Weight of Evidence for the Fleischmann-Pons Effect

Rodney Johnson and Michael Melich
W. E. Meyer Institute of Systems Engineering,
Naval Postgraduate School, Monterey, CA[1]

Abstract

D. Cravens and D. Letts [1] have analyzed a portion (167 papers) of the published literature reporting on D_2O electrolysis experiments such as Fleischmann and Pons's (FP). They identify four criteria for what constitutes a "proper" FP experiment and state that experiments that satisfy all four criteria are likely to succeed in producing excess heat, while those that do not are likely to fail. This paper presents results of using a Bayesian network for probabilistic analysis of this claim. Consideration of a small subset of the papers (eight) is sufficient to give a likelihood ratio of about 10 to 1 in favor, and this number appears to grow generally rapidly, though not monotonically, as more papers are added to the set.

1. Introduction

Some of us, when asked why we tend to accept the reality of the Fleischmann-Pons effect, reply with the statement:

> *"It's not any one experiment; it's the number and variety of confirmations by independent researchers around the world."*

More generally, independent replication is considered an important step in acceptance of new experimental results. This paper reports an attempt to model the situation using a Bayesian network: a proposition ("The effect is real") with a number of pertinent reports, each open to doubt, but collectively sufficient to convert initial doubt (a low prior probability) into acceptance (a high posterior probability, conditional on the evidence).

1.1. Cravens-Letts database

D. Cravens and D. Letts [1] report a study of 167 selected papers concerning heat generation in electrochemical systems of the "classical" Fleischmann-Pons type: electrolytic cells with Pd cathodes in D_2O-based electrolyte. The list spans the years 1989–2007 and is non-exhaustive mainly because papers were included only if available in digital form. The authors rated the papers, when possible, according to four *yes/no* "enabling criteria," related to (1) cathode loading, (2) good chemical procedures, (3) operating current densities, and (4) non-equilibrium operation. (See the paper [1] for detailed statements of how the criteria were assigned.) In addition they assigned a *yes/no* value according to whether excess power was reported or not.

---

[1] The views expressed herein are those of the authors and not necessarily those of the U.S. Government, Department of Defense, Department of the Navy, or the Naval Postgraduate School.

They succeeded in rating 122 of the 167 papers and, after statistical analysis, concluded that production of excess power was highly correlated with the number of criteria satisfied—very likely if all four were met and less likely if fewer were met. We here report on a probabilistic test of that claim by use of a Bayesian network.

1.2. What is the problem?

We are interested in questions such as:

> "*Given* that in paper #1, where all 4 criteria were met, heat was observed, *and* in paper #2, where only 2 criteria were met, no heat was observed, *and . . .* in paper #167, . . . heat was observed, *then* what can we say about the probability that the FP effect is "real"? And what is the probability that a new experiment satisfying all 4 criteria will produce excess heat?"

In condensed-matter nuclear science in general we face multiple observations and experimental results, and multiple conjectures and hypotheses that might explain them.

To illustrate, consider the propositions:

> *A*: Nuclear reactions occur at low temperature in solids.
>
> *B*: Excess heat is observed.
>
> *C*: Helium production is observed.
>
> *D*: Emission of energetic particles is observed.

Then *B, C,* and *D* are observations that can serve as evidence in support for *A*, considered as a hypothesis. Likewise, consider propositions:

> *E*: Known nuclear reactions & quantum many-body effects.
>
> *F*: "New physics".[1]
>
> *G*: Error / deception.
>
> *H*: Excess heat is reported.

Then *E, F,* and *G* are alternative hypotheses that might explain observation *H.* The relations between the propositions are shown schematically in Figs. 1 and 2. These are simple examples of *Bayesian networks,* which are discussed in Section 2.2 below.

---

[1] . . . whatever we might choose to mean by the phrase. The propositions listed here are informal, abbreviated, and intended primarily as illustration.
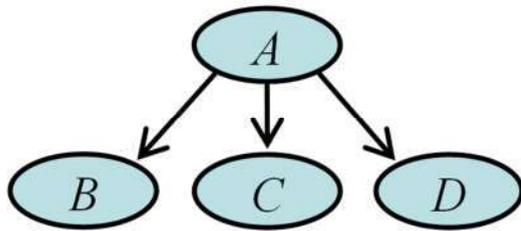
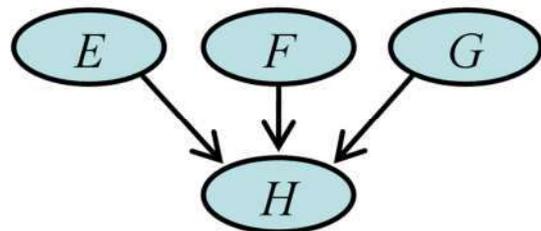Figure 1. Multiple support for a hypothesis   Figure 2. Alternative explanations

2. Bayesian methods

In general there may be more complicated interrelations (as in Fig. 3 further down). We need help in thinking quantitatively about such problems, and probability theory provides tools for doing so. *Bayes's rule* (or *Bayes's theorem*) is a fundamental rule of probability, used in updating the probability of a proposition in the light of new information. There are various methods based on it (called "Bayesian"), including *Bayesian networks,* which allow representing complex relations between propositions and making inferences concerning their probabilities.

2.1. Rules for probability

The degree of credence we accord to a proposition is (or should be) subject to change when we learn new relevant information. In quantitative terms, if A is a proposition to which we have initially assigned a probability $P(A)$, and we then obtain new information in the form of a proposition B, we update the probability of A to a quantity $P(A \mid B)$, the *conditional probability of A, given B*. One also uses the terms *prior* and *posterior* probabilities for $P(A)$ and $P(A \mid B)$, respectively. The process could continue, of course. Obtaining further new information, say C, leads to $P(A \mid BC)$, and so on. In this section we collect some basic rules, prominent among them Bayes's theorem, for dealing with conditional probabilities. We recommend the textbook by Jaynes [2] for (along with much else) a thorough discussion of what we here touch on lightly.

