



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Using Dempster-Shafer theory to assess explanations for the Zimbabwe anthrax outbreak of 1978-79

S. P. Velsko

May 8, 2015

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Using Dempster-Shafer theory to assess explanations for the Zimbabwe anthrax outbreak of 1978-9

Stephan P. Velsko
Lawrence Livermore National Laboratory
April 30, 2015

Summary

I use a simple version of Dempster-Shafer theory (DST) to evaluate evidence from the anthrax outbreak in Zimbabwe (Rhodesia) during 1978-9 and assess the question of whether the outbreak arose naturally or was caused deliberately. The method is derived from a DST-based “opinion calculus” developed by Josang and others in which each item of evidence is used to elicit an opinion about the focal hypothesis, in this case the assertion that the outbreak was deliberately generated. I use a verbal probability expression mapping method to elicit opinions as vectors expressing degrees of belief, disbelief and uncertainty. Two rules for combining the evidence to produce an overall opinion are compared. Both rules result in an opinion favoring the hypothesis that the outbreak was natural in origin. Many items of evidence that have been proffered over the years are shown to be nearly irrelevant to the final conclusion, while the absence of certain expected types of evidence plays a critical role in the assessment.

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Auspices Statement

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

1. Introduction

Dissatisfaction with various aspects of Bayesian inference led to the invention of a variety of alternative theories for assessing the way that bodies of evidence provide support to an assertion or hypothesis. These theories are concerned with how to handle subjective or difficult-to-quantify levels of uncertainty about the veracity or validity of non-statistical or “soft” evidence. In addition they are applicable to inferences based on incomplete statistical evidence where Bayesian inference cannot be rigorously applied. One of the most extensively studied alternatives to Bayesian inference is Dempster-Shafer theory (DST).¹ Scholars have developed many variants and refinements of the original theory over the last four decades, and applied it to problems in engineering and artificial intelligence.² DST methods represent a “step up” from more qualitative methods such as the Analysis of Competing Hypotheses (ACH)³, and mitigate several problems with that methodology.^{4,5} In this note I show how a simple version of DST can be used to assess evidence from a past investigation of an anthrax outbreak where there was controversy about whether the outbreak arose naturally or was caused deliberately.

The version of DST used in this report is primarily concerned with evidence in the form of assertions from sources with less-than-perfect credibility. In forensic contexts these sources may be witnesses whose reliability is questionable, written reports whose conclusions are not well supported or are ambiguous, or official documents that may or may not be genuine. In addition, DST can also capture “pseudodiagnostic” forensic evidence based on incomplete statistical inference and “absence of evidence” arguments as well. The theory addresses several issues:

- (1) How should the uncertainty associated with unreliable sources or incomplete statistical evidence be expressed?
- (2) How should the uncertainties from multiple items of this type of evidence be combined to form an overall uncertainty assessment for a hypothesis?
- (3) How should this type of uncertain evidence be combined with (complete) statistical evidence where uncertainty can be rigorously expressed in terms of likelihoods or probabilities?

It has long been understood that formalized inferences based solely on such evidence are often not convincing because they are based on an attempt to explicitly *quantify* subjective assessments of uncertainty.⁶ This is almost certainly one reason why DST has never become widely used among investigative or analytic practitioners in spite of many years of “technology push”.^{5,6} (Of course, there has been widespread and persistent resistance to the adoption of *any* quantitative structured analysis techniques among practitioners for many years.) However, advocates argue with some justification that DST is nonetheless useful as a framework for organizing evidence, assessing the strength of individual pieces of evidence, and determining the sensitivity of conclusions to the assessed reliability of

each evidence item. Wildly different subjective assessments from different analysts – once they are explicitly quantified – can be valuable indicators of bias, misunderstanding, or hidden assumptions that might otherwise go unrecognized. Moreover, DST provides a consistent and transparent basis for combining evidence into an *overall* assessment that otherwise would be determined subjectively without regard for consistency or transparency.

Of course, similar claims may be made for alternative methods such as Bayesian nets⁷ and Wigmore charting.⁸ By choosing to illustrate the use of DST for analyzing a controversial historical event I do not claim that it is necessarily superior to other modes of analysis. Comparisons of a variety of different systems including DST tend to conclude that each has its advantages. The important question, of course, is whether one can gain any insights from the analysis that are useful and non-obvious.

In choosing the Zimbabwe outbreak as an exemplar I was motivated by the rich variety of evidence available in the open literature and the enduring controversy surrounding the event. I do not suggest that the analysis here will do anything to resolve the controversy.

