J. Condensed Matter Nucl. Sci. 4 (2011) 225–240

Research Article

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 93943, USA

Abstract

Cravens and 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 (12) is sufficient to give a likelihood ratio of about 28 to 1 in favor, and this number appears to grow generally rapidly, though not monotonically, as more papers are added to the set.

Keywords: Bayesian networks, Cold fusion, Fleischmann–Pons effect, Weight of evidence

PACS: 01.30.Rr, 02.50.Cw, 25.45, 82.45.

1. Introduction

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

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

Independent replication is considered as an important step in acceptance of new experimental results. We here report an attempt to model in formal terms the intuition that[1] if a number of published reports have a significant probability of being right about excess heat, then "having a lot dramatically increases the probability."

We use Bayesian probability theory, a discipline with a long history of use in dealing precisely and systematically with uncertain information. (This dates back at least to its use by Laplace in comparing imperfect astronomical

*The views expressed herein are those of the authors and not necessarily those of the US Government, Department of Defense, Department of the Navy, or the Naval Postgraduate School.

†E-mail: melich@alumni.rice.edu

[1] . . . to paraphrase an anonymous reviewer of this article.

observations in celestial mechanics with Newton's gravitational theory. We are persuaded in part by the arguments of Jaynes [2].[2]) Specifically we use a *Bayesian network* (see the subsection with that title) a tool for dealing with complex sets of interrelated propositions. Here we have a proposition, "the FPE is real" with other propositions concerning a number of pertinent reports, each singly open to doubt but together sufficient in some cases to convert initial skepticism into acceptance.

We summarize results in terms of the *weight of evidence* (see the subsection with that title) a quantity that indicates how one should modify a probability assessment in the light of new evidence. Readers will each bring their own prior information to the table. Those who give a sufficiently low prior probability to the FPE will remain unpersuaded after taking the contents of the literature into account, though they should be less skeptical than before. Those who give a probability of flat-out 0 will retain that value, having precluded the possibility of learning. Those who already accept the FPE will find confirmation of their assessment. And those who are not sure of their probability assignment may find help in giving it a numerical value (see subsection "Weight of Evidence" and the references there to Jaynes and I.J. Good.)

1.1. Cravens—Letts database

Cravens and Letts [1] report a study of 167 selected papers concerning heat generation in "classical" Fleischmann–Pons electrolytic systems: 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 details.) In addition they assigned a *yes/no* value according to whether excess power was reported. They successfully rated 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.

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'?"

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:

[2] "In this workshop we are venturing into a smoky area of science where nobody knows what the real truth is. [Always in such fields] supreme self-confidence takes the place of rational arguments. Therefore we will try to avoid dogmatic assertions . . . We think that the original viewpoint of James Bernoulli and Laplace [about probability] offers some advantages today in both conceptual clarity and technical results for currently mysterious problems."–Jaynes [3]

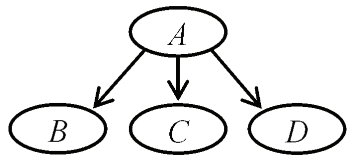

Figure 1. Multiple support for a hypothesis.

E: Known nuclear reactions & quantum many-body effects.
F: "New physics".[3]
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 *Bayesian networks*; see the subsection below with that title.

2. Bayesian Methods

In general there may be more complicated interrelations (as in Fig. 3). 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 formal 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 $P(A|B)$, the *conditional probability* of A, given B. One also uses the terms *prior* and *posterior* probabilities for $P(A)$ and $P(A|B)$, respectively. The process could continue, of course. Obtaining further new information, say C, leads to $P(A|BC)$, 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:

[3] . . . whatever we might choose to mean by the phrase. The propositions listed here are informal, abbreviated, and intended primarily as illustration.

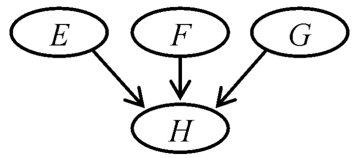

Figure 2. Alternative explanations.