2.1.1. Bayes example problem

It is common in textbooks to introduce Bayes's theorem with an example: medical screening. Say you are a doctor screening for an uncommon but serious disease, where "uncommon" means

> 1% of people in the general population have the disease.

Also suppose there is a quite reliable test for the disease:

> 98% of people with the disease will test positive;
> 95% of those without the disease will test negative.

You give one of your patients the test as part of a routine physical, and the results come back positive. Do you tell the patient: "There is a 98% chance that you have a serious disease"?

We will see that the probability is actually closer to 16.526%, or about one chance in six, not 98%. Your patient is probably healthy. (Expensive or risky treatment is unjustified. But more testing is mandatory; ignoring a 1 in 6 chance amounts to Russian roulette.)

We express the given information symbolically:[1]

D: disease T: test positive
D': no disease T': test negative

$P(D) = 0.01$: probability of having the disease in the absence of test results
$P(D \mid T)$: the conditional probability of having the disease, given positive test results.

We want $P(D \mid T)$. We have $P(D)$ and two other conditional probabilities:

$P(T \mid D) = 0.98$: probability of a positive test, given that the disease is present;
$P(T' \mid D') = 0.95$: probability of a negative test, given that the disease is absent.

2.1.2 Rules

Product rule: probability that A and B are both true

$$P(AB) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B)$$

Bayes's rule:

$$P(A \mid B) = P(B \mid A)\,P(A)\,/\,P(B)$$

Sum rule:

$$P(B) = P(B \mid A)\,P(A) + P(B \mid A')\,P(A') + P(B \mid A'')\,P(A'') + \ldots$$

where A, A', A'', … are an exhaustive set of mutually exclusive propositions— that is, one must be true, but no two can be true at once.

Bayes's theorem follows directly from the product rule: divide by $P(B)$. The sum rule is useful for evaluating the denominator $P(B)$ on the right-hand side of Bayes's rule. Some variants of these rules can be useful; we may use

$$P(A \mid B) = P(AB)\,/\,P(B) \tag{1}$$

in place of Bayes's rule as just given, and we may use the sum rule in the form

$$P(B) = P(AB) + P(A'B) + P(A''B) + \ldots \tag{2}$$

2.1.3 Solution of example problem

Applying Bayes's rule to the previously given probabilities gives

---

[1] In general (as in the "sum rule" of Section 2.1.2) we use a notation such as A, A', A'', … to denote a set of propositions exactly one of which is true. Here we assume that D and D' are such a set (one either has the disease or one doesn't) and likewise for T and T' (only two test results are possible: positive and negative). In this special case of just two alternatives, one can read the prime symbol as logical negation: *not-D* for D' and *not-T* for T'.

$$P(D \mid T) = P(T \mid D) \, P(D) \, / \, P(T)$$
$$= 0.98 \times 0.01 \, / \, P(T)$$

and the sum rule gives

$$P(T) = P(T \mid D) \, P(D) + P(T \mid D') \, P(D')$$
$$= 0.98 \times 0.01 + 0.05 \times 0.99$$
$$= 0.0098 + 0.0495 = 0.0593$$

where we have used the fact that $P(T \mid D') = 1 - P(T' \mid D') = 1 - 0.95 = 0.05$. Finally,

$$P(D \mid T) = 0.98 \times 0.01 \, / \, 0.0593 = 0.16526$$

which is the stated result of 16.526%, or about one chance in six.

2.2. Bayesian networks

A Bayesian network is a graphical representation of complex relations between propositions; it allows inferences concerning their probabilities. Figure 3 shows an example slightly more general than the ones shown in Figs. 1 and 2.
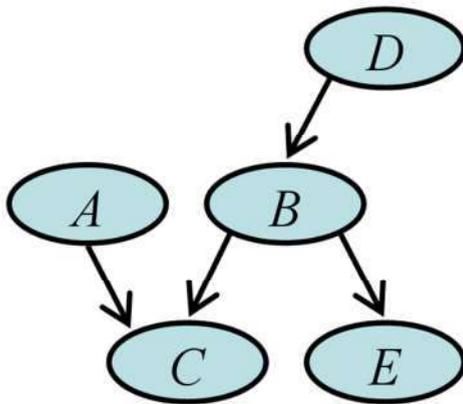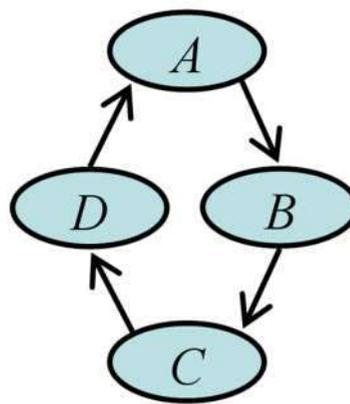

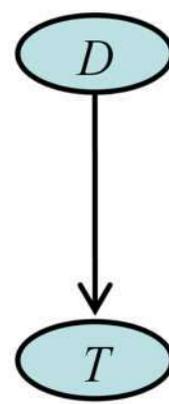**Figure 3. Bayesian network.**   **Figure 4. Loop (not allowed)**   **Figure 5. Medical screening example**

A Bayesian network consists of nodes connected by arrows. Loops, as in Fig. 4, can lead to contradictions and are not allowed. (This means that the network is a *directed acyclic graph*.) With each node is associated a "random variable" (such as A, B, C, … in Fig. 3). By calling a variable such as A "random" we mean simply that:

(1) There is a set of possible values $\{a_1, a_2, \dots, a_n\}$, so that the propositions $A = a_1$, $A = a_2$, …, $A = a_n$ form an exhaustive set of mutually exclusive propositions; and

(2) We can talk about probabilities (perhaps conditional) of these propositions, e.g. $P(A = a_i)$, $P(B = b_j \mid A = a_i)$.

True-false proposition, such as *D* and *T* of the medical screening example, are included (see Fig. 5); the set of values is just {*true*, *false*}.

