

The Perils of Post-Hockery

.....

When we engage in post-hoc reasoning, we blur the distinction between the contexts of discovery and verification and risk misplacing our faith in unwarranted beliefs. Some simple steps can avoid the seductive traps of post-hockery.

JOHN RUSCIO

How do psychologists come to believe in the diagnostic utility of responses to a series of abstract inkblots? How can each student in an entire class believe that a generic character sketch is a custom-tailored description of his or her personality? Individuals may cling tenaciously to these beliefs, despite the scientific and logical evidence that refutes them. Unfortunately, it is relatively simple for anyone with an active imagination, access to a steady supply of ambiguous data, and a willing suspension of critical thinking to “discover” and “validate” personally meaningful relationships that do not actually exist. The key is to focus on post-hoc (“after the fact”) interpretations of alleged phenomena that selectively admit evidence.

Strengths and Shortcomings of Human Judgment

"Post-hockery" takes advantage of several interrelated strengths and shortcomings of human judgment. We often blur the crucial distinction between the generation and testing of hypotheses. In the *context of discovery*, our outstanding pattern-recognition and reasoning abilities are indispensable. We can detect potentially relevant environmental cues and formulate sophisticated hypotheses about underlying causal relationships. However, unaided judgment can fail us in the *context of verification*. We do not routinely subject our cherished beliefs to rigorous tests, and we often accept the first preferred explanation as fact. Particularly if this explanation is interesting or entertaining, confidence may become unshakable.

One of the surest signs of post-hockery is the inability of its practitioners to predict future events. Although successful prediction does not necessarily imply a deep understanding, the reverse is true. One who truly understands a relationship can use this knowledge to make valid predictions, whereas the post-hoc reasoner can only "explain" past events. Three examples serve to illustrate the importance of prediction in testing claims to knowledge.

The Bible Code

The recent controversy over *The Bible Code* (Drosnin 1997) stems from post-hoc reasoning. The alleged discovery of God's hidden messages actually takes advantage of post-hockery in two problematic ways. As David Thomas (1997, 1998) has exquisitely detailed, the reported probabilities of finding the messages are computed in a wildly inappropriate manner. If you look in a fantastic number of places, and count anything that you stumble upon as supportive evidence, you are guaranteed to discover meaning where none exists. That the hidden messages of the Bible Code can be found with statistically predictable regularity in a wide array of texts speaks strongly against their profundity.

A more subtle problem with the code is of special relevance here. Important messages are supposedly revealed through the juxtaposition of words such as "Hitler" and "Nazi," "Roswell" and "UFO." Here is a challenge to those who believe that the Bible contains hidden messages: *predict something*. Promoters of the Bible Code tell us that it can reveal the future, so by all means reveal it to us! If coincidental text alignments are the culprit, however, then "meaning" will be found only *after* events have transpired, and the predictive failure of the Bible Code will reveal a message that is certainly not hidden.

The "Hot Hand" in Basketball

A second example of post-hockery is the widespread belief in the streak shooting of basketball players. You would be hard-pressed to watch an entire televised game without at least once being told that so-and-so has the "hot hand." There are two distinct ways to understand this remark. It might simply indicate that a player has made a series of shots. Nobody would take exception with this descriptive use of the term. Fans, coaches, and players, however, agree that this is not the intended meaning. The term "hot hand" is used and under-

stood in a predictive sense: the player with the "hot hand" is engaged in streak shooting that is expected to continue.

Unfortunately for its adherents, the "hot hand" label holds no predictive value. Research shows that the "hot hand" marks past success but predicts nothing (Gilovich, Vallone, and Tversky 1985). This unwarranted belief is likely due to false pattern recognition. There does seem to be something magical about a player who has made several successive shots. Fans, coaches, and players alike feel compelled to create a causal explanation rather than to accept the mundane statistical fact that when outcomes are distributed probabilistically, streaks occur by chance far more often than we imagine.¹