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:[4]

D: disease T: test positive
D': no disease T': test negative
$P(D) = 0.01$: probability of disease in the absence of test results
$P(D|T)$: the conditional probability of disease, given positive test results

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

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

2.1.2. *Rules*

We collect here some basic rules of probability theory. These will be used in evaluating $P(D|T)$.

Product rule: probability that A and B are both true

$$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

Bayes's rule:

$$P(A|B) = P(B|A)P(A)/P(B)$$

Sum rule:

$$P(B) = P(B|A)P(A) + P(B|A')P(A') + P(B|A'')P(A'') + \cdots$$

where

$$A, A', A'', ...$$

[4] In general (as in the "sum rule" of the next sub-subsection) 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'.

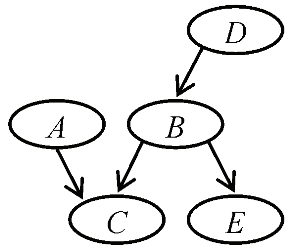

Figure 3. Bayesian network.

are an exhaustive set of mutually exclusive propositions – that is, one must be true, but no two can be true at once.

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

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

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

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

2.1.3. *Solution of example problem*

Applying Bayes's rule to the previously given probabilities gives

$$P(D|T) = P(T|D)P(D)/P(T) = 0.98{\times}0.01/P(T)$$

and the sum rule gives

$$\begin{aligned} P(T) &= P(T|D)P(D) + P(T|D')P(D') \\ &= 0.98 \times 0.01 + 0.05 \times 0.99 \\ &= 0.0098 + 0.0495 = 0.0593 \end{aligned}$$

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

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

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.

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 $\{a_1, a_2, \ldots, a_n\}$, 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. $P(A = a_i)$, $P(B = b_j|A = a_i)$.

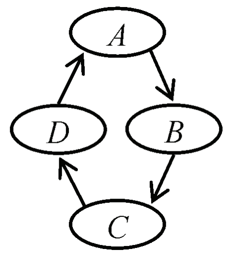

Figure 4. Loop (not allowed).

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 $P(C = c|A = a, B = b)$ 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 the medical screening example. There, we learned the value $T = \textit{true}$ for the test result, making it no longer uncertain. Consequently we could update our probability for D, disease, from the prior value $P(D)$ to the posterior value $P(D|T = \textit{true})$. Analogously, if we learn values for some of the variables, say C and E, in the more elaborate network of Fig. 3, we can ask how the new information affects the probabilities for the values of other 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:

$$P(ABCDE) = P(A)P(B|D)P(C|AB)P(D)P(E|B). \tag{3}$$

In general there is one factor for each node, consisting of the associated probability expression (conditional or unconditional).

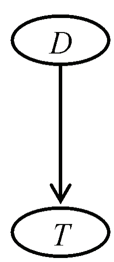

Figure 5. Medical screening example.

From this we can calculate other conditional probabilities such as $P(B|CE)$, for example: the updated probabilities for B, given that we have learned values for C and for E. By Eq. (1), the alternative form of Bayes's rule given above, we can write:

$$P(B|CE) = P(BCE)/P(CE). \tag{4}$$

We can get the numerator, $P(BCE)$, by using essentially the alternative form of the sum rule, Eq. (2) above: 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 [4], for example. There are also useful on-line tutorials by Breese and Koller [5] and by Murphy [6].

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. [7] at the University of British Columbia and made available for download. 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.

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 $P(B|A)$, $P(B|A')$, $P(B|A'')$, ... that occur in the sum rule (under "Rules" above) are called the *likelihoods* of $A, A', A'', ...$[5] For a pair of alternatives, A and A', the quotient $P(B|A)/P(B|A')$ 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 $O(A|B) = P(A|B)/P(A'|B)$.

