

Critical Thinking in Anesthesia

Eighth Honorary FAER Research Lecture

Steven L. Shafer, M.D.*

SCIENTIFICALLY based professionals face the dilemma described by James Oberg, National Aeronautics and Space Administration engineer and space journalist: "You must keep an open mind, but not so open that your brains fall out."¹ Startling advances, such as the role of *Helicobacter pylori* in the etiology of stomach ulcers,² remind us of the need to be receptive to unexpected discoveries that challenge our beliefs.

As practitioners of a science-based profession, physicians have a responsibility to patients to (1) critically interpret their medical history, (2) critically read and evaluate the scientific literature, and (3) critically plan their therapy using evidenced-based medicine. How well do we meet this standard?

Anesthesiologists are arguably the most experienced practitioners of cardiopulmonary resuscitation. For example, my general anesthetics typically start with a ventilatory arrest (usually by intent), followed by need for cardiovascular support. The 2005 American Heart Association guidelines for cardiopulmonary resuscitation assess the evidence for "best practices" in resuscitation.³ Class I recommendations describe interventions that unambiguously improve outcome and should always be provided. These include defibrillation, tracheal intubation, and (remarkably) taping the endotracheal tube. Class IIa recommendations, whose benefits almost always exceed risks, include many widely accepted interventions, such as administering oxygen, confirming carbon dioxide returning from the airway, the use of various airway maintenance devices, proper use of a defibrillator, magnesium for the treatment of torsade de pointes, fibrinolysis for suspected

pulmonary embolus, and use of the impedance threshold device to improve recovery.

In all probability, you have never heard of the impedance threshold device, also a class IIa recommendation, despite the fact improved outcome has been demonstrated in more than 50 animal and clinical trials of resuscitation. The device is a special valve that attaches to a facemask or advanced airway device (fig. 1). In the absence of a positive-pressure breath, it seals off the airway when the chest recoils during the decompression phase of cardiopulmonary resuscitation, thereby generating more negative pressure in the thorax to draw more blood into the great vessels and heart. This also decreases intracranial pressure. The device doubles cardiac output during cardiopulmonary resuscitation. According to a recent meta-analysis, the impedance threshold device nearly doubled short- and long-term survival after cardiac arrest, at a cost of approximately \$99 per unit.

The American Heart Association Guidelines characterize epinephrine and amiodarone as class IIb recommendations, meaning that benefit may exceed risk. Lidocaine and atropine are class III recommendations, indicating insufficient evidence. What does it mean that few anesthesiologists have ever used the impedance threshold device, which roughly doubles survival at nominal cost, and has stronger evidence for benefit than epinephrine, amiodarone, lidocaine, and atropine, mainstays of current practice? In my view, it means that we are not as scientifically based in our practice as we would like to believe. Instead, our clinical practice reflects traditions and beliefs instilled during training, which only gradually adapt to scientific progress.

As practitioners of a science-based profession, we also have a responsibility to the society in which we live to function as role models for critical thinking. The public is surprisingly naive about the scientific method. How else can one explain the marketing of such unlikely marvels as running your car on water,[†] a heater that produces more energy than used,[‡] or curing cancer with magnets?[§]

Consider the following urban legend: A businessman in New Orleans meets a prostitute in a bar. They go to her hotel room, where he blacks out. The next morning he awakens in a bathtub, with a note telling him to call 911. The dispatcher tells him to feel for a tube protruding from his lower back. He finds one, and is told his kidney has been stolen.

Clinicians will readily dismiss this story because of the impossibility of performing a donor nephrectomy in a hotel room, and the sheer weirdness of awakening from surgery and anesthesia in a bathtub with a tube in your back.

This article is accompanied by an Editorial View: Eisenach JC: Curiosity, discipline, and joy; and the greatest of these is joy. ANESTHESIOLOGY 2009; 110:712-3.

* Professor of Anesthesiology.

Received from Columbia University, New York, New York. Submitted for publication November 17, 2008. Accepted for publication December 15, 2008. Dr. Shafer receives salary support from the International Anesthesia Research Society for his work as Editor-in-Chief of *Anesthesia & Analgesia*. Presented at the 2008 Annual Meeting of the American Society of Anesthesiologists, Orlando Florida, October 20, 2008.

Address correspondence to Dr. Shafer: Department of Anesthesiology, Columbia University Medical Center, 630 West 168th Street, P&S Box 46, New York, New York 10032. sshafer@columbia.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

† <http://convert-car-to-hydrogen.cn/>. Accessed November 14, 2008.

‡ <http://www.dailymail.co.uk/sciencetech/article-481996/How-12inch-miracle-tube-half-heating-bills.html>. Accessed November 14, 2008.

§ <http://www.stopcancer.com/magnetswollenrpt.htm>. Accessed November 14, 2008.