Arrows indicate conditional dependence. If there is an arrow from a node *X* to a node *Y*, we call *X* a *parent* of *Y*. Thus the parents of *C* in Fig. 3 are *A* and *B*. A variable has a probability distribution conditional on its parents. In the case of *A*, *B*, and *C*, this means that conditional probabilities $P(C = c \mid A = a, B = b)$ are given for all values *a*, *b*, and *c* in the value sets of *A*, *B*, and *C*, respectively. This generalizes in a straightforward way to any number of parents. For a node without parents, such as *A*, we require the unconditional probabilities $P(A = a)$ for each *a*.

Bayesian networks can be used for updating our probabilities for values of some variables when we obtain new information in the form of values for other variables. This generalizes what we did in the medical screening example. There, we learned the value *T = true* for the test result, making it no longer uncertain (or "random"). Consequently we were able to update our probability for *D*, disease, from the prior value $P(D)$ to the posterior value $P(D \mid T = true)$. Analogously, we could suppose we learn values for some of the variables, say *C* and *E*, in the more elaborate network of Fig. 3, and we could ask how the new information affects the probabilities for the values of some other variable or variables, such as *B*.

To begin, in terms of the conditional and unconditional probabilities associated with the nodes, we can write an expression for the *joint* probability distribution for the entire set of variables; for the illustrative network of Fig. 3, this is the set of probabilities $P(A = a, B = b, C = c, D = d, E = e)$ that *A = a and B = b and C = c and D = d and E = e*, where *a*, *b*, *c*, *d*, and *e* range over their respective value sets. We show this in a shorthand notation, writing *A* for *A = a*, *B* for *B = b*, etc., so that the desired set of probabilities is denoted by $P(ABCDE)$; they are then given by:

$$P(ABCDE) \;=\; P(A)\, P(B \mid D)\, P(C \mid AB)\, P(D)\, P(E \mid B) \tag{3}$$

In general there is one factor for each node, consisting of the associated probability expression (conditional or unconditional). From this we can calculate other conditional probabilities such as $P(B \mid CE)$, for example: the updated probabilities for *B*, given that we have learned values for *C* and *E*.

By equation (1), the alternative form of Bayes's rule from §2.1.2, we can write:

$$P(B \mid CE) \;=\; P(BCE) \,/\, P(CE) \tag{4}$$

We can get the numerator, $P(BCE)$, by using essentially the alternative form of the sum rule, equation (2) from §2.1.2: sum (3) over the variables that do not occur in $P(BCE)$:

$$P(BCE) \;=\; \sum_{a,d} P(A = a, B, C, D = d, E)$$

Likewise we get the denominator by summing over *B* as well:

$$P(CE) \;=\; \sum_{a,b,d} P(A = a, B = b, C, D = d, E) \;=\; \sum_{b} P(B = b, C, E)$$

And the last two equations allow us to compute the desired quotient in (4).

For more information about Bayesian networks, see the textbook by Jensen [3], for example. There are also useful on-line tutorials by Breese and Koller [4] and by Murphy [5].

Software support is necessary for work with networks of any substantial size. For the work reported here we took advantage of a Java applet written by Yap, Santos, et al. [6] at the University of British Columbia and made available for download.
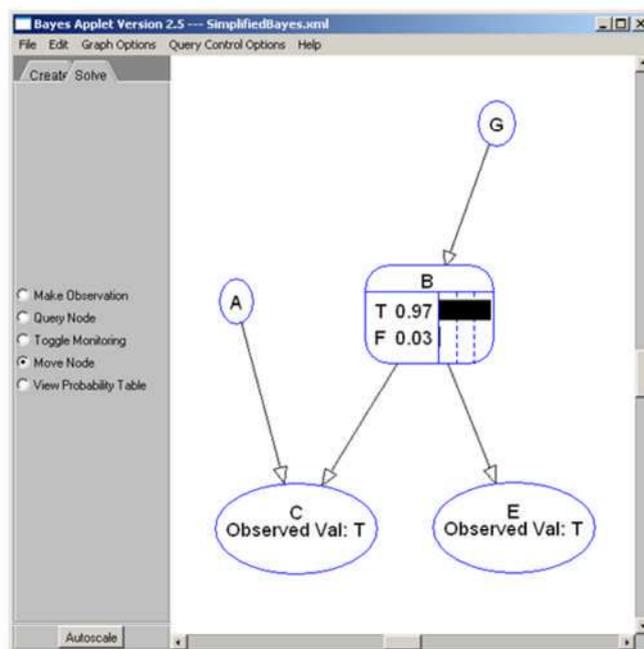

Figure 6. Bayesian network applet

This allows one to draw a network by means of a graphical interface, enter conditional probabilities in tabular form, set observed values for selected nodes, and display the resulting probabilities for other nodes. Figure 6 shows the display for a network similar to the one in Fig. 3.

2.3. Weight of evidence

For inference about a *yes/no* proposition, a formulation of Bayes's theorem in terms of *odds* and *likelihoods ratios* can be useful. First, a bit of terminology: The quantities $P(B \mid A)$, $P(B \mid A')$, $P(B \mid A'')$, … that occur in the sum rule (§2.1.2) are called the *likelihoods* of *A, A', A'',* … .[1] For a pair of alternatives, *A* and *A'*, the quotient $P(B \mid A) / P(B \mid A')$ is called the likelihood ratio. When these are the only alternatives, we have $P(A) / P(A') = P(A) / (1 - P(A))$; this quantity is the (prior) odds for *A* and denoted by $O(A)$. Similarly, the posterior odds for *A* are $O(A \mid B) = P(A \mid B) / P(A' \mid B)$.