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

$$P(A|B) = P(A)P(B|A)/P(B)P(A'|B) = P(A')P(B|A')/P(B)$$

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

$$P(A|B)/P(A'|B) = [P(A)/P(A')][P(B|A)/P(B|A')].$$

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:

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

which we can state as:

[5] 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.

"posterior odds = prior odds × likelihood ratio"

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

$$O(A|B_1 B_2, ...) = O(A)[P(B_1|A)/P(B_1|A')][P(B_2|A)/P(B_2|A')]...$$

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

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

which is called the *weight of evidence* for A provided by B_i (see [8, 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 $B_1, B_2, \ldots$ 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 [8, 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 $P(T|D) = 0.98$, $P(T'|D') = 0.95$, $P(D) = 0.01$. In practice such numbers are often gotten from a study: 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 $P(T|D)$ is in a certain range with a certain probability. What do we mean by the probability of a statement about other probabilities?[6].

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,[7] 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 circle (Fig. 6). 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. 7). 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 $\sum_p P(H_p) = 1$.

[6] The need to take systematic account of uncertainties in our information is ubiquitous and has a long history–cf. our previous mention of Laplace ([9, 2, Ch. 5])

[7] 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 [2, Chapter 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.

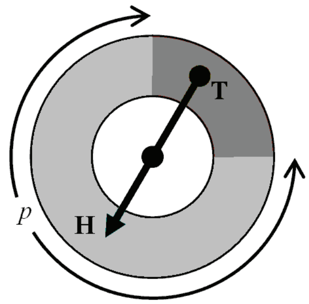

Figure 6. "Biased coin" spinner.

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

$$P(H_p|E_{11}) = P(E_{11}|H_p)P(H_p)/P(E_{11}) = p/P(E_{11})$$

Here $P(E_{11}|H_p)$ 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 ("Rules" above) gives

$$P(E_{11}) = \int_0^1 P(E_{11}|H_p)P(H_p)dp = \int_0^1 pdp = 1/2$$

$$P(H_p|E_{11}) = 2p \tag{5}$$

as in Fig. 8.

Now the probability of heads on the next trial is:

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

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

$$P(\text{"one more head"}|E_{11}) = \int_0^1 2p^2 dp2/3$$

We can continue making trials and updating our probability distribution for H_p. With the notation E_{mn} = "m heads observed in n trials," we obtain the general formula:

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

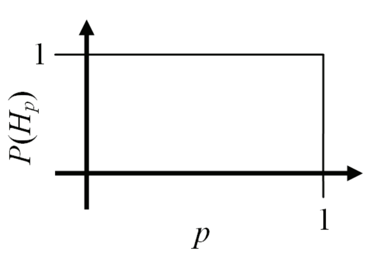

Figure 7. Uniform prior $P(H_p)$.

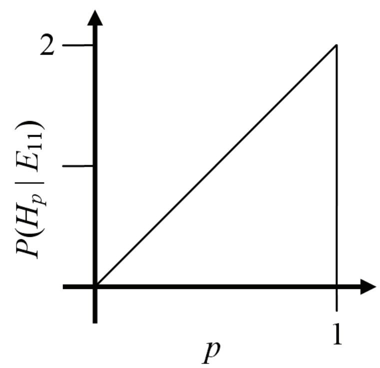

Figure 8. One head in one trial observed.

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

$$P(\text{“one more head”}|E_{mn}) = (m+1)/(n+2).$$

This is Laplace's *rule of succession*: with 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. 9). 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 Eq. (6.35) in [2].

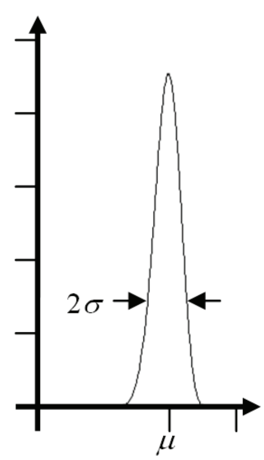

Figure 9. Peak shape.