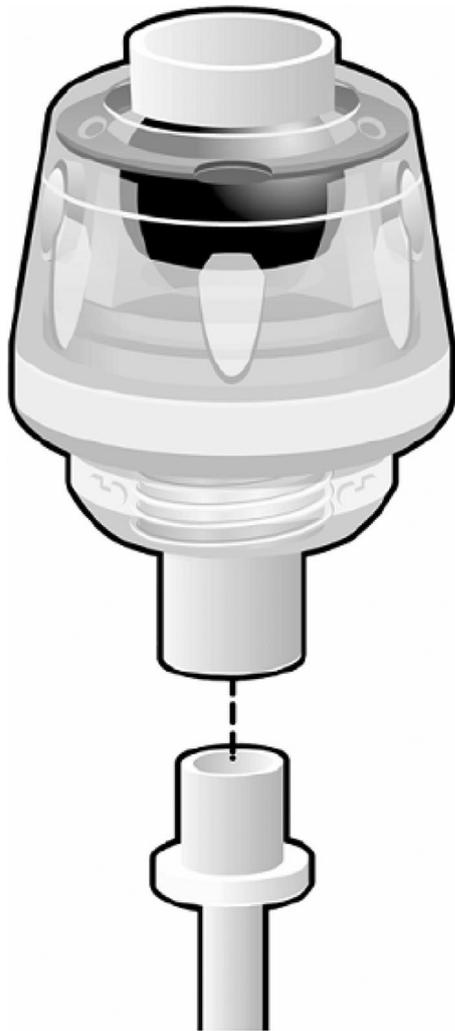


Fig. 1. Impedance threshold device for increasing cardiac output during cardiopulmonary resuscitation.

However, the claim has been so widely circulated that the National Kidney Foundation issued a formal repudiation.^{||} However, when this “innocent” urban myth circulated in South America, there was a 90% decrease in *cadaveric* kidney donations.⁴ Considering that many individuals waiting for transplants die before receiving a transplant, this “harmless” urban legend may have resulted in significant morbidity and mortality. Ignorance has consequences!

You have likely received e-mails promising fortunes in exchange for some nominal effort, such as helping transfer money from Nigeria or contacting an agency that has chosen your e-mail address in a previously unknown lottery. When the victim attempts to procure the money, he or she is instructed to pay increasingly large advance fees to facilitate the transaction. Such transparently bogus schemes net billions of dollars every year from gullible individuals.[#]

|| <http://www.kidney.org/transplantation/donorFamilies/shareDonation.cfm>. Accessed November 14, 2008.

http://ultrascan.nl/html/419_statistics.html. Accessed November 11, 2008.

Table 1. Taxonomy for Reasoning from Evidence

1. Falsifiability: Could an experiment be designed that could disprove (“falsify”) the hypothesis?
2. Logic: Is the argument logically sound, including the statistical methods of inference?
3. Comprehensiveness: Has all the available evidence been considered?
4. Honesty: Has the evidence been considered honestly, without misrepresentation, without self-deception, and with willingness to accept whatever conclusion is supported by the weight of the evidence?
5. Replicability: Is the conclusion based on evidence that can be reproduced by multiple investigators?
6. Sufficiency: Is the weight of the evidence adequate to support the claim (e.g., evidence based on authority or testimony is inadequate, extraordinary claims require extraordinary evidence)?

From Lett.⁵

Millions of gullible individuals firmly believe in alien abductions, crystal healing, channeling, psychic power, touch therapy, and implausible dietary wonders because they do not have the intellectual tools to evaluate fraudulent claims. More broadly, our political process, our allocation of national resources, our foreign policy, and our stewardship of the planet all suffer from an endemic resistance to think critically as a society. As physicians, we need to lead by example.

If we are to be role models of critical thinking, we need to understand how to evaluate claims based on evidence. James Lett, Professor of Anthropology at Indian River Community College, proposed a taxonomy of six essential elements for reasoning from evidence: falsifiability, logic, comprehensiveness, honesty, replicability, and sufficiency (table 1).⁵

Falsifiability

Falsifiability refers to whether a claim can be disproved by evidence. If the truth of a claim is completely immune from all evidence against it, the claim cannot be evaluated scientifically. For example, some cultures believe that everything has a spirit, including the rocks, the trees, and even the wind. This belief does not yield any testable predictions. How can you disprove that the wind has a spirit? As these beliefs cannot be proved or disproved from evidence, they are simply beyond the realm of science.

A lot of “new age” thinking suffers from lack of falsifiability. Can people really channel their thoughts through crystals? If this leads to testable predictions, then it can be prospectively evaluated. If not, the claim is scientifically meaningless. The same is true for many claims from practitioners of “alternative medicine.” How can one evaluate a treatment that promises to “improve your well-being”? If that cannot be defined, it cannot be tested.

Infinite excuses are often used to protect pseudoscientific viewpoints from falsifiability. For example, creation “scientists” hold that God created the earth and its life forms based on literal reading of the Bible. The mountain of evidence for evolution through natural selection, drawn from epidemiology, evolutionary biology, comparative biology, geology, statistical modeling, paleontology, molecular biology, and genetics, including the sequencing of the genomes of many species, is entirely dismissed. Why are dinosaur bones buried in the earth? “God wanted it that way!” Not only is creation “science” devoid of falsifiable propositions, but infinite excuses preclude rational debate about the merits of creationism.