Now write Bayes's rule for *A* and for *A'*:

---

[1] Recall that *A, A', A'',* . . . form an exhaustive set of mutually exclusive propositions. "Likelihood" is used in a technical sense. The terminology is unfortunate because it may give the impression that the likelihoods are conditional probabilities of *A, A', A'',* . . . , which they are not; in particular they need not sum to 1.

$$P(A \mid B) = P(A)\, P(B \mid A) / P(B)$$
$$P(A' \mid B) = P(A')\, P(B \mid A') / P(B)$$

and divide the first equation by the second. The factors of $P(B)$ cancel, and we get:

$$P(A \mid B) / P(A' \mid B) = [P(A) / P(A')]\, [P(B \mid A) / P(B \mid A')]$$

The left-hand side is the posterior odds for A, the first factor on the right is the prior odds, and the second factor is the likelihood ratio. Thus:

$$O(A \mid B) = O(A)\, [P(B \mid A) / P(B \mid A')]$$

which we can state as:

"posterior odds $=$ prior odds $\times$ likelihood ratio"

If the "evidence" B consists of several observations B_1, B_2, … that are *independent* in the sense that $P(B_1 B_2, \ldots \mid A) = P(B_1 \mid A)\, P(B_2 \mid A) \ldots$ and $P(B_1 B_2, \ldots \mid A') = P(B_1 \mid A')\, P(B_2 \mid A') \ldots$, then the equation generalizes to

$$O(A \mid B_1 B_2, \ldots) = O(A)\, [P(B_1 \mid A) / P(B_1 \mid A')]\, [P(B_2 \mid A) / P(B_2 \mid A')] \ldots$$

Taking logs of all the factors gives an additive version. Thus taking a new piece of independent evidence B_i into account just increments the log of our odds for A by

$$\log\, [P(B_i \mid A) / P(B_i \mid A')]$$

which is called the *weight of evidence* for A provided by B_i. (See Good [7] and Jaynes [2, pp. 91 ff.].)

If one starts with noncommittal prior odds of 1:1, evenly balanced between acceptance and rejection of a proposition, then the likelihood ratio of the evidence gives ones posterior odds. On the other hand, one can view the reciprocal of the likelihood ratio as a "critical prior": the prior odds such that the evidence would bring us to posterior odds of 1:1. In this latter role, the likelihood ratio can help us in assigning a numerical value to our prior odds for a preposition; imagine a successions of independent repetitions B_1, B_2, … of an experiment with a given likelihood ratio and ask how many successful outcomes would bring us to a state of uncertainty, poised between acceptance and rejection. (See Good [7] and Jaynes [2, Ch. 5].)

Our task will be to evaluate the likelihood ratio (equivalently, the weight of evidence) for the proposition that "the FP effect is real" provided by Cravens and Letts's ratings of a subset of the papers in their database.

2.4. Estimating probabilities

In the medical example we were *given* the values $P(T \mid D) = 0.98$, $P(T' \mid D') = 0.95$, $P(D) = 0.01$. In practice such numbers are commonly gotten from a study, e.g. give the test to some people known to have the disease, and observe that about 98% test positive. The numbers are known only with some uncertainty, e.g. "The fraction of people with the disease who test

positive is in the range 0.980 ± 0.002 with probability 68%. This seems to be saying that $P(T \mid D)$ is in a certain range with a certain probability. What do we mean by the probability of a statement about other probabilities?[1]

Our treatment of Cravens and Letts's evidence will involve probabilities that are not known in advance but are estimated from the data. To illustrate the considerations involved, we present a simple problem.

The "biased coin" problem concerns a coin for which the probability p of heads is some arbitrary number between 0 and 1, not known to us and not necessarily 0.5. It is not at all clear how one could construct such an object in practice,[2] so it may be better to think of a game spinner with two sectors, marked H and T, with H containing a fraction p of the full circle (Fig. 7).
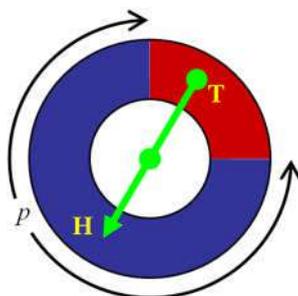

Figure 7. "Biased coin" spinner

If we spin so that the probable location of the pointer is uniformly distributed over the circle, the probability of its showing heads is p. Now write H_p for the proposition that the size of the H sector is p, and suppose that this unknown size was chosen at random (uniformly) between 0 and 1 (Fig. 8). We are now dealing with continuous probability distributions; $P(H_p)$ is a probability *density,* not a discrete value, and satisfies $\int P(H_p)\,dp = 1$ rather than $\sum_p P(H_p) = 1$.

Suppose we spin once and observe a head. What is our revised probability for H_p, given E_{11}: one head in one trial? Bayes's rule for continuous probability distributions gives:

$$P(H_p \mid E_{11}) = P(E_{11} \mid H_p)\,P(H_p) / P(E_{11})$$
$$= p / P(E_{11})$$

---

[1] The need to take systematic account of uncertainties in our information is ubiquitous and has a long history. It played a central role in Laplace's comparison of imperfect astronomical observations with Newtonian gravitational theory; see Jaynes [8; 2, Ch. 5].

[2] We might try loading a coin by making it of two layers with lead on one side and aluminum on the other. This turns out not to be effective; see Jaynes [2, §10.3], "How to cheat at coin and die tossing." Jaynes shows in fact that the probability of heads is not just an intrinsic physical property of the coin and may have little to do with quantities such as the displacement of the center of gravity of the coin from its geometrical center.
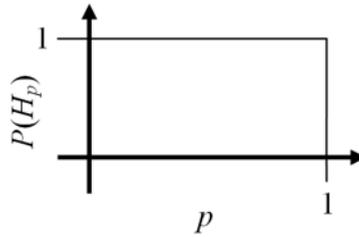
Figure 8. Uniform prior $P(H_p)$

Here $P(E_{11} \mid H_p)$ is p, because that's what H_p says: the probability of getting a head is p. And $P(H_p)$ is 1 by assumption.

The continuous version of the sum rule (§2.1.2) gives

$$P(E_{11}) = \int_0^1 P(E_{11} \mid H_p) P(H_p) \, dp$$

$$= \int_0^1 p \, dp = 1/2$$

$$P(H_p \mid E_{11}) = 2p \tag{5}$$
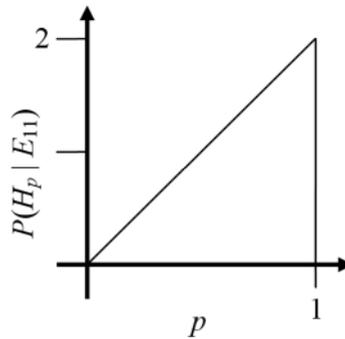
as in Fig. 9.