2. The DST formalism

In this report I use the DST-based framework and notation developed by Josang⁹⁻¹¹, Haenni and Hartmann¹² and Haenni¹³. In this framework an opinion about an assertion or proposition is expressed as a vector $\hat{\omega} = (b, d, i)$ where:

- b = degree of belief in the assertion
- d = degree of belief in the assertion's negation (“disbelief”)
- i = degree of belief not assigned to either the assertion or its negation, but is withheld as “ignorance”

Assertions are either hypotheses whose truth is in contention, evidence items whose validity may be questioned, or judgments about how an evidence item increases or decreases our degree of belief in a hypothesis. In the simplest version of this framework there is a single hypothesis H and its negation \bar{H} , and at least one evidence item E . Associated with E is an “opinion” about its validity $\hat{\omega}(E)$, and an opinion about the degree to which E implies H , $\hat{\omega}(H|E)$. The opinion $\hat{\omega}(H|E)$ is a measure of the *probative value* of E , while $\hat{\omega}(E)$ is the *reliability* of E . We say that each evidence item E induces a certain degree of belief b that H is true. The “disbelief” d is the likelihood that \bar{H} is true induced by the same evidence. As will be discussed below, belief and disbelief depend on both the probative value and reliability of the evidence.

Unlike Bayesian inference, DST does not require that any belief not assigned to an assertion be assigned to its negation. Thus, if an item of evidence is considered unreliable or ambiguous, we may withhold some belief from both E and \bar{E} . This “withheld belief” is assigned to the state of ignorance $\{E, \bar{E}\}$, meaning that the

evidence might or might not be true and we cannot say which. The values of b , d and i are determined subjectively, explained in more detail in section 3.

The values of b , d , and i are expressed as numbers between 0 and 1. The DST framework requires that:

$$b + d + i = 1 \quad (1)$$

As explained in Appendix A this relation implies that opinions can be expressed more compactly as two-dimensional vectors. However, it is convenient for clarity and other reasons to retain the three dimensional representation.

Given each evidence item E , an opinion about the truth of H can be obtained from combining $\hat{\omega}(E)$ and $\hat{\omega}(H|E)$ using a rule devised by Josang^{9,10} for the “discounting” (denoted \otimes) of a conditional opinion by the opinion about its antecedent:

$$\hat{\omega}(E) = \begin{pmatrix} b_1 \\ d_1 \\ i_1 \end{pmatrix} \quad (2a)$$

$$\hat{\omega}(H|E) = \begin{pmatrix} b_2 \\ d_2 \\ i_2 \end{pmatrix} \quad (2b)$$

$$\hat{\omega}(H, E) = \hat{\omega}(H|E) \otimes \hat{\omega}(E) = \begin{pmatrix} b_3 \\ d_3 \\ i_3 \end{pmatrix} \quad (2c)$$

$$b_3 = b_1 b_2 \quad (3a)$$

$$d_3 = b_1 d_2 \quad (3b)$$

$$i_3 = d_1 + i_1 + b_1 i_2 \quad (3c)$$

In addition, DST provides rules for combining the belief vectors for H deduced from two pieces of evidence to produce an overall belief vector. In formal terms, if E_1 and E_2 are two evidence items, and we have calculated $\hat{\omega}(H, E_1)$ and $\hat{\omega}(H, E_2)$ using equations (3), then we can combine the two opinion vectors to obtain $\hat{\omega}(H, E_2, E_1)$. For a set of many evidence items this can be iterated until we have the opinion for H induced by all the evidence.

Various combination rules have been suggested based on intuitions about desirable properties that such a rule should possess.^{14,15} One such rule is “Dempster’s rule”.

According to this rule two opinion vectors (b_1, d_1, i_1) and (b_2, d_2, i_2) combine to give:

$$(b_3, d_3, i_3) = \left(\frac{b_1 b_2 + b_1 i_2 + i_1 b_2}{1 - b_1 d_2 - d_1 b_2}, \frac{d_1 d_2 + d_1 i_2 + i_1 d_2}{1 - b_1 d_2 - d_1 b_2}, \frac{i_1 i_2}{1 - b_1 d_2 - d_1 b_2} \right) \quad (4)$$

Another possible combination rule is the ‘‘Bayesian consensus’’ rule introduced by Josang:^{9,10}

$$(b_3, d_3, i_3) = \left(\frac{b_1 i_2 + i_1 b_2}{i_1 + i_2 - i_1 i_2}, \frac{d_1 i_2 + i_1 d_2}{i_1 + i_2 - i_1 i_2}, \frac{i_1 i_2}{i_1 + i_2 - i_1 i_2} \right) \quad (5)$$

Both of these rules are commutative and associative, which means that if there are more than two evidence items, the opinion vectors can be combined by successive pairwise combination in any order and the result will be the same.

Another important quantity in the DST framework is the *plausibility* of a hypothesis (or evidence item, or other assertion) defined by:

$$Pl(H) = 1 - d = b + i \quad (6a)$$

$$Pl(\bar{H}) = 1 - b = d + i \quad (6b)$$

An important feature of DST is that Bayesian likelihoods are bounded by belief and plausibility, for example:

$$b(H) \leq P(H) \leq Pl(H) \quad (7a)$$

$$d(H) \leq P(\bar{H}) \leq Pl(\bar{H}) \quad (7b)$$

The relationship between DST and probability, and the origin of equations (7) are discussed in Appendix B while Appendix A explains further how opinions can be regarded as expressing probability intervals. A practical consequence is that probability intervals derived from verbal probability expressions can be used to elicit opinion vectors, as discussed in section 3.