Similarly, believers that the earth has been visited by extraterrestrials answer all criticism regarding the complete lack of evidence for extraterrestrial visitation on the basis that the evidence is suppressed by the government. It is all kept under lock and key in Area 51! This permits infinite excuses and prevents falsifiability of the claims. This is the essence of junk science.

Logic

Logic involves valid methods of inference. Some methods are self-evident. If I say that all dogs have fleas, and Roxy is a dog, it follows that Roxy has fleas. However, if I say that all dogs have fleas, and Roxy has fleas, it does not follow that Roxy is a dog. Roxy could be the neighbor’s kid.

Logic can be devilishly confusing. Consider the 1960s television game show *Let’s Make a Deal* hosted by Monty Hall. This show was based on a stage with three doors. Behind one of the doors was a valuable prize, frequently “Monty’s Cookie Jar” stuffed with money. Behind the other two doors were joke prizes, typically goats. The contestant was instructed to pick one of the doors. Let’s assume the contestant picks door 1.

At this point in the show, Monty would typically open one of the two doors not chosen by the contestant, invariably revealing a goat to howls of inexplicable laughter from the audience. (Since Monty knew where the cookie jar was, did they think that Monty would reveal the cookie jar and spoil the game?) Let’s assume that Monty has opened door 3, revealing the goat. Ha ha! At this point the contestant is offered a choice: stay with door 1, the original selection, or switch to door 2. Nearly all contestants chose to stay with door 1.

The logic could not be simpler: There are just three doors, and one question: stay or switch? It is remarkably nonobvious that the probability of winning Monty’s Cookie Jar is doubled by switching to door 2. The reason

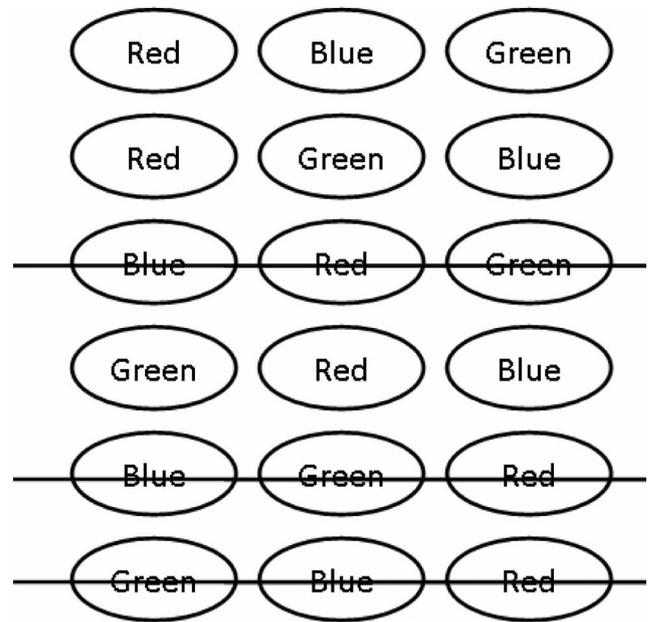


Fig. 2. The six possible choices for color preference of red, blue, and green M&Ms. Lines show the choices removed when the monkey indicates a preference for red over blue. Among the remaining three choices, there is a two thirds probability that the monkey will pick green over blue.

is that in opening door 3, Monty has told the contestant nothing about door 1. There was a one-third probability that this was the correct choice when the choice was made, and there is still a one-third probability after door 3 is opened. There was a two-thirds probability that Monty’s Cookie Jar was behind doors 2 or 3, and that is still the case. However, because there is zero chance that Monty’s Cookie Jar is behind door 3, there is a two-thirds probability that it is behind door 2. Therefore, the probability of winning Monty’s Cookie Jar is doubled by switching.**

The same problem has confounded research in psychology for several decades. In studies of cognitive dissonance, a monkey is offered a choice between a red and a blue M&M. Let us assume that the monkey selects red. The monkey is then given a choice between a blue and a green M&M. The monkey is approximately twice as likely to pick green. Assuming that there was a 50:50 chance that the monkey would pick either color, the preference for green was assumed to represent “cognitive dissonance,” psychological angst over the selection of the previously rejected blue M&M. This has been the basis of decades of research in cognitive dissonance.

However, there is *not* a 50:50 chance of the monkey picking a green M&M *versus* blue M&M. As pointed out by Keith Chen,⁶ this is the Monty Hall problem. Assume that the monkey has a true preference among the three colors. Figure 2 shows the six possible preferences. The monkey’s preference for the red M&M over the blue M&M rejects half of the possible preferences, shown as

** http://en.wikipedia.org/wiki/Monty_hall_problem. Accessed November 15, 2008.

a line through three of the six possible preferences in figure 2. Among the remaining three possible rank orders, green is preferred over blue in two cases, and blue is preferred over green in one case. Therefore, the animal is twice as likely to pick the green M&M as to pick the blue M&M, not 50:50 as was assumed. The Monty Hall and monkey M&M choice problems are striking for the nonobviousness of the correct answer, which is particularly remarkable given the exceedingly simple structure: three doors or three M&Ms.

Logic can be distorted by the ambiguity in our language. For example, many would agree that a ham sandwich was better than nothing. It is similarly evident that nothing is better than eternal happiness. If both statements are true, does it not logically follow that a ham sandwich is better than eternal happiness?