Atheoretical Statistical Analyses

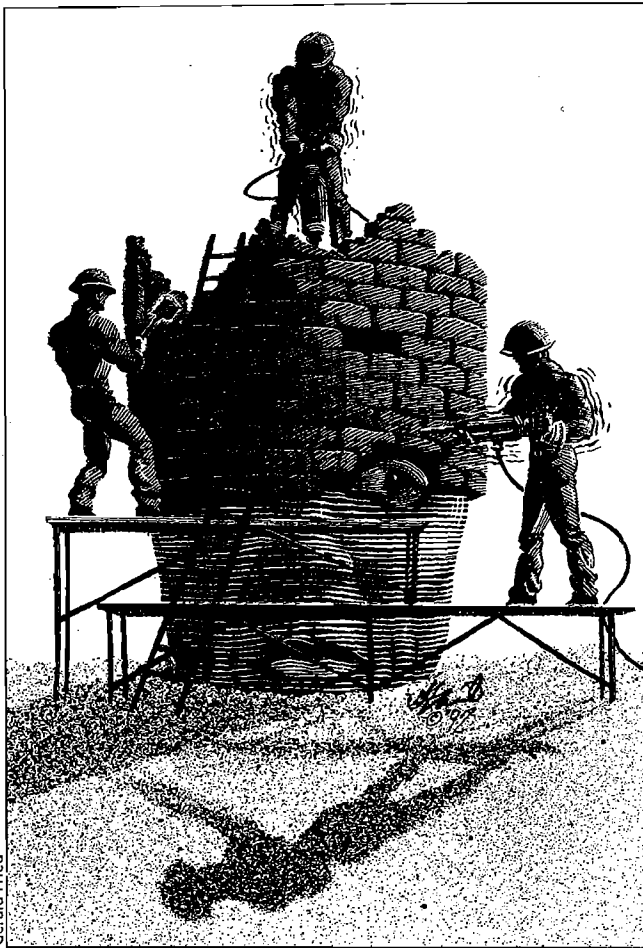
A third example of post-hockery stems from statistical analysis methodologies used by many researchers. Computerized data analysis, which has greatly accelerated the pace of scientific progress, provides massive input to post-hoc analysis. The enormous difficulty involved in hand calculations once restricted complex statistical analysis to explicit tests of hypotheses. These days, a few keystrokes can generate as much statistical output as an army of graduate students operating their pocket calculators all summer long. Correlating just a couple dozen variables with one another will produce a matrix containing nearly 300 correlation coefficients. By convention, results that occur at a level expected by chance just 5 percent of the time are called "statistically significant." We can therefore expect about fifteen spuriously significant correlations within every matrix of 300. Each spurious correlation is grist for the master of post-hockery.

Just as the Bible Code may seem impressive when you consider only the successful searches, reporting statistically significant results without mentioning the sheer number of tests conducted can have a powerful impact on readers. Moreover, our pattern-recognition abilities are rivaled or surpassed by the ease with which we can concoct an explanation for anything, after the fact. These fifteen spurious results will be couched in theoretical terms that convey an impressive degree of insight or understanding when they emerge in published form. The temptation to seek out and report statistically significant results is more powerful than many individuals can resist, owing both to their innate curiosity and the precarious "publish-or-perish" culture of researchers.

The Fallacy of "Personal Validation"

So the discovery of potential relationships is easy. If the distinction between contexts of discovery and verification was clear, we would not get into trouble. Research on human judgment, however, has shown how feebly we subject our beliefs to tests, if indeed we test them at all. "Personal validation" refers to our tendency to seek weak, experiential support for our

John Ruscio, Ph.D., is a social psychologist in the Department of Psychology (MS 062), Brandeis University, Waltham, Massachusetts 02254-9110. His research interests are in the areas of statistical methodology in behavioral research and the psychology of judgment and decision making. Email: ruscio@brandeis.edu.



Gerald Fried

hypotheses in lieu of a more balanced pursuit of persuasive evidence. Although there is substantial overlap in many of our thinking habits, some main influences can be teased out.

Confirmation Bias

The selective search for confirmatory information shields us from relinquishing many of our cherished beliefs. Given enough ambiguous information, we can almost always ferret out some support for our beliefs. For example, consider the second question posed at the beginning of this article: How can each student in an entire class believe that a generic character sketch is a custom-tailored description of his or her personality? The trick is to use descriptions so vague that they apply to most anyone. Forer (1949) composed this character sketch, which illustrates the technique marvelously:

You have a strong need for other people to like you and for them to admire you. At times you are extroverted, affable, and sociable, while at other times you are introverted, wary, and reserved. You have a great deal of unused energy which you have not turned to your advantage. While you have some personality weaknesses, you are generally able to compensate for them. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations. You pride yourself on being an independent thinker and do not accept other opinions without satisfactory proof. You have a tendency to be critical of yourself. Some of your aspirations tend to be pretty unrealistic.

These statements do capture our personalities well, and we can find some anecdotal evidence to support all of the statements if we dig deep enough into our memories. What our evaluations lack is a consideration of the base rate likelihood that this sketch also fits *any* randomly selected person. In this case, the likelihood is quite high. Rational thought is overwhelmed by the bias toward confirmation.