Finally, DST also quantifies the *conflict* between any two evidence items E_i and E_j with respect to H by:

$$\kappa_{ij} = b_i d_j + b_j d_i \quad \text{for } i \neq j \text{ and } \kappa_{ii} = 0 \quad (8)$$

A collection of N evidence items relevant to H thus defines a symmetric $N \times N$ matrix of κ values, characterizing the degree of conflict among the entire body of evidence. Some matrix norm such as the Frobenius norm could be used to reduce this measure of conflict to a single number. The conflict matrix also defines a weighted graph connecting the evidence items, so that in very large unorganized bodies of

evidence, algorithms for finding graph community structure could be used to find subsets of consistent evidence.

3. Mapping verbal probability expressions or verbal odds expressions to opinions

One of the simplest ways to elicit b , d and i values regarding evidence items and inferences drawn from them is to use verbal probability expression mapping. There are a variety of references that provide numerical interpretations of verbal probability expressions, many of which are very similar.¹⁶⁻¹⁹ Each probability term corresponds to a range of probability values defined by an upper and lower bound. A particularly elaborate example is given in Table 1, which was derived from a report of the Australian Transport Safety Bureau.²⁰

Table 1. Verbal probability terms, corresponding upper and lower probability values, and derived values of (b,d,i) ; based on values of P_{upper} and P_{lower} given in reference 20.

Term	P_{upper} Pl	P_{lower} b	d	i
Certain	1	1	0	0
Virtually certain	1	0.98	0	0.02
Extremely likely	1	0.95	0	0.05
Very likely	1	0.9	0	0.1
Moderately likely	0.9	0.67	0.1	0.23
Likely (probable)	1	0.67	0	0.33
Somewhat likely	0.67	0.5	0.33	0.17
More likely than not	1	0.5	0	0.5
About as likely as not	0.67	0.33	0.33	0.34
Not probable	0.5	0	0.5	0.5
Somewhat unlikely	0.5	0.33	0.5	0.17
Unlikely	0.33	0	0.67	0.33
Moderately unlikely	0.33	0.1	0.67	0.23
Very unlikely	0.1	0	0.9	0.1
Extremely unlikely	0.05	0	0.95	0.05
Exceptionally unlikely	0.02	0	0.98	0.02
Impossible	0	0	1	0

The connection between verbal probability terms and the bounds on likelihood indicated by equations (7) provide a method of eliciting subjective estimates of the opinion vector. For a given assertion we ask *what is the best verbal probability term that should be assigned to this assertion?* As explained in Appendix A, from the

correspondence between the verbal term and the upper and lower probability limits P_{upper} and P_{lower} , the opinion vector is given by:

$$b = P_{lower} \tag{9a}$$

$$d = 1 - P_{upper} \tag{9b}$$

$$i = P_{upper} - P_{lower} \tag{9c}$$

This elicitation technique may be aided by the use of a visual representation of the probability ranges corresponding to the verbal terms as shown in Figure 1.