Logic also involves statistics. Statistics can be thought of as having two flavors: Fisherian and Bayesian. Fisherian statistics involves the determination of the probability of nonrandom events in a population based on known distributions of random events. Most statistics in scientific journals are Fisherian. The problem with Fisherian statistics is that they are unintuitive. For example, assume you conduct a clinical trial that does not reach its primary endpoint but achieves a statistically significant ($P = 0.05$) secondary endpoint. What is the probability that an absolutely identical repeat of the trial will reproduce the secondary endpoint at $P < 0.05$? I would have thought that if $P < 0.05$ in the first trial, there is a 19 in 20 probability of achieving a statistically significant result in a repeat of the trial. However, that is not correct. As pointed out by Goodman⁷ and subsequently by O'Neill,⁸ there is only a 57% probability of reproducing a $P < 0.05$ result in a second trial. Indeed, to have a 90% probability of reproducing the finding at $P < 0.05$, the P value in the first trial must be 0.001 or less.

Bayesian statistics go back 200 yr to the work of Reverend Thomas Bayes (fig. 3), a Presbyterian minister with an interest in mathematics. Reverend Bayes worked out the fundamental relation between evidence and a conclusion now called *Bayes' theorem*.^{††} Reverend Bayes' famous theorem was published posthumously from mathematical folios found in his home. His intentions in developing his theory of probability remain a mystery.

Bayes' theorem relates a hypothesis, H, to the evidence for the hypothesis, E, and the setting, I. Specifically, $p(H | E, I) = p(H | I) \times \frac{p(E | H, I)}{p(E | I)}$, which can be restated as the posterior probability that hypothesis H is true, given that the evidence E in setting I, $p(H | E, I)$, is the prior probability of the hypothesis being true in the setting, $p(H | I)$, times the probability of the evidence, if the hypothesis is true, in the setting, $p(E | H, I)$, normal-



Fig. 3. Thomas Bayes, 1702–1761.

ized to the probability of the evidence in the setting, $p(E | I)$. Bayes' theorem can be restated in the following terms: The posterior probability (*i.e.*, your conclusion after considering the data) is based on the prior probability (*i.e.*, how likely the conclusion was before the new data were introduced) and on the quality of the new data that support the conclusion.

Bayes' theorem quantitates common sense. If a particular conclusion is unlikely, then it takes powerful evidence to change our thinking. However, if a particular conclusion is likely, ordinary evidence provides adequate support. By way of example, it would be an extraordinary claim that a new opioid did not cause ventilatory depression. A study demonstrating this at $P < 0.05$ would not be adequate to support this conclusion. However, the notion that a new 5-hydroxytryptamine type 3 antagonist was effective for treating postoperative nausea and vomiting is a very ordinary claim. A reduction in postoperative nausea and vomiting at $P < 0.05$ would provide adequate evidence of the efficacy of a novel 5-hydroxytryptamine type 3 antagonist.

Bayes' theorem demonstrates that not all evidence is the same. Medical claims published in *Science* and *Nature* have considerably more weight than medical claims published in weekly tabloids. Physicians have formal systems for evaluating evidence, such as the Jadad score⁹ to evaluate the quality of clinical trials. Bayes' theorem provides a quantitative method to incorporate the strength of evidence on the probability of the conclusion.

Bayes' theorem describes the logic we typically refer to as clinical judgment. If you see a 20-yr-old patient with chest pain, you would not consider a heart attack very

^{††} http://en.wikipedia.org/wiki/Bayes_theorem. Accessed November 15, 2008.

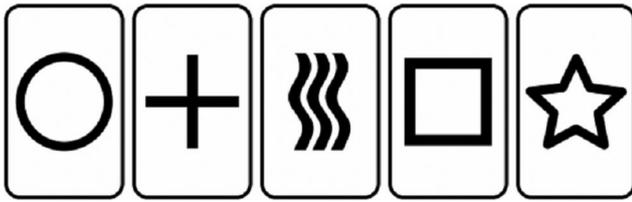


Fig. 4. Zener cards, used by Joseph B. Rhine in studies of extrasensory perception.

likely. However, the same chest pain in a 70-yr-old diabetic smoker would immediately trigger concern about a heart attack. This reflects sound clinical judgment: Young people are not very likely to have coronary artery disease. However, in Bayesian terms, we have a powerful prior (elderly smokers are much more likely to have coronary artery disease than are young healthy individuals) and low-quality evidence (nonspecific chest pain). Bayes' theorem also shows why high-quality evidence, such as elevated troponin levels in the young patient, can readily overpower the prior expectation.

Fifteen years ago, an extraordinary claim was made: Stomach ulcers were caused by bacterial infection.² This was dismissed because the prevailing knowledge was that ulcers were caused by stress. This extraordinary claim was followed by extraordinary evidence,¹⁰ including consumption of a beaker of *H. pylori* by one of the investigators,^{‡‡} who went on to share the 2005 Nobel Prize in Physiology or Medicine. As Bayesian thinkers, we changed our beliefs about ulcers when confronted with the extraordinary evidence that stomach ulcers really were a result of *H. pylori* bacterial infection.