When a belief is interesting or entertaining, we are especially unlikely to seek evidence that might refute it. In such cases, contradictory evidence may be perceived as less interesting and attract less attention. To demonstrate this bias, try mentioning the New York City Blackout of 1965 to someone and asking what comes to mind. In my experience, most people believe that this is when a lot of babies were conceived. There was a widespread human-interest story maintaining that a spike in the New York City birth rate occurred on Monday and Tuesday of the week nine months after the blackout. This is true, but a similar spike actually occurred *every* Monday and Tuesday. The cause of this anomaly: doctors preferred not to work on weekends. Induced labor and Caesarian section births were scheduled for the beginning of the week. Although this was discovered relatively shortly after the original story broke, the true explanation was noticed by few. It conflicted with post-hockery that was more interesting and more entertaining.

Employing a double standard of evidence is another common method by which evidence is biased toward confirming our beliefs. We assign great weight to evidence that favors our beliefs and ignore or discredit contradictory evidence. Consider the first question asked at the beginning of this article: Why do some psychologists believe in the diagnostic utility of responses to a series of inkblots? Gilovich and Savitsky (1996) describe the formation of pseudoscientific beliefs, such as those held by interpreters of the Rorschach inkblot test, through representative thinking: "like goes with like." In Chapman and Chapman's (1967) classic research on the use of projective tests, clinicians' interpretations were strongly biased in the direction of preconceived notions. Do all psychologists recognize the weak, experiential support for their reliance on the Rorschach?

No. Those who routinely employ the Rorschach justify this practice on the basis of extensive personal experience while rejecting a large body of empirical evidence indicating that this method has little or no diagnostic reliability or validity (Dawes 1994). Personal experience is accepted by those who believe in the efficacy of the Rorschach, whereas the experience of others or contradictory scientific evidence is grossly underweighted.

Absent or Misleading Feedback

We receive little, if any, objective feedback with which to test our beliefs. Imagine the case of a clinician who uses the Rorschach for assessment. The test is administered, responses are interpreted, and a diagnosis is reached. For a practitioner subscribing to one school of psychotherapy, a diagnostic mistake may have no consequence for the course of treatment. If therapy is held constant across patients, how can a diagnostic error ever be detected? Belief in the utility of the Rorschach

becomes invulnerable due to a lack of corrective feedback.

Suppose, instead, that successful treatment hinges on a correct diagnosis. Will a clinician receive corrective feedback even when an incorrect diagnosis is made, and an ineffective therapeutic technique used? Given the usual reasons why patients seek counseling and the dynamics of the patient-therapist relationship, it should come as no surprise that patients tend to accept and agree with their therapists' diagnoses and advice. Incorrectly diagnosed individuals will often believe that they are receiving the best available treatment. The lack of treatment success is a fact that emerges so slowly and gradually that it may never arouse suspicion. Even if it does, neither the therapist nor the patient is likely to question the validity of the original diagnosis, much less its evidential basis. There are many ready-made alternatives. Perhaps this is an atypical case. Perhaps therapy did not begin early enough. Perhaps the patient is resistant to the therapist (or vice versa). We can generate countless alternative explanations for disappointing outcomes, none of which will call into question a cherished belief.

The feedback that we do obtain, as a rule, is systematically misleading. Many types of feedback are exclusively positive in nature or easily distorted to become so. How do we evaluate our teaching skill? Standardized rating forms can be dismissed as worthless, thrown on the same heap with all else that is standardized—just ask any poor teacher. We learn from experience. Which of our former students will keep in touch with us? Which students ask for letters of reference? If a student enrolls in another of our courses, this is clear confirmation of teaching skill (never mind that we teach required courses). The student who does not do so, in the unlikely event that we notice, is assumed to have had other commitments that created a scheduling conflict.

We selectively expose ourselves to situations and environments containing only favorable information. We tend to associate with people who think the way that we do, read books and articles that support our views and discard those that do not, join clubs and organizations to interact with others who share our interests, and so forth. With all of these defensive measures in place, where will corrective feedback come from?

Hindsight Bias and Overconfidence

Once we have discovered a personally meaningful relationship, unearthed confirmatory evidence, and received flattering feedback, 20/20 hindsight monopolizes our vision. We begin to feel that we "knew it all along" (Fischhoff 1975), replacing speculation with certitude through a process of "creeping determinism" (Fischhoff 1980). We come to hold undue confidence in our beliefs. Because most everyone is overconfident much of the time—despite feelings to the contrary—perhaps the best way to demonstrate this point is for you to test your own confidence.