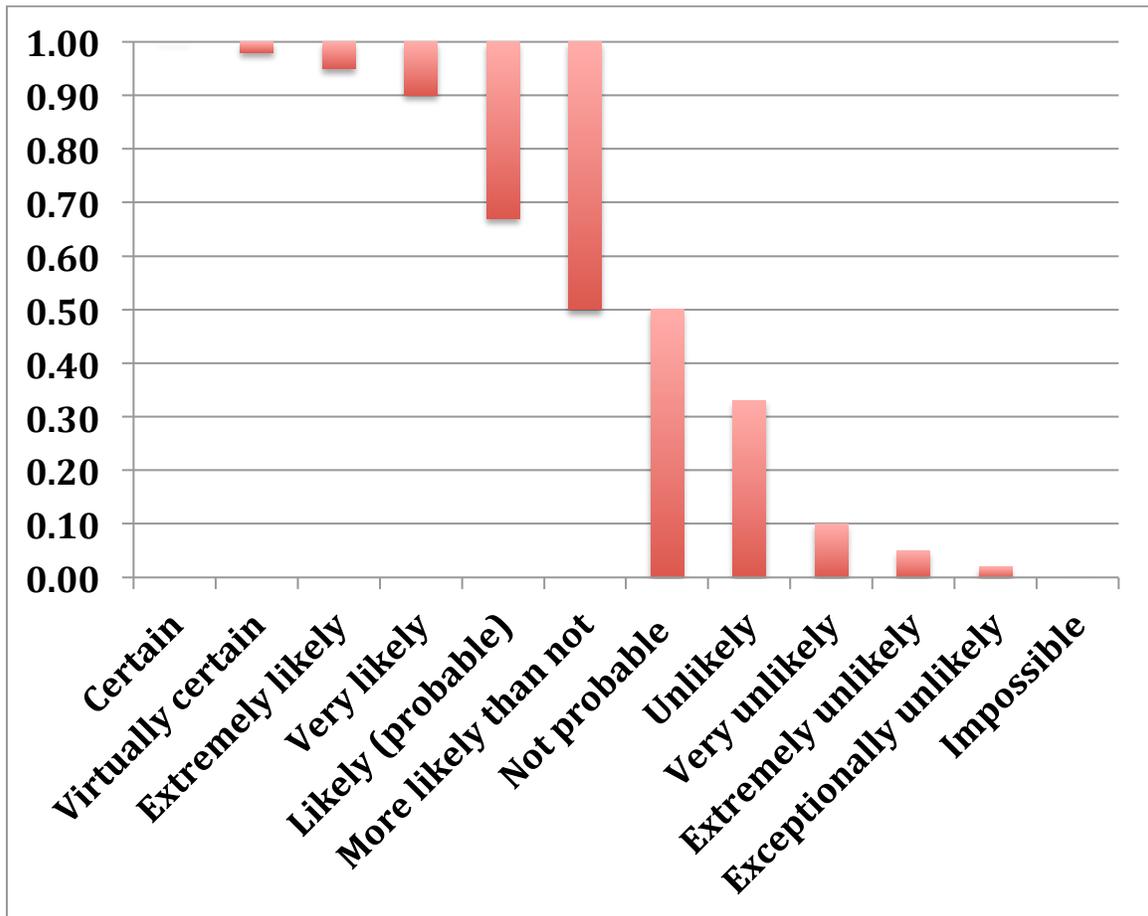


Figure 1. Probability ranges associated with selected verbal expressions from table 3.

The verbal probability scale can be extended to opinions expressed as odds. In this scheme one asks “what is the smallest number n that would express your estimate of the odds in favor or against this assertion?” In other words one elicits an opinion expressed as “at least n to 1 in favor of” or “at least n to 1 against” the truth of a given assertion. Given n , the values of P_{lower} and P_{upper} (Pl) are given by table 2.

Figure 2 shows the corresponding probability ranges for odds estimates elicited this way. Note that “even” (1:1) odds are taken as an expression of complete uncertainty (“complete ignorance”, $i=1$) rather than an expression of a sharp 50% probability.

Table 2. Probability bounds for odds elicitations of opinion vectors.

Probability bound	At least n to 1	
	In favor of	Against
P_{lower}	$1 - 1/n$	0
$P_{upper} = Pl$	1	$1/n$

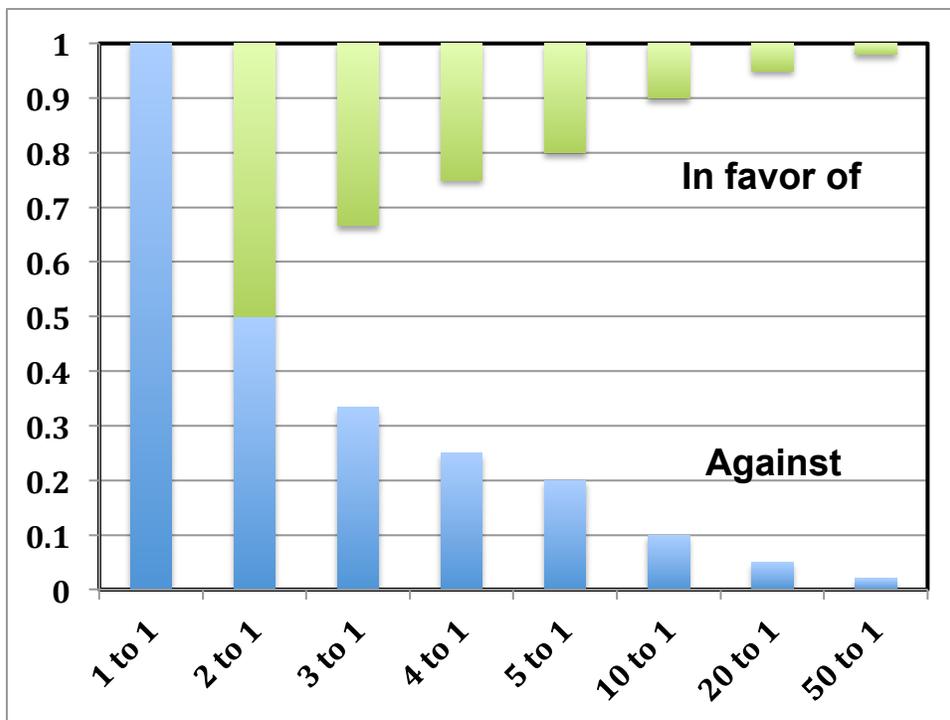


Figure 2. Probability ranges associated with opinions expressed as minimum odds in favor or against an assertion.