The daily miracle of anesthesia is a Bayesian experiment. Our drugs are reliable, so we know that a patient given a reasonable dose of propofol, fentanyl, and sevoflurane has an enormously high likelihood of being unconscious. How likely is it that the patient is asleep if the observed response to surgical incision is

- A. the blood pressure goes up?
- B. the heart rate goes up?
- C. the patient moves?
- D. the Bispectral Index goes from 60 to 75?
- E. the Bispectral Index goes from 60 to 95?
- F. the patient opens his eyes?
- G. the patient reaches up, extubates herself, and screams in pain?

The progression from A to G shows increasingly powerful evidence of wakefulness. For A through D, the likely conclusion is that the patient is unconscious. However, it would be very hard to dismiss the evidence in

‡‡ http://en.wikipedia.org/wiki/Helicobacter_pylori. Accessed November 15, 2008.

§§ <http://www.skepdic.com/essays/psihistory.html>. Accessed November 15, 2008.

||| <http://www.gentlebirth.org/archives/birthjoy.html>. Accessed November 16, 2008.

case G, no matter how much drug the patient apparently received. Perhaps it all went into the sheets.

Bayesian inference helps us resolve Oberg's riddle: how to keep an open mind without letting your brains fall out. No scientific doctrine is known with such certainty that it is immune to evidence. However, extraordinary claims require extraordinary evidence. For those that claim that the earth is flat, the scientific response is "please show me extraordinary evidence to back up your extraordinary claim of a flat Earth." The same response, *mutatis mutandis*, applies to those who claim that the story of creation in Genesis is true, those who claim that psychics can see into the future, and those who claim to have been abducted by aliens.

Comprehensiveness

Comprehensiveness means that the evidence used to evaluate the truth of something must be exhaustive. There is a tendency to ignore evidence that disagrees with one's beliefs. For example, in the 1920s, Joseph Banks Rhine set up a laboratory at Duke University to study extrasensory perception. Subjects were asked to identify which of five "Zener cards" (fig. 4) the investigator was looking at. Some subjects did better than expected, and some did worse. In Dr. Rhine's view, subjects who guessed far more accurately than expected demonstrated extrasensory perception—as did subjects who guessed far less accurately than expected!^{§§} Indeed, Dr. Rhine was convinced that those who did much worse than expected were malicious, using their extraordinary sensory powers to undermine his research. He excluded them from his analysis.¹¹ When he analyzed the remaining subjects the overall result was better than expected, providing evidence of extrasensory perception. Far from demonstrating psychic power, his results are a demonstration of the power of selective data analysis.

We occasionally hear claims related to our profession that violate the requirement of comprehensiveness. For example, it is sometimes claimed that "women have been doing it [giving birth] all the way back to Eve"^{|||} without anesthesia services, so such services are unnecessary. That ignores the enormous changes in maternal and neonatal survival, changes that reflect modern obstetric and anesthesia practice. We sometimes overhear that general anesthesia is so safe and routine that no special skill is required. That ignores the years of rigorous training now expected of anesthesia providers, training that permits an illusion that delivery of anesthesia is safe and routine.

Honesty

Honesty means that the evidence must be evaluated without self-deception. It also means that if the evidence disproves something you believe, you must change your

beliefs. For example, Jewish tradition holds that God placed all knowledge in the Torah. Because the Torah does not discuss molecular biology, plate tectonics, and quantum mechanics, scientists have used computers to search for hidden knowledge in the Torah. One technique is called “equidistant letter sequences,” the juxtaposition of words spelled out in letters spaced at fixed intervals. For example, in Genesis 31:28 of the King James version of the Bible, we find the text: “t e R s t h O u h a S t n o W d o n E f o o L i s h L y i n s o d o.” Note that every 5th letter, capitalized, spells out “ROSWELL,” the site of a supposed UFO landing in New Mexico. Every 12th letter, italicized, spells out “UFO.” By the logic of equidistant spacing, this demonstrates that the Bible predicted the landing of a UFO in Roswell, New Mexico!

In 1997, Michael Drosnin published *The Bible Codes*, which claimed that the Torah predicted the assassination of Israeli Prime Minister Yitzhak Rabin, including the date and name of the assassin, and a host of other contemporary events.¹² He challenged skeptics by saying, “When my critics find a message about the assassination of a prime minister encrypted in *Moby Dick*, then I’ll believe them.”^{##} Using equidistant letter spacing, one can find detailed predictions in *Moby Dick* of the deaths of Indira Gandhi, Leon Trotsky, and John F. Kennedy.^{***} In fact, equidistant letter spacing provides a very detailed description of the death of Michael Drosnin.^{†††} The honest scientist accepts overwhelming evidence and changes his or her beliefs. In the case of *The Bible Codes*, that has not happened. Instead, Drosnin has gone on to publish two more books: *The Bible Code II: The Countdown*, which predicted that we would all perish in 2006, and (when that did not happen) *The Bible Code: The Quest*, which says that the extraterrestrials left a code that reveals a myriad of secrets in a steel obelisk. Please check your basement.