Figure 9. One head in one trial observed

Now the probability of heads on the next trial is:

$$P(\text{"one more head"} \mid E_{11}) = \int_0^1 P(\text{"one more head"} \mid E_{11} H_p) \, P(H_p \mid E_{11}) \, dp$$

The first factor in the integrand is p by definition of H_p, and equation (5) gives the second. So

$$P(\text{"one more head"} \mid E_{11}) = \int_0^1 2p^2 \, dp = 2/3$$

We can continue making trials and updating our probability distribution for H_p. Possible results are shown in Figs. 10–12.

With the notation E_{mn} = "m heads observed in n trials," these represent:

$$P(H_p \mid E_{22}) = 3p^2 \qquad P(H_p \mid E_{12}) = 6p(1-p) \qquad P(H_p \mid E_{24}) = 30\, p^2(1-p)^2$$
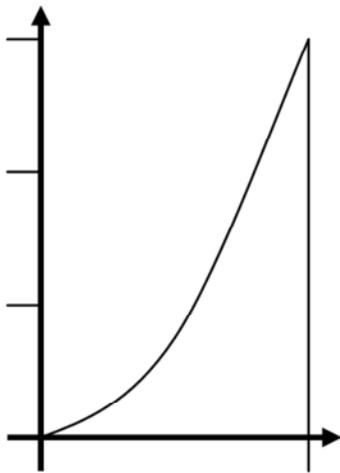

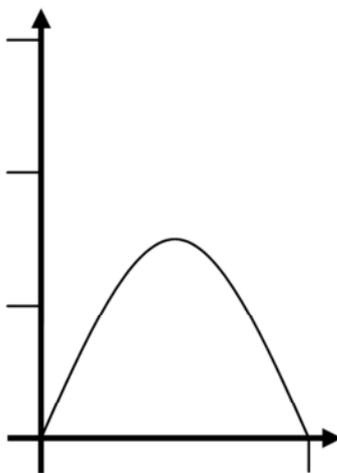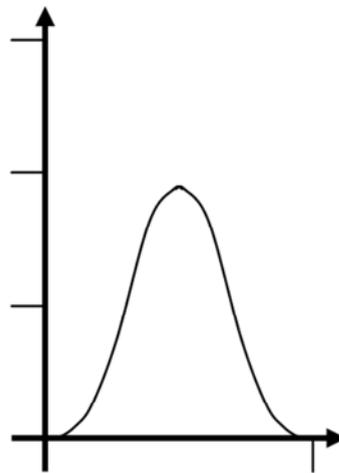**Figure 10. 2 heads in 2 trials**   **Figure 11. 1 heads in 2 trials**   **Figure 12. 2 heads in 4 trials**

The general formula for m heads out of n trials is:

$$P(H_p \mid E_{mn}) = [(n+1)! \,/\, m!(n-m)!]\, p^m(1-p)^{n-m}$$

and the formula for the probability of heads on the next trial is:

$$P(\text{"one more head"} \mid E_{mn}) = (m+1)\,/\,(n+2).$$

This is Laplace's *rule of succession*: assuming a uniform prior for H_p and m "successes" out of n independent trials, the probability μ of success on the next trial is given by

$$\mu = (m+1)\,/\,(n+2)$$

The successive posterior distributions peak up more and more sharply as the number of trials increases (Fig. 13). The width of the peak is 2σ, where the standard deviation σ is given by

$$\sigma = \sqrt{\mu(1-\mu)/(n+3)}$$

and the mean μ is as just given. For a derivation of σ, see equation (6.35) in Jaynes [2, ch. 6].

The assumption of a uniform prior may or may not be justified, depending on available information. But if the prior is continuous and non-zero near μ, the shape of the posterior will often be found to resemble Fig. 13.
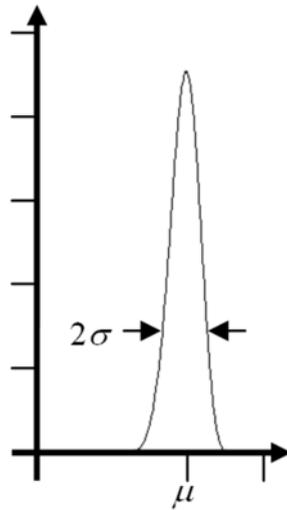
Figure 13. Peak shape

3. Problem setup

The network we developed is shown in Fig. 14. Node R is the proposition of interest—roughly speaking, "is the FP effect 'real'?" The nodes E_2, E_8, … , E_{28} refer to the results published in the set of papers selected for initial consideration; the subscripts are index numbers of the papers in the Cravens–Letts [1] database. The other "E" nodes are auxiliary nodes associated with the papers, and the "P" nodes are various probabilities to be estimated from the data by means illustrated in §2.4.

Figure 14. Network for eight selected papers (initial configuration)

3.1. Selected papers

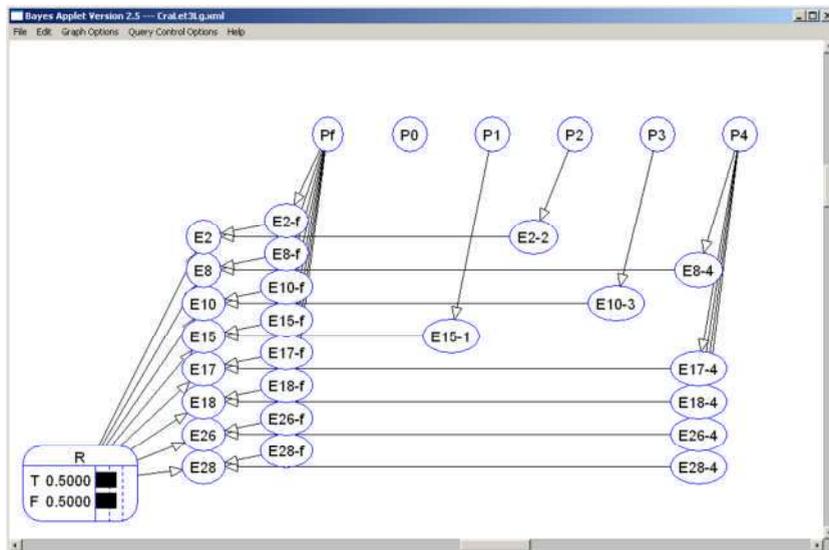Cravens and Letts [9] suggested the following eight papers for initial consideration:

#	Cri	Heat	Citation
2	2	No	R. D. Armstrong et al., *Electrochimica Acta* **34** (9) 1319–1322 (Sep. 1989).
8	4	Yes	R. C. Kainthla et al., *Electrochimica Acta* **34** (9) 1315–1318 (Sep. 1989).
10	3	No	N. S. Lewis et al., *Nature* **340** (6234) 525–530 (Aug. 17, 1989).
15	1	No	D. E. Williams et al., *Nature* **342** (6248) 375–384 (Nov. 23, 1989).
17	4	Yes	A. J. Appleby et al., *Proc. First Ann. Conf. Cold Fusion*, 32–43 (Mar. 1990).
18	4	Yes	Y. Arata & Y.-C. Zhang Proc. *Japan Acad. B* **66** (1) 1–6 (1990).
26	4	Yes	S. Guruswamy & M. E. Wadsworth, *Proc. First Ann. Conf. Cold Fusion*, 314–327, (Mar. 1990).
28	4	Yes	T. Lautzenheiser & D. Phelps, Amoco Production Company Research Report T-90-E-02, 90081ART0082 (Mar. 1990).

The numbers under "#" are the index numbers of the papers in Cravens and Letts's database. The numbers under "Cri" give the number of enabling criteria satisfied by the paper. A Yes or No under "Heat" indicates whether excess heat was reported.

3.2 Network propositions

Proposition R can also be phrased as "the experimental treatment makes a difference". We consider two alternatives:

- $R = false$:[1] the probability of observing excess heat is the same (P_f) regardless of whether all, some, or none of Cravens and Letts's enabling criteria are satisfied. This would imply that reported observations of excess heat are the result of error, deception, or extraneous factors.

- $R = true$: the probability of observing excess heat has one of several values ($P_0, \ldots, P_4$), depending on the number of enabling criteria that are satisfied.

E_i states that excess heat was reported in paper number i of the data base.

E_{if} states that excess heat was reported in paper number i in case $R = false$. Its truth value is irrelevant in case $R = true$. Its conditional probability is simply the value of P_f.

E_{in} states that excess heat was reported in paper number i in case $R = true$, where n is the number of enabling criteria met by the paper. Its truth value is irrelevant in case $R = false$. Its conditional probability is simply the value of P_n.

Nodes E_{if} and E_{in} exist to simplify the expression of the conditional probabilities of E_i, rather than for any intrinsic interest of their own. E_i is *true* if either (1) R and E_{in} are both *true* or (2) R is *false* and E_{if} is *true*; E_i is *false* otherwise. The E_{if} and E_{in} nodes could be eliminated and E_i made directly dependent on R, P_f, and P_n at the expense of expanding Table 2 below to a table with 50 rows.

---

[1] A statistician of the orthodox persuasion might call this alternative a "null hypothesis."

3.3   Network variables

P_f is the probability of excess heat being reported in case $R = false$.

P_n is the probability of excess heat being reported in an experiment satisfying n of the enabling criteria ($n = 0, \ldots ,4$) in case $R = true$.

P_f and $P_0, \ldots , P_4$ are probabilities to be estimated from the data by means illustrated in §2.4. Ideally they would each be described by a continuous probability density on the interval from 0 to 1. Because of practical limitations of the software, we used fairly coarse discrete approximations.

3.4. Probability tables

The prior and conditional probabilities for the nodes of the network are specified in tabular form.

We set the prior probability of R equal to 0.5, as shown in Table 1, giving prior odds of 1. Consequently the posterior odds are equal to the likelihood ratio. (See §2.3.) This makes it easy to determine the weight of evidence from the program outputs.

Table 1. $P(R)$

R	
true	false
0.5	0.5

The conditional probability of E_i is specified as in Table 2. This simply makes E_i agree with E_{if} when R is false and with E_{in} when R is true. The actual probability values are those of E_{if} in the first case and E_{in} in the second.

Table 2. $P(E_i \mid R\ E_{if}\ E_{in})$

R	E_{if}	E_{in}	E_i true	E_i false
true	true	true	1	0
true	true	false	0	1
true	false	true	1	0
true	false	false	0	1
false	true	true	1	0
false	true	false	1	0
false	false	true	0	1
false	false	false	0	1

The conditional probabilities of E_{if} and E_{in} are given in Tables 3 and 4. The probability of E_{if}, given P_f, is by definition simply the value of P_f; and the probability of E_{in}, given P_n, is the value of P_n.

Table 3. $P(E_{if} \mid P_f)$

	E_{if}	
P_f	true	false
0.1	0.1	0.9
0.3	0.3	0.7
0.5	0.5	0.5
0.7	0.7	0.3
0.9	0.9	0.1

Table 4. $P(E_{in} \mid P_n)$

	E_{in}	
P_n	true	false
0.1	0.1	0.9
0.3	0.3	0.7
0.5	0.5	0.5
0.7	0.7	0.3
0.9	0.9	0.1

The prior probabilities of P_f and P_n ($n = 0, \ldots ,4$) are shown in Tables 5 and 6. They are all the same: a coarse discrete approximation to a uniform distribution on the unit interval.