The ultimate justification for this heuristic elicitation scheme comes from the work of Josang, who has demonstrated a rigorous equivalence between the opinion calculus and the logic of combining uncertain probability distributions – i.e. probability distributions of probabilities.^{9,10} While actual distributions of probabilities have the form of beta functions, the verbal probabilities can be thought of as simple box-like representations of these distributions. The range of probability indicated by each verbal expression represents an uncertainty in the actual probability, which is mapped to the “ignorance” component of the opinion. A slightly more complicated mapping between verbal probabilities and opinion vectors was proposed by Pope and Josang.⁵

4. Incomplete statistical evidence as opinion

Two types of evidence derived from molecular microbiology are commonly found in investigations of disease outbreaks. One is detection of a pathogen by means of some assay in a location where production, handling, or dissemination of a biological agent is suspected.²¹ The support such evidence gives to hypotheses about the origin of the outbreak in question (i.e. whether it is H = man-made or \bar{H} = natural) depends on the likelihoods of observing a positive result in the location in question under the two hypotheses. Usually it is assumed that the likelihood of observing a positive detection at such a location in the case that the outbreak is man-made is near certain, i.e. $P(+|H) \sim 1$. On the other hand, the likelihood of observing a positive detection when the origin is natural depends on the “background” level of the pathogen or its genetic near-neighbors, and on the likelihood of laboratory contamination of evidence samples. In any case, this likelihood, $P(+|\bar{H})$, is often assumed on intuitive grounds to be smaller than the likelihood under H , although no explicit statistical evidence is available to support it.

The second type of molecular evidence involves the comparison of the genetic sequences of two pathogen isolates, one sampled from the outbreak in question and the other a “reference” isolate taken from a suspected or potential source of the outbreak strain.²² In this case E is a measure of sequence similarity. A high degree of similarity between the genetic sequences is intuitively taken as evidence supporting the source hypothesis even though explicit statistical evidence of the likelihood of observing that degree of similarity with isolates from other sources is not available. More precisely, there is a strong intuition that the likelihood of observing such a high degree of genetic similarity between the two isolates if they come from the same source is higher than the likelihood of observing that degree of similarity if the outbreak strain actually originated from a different source.

Statistical evidence where $P(E|H)$ is known (or can be convincingly estimated) but $P(E|\bar{H})$ is not known can be called “pseudodiagnostic”.²³ The intuition that

$$0 \leq P(E|\bar{H}) \leq P(E|H), \quad (13)$$

when combined with Bayes’s rule, leads to an interval estimate of the posterior probability:

$$P(H) \leq P(H|E) \leq 1 \quad (14)$$

Thus, by arguments in the previous sections, we can identify inferences based on such evidence as *opinions* with

$$b(H) = P(H) \quad (15a)$$

$$d(H) = 0 \quad (15b)$$

$$i = P(\bar{H}) = 1 - P(H) \tag{15c}$$

Note that the opinion depends only indirectly on the observed evidence E through equation (13), but directly on the subjective prior likelihood of H . Authors who have used pseudodiagnostic genetic similarity evidence to draw inferences about sources often argue that the weight given to such an inference must rely on its consistency with “epidemiological evidence.”²⁴ If we take $P(H)$ to be a reflection of such evidence, then equations (15) are the quantitative expression of this point of view.

In practice, of course, we do not directly elicit a “prior” for H when assessing an opinion about incomplete statistical evidence. Instead, the subjective opinion about the belief in H induced by (say) the molecular biological evidence is first expressed as a verbal probability term such as those in table 3 which have $d = 0$, e.g. “likely”, “extremely likely”, or “virtually certain”, or use an odds-in-favor elicitation as in Figure 2. These translate directly into values for b and i (d being fixed at zero) which express the degree of belief.

5. Expressing absence-of-evidence arguments as opinions

Judgments about many historical outbreaks where there are allegations about deliberate use of biological agents often rest on the credibility that, after time has passed no decisive evidence has emerged in spite of compelling reasons to expect it should, if the allegations were true. We can develop a very similar argument to the one in section 4 to express the opinion induced by an *absence of evidence* in such cases.²⁵

Let N = “no credible witness has emerged” or “no documents have been found” or some similar assertion regarding evidence that would be expected, in one’s opinion. Let H_n be the hypothesis that the outbreak is natural, and H_m be the hypothesis that the outbreak was man-made. Then, thinking in terms of likelihoods, we have:

$$P(N|H_n) = 1 \text{ (If it were natural, then no witness to the contrary would exist.)}$$

$$P(N|H_m) = \text{some number between 0 and 1; we don’t know what it is.}$$

$$P(H_n|N) = \frac{P(N|H_n)P(H_n)}{P(N|H_n)P(H_n)+P(N|H_m)P(H_m)} \tag{16}$$

Replacing our unknown likelihood by x we have:

$$P(H_n|N) = \frac{P(H_n)}{P(H_n)+xP(H_m)} \tag{17}$$

When $x = 0$, then $P(H_n|N) = 1$ and when $x = 1$ then $P(H_n|N) = P(H_n)$, because either H_n or H_m is true, so $P(H_n) + P(H_m) = 1$.