Scientific honesty requires a scientific worldview, which holds that evidence is paramount. By contrast, a faith-based worldview starts with a proposition based on faith, not based on evidence. The evidence is then selectively evaluated, and ignored or discounted if it does not support the proposition. A scientist in the former Soviet Union, Trofim Lysenko, held that acquired traits could be genetically passed to offspring.¹³ Lysenko’s view gained the favor of Stalin, who believed that Lysenkoism affirmed the inevitable triumph of the proletariat.

<http://www.newsweek.com/id/95751/page/2>. Accessed November 15, 2008.

*** <http://cs.anu.edu.au/bdm/dilugim/moby.html>. Accessed November 15, 2008.

††† <http://cs.anu.edu.au/bdm/dilugim/drosnin.html>. Accessed November 15, 2008.

‡‡‡ <http://www.answersingenesis.org/articles/nab/does-c14-disprove-the-bible>. Accessed November 13, 2008.

§§§ www.intelligentdesignnetwork.org. Accessed November 14, 2008.

Biologists who held to a Darwinian view of inheritance were purged from Soviet academic ranks and sent to the Siberian Gulag. It did not matter that the scientific evidence directly contradicted Lysenko’s view of inheritance. His view was incorporated into Marxist doctrine, and was held to be true as a matter of government imposed faith. Lysenkoism was abandoned in the Soviet Union in the mid-1960s, but Russian biology has never recovered.

Evolution continues to be a lightning rod for attacks by those holding faith-based worldviews, most recently in the canard “intelligent design.” The debate over evolution is not carried out in the scientific literature, because the evidence for intelligent design does not stand up to the scrutiny of scientific peer review. Instead, the arguments are advanced primarily on the Internet, where claims can be posted without the imprimatur of peer review.

Consider the following text from a Creationist Web site^{†††}:

When a scientist’s interpretation of data does not match the clear meaning of the text in the Bible, we should never reinterpret the Bible. God knows just what He meant to say, and His understanding of science is infallible, whereas ours is fallible. So we should never think it necessary to modify His Word. Genesis 1 defines the days of creation to be literal days (a number with the word “day” always means a normal day in the Old Testament, and the phrase “evening and morning” further defines the days as literal days). Since the Bible is the inspired Word of God, we should examine the validity of the standard interpretation of ¹⁴C dating by asking several questions . . .

This is an unambiguous representation of a faith-based worldview: The Bible is inerrant. Any claims of science, including the claim that the earth is more than a few thousand years old, are necessarily wrong. The Web site goes on to attack the scientific basis of carbon dating, based on analyses by “experts” who do not have the imprimatur of scientific peer review. The typical individual perusing this site (including me) does not have the scientific training to evaluate the evidence that supposedly disproves carbon dating. That is why the argument might seem persuasive to the reader who does not distinguish between peer-reviewed evidence and evidence posted on an Internet site that claims the Bible is literally true and inviolably inerrant.

One can find many Web sites devoted to intelligent design.^{§§§} However, the story in the peer-reviewed literature is quite different. Of 99 articles identified by a PubMed search of *intelligent design* (on November 14, 2008), the majority are defenses of evolution against claims of intelligent design. Not appearing in the search is the single “scientific” article supporting the claims of intelligent design,¹⁴ written by Stephen Meyer of the Discovery Institute. This article was published without peer review in a nonindexed journal and was subse-

quently retracted by the journal for insufficient scientific merit.||||

Despite the total lack of scientific support, proponents of intelligent design argue through courts, school boards, and the occasional senator### that intelligent design should be taught in science classes as an alternative to evolution. That is dishonest. As noted by Sober, “in all its forms, intelligent design fails to constitute a serious alternative to evolutionary theory.”¹⁵ Science classes must teach critical aspects of Bayesian inference, including the need to critically assess scientific evidence. Above all else, science classes must teach the fundamental tenet of science: Data trumps theory. Given the complete lack of peer-reviewed data supporting intelligent design, it is disingenuous to represent intelligent design as a scientific alternative to evolution, which is supported by a staggering compilation of observations and scientific synthesis¹⁶ and is the fundamental organizing principle the biologic sciences. To quote Judge John E. Jones in his 2005 decision in *Kitzmiller vs. Dover Area School District*,

Intelligent design is not science. We find that ID fails on three different levels, any one of which is sufficient to preclude a determination that ID is science. They are (1): ID violates the centuries-old ground rules of science by invoking and permitting supernatural causation (2); the argument of irreducible complexity, central to ID, employs the same flawed and illogical contrived dualism that doomed creation science in the 1980s; and (3) ID’s negative attacks on evolution have been refuted by the scientific community.

It is possible that the correct worldview is the faith-based view. If so, perhaps intelligent design is true after all. “Truth” is not the subject of this article. This article is about reasoning from evidence. A worldview that uniformly discounts evidence that does not support a one’s beliefs is not scientific.

