a piece of evidence too fascinating to ignore, too clear to misperceive, too painful to deny, which makes vivid still other signals they do not want to see forcing them to alter and surrender the worldview they have so meticulously constructed.*

Jean Piaget, the great developmental psychologist, distinguished between development and learning. Development was change to the entire cognitive structure used to observe and interpret experience. Learning was concerned with a specific situation or problem and dealt with only specific cognitive structures.[9]

So we see that there is a complex set of forces, cognitive and psychological as well as logical, that may lead us to unwarranted induction. It is important to be aware of this and try to practice a discipline of being careful about our inferences. It is also useful to keep in mind that empirical generalizations are probable only.[10] The next observation may overturn the generalization. As Deming said, "No matter

[8] Kuhn, Thomas S., *The Structure of Scientific Revolutions, 2nd Ed*., University of Chicago Press, 1970.
[9] Piaget identified stages of cognitive development. Only in the highest stage of development are we able to reflect on our own thought processes. Bybee, R.W. and Sund, R.B., *Piaget for Educators*, Waveland Press, 1990.
[10] See C.I. Lewis, *Mind and the World-Order*, Scribner's, 1929.

how strong be our degree of belief, we must always bear in mind that empirical evidence is never complete."[11]

Limits of Applicability of a Theory

In January, 2003, the space shuttle Columbia was launched. During the launch, a block of foam insulation struck the leading edge of the left wing of Columbia. This strike produced damage to Columbia's thermal protection system that led to disintegration of the shuttle as it re-entered the earth's atmosphere. In work done to try to assess the potential damage done while the shuttle was still in orbit, analysts used a mathematical modeling tool called Crater to predict the depth of possible damage. Crater was normally used to predict whether small debris, such as ice on the External Tank that held fuel for launch, would pose a threat during launch. Crater had been calibrated with testing done on small debris – on the order of three cubic inches. Crater was judged to be a conservative tool; that is, it tended to predict more damage than actually occurred.

In the discussion of the Crater model, the Columbia Accident Investigation Board stated:

> *Although Crater was designed, and certified, for a very limited set of impact events, the results from Crater simulations can be generated quickly. During STS-107* [the Columbia mission]*, this led to Crater being used to model an event that was well outside the parameters against which it had been empirically validated. As the accompanying table shows, many of the STS-107 debris characteristics were orders of magnitude outside the validated envelope. For instance, while Crater had been designed and validated for projectiles up to 3 cubic inches in volume, the initial STS-107 analysis estimated the piece of debris at 1200 cubic inches – 400 times larger.*

A table that appears in Chapter 6 of the Investigation Board's report is shown below:

Crater parameters used during development of experimental test data versus STS-107 [Columbia] *analysis:*

Test Parameter	*Test Value*	*STS-107*
Volume	*Up to 3 cu.in*	*10" x 6" x 20" = 1200 cu.in.**
Length	*Up to 1 inch*	*~ 20 inches **
Cylinder Dimensions	*<= 3/8" dia x 3"*	*6"dia x 20"*
Projectile Block Dimensions	*<= 3" x 1" x 1"*	*6" x 10" x 20" **
Tile Material	*LI-900 "acreage" tile*	*LI-2200 * and LI-900*
Projectile Shape	*Cylinder*	*Block*

** Outside experimental test limits*

The Board concluded that "the use of Crater in this new and very different situation compromised NASA's ability to accurately predict debris damage in ways that Debris Assessment Team engineers did not fully comprehend."

[11] Deming, W. Edwards, *Out of the Crisis*, MIT Center for Advanced Engineering Study, 1986.