So $P(H_n) \leq P(H_n|N) \leq 1$, and we can set

$$b = P(H_n)$$

$$d = 0$$

$$i = P(H_m)$$

In other words, $\hat{\omega}(H_n) = \begin{pmatrix} P(H_n) \\ 0 \\ P(H_m) \end{pmatrix}$.

Similarly one can show: $\hat{\omega}(H_m) = \begin{pmatrix} 0 \\ P(H_n) \\ P(H_m) \end{pmatrix}$.

Note that, just as in section 4, to express beliefs based on absence of evidence we elicit verbal probability terms that have the prescribed form of the opinion vector— $P(H_n)$ and $P(H_m)$ are not treated as actual *a priori* or base-rate probabilities.

6. Illustrating the formalism: the Zimbabwe anthrax outbreak

This section will apply the DST formalism outlined in the previous sections to the example of the Zimbabwe anthrax outbreak of 1978-1979. This outbreak took place during the final years of the effort by black nationalist insurgents to overthrow white rule in then Rhodesia and resulted in a very large number of human cutaneous anthrax cases. The two basic hypotheses regarding the origin of the outbreak were:

H_m (man-made): The outbreak was the result of deliberate infection of cattle with anthrax by elements of the white Rhodesian counter-insurgency forces.²⁶

H_n (natural): The outbreak was a re-emergence of endemic anthrax, exacerbated by the practice of communal butchering of infected cattle and trading of meat among tribal peoples, and the difficulty of veterinary control efforts because of the fighting.

I have identified 17 items of evidence relevant to this event. Much of the evidence originally gathered in support of the “man-made” hypothesis for the Zimbabwe outbreak comes from papers by Nass.^{27,28} The five major arguments made by Nass are embodied in evidence items 1-5. Items 6-8 represent items of evidence provided by scientific papers about Anthrax outbreaks in Zimbabwe over time or prior to the 1978-1979 outbreak. Items 9-12 represent eyewitness or documentary evidence about white Rhodesian chem-bio operations other than anthrax, and the *absence* of such evidence for the anthrax outbreak from sources where it might have been expected. Finally, items 13-17 are testimonial evidence alleging specific anthrax spreading operations provided by former Rhodesian counterinsurgency

operatives.

For each evidence item I evaluate $\hat{\omega}(E)$ (the credibility of E) based on the verbal probability expression from table 1 that I consider best expresses the degree of reliability of that piece of evidence. Similarly a verbal probability expression is chosen to express my opinion about the probative value $\hat{\omega}(H_m|E)$. A summary of these evaluations for the 17 individual evidence items is provided in Table 3. Next, for each evidence item the quantities $\hat{\omega}(E)$ and $\hat{\omega}(H_m|E)$ are combined using Josang's discounting combination rule to form $\hat{\omega}(H_m, E)$. The 17 individual opinions are then combined using both the Dempster rule and Josang's consensus rule. I examine the contribution to the overall opinion from various sub-groupings of evidence items.

6.1 The individual evidence items

1. *Anthrax outbreaks were uncommon in Zimbabwe prior to 1978.*

This was asserted by Nass on the basis of historical records²⁷, but such records are arguably incomplete. In their study of historical outbreak trends Chikerema, et. al. note the unreliability of reporting in pre-independence Rhodesia:²⁹

“During the pre-independence era (before 1980), the incidence of anthrax could not be accurately determined as data on anthrax and other diseases could not be collected systematically, especially in rural areas. ... In most instances, many anthrax outbreaks, particularly in rural areas, were unlikely to have been diagnosed, reported, and forwarded to the central data collection centers.”(ref. 29, page 66)

In addition, in other countries where anthrax is endemic it is not unusual to find that there are very long gaps between recorded outbreaks.³⁰ For example Stein states:

“In the United States anthrax usually occurs in epizootic form in regions where the disease has existed for long periods. However, it may occur sporadically anywhere at any time and thus may appear where previously not identified or where it has been quiescent for a long period.” (Ref. 30, pages 347-348)

If Nass's assertion were based on very good record keeping over the previous century and her assertion were true one could generously conclude that the probative value of such evidence $\hat{\omega}(H_m|E)$ is “moderately likely”; however, the unreliability of the historical record and its short duration lead one to conclude that $\hat{\omega}(E)$ for this item of evidence is “moderately unlikely”.

2. *The outbreak was unusually large*

Nass points out that the reported incidence of human cases during the outbreak “was more than 400 times the average incidence over the previous 29 years.”²⁷ However, the distribution of *natural* anthrax outbreak sizes exhibits “fat tailed” behavior, so the likelihood of a large outbreak is much greater than one might intuit from the average behavior.³¹ Thus, although the observation itself has high credibility and $\hat{\omega}(E)$ = “highly likely”, the observation of an unexpectedly large outbreak does not provide particularly strong evidence for a man-made outbreak, and $\hat{\omega}(H_m|E)$ is assigned “about even odds.”