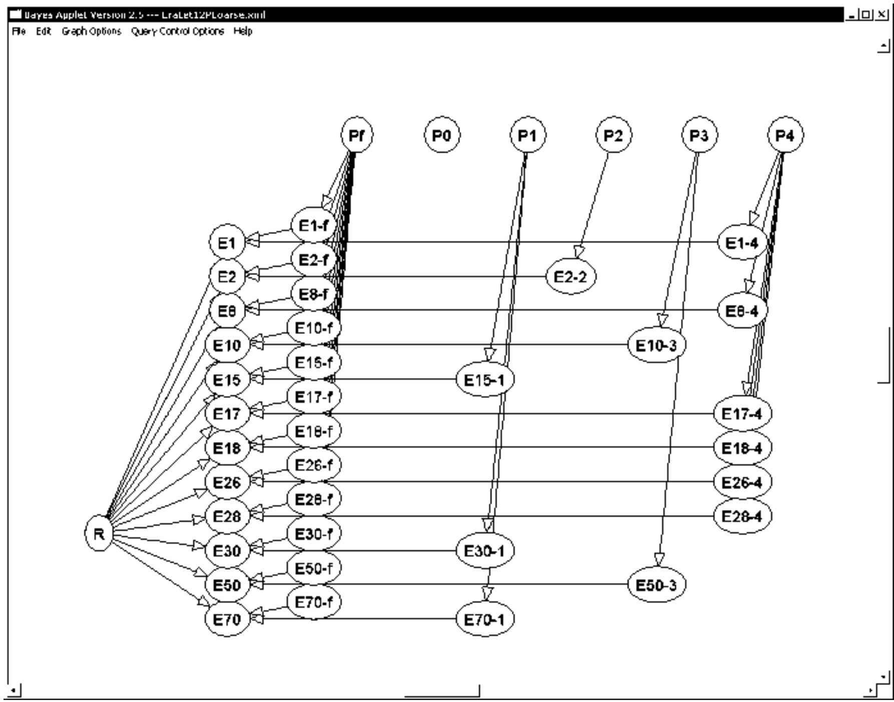

Figure 10. Network for twelve selected papers (initial configuration).

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. 9.

3. Problem Setup

The network we developed is shown in Fig. 10 as displayed by the applet [7] that was mentioned at the end of the subsection "Bayesian Networks". Node R is the proposition of interest – roughly speaking, "is the FP effect 'real'?" The nodes E_1, E_2, …, E_{70} 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 the previous subsection.

Table 1. $P(R)$

R	
True	False
0.5	0.5

3.1. Selected papers

Cravens and Letts [10] suggested eight papers for initial consideration: those numbered 2 through 28 in the following list; these were deemed to represent particularly significant or influential early positive and negative results. To these we added the initial Fleischmann–Pons report (#1) and three arbitrarily selected later papers (#30, #50, #70).

#	Cri	Heat	Citation
1	4	Yes	M. Fleischmann & S. Pons, J. Electroanal. Chem. 261 (2, part 1) 301–308 (Apr. 10, 1989).
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).
30	1	No	G. R. Longhurst et al., J. Fusion Energy 9 (3) 337–343 (Sep. 1990)
50	3	Yes	V. C. Noninski & C. I. Noninski, Fusion Technology 19 (2) 364–368 (Mar. 1991)
70	1	No	T. I. Quickenden & T. A. Green, J. Electroanal. Chem. 344 (1–2) 167–185 (Jan. 15, 1993).

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*: 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, …, 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 to a table with 50 rows.

Table 2. $P(E_i|RE_{if}E_{in})$

			E_i	
R	E_{if}	E_{in}	True	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

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 under "Estimating Probabilities" above. 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 subsection "Weight of Evidence".) This makes it easy to determine the weight of evidence from the program outputs.

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.

The conditional probabilities of E_{in} are given in Table 3; the same values apply also for E_{if}. The probability of E_{in}, given P_n, is by definition simply the value of P_n; and the probability of E_{if}, given P_f, is the value of P_f.