Usually, experimentation that is used to develop and test theories is limited to the circumstances under which the theory is expected to be applied. In the case of Crater, the test parameters were limited to values that included the conditions of expected application. It is advisable to avoid use of a theory or a model to apply to a set of circumstances that are outside the boundaries of test. To aid in avoidance, it would be a good idea to carefully examine the conditions of test before using a theory in a new, possibly untested, circumstance. According to Deming,

> *Any theorem is true in its own world. But which world are we in? Which of several worlds makes contact with ours? That is the question.* [12]

My friend Ian Bradbury pointed out that we face an inferential gap in cases such as validation testing with prototype parts (differing materially from the intended method of production) and test conditions that differ materially from intended use. Ian stated, "Deliberate introduction of variability in a designed way based upon subject matter knowledge can help reduce the inferential gap. Deliberate consideration of the theory behind the Crater model with subject matter knowledge could have assisted in the judgment of likely effectiveness for the intended inference. Efforts previously to empirically test the limits of model applicability against predictions from underlying theory would have been better still."[13]

Theories and Assumptions Not Made Explicit

Dietrich Dörner writes as follows:

> *If we want to operate within a complex and dynamic system, we have to know not only what its current status is but what its status will be or could be in the future, and we have to know how certain actions we take will influence the situation. For this, we need "structural knowledge," knowledge of how the variables in the system are related and how they influence one another… The totality of such assumptions in an individual's mind – assumptions about the simple or complex links and the one-way or reciprocal influences between variables – constitute what we call that individual's "reality model." A reality model can be explicit, always available to the individual in a conscious form, or it can be implicit, with the individual himself unaware that he is operating on a certain set of assumptions and unable to articulate what those assumptions are. Implicit knowledge is quite common… An individual's reality model can be right or wrong, complete or incomplete… The ability to admit ignorance or mistaken assumptions is indeed a sign of wisdom…*[14]

In the example of the use of Crater, it appears that an implicit assumption was present that the model performed the same as the conditions of test in a situation in which the piece of debris striking the Columbia was 400 times larger. Had this assumption been made explicit, perhaps someone would have

---

[12] Deming, *The New Economics, 2nd Ed.*
[13] Ian S. Bradbury, personal communication.
[14] Dörner, Dietrich, *The Logic of Failure*, Metropolitan Books, 1996.

asked for an explanation of the rationale behind the assumption. Any conclusions drawn from the analysis should have been qualified by a statement of the conditions under which Crater was tested.

The kinds of over-adjustment that Deming described in his funnel experiment involve an implicit assumption that every outcome of a process is produced by a special cause that must be addressed by adjustment of the aim of the funnel. There are numerous opportunities to use this assumption in practice to the detriment of future outcomes.

There are implicit theories in use in less technical circumstances. For example, Douglas McGregor articulated Theory X and Theory Y to describe the views that were held about the nature of people in the workplace. In the management of organizations, implicit theories about the value and effects of competition, the nature and sources of motivation, the relationships among organizational components, and many other aspects of the organization's functioning govern the actions of management every day. One of my favorite quotes is from John Maynard Keynes: "Practical men who believe themselves to be devoid of any intellectual influences are usually the slaves of some defunct economist."[15] Individuals and organizations could learn and improve by bringing their assumptions and theories to the surface and questioning them.

Doubtless there are other blind spots that can introduce difficulties into learning and inference, some of which may have been exemplified in the preceding discussion. The more we work at trying to identify the traps awaiting us in learning and inference, the more adept we may become at both of these endeavors.

[15] Keynes, J.M., *The General Theory of Employment, Interest and Money*, 1936.

Biography

Gipsie Ranney is an international consultant to organizations on management, quality improvement and statistical methodology. She was a member of the faculty of the Department of Statistics at the University of Tennessee, Knoxville for fifteen years. She was a co-founder of the University of Tennessee's Institute for Productivity through Quality, and she developed and conducted numerous seminars on quality improvement. She served as Director of Statistical Methodology for General Motors Powertrain Group from 1988 to 1992. She co-authored *Beyond Total Quality Management: Toward the Emerging Paradigm*, published by McGraw-Hill, and contributed to *Competing Globally Through Customer Value*, published by Quorum. She has published papers on quality improvement and statistical methods. The American Society for Quality awarded her the Deming Medal for 1996, "for outstanding contribution in advancing the theory and practice of statistical thinking to the management of enterprises worldwide." Gipsie holds a B. S. in Mathematics from Duke University and a Ph.D. in Statistics from North Carolina State University.