Plous (1993) has devised a simple test of factual knowledge to evaluate confidence. For each of ten questions, you construct an interval such that you are 90 percent confident it will contain the correct answer. Thus you might say, for Question 1 below, that you are 90 percent confident that Martin Luther King's age at the time of his death was somewhere between

___ years and ___ years (you fill in the blank). The key to this assessment is not whether you know the exact answers—although you may for some questions—but whether you can construct 90 percent confidence intervals that are actually correct 90 percent of the time. Here, now, are the ten questions:

1. What was Martin Luther King, Jr.'s age at death?
2. What is the length of the Nile River, in miles?
3. How many countries belong to OPEC?
4. How many books are there in the Old Testament?
5. What is the diameter of the moon, in miles?
6. What is the weight of an empty Boeing 747, in pounds?
7. In what year was Mozart born?
8. What is the gestation period of an Asian elephant, in days?
9. What is the air distance from London to Tokyo, in miles?
10. What is the deepest known point in the ocean, in feet?

The answers to these questions, as of the 1993 publication of Plous's book, appear in Note 2. If you got all ten questions correct in this way, you are *underconfident*, not to mention a member of a breathtakingly small minority! If you got exactly nine questions correct, your confidence is *properly calibrated* to your accuracy level; this, too, is exceedingly rare. If you got fewer than nine questions correct, you are *overconfident*, like the vast majority of us. Although this quiz may seem trivial, it is just one method of eliciting confidence judgments. Overconfidence is in fact a pervasive problem across many types of beliefs, tasks, and situations (Alpert and Raiffa 1982; Oskamp 1965), and it is every bit as common among experts as laypersons (Dawes 1994; Faust and Ziskin 1988).

Foiling Post-Hockery: Some Advice

If post-hockery is the disease, then what is the cure? Scientific psychology documents flaws in our reasoning, but it also offers suggestions for improvement. Healthy thinking habits do take some practice, but the payoff is a more rational belief system. Here are three concrete strategies for reducing susceptibility to post-hockery.

Formulate Multiple Alternative Hypotheses. Throughout his illustrious career, Carl Sagan was an advocate of skepticism, wonder, and the proper balance between the two. In the context of discovery, exercise a great deal of wonder. The best known method for minimizing the judgmental shortcomings associated with post-hockery is the consideration of alternative hypotheses (Arkes 1991, McDonald 1998). Confidence in any particular belief is appropriately reduced by the existence of several alternative hypotheses. Stronger evidence must be sought to evaluate competing theories, and the correct hypothesis is more likely to be included in a larger pool of possibilities.

Apply Consistent Standards of Evidence. In the context of verification, exercise a great deal of skepticism. Subject your beliefs to the same rigorous tests that you demand of others. Those with the toughest standards of evidence will acquire and retain the most accurate beliefs. Be open to the possibility that the pattern you believe you are observing may in fact be random fluctuation.

tuations. In his book *Statistics as Principled Argument*, Robert Abelson (1995) sums this up nicely by urging readers to “give chance a chance.” When you find yourself proclaiming something as meaningful and important, ask yourself whether you would accept someone else’s idea on the same evidentiary basis.

Ask What Can Be Predicted. When you can only see meaning in data by looking backward, post-hockery is strongly indicated. You do not truly understand a phenomenon until you can generate testable predictions. Every one of the examples of unwarranted belief cited in this article could have been avoided if its creator had been challenged to make predictions. When someone makes a claim to knowledge ask, by way of evidence, “What can you predict?” Adding this simple habit to one’s repertoire of critical thinking skills will help to clearly differentiate the context of discovery from that of verification. The ability to spot post-hockery in an argument, one’s own or another’s, is well worth a bit of practice.

References

- Abelson, Robert P. 1995. *Statistics as Principled Argument*. Mahwah, NJ: Erlbaum.
- Alpert, Marc, and Howard Raiffa. 1982. A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Arkes, Hal R. 1991. Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin* 110: 486–498.
- Chapman, Loren J., and Jean Chapman. 1967. Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology* 72: 193–204.
- Dawes, Robyn M. 1994. *House of Cards: Psychology and Psychotherapy Built on Myth*. New York: Free Press.
- Drosnin, Michael. 1997. *The Bible Code*. New York: Simon and Schuster.
- Faust, David, and Jay Ziskin. 1988. The expert witness in psychology and psy-

chiatry. *Science* 241: 31–35.