3. *The outbreak was multi-focal, "jumping" from district to district*

In her assessment of the outbreak Nass asserts:

“Most [anthrax] outbreaks are characterized by a high degree of focality. Cases occur in limited areas only. Yet in Zimbabwe from 1978 to 1980, the disease spread from area to area, until six of the eight provinces were affected.” (ref. 27, page 199)

This assertion is a critical element of Nass’s case for a man-made origin of the outbreak, because it is difficult to explain by some natural mechanism. The notion that the outbreak was multi-focal, jumping from one district to another appears to originate from papers by Davies,^{32,33} and repetitions of the same assertion in subsequent papers by other authors. Davies paints the following picture of anthrax “spreading” from one area to another:

“If the spread to Mzingwane, Bembezi, and Filabusi (November 1979) is ignored it is easy to picture the epidemic spreading from Nkai westward to Lupane, Wankie (Hwange), Tjolutjo (Tsholotsho) and Nyamandlovu Districts, south to Inyati and Bubi, southwest to Plumtree, Motopos, Kezi, Gwanda and Beitbridge” (Ref. 32, page 294)

In his book *Assignment: Selous Scouts*,³⁴ Jim Parker uses a map to detail this supposed sequence of anthrax outbreaks with time, noting “the haphazard way the outbreaks occurred.” If this picture were accepted as true, it would immediately lead to suspicion since anthrax is known not to be communicable among cattle or humans. This, in turn, leads to the inference that anti-insurgency forces were actually spreading the disease.

However, Table 3 in Davies own paper (reference 32) places the “spread” of anthrax in perspective: most (86%) human cases occurred in Nkai district (Matabeleland province) with very few patients originating in other districts. Similarly, Table 4 in the same paper shows that the vast majority of cases in the Midlands province occurred in the Que Que (Kwekwe) district, adjacent to Nkai. So this would be more accurately described as a large outbreak in one geographic area with a small number of cases originating in adjacent and more remote areas.

It is important to note that all of these papers describe human cases of anthrax, and there is no actual description of the statistics and spatial distribution of cases among cattle. The data would be consistent with a picture of a primary cattle outbreak localized to a single geographic area with occasional random transport of meat to places where there are only few if any infected cattle. Consistent with this view, Davies quotes a “Senior Medical Assistant” (Mr. J.M. Gwaza) as writing:

“One contributing factor was the meat of the carcasses infected by anthrax being sold around to the public as meat slaughtered at home. So the disease spread rapidly as the meat was carried on scotch carts [sic] to distant places in the [Tribal Trust Land or Communal Farming Area].” (Ref. 32, page 293)

Davies also quotes a Mr. P.J. Mutangiri as writing:

“Due to lack of food at the time people ate the infected meat and the disease quickly spread to almost all the villages as meat had become the main food.” (Ref. 32, page 294)

In summary, the notion that the outbreak “spread” geographically across the country is not accurate. Thus for this item of evidence we can conclude that while $\hat{\omega}(H_m|E)$ is “highly likely”, actual data indicates that the outbreak was a single focus event, and $\hat{\omega}(E)$ for this item of evidence is “unlikely”.

4. *Mostly black-owned cattle were affected*

During the insurgency, veterinary services could not be delivered to the tribal areas because of insurgent violence:

“...reports began to come in about cattle deaths. A veterinary team arrived to begin a vaccination campaign, but was ambushed. The team managed to vaccinate almost 8000 cattle and obtained good coverage in neighboring commercial farms, where no cases of cattle or human anthrax occurred.” (Ref. 35, page 32)

This would indicate that tribal cattle would likely not have been vaccinated regardless of whether the outbreak was deliberate or natural, and a higher fraction of black-owned cattle would become infected regardless. Thus, although the observation has high credibility $\hat{\omega}(E)$ = “highly likely”, $\hat{\omega}(H_m|E)$ is assigned “about even odds.”

5. *The outbreak stopped at Zimbabwe’s borders*

This is cited by Nass without any actual evidence except the absence of any contemporary reports of anthrax outbreaks in neighboring countries.²⁷ It is certainly true that if there were sharp national boundaries between outbreak and

non-outbreak areas it would be unexpected on the basis of ecological or animal travel considerations and could be an indication that the outbreak was artificially being restricted to one country. Thus such an observation, *if true*, would increase belief in H_m – arguably making $\hat{\omega}(H_m|E)$ = “likely”. However, no actual evidence for such a sharp artificial boundary is provided by Nass. Moreover, Rhodesian counterinsurgency operations (including chem and bio attacks) were regular occurrences in at least two bordering countries – Zambia and Mozambique, and there is no sensible reason why the Rhodesians would have hesitated to use anthrax in those areas.³⁶ In fact, one could argue that anthrax use *would have been expected* in these countries if they were being carried out in Zimbabwe. Therefore $\hat{\omega}(E)$ has been rated “unlikely” for this piece of evidence.