The Internet has several resources that are models of objectivity. Specifically, PubMed, sponsored by the National Library of Medicine, is a wonderful resource for rapidly accessing the peer-reviewed literature in the medical and biologic sciences. Urban myths are frequently evaluated, and either supported or debunked, by www.snopes.com. The claims of U.S. politicians of all political stripes are objectively vetted at www.factcheck.org, run by the Annenberg School of Public Policy at the University of Pennsylvania. Last, the open-access Wikipedia has proven itself to be a remarkably robust forum for distilling accurate information from the combined wisdom of millions of motivated authors.

|||| http://www.biolosocwash.org/id_statement.html. Accessed November 14, 2008.

http://en.wikipedia.org/wiki/Santorium_Amendment. Accessed November 14, 2008.

**** <http://partners.nytimes.com/library/national/science/050399sci-cold-fusion.html>. Accessed November 14, 2008.

†††† Feynman RP: Rogers Commission Report. Appendix F: Personal Observations on the Reliability of the Shuttle. Available at: <http://science.ksc.nasa.gov/shuttle/missions/51-l/docs/rogers-commission/Appendix-F.txt>. Accessed November 15, 2008.

Replicability

Replicability is a fundamental part of the scientific method. A single experiment is never adequate. Every scientist has his or her biases. Every experimental result, no matter how well conducted, may be based on fraud, bias, or undetected error.

The “cold fusion” saga highlights the importance of replicability. On March 23, 1989, Professors Pons and Fleishman at the University of Utah announced the production of significant heat by electrolysis in deuterium oxide, using electrodes of palladium and platinum.¹⁷ The story was front-page news around the world. Three months later, three physicists at Cal Tech, Stephan Koonin, Nathan Lewis, and Charles Barnes, told the American Physical Society that their analysis found numerous errors in methods and results, that the results of Pons and Fleishman were inconsistent with fundamental physics theory, and that the experiments were not reproducible.**** Of these, the truly damning finding was that the experiments were not reproducible. Had the experiments been reproducible, physicists would be forced to bring fundamental theory into agreement with the experimental results.

Critical thinking is driven by evidence. As humble observers, we must always adjust our beliefs to the evidence supplied by nature. Nature doesn’t give a hoot what we think. Nature won’t adapt the evidence to conform with our beliefs. As Richard Feynman said in his critique of the Challenger shuttle explosion, “nature cannot be fooled.”††††

Sufficiency

Last, we turn to sufficiency. Is the evidence offered sufficient to support a claim or belief? The burden of proof is on the claimant. Years ago, a colleague told me that his friends saw a flash through the sky, followed by a crash, during a camping trip in Mexico. They went to investigate, and (no kidding) saw small glowing extraterrestrial creatures. I didn’t believe it. He asked me, “Well, what is your explanation?” I can think of dozens of explanations: His friends were playing a joke on him, his friends were drunk, his friends were trying out local hallucinogens. It doesn’t matter. I’m not making the claim, so it is not up to me to explain the findings. All I know is that extraterrestrial visitation is the most extraordinary of claims and is typically supported by completely insufficient evidence. As Bayes demonstrated, extraordinary claims require extraordinary evidence.

A week ago, I spoke to a group of anesthesiologists in Plymouth, England. Afterward, an anesthesiologist related that in his youth he twice profoundly affected the rolling of dice by mental effort. He wanted to know what I thought of this. Why should I think anything? I’m not the one making a claim of extraordinary power. However, I

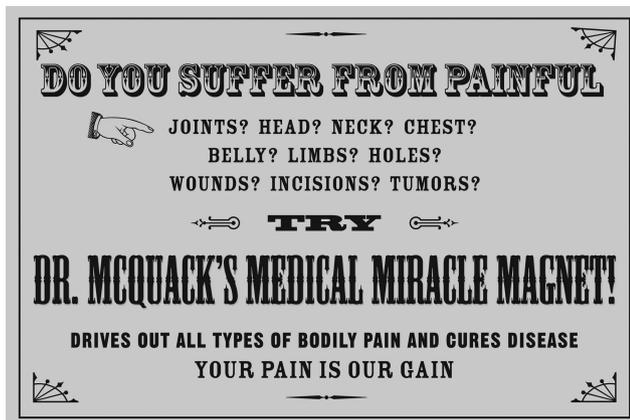


Fig. 5. Static magnets for pain management were lampooned on the February 2007 cover of *Anesthesia & Analgesia*. Artwork by Jan Hurst; used with permission of Lippincott Williams & Wilkins.

pointed out that his claim of seemingly magical abilities was backed up by the shoddiest of evidence: recollections from his childhood, decades ago, with no objective documentation. Although he was convinced of his claim, from my perspective the evidence was wholly insufficient.

Static magnets for pain management is a \$500 million per year business supported by insufficient evidence.†††† For example, Dr. Alvin Bakst claims to treat multiple types of pain with “unipolar neodymium magnets strategically and anatomically placed to reach the affected nerve causing the pain” (despite the fact that unipolar magnets do not exist).§§§§ A recent randomized, sham-controlled study identified no efficacy for static magnets in postoperative pain,¹⁸ an expected conclusion,¹⁹ highlighted with a satirical cover for *Anesthesia & Analgesia* (fig. 5). After publication, the manufacturer contacted me to complain about the article, citing the authors’ failure to measure the magnetic field in the skin. In response, I requested that the manufacturer identify a single reproducible model in which static magnets reduced pain. No such model was offered, nor does any such model exist.|||||

Our lay colleagues place great faith in the testimony of experts. Indeed, expert testimony is enshrined in our legal system. My advice is to never believe anything based on “expert” testimony alone. For example, I am reasonably expert on the clinical pharmacology of opioids. I also believe that morphine is a superior drug to meperidine. The burden of proof for this claim is on me.