Table 5. $P(P_f)$

P_f				
0.1	0.3	0.5	0.7	0.9
0.2	0.2	0.2	0.2	0.2

Table 6. $P(P_n)$

P_n				
0.1	0.3	0.5	0.7	0.9
0.2	0.2	0.2	0.2	0.2

4. Results

After entering the probability tables in the nodes of the network of Fig. 14, we successively declared "observed" values for the nodes E_i, starting with *false* for E_2 and finishing with *true* for E_{28}. The final state of the network is shown in Fig. 15, in which display of the probability distributions of the nodes P_f, P_0, … , P_4, has been enabled.
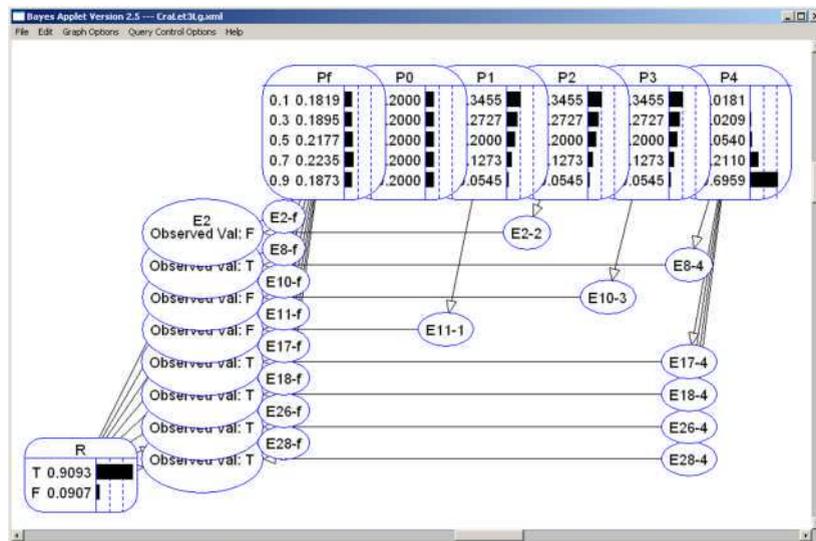

Figure 15. Final configuration of network for eight selected papers

The posterior probabilities for $R = true$ and $R = false$ are 0.9093 and 0.0907, giving posterior odds of 10.025. This is also the final value of the likelihood ratio, since we started with prior

odds of 1.0. The value of the likelihood ratio is plotted in Fig. 16 as a function of the number of papers taken into account, from 1 paper (#2 only), 2 papers (#2 and #8), through 8 papers.

The likelihood ratio for R, give 1 paper, is 1.0, exactly equal to the prior value of 1.0 with no papers at all (not plotted). With one paper, the distributions of P_f and P_2 were identical—a bit biased toward the low values, as the first paper (#2) reported no heat. There was not yet a basis for choosing between the two. Adding a second paper (#8, reporting heat) increased the ratio to about 1.47, and adding a third (#10, reporting no heat) made no difference. The next four, consistently showing heat with four criteria satisfied, brought a steep increase in the likelihood ratio to 10.025.

At that point the posterior distribution for P_4 was strongly biased toward high values, as shown in Fig. 15, and as one would expect. P_1, P_2, and P_3 showed weaker and identical biases toward low values, since in each case (1, 2, or 3 criteria met) one instance was observed, reporting no heat. P_0 was flat, unchanged from its prior, as no evidence was included bearing on the case of 0 criteria met. The distribution of P_f was relatively flat, close to its prior, as the probability for $R = false$ was at that point estimated as being rather small. Note that if R were definitely known to be *true*, the value of P_f would be irrelevant, and we would expect it to be no different from its prior. As an experiment, we changed the prior value for $R = false$ to 0.99, sufficiently skeptical that the posterior value came to about 0.91; in that case, the posterior distribution for P_f showed an apparent peak between 0.5 and 0.6, about where the rule of succession would predict for 5 successes (heat observed) out of 8 trials.
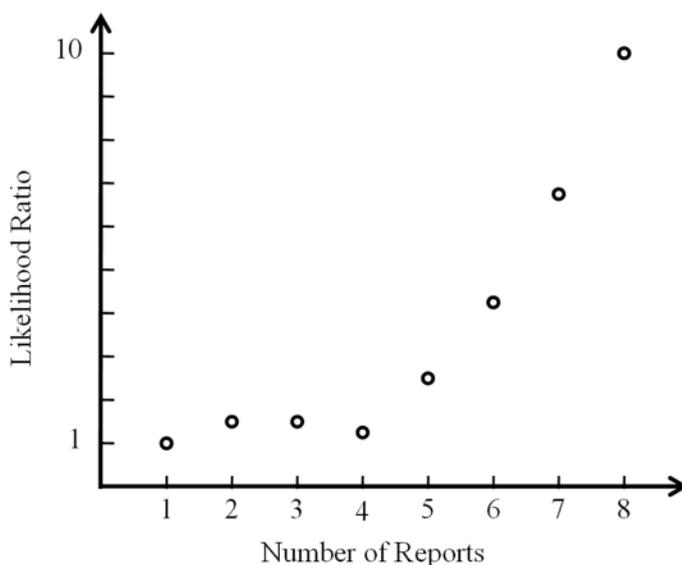

Figure 16. Change in likelihood ratio as more and more papers are taken into account

Eight papers is a small enough sample that no particular significance should be attached to the particular final numerical value of 10.025 for the likelihood ratio for R, though the qualitative behavior of Fig. 16 is suggestive. Moreover, the set of eight is not a representative sample of the data base; some were selected for historical significance. In particular, papers #10 and #15 are accounted by Cravens and Letts [1] as "the most important papers in the field of

Condensed Matter Nuclear Science" for their early and lasting negative impact. On the other hand the announcement by Fleischmann and Pons (#1 in the database) was omitted.

In addition, the model presented here does not address the important question of publication bias. We have looked at toy models for experimenter bias, but incorporating such factors in the present model is a matter for future work. The problem is less severe for the period immediately following the Fleischmann–Pons announcement, when a number of researchers were quite willing to report negative results. However, the more recent papers in the Cravens–Letts database are preponderantly positive, and incorporating them all in the computation would risk producing inflated values for the likelihood ratio for R unless the tendency not to report failures were taken into account.

It was nevertheless felt desirable to include a larger number of papers. Attempts to increase the number of papers from 8 to 12, while successful (Johnson and Melich [10]), made it clear that we were approaching the computational limits of the Java applet. Moreover the scheme we have described lumps together all papers that meet the same number of criteria; those meeting the first two enabling criteria are counted together with those meeting the last two. It was thought to be also desirable to consider particular subsets of the four criteria, rather than simply the count, expanding the number of cases from 5 to 16.