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

Table 3. $P(E_{in}|P_n)$

	E_{in}	
P_{in}	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. $P(P_n)$

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. 10, we successively declared "observed" values for the nodes E_i, starting with *true* for E_1 and finishing with *false* for E_{70}. The final state of the network is shown in Fig. 11. Display of the probability distributions of the nodes R and P_f, P_0, …, P_4, has been enabled.

The posterior probabilities for $R =$ *true* and $R =$ *false* are 0.9655 and 0.0345, corresponding to posterior odds of 27.99. This then is the final value of the likelihood ratio, since we started by setting the prior odds to 1.0. The value of the likelihood ratio is plotted in Fig. 12 as a function of the number of papers taken into account, from 1 paper (#1 only), through 12 papers. The notations across the top show for each point the number of enabling criteria met and whether excess heat was observed.

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_4 were identical—a bit biased toward high values, as the first paper (#1) reported heat. There was not yet a basis for choosing between the two. Adding a second paper (#2, reporting no heat) increased the ratio to about 1.47. Thereafter the trend is generally upward with increasing steepness, but with a conspicuous glitch at the 11$^{\text{th}}$ paper (#50). The two neighboring papers, #30 and #70, reported *no* for excess heat, yet

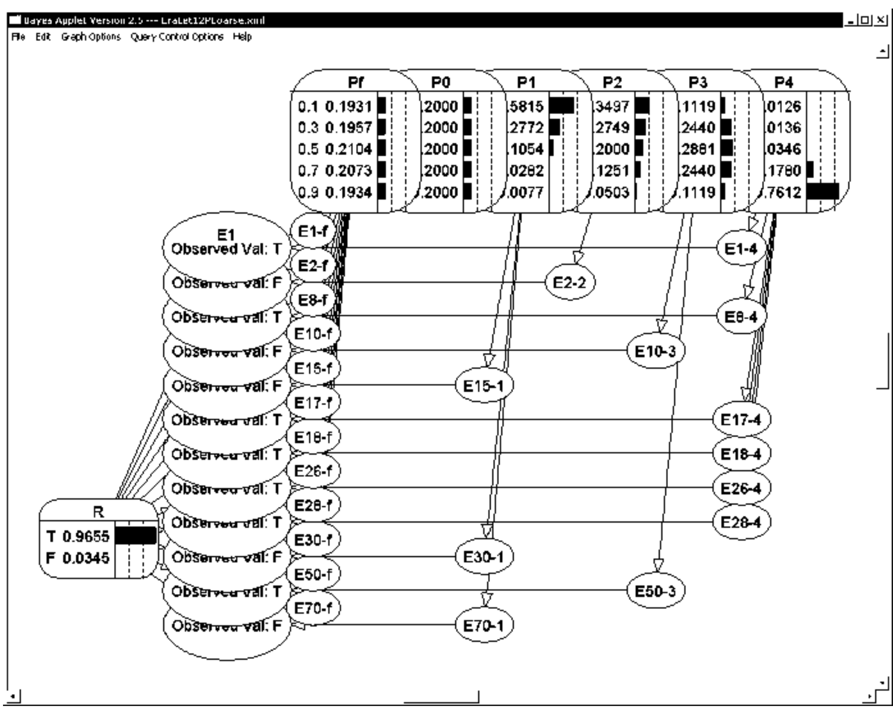

Figure 11. Final configuration of network for twelve selected papers.