†††† <http://nccam.nih.gov/health/magnet/magnet.htm#use>. Accessed November 14, 2008.

§§§§ <http://www.drakstmagnetics.com/>. Accessed November 15, 2008.

||||| <http://nccam.nih.gov/health/magnet/magnet.htm>. Accessed November 14, 2008.

http://en.wikipedia.org/wiki/Abduction_phenomenon. Accessed November 15, 2008.

***** http://en.wikipedia.org/wiki/Day_care_sexual_abuse_hysteria. Accessed November 15, 2008.

††††† <http://everything2.com/e2node/Shrine%2520of%2520the%2520Holy%2520Tortilla>. Accessed November 15, 2008.



Fig. 6. “If you don’t get the facts, the facts will get you.”—Robert L. Shafer. Photograph used with permission.

It is an ordinary claim, so ordinary evidence should suffice. However, the fact that I claim that morphine is a better drug than meperidine is not evidence of anything. My opinion is irrelevant. All that matters is the evidence.

At an international meeting on opioid pharmacology, an “expert” claimed that remifentanyl provided profound analgesia without causing respiratory depression. This is surely an extraordinary claim, and begs for extraordinary evidence. The “evidence” was that this “expert” could talk to patients after they received remifentanyl for cardiac surgery and they seemed to be comfortable and were breathing OK. He supported his extraordinary claim with grotesquely inadequate evidence.

Conclusion

We live in a world that is swamped with uncritical acceptance of medical quackery, junk science, and unsupported beliefs. Many people do not have the training to analyze whether aliens abduct people to probe their genitals,#### young children engage in satanic rituals at their local nursery school and return home as the same happy-go-lucky kids,*****²⁰ and diseases can be healed by a corn tortilla bearing the face of Jesus.††††† We owe it to ourselves, our patients, our scientific colleagues, and the society in which we live to use critical thinking as a fundamental part of who we are and how we approach our personal and professional lives. Our personal and professional beliefs must be based on critical evaluation of data, because “if you don’t get the facts, the facts will get you” (fig. 6).

References

1. Shafer SL: Did our brains fall out? *Anesth Analg* 2007; 104:247-8
2. Marshall BJ, Warren JR: Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1984; 1:1311-5
3. ECC Committee, Subcommittees and Task Forces of the American Heart Association: 2005 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation* 2005; 112 (suppl): IV1-203
4. Radford B: Bitter harvest: The organ-snatching urban legends. *Skeptical Inquirer* 1999; 23:34-9
5. Lett J: A field guide to critical thinking. *Skeptical Inquirer* 1990; 14:153-60
6. Chen KM: Rationalization and Cognitive Dissonance: Do Choices Affect or Reflect Preferences? Cowles Foundation Discussion Paper No. 1669. 2008
7. Goodman SN: A comment on replication, P values and evidence *Stat Med* 1992; 11:875-9
8. O'Neill RT: Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. *Control Clin Trials* 1997; 18:550-6
9. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S: Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Control Clin Trials* 1995; 16:62-73
10. Marshall BJ, Goodwin CS, Warren JR, Murray R, Blincow ED, Blackburn SJ, Phillips M, Waters TE, Sanderson CR: Prospective double-blind trial of duodenal ulcer relapse after eradication of *Campylobacter pylori*. *Lancet* 1988; 8627:1437-42
11. Park R: *Voodoo Science: The Road from Foolishness to Fraud*. Oxford, United Kingdom, Oxford University Press, 2000, pp 40-3
12. Drosnin M: *The Bible Code*. New York, Simon and Schuster, 1997
13. Soyfer VN: The consequences of political dictatorship for Russian science. *Nat Rev Genet* 2001; 2:723-9
14. Meyer SC: The origin of biological information and the higher taxonomic categories. *Proc Biol Soc Washington* 2004; 117:213-39
15. Sober E: What is wrong with intelligent design? *Q Rev Biol* 2007; 82:3-8
16. Kutschera U, Niklas KJ: The modern theory of biological evolution: An expanded synthesis. *Naturwissenschaften* 2004; 91:255-76
17. Fleischmann M, Pons S: Electrochemically induced nuclear fusion of deuterium. *J Electroanalytical Chem* 1989; 261:301-8
18. Cepeda MS, Carr DB, Sarquis T, Miranda N, Garcia RJ, Zarate C: Static magnetic therapy does not decrease pain or opioid requirements: A randomized double-blind trial. *Anesth Analg* 2007; 104:290-4
19. Flamm BL: Magnet therapy: Healing or hogwash? *Anesth Analg* 2007; 104:249-50
20. Garven S, Wood JM, Malpass RS, Shaw JS III: More than suggestion: The effect of interviewing techniques from the McMartin Preschool case. *J Appl Psychol* 1998; 83:347-59