We have subsequently obtained an exact analytical expression for the likelihood ratio. This not only circumvents the computational limits, but it uses exact uniform priors for the P's (P_f, P_0, P_1, P_2, P_3, P_4) rather than the coarse discrete approximations of Tables 5 and 6. And it generalizes easily from 5 classes of papers (0, 1, … , 4 criteria satisfied) to 16 classes (all distinct subsets of the 4 criteria). Here is a sketch of the results.

To begin, remove the intermediate nodes E_{if}, E_{ik} in Fig. 14 so that, e.g., E_2 depends directly on P_f and P_2. The joint probability for the network becomes:

$$P(R)\, P(P_f)\, P(P_1) \ldots P(P_4)\, P(E_2 \mid R\, P_f P_2)\, P(E_8 \mid R\ P_f P_4) \ldots P(E_{28} \mid R\, P_f P_4)$$

and dividing by $P(R)$ gives the probability conditional on R:

$$P(\, \cdots \mid R)\, =\, P(P_f)\, P(P_1) \ldots P(P_4)\, P(E_2 \mid R\, P_f P_2)\, P(E_8 \mid R\ P_f P_4) \ldots P(E_{28} \mid R\, P_f P_4)$$

where $\cdots$ stands for all the variables other than R. The priors for the P's are uniform, i.e. 1. The conditional probabilities are given by, e.g.

Table 7. $P(E_2 \mid R\, P_f P_2)$

	E_2	
R	True	false
true	P_2	$1 - P_2$
false	P_f	$1 - P_f$

for $P(E_2 \mid R\, P_f P_2)$. Thus the joint probability is

$$P(\, \cdots \mid R = false)\, =\, (1 - P_f)\, P_f\, (1 - P_f)\, (1 - P_f)\, P_f\, P_f\, P_f\, P_f$$
$$P(\, \cdots \mid R = true)\, =\, (1 - P_2)\, P_4\, (1 - P_3)\, (1 - P_1)\, P_4\, P_4\, P_4\, P_4$$

for the case $E_2 = false$, $E_8 = true$, $E_{10} = false$, $E_{11} = false$, $E_{17} = true$, $E_{18} = true$, $E_{26} = true$, $E_{28} = true$. (For the subscripts in the second expression, recall that the values of k, number of criteria satisfied, are 2, 4, 3, 1, 4, 4, 4, 4 for the respective papers numbered 2, 8, …, 28.)

For a general collection of papers, not necessarily those of Fig. 14, the expressions are

$$P(\cdots|R = false) = P_f^M (1 - P_f)^{N-M}$$

$$P(\cdots|R = true) = \prod_{k=0}^{4} P_k^{m_k} (1 - P_k)^{n_k - m_k}$$

where N is the total number of papers, M is the number reporting excess heat, n_k is the number of papers in class k (i.e. k enabling criteria satisfied) and m_k is the number of those reporting excess heat. Integrating out the P's gives the probability of the observed quantities (criteria satisfied, heat reported) conditional on R:

$$P(\text{observed} \,|\, R = false) = M!(N - M)!/(N + 1)!$$

$$P(\text{observed} \,|\, R = true) = \prod_{k=0}^{4} m_k!(n_k - m_k)!/(n_k + 1)!$$

and the likelihood ratio is the second of these quantities divided by the first.

For the 8 selected papers,

$$n_0 = 0, \quad n_1 = 1, \quad n_2 = 1, \quad n_3 = 1, \quad n_4 = 5 \quad N = 8$$
$$m_0 = 0, \quad m_1 = 0, \quad m_3 = 0, \quad m_4 = 0, \quad m_5 = 5 \quad M = 5$$
$$P(\text{observed} \,|\, R = false) \;=\; 1/504$$
$$P(\text{observed} \,|\, R = true) \;=\; (1/1)(1/2)(1/2)(1/2)(1/6) \;=\; 1/48$$

The likelihood ratio is therefore 504/48, or 10.5 (exactly). This is to be compared with the value of 10.025 obtained numerically with the applet, using a 5-point discrete approximation to a uniform prior for the P's. It appears that the use of such a discrete approximation has not resulted in serious error in the computed likelihood ratio.

There is no problem in generalizing from 5 classes of papers (0, 1, … , 4 criteria satisfied) to 16 classes (all distinct subsets of the 4 criteria): just let k above index the 16 classes and adjust the limits of k in the products accordingly.

Acknowledgements

References

1. D. Cravens and D. Letts, "The enabling criteria of electrochemical heat: beyond a reasonable doubt," *Proc. 14th International Conference on Condensed Matter Nuclear Science (ICCF-14)*, Washington, D.C., August 10–15, 2008
2. E. T. Jaynes, *Probability Theory: the Logic of Science*, Cambridge University Press, Cambridge, UK, 2003
3. F. V. Jensen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York, 2001.

4. J. Breese and D. Koller, "Tutorial on Bayesian networks," <http://ai.stanford.edu/~koller/BNtut/BNtut.ppt>, 1997

5. K. Murphy, "A brief introduction to graphical models and Bayesian networks," <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>, 1998

6. A. Yap, J.R. Santos et al. (CIspace Group, Laboratory for Computational Intelligence, University of British Columbia), available for download under "Belief and Decision Networks" at <http://www.aispace.org/downloads.shtml>; see also "Terms of Use" at <http://www.aispace.org/about.shtml>

7. I. J. Good, *Probability and the Weighing of Evidence*, C. Griffin, London, 1950

8. E. T. Jaynes, "Confidence intervals vs. Bayesian intervals," in W. L. Harper and C. A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. II, 175–257, D. Reidel Publishing Company, Dordrecht, Netherlands, 1976

9. D. Cravens and D. Letts, personal communication

10. R. Johnson and M. Melich, "Weight of Evidence for the Fleischmann-Pons Effect," in J. Marwan (ed.), *Low-Energy Nuclear Reactions Sourcebook*, American Institute of Physics, in press