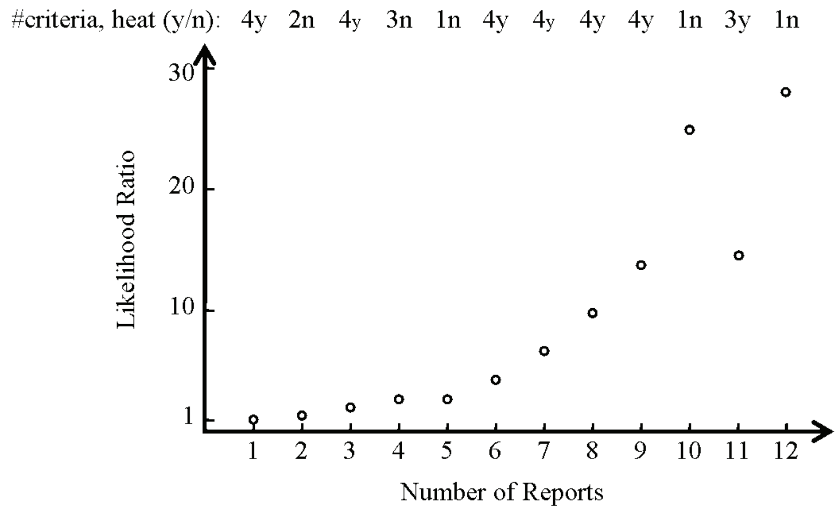

Figure 12. Change in likelihood ratio as more and more papers are taken into account.

their inclusion increased the likelihood ratio for R. On the other hand, paper #50, though positive for heat, nevertheless decreased the likelihood ratio for R. A possible explanation is that only one previous paper had met exactly 3 of the 4 criteria, and that one was negative for heat. This disagreement, one *no* and one *yes* for heat, made the case "3 criteria met" appear "random" and so apparently decreased the likelihood ratio for R. This underscores the fact that R is asking whether *the experimental treatment makes a difference*. The observation of no heat when some of the criteria are not met can serve as evidence for R just as well as the observation of heat when all are met.

In the final configuration, the posterior distribution for P_4 was strongly biased toward high values, as shown in Fig. 11, and as one would expect. P_3 was symmetrically distributed, somewhat peaked near 0.5, as for that case the evidence was balanced, with one paper positive for heat and one negative. P_1 and P_2 were biased toward low values, again as one would expect. 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 quite flat, close to its prior, as the probability for $R = false$ was 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 equal to its prior.

Twelve papers is a small enough sample that no particular significance should be attached to the particular final numerical value of 27.99 for the likelihood ratio for R, though the qualitative behavior of Fig. 12 is suggestive. Moreover, the set of 12 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.

It would be desirable in the future to include substantially more papers – ideally all that were successfully rated according to criteria met and presence of heat. The present scheme lumps together all papers that meet the same number of criteria. It would be desirable to consider particular subsets of the four criteria, rather than simply the count, expanding the number of cases from 5 to 16. The ability to handle substantially more papers might make that feasible.

Acknowledgments

The authors gratefully acknowledge the support of the Defense Threat Reduction Agency for this work.

References

[1] D. Craven, D. Letts, The enabling criteria of electrochemical heat: beyond a reasonable doubt, in *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, UK: Cambridge University Press, 2003.

[3] E.T. Jaynes, Probability in quantum theory, in *Complexity, Entropy, and the Physics of Information, SFI Studies in the Sciences of Complexity*, vol. VIII, W.H. Zurek (ed.), Addison-Wesley, 1990, pp. 381–403.

[4] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York, 2001.

[5] J. Breese, D. Koller, Tutorial on Bayesian networks, <http://ai.stanford.edu/∼koller/BNtut/BNtut.ppt>, 1997.

[6] K. Murphy, A brief introduction to graphical models and Bayesian networks, <http://www.cs.ubc.ca/∼murphyk/Bayes/bnintro.html>, 1998.

[7] 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>.

[8] I.J. Good, *Probability and the Weighing of Evidence*, London: C. Griffin, 1950.

[9] E.T. Jaynes, Confidence intervals vs. Bayesian intervals, in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. II, edited by W. L. Harper and C. A. Hooker, Dordrecht, Netherlands: D. Reidel Publishing Company, 1976, pp. 175–257.

[10] D. Craven, D. Letts, Private communication.