- Forer, Bertram R. 1949. The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology* 44: 118–123.
- Fischhoff, Baruch. 1975. Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1: 288–299.
- . 1980. For those condemned to study the past: Reflections on historical judgment. In R. A. Sweder and D. W. Fiske (Eds.), *New Directions for Methodology of Behavioral Science: Fallible Judgment in Behavioral Research*. San Francisco: Jossey-Bass.
- Gilovich, Thomas, and Kenneth Savitsky. 1996. Like goes with like: The role of representativeness in erroneous and pseudoscientific beliefs. *SKEPTICAL INQUIRER* 20(2), March/April: 34–40.
- Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17: 295–314.
- McDonald, John. 1998. 200% probability and beyond: The compelling nature of extraordinary claims in the absence of alternative explanations. *SKEPTICAL INQUIRER* 22 (1), January/February: 45–49, 60.
- Oskamp, Stuart. 1965. Overconfidence in case-study judgments. *Journal of Consulting Psychology* 29: 261–265.
- Plous, Scott. 1993. *The Psychology of Judgment and Decision Making*. Philadelphia: Temple University Press.
- Thomas, David E. 1997. Hidden messages and the Bible Code. *SKEPTICAL INQUIRER* 21(6), November/December: 30–36.
- Thomas, David E. 1998. Bible Code developments. *SKEPTICAL INQUIRER* 22(2), March/April: 57–58.

Notes

1. This by no means implies that the outcomes of shots are determined by chance. In fact, nothing could be further from the truth. A wide range of factors contribute to success: a player’s baseline skill level, the difficulty of the attempted shot, the strength of the defense, and so forth. What the research does indicate is simple: if you had a complete listing of the causal factors determining the outcome of an attempted shot, the success of previous shots would not appear on it.

2. (1) 39 years. (2) 4,187 miles. (3) 13 countries. (4) 39 books. (5) 2,160 miles. (6) 390,000 pounds. (7) 1,756. (8) 645 days. (9) 5,959 miles. (10) 36,198 feet. □

INVESTIGATIVE FILES

from page 25

she stated, “It still functions, though people aren’t allowed to go up it very often” (Bobbin 1998). It would thus appear that the Loretto staircase is subject to the laws of physics like any other.

The other mysteries that are emphasized in relation to the stair are the identity of the carpenter and the type of wood used. It seems merely mystery mongering to suggest that there is anything strange—least of all evidence of the supernatural—in the failure to record the name of an obviously itinerant workman.

As to the wood, that it has not been identified precisely means little. The piece given to a forester for possible identification was exceedingly small (only about 3/4-inch square by 1/8-inch thick) whereas much larger (six-inch) pieces are preferred by the U.S. Forest

Service’s Center for Wood Anatomy (which has made many famous identifications, including artifacts taken from King Tut’s tomb and the ladder involved in the Lindbergh kidnapping) (Knight 1997). The wood *has* reportedly been identified as to family, *Pinaceae*, and genus, *Picea*—i.e., spruce (Easley 1997), a type of “light, strong, elastic wood” often used in construction (“Spruce” 1960). But there are no fewer than thirty-nine species—ten in North America—so that comparison of the Loretto sample with only two varieties (Easley 1997) can scarcely be definitive.

In the final analysis the “mysteries” of the spiral staircase at the Loretto Chapel are evidence, not of its miraculous production but instead of its human—quite fallibly human—manufacture.

References

- Albach, Carl R. 1965. Miracle or wonder of con-

struction? reprint from *Consulting Engineer*, December, n.p.

- Bobbin, Jay. 1998. “The Staircase.” Review in *TV Topics*, *The Buffalo News*, April 12, pp. 1, 24–25.
- Bullock, Alice. 1978. *Loretto and the Miraculous Staircase*. Santa Fe, N.M.: Sunstone Press.
- Dietz, Albert G. H. 1991. *Dwelling House Construction* 5th ed. Cambridge, Mass.: The MIT Press.
- Easley, Forrest N. 1997. *A Stairway from Heaven?* Privately printed.
- “Interior Decoration.” 1960. *Encyclopaedia Britannica*.
- Knight, Christopher. 1997. “Just What Kind of Wood . . . ?” *Wall Street Journal*, October 22.
- Locke, Jim. 1992. *The Well-Built House*, revised ed. New York: Houghton Mifflin Co.
- Loretto Chapel. N.d. Text of display card, photographed by author, 1993.
- “Spruce.” 1960. *Encyclopaedia Britannica*.
- “Stair.” 1960. *Encyclopaedia Britannica*.

Acknowledgments

Once again I am grateful to Tim Binga, Director of the Center for Inquiry Libraries, for research assistance and to Ranjit Sandhu for manuscript preparation. □