6. The outbreak was “in season”

Chikerema, et. al. found that when the number of outbreaks per seasonal period were aggregated over the period 1995 – 2005 in Zimbabwe, the disease was approximately 3 times more likely to occur during the September – November period compared to other seasons.²⁹ The peak is in October, but November is second-highest. Thus the Zimbabwe outbreak was “in season”. Clearly this is only weak evidence in favor of H_n , making $\hat{\omega}(H_m|E)$ = “moderately unlikely” although $\hat{\omega}(E)$ may be considered “virtually certain.” See Appendix D for an alternative argument with a similar conclusion.

7. Observations from the 1974 Rhodesian anthrax outbreak

A 1975 paper by two Rhodesian veterinarians describes a number of human cutaneous anthrax cases from a small outbreak in Zimbabwe (Rhodesia) in 1974.³⁷ Of primary interest is the observation that at least five cases arose from the handling of uncooked meat from a single cow that had contracted anthrax. One of the victims was a butcher who lived 8 km distant from the site where the carcass was butchered.

A strong implication of this paper is that cultural practices such as communal butchering and sharing of anthrax infected meat can amplify the number of human cases, and can account for wide geographical dispersion of cases even when the actual animal outbreak is highly localized. It is likely that these phenomenon were amplified during the insurgency period where tribal peoples were often displaced and normal food supplies were disrupted. Thus, there is no need to invoke deliberate spreading to explain the “unusual” features of the 1978-1979 outbreak such as the large number of human cases.

Since this information appears in a peer-reviewed scientific paper published prior to 1978, $\hat{\omega}(E)$ may be considered “highly likely.” However, this information clearly weakens the implication of H_m over H_n , making $\hat{\omega}(H_m|E)$ = “about even odds”.

8. Other diseases besides anthrax increased as well

The disruption of veterinary services during the seven years of conflict in the Rhodesian insurgency 1972-1979 gave rise to increases in many animal diseases besides anthrax, including trypanosomiasis, foot-and-mouth disease, and rabies.³⁸ This information appears in a peer-reviewed scientific paper published near the end of the insurgency, so $\hat{\omega}(E)$ may be considered “highly likely,” but it clearly weakens the implication of H_m over H_n by supporting the role of reduced veterinary services in increasing the severity of natural outbreaks, making $\hat{\omega}(H_m|E) =$ “about even odds”.

9. Other incidents of CB use by white Rhodesian forces occurred

A variety of operations using chemical or biological agents have been described by Rhodesian government “insiders” and documented in various ways. These include government documents and independent medical reports.^{39,40} Many of these descriptions have the “ring of truth”, such as unsuccessful attempts to infect insurgents in Mozambique with cholera. This evidence item is therefore assigned a high credibility rating with $\hat{\omega}(E)$ considered to be “highly likely.” Moreover, if these operations occurred it arguably lends credibility to the use of anthrax as well. Therefore I have assigned $\hat{\omega}(H_m|E) =$ “highly likely”.

10. No credible tribal witnesses

A variety of means have been suggested for the mechanism by which white Rhodesian operatives infected cattle with anthrax. These include feeding cattle infected grain⁴¹, using “cattle cakes”⁴², and disseminating unspecified material from an airplane.²⁷ To cause a large outbreak the operation must have been extensive. The cattle owned by tribal people are an important part of their wealth⁴³, and vigilance against predators and poaching was likely to be fairly rigorous. It seems highly unlikely that infection operations would go completely un-noticed. Yet Nass, who spent time investigating the event in Zimbabwe presented no witnesses from among the tribal peoples. Given the degree of scrutiny given to this event by advocates of the man-made hypothesis at that time (1989-1992), it is unthinkable that such witnesses would not be produced if they existed. Only in 2005, well after Nass’s papers were published and other derivative publications repeated her claims did any mention of tribal witnesses emerge.⁴⁴ Even then it is not clear if these witnesses, who claimed to have observed aircraft dispersing “white powder” over pastures, were not prompted by the interviewers. Other witnesses interviewed at the same time⁴⁴ provided much less credible stories about how the Rhodesian forces dispersed anthrax. Therefore if $E =$ the absence of credible witnesses, then $\hat{\omega}(E)$ is rated “virtually certain”; this “absence of expected evidence” induces an opinion that H_m is not likely, i.e. $\hat{\omega}(H_m|E) =$ “unlikely”.

11. No government documents

Mark Wheelis has noted:

“... the lack of detailed documentation since the collapse of White minority governments in Rhodesia and South Africa undermines the case for deliberate instigation. In order to deliberately create such a large and widespread outbreak, many separate attacks, using large amounts of material distributed over large areas, would have been necessary. It seems unlikely that a program of such a magnitude would not have been better documented by now.” (Ref. 45, page 22)

After the fall of the apartheid government in South Africa in 1993, investigations by the new government revealed extensive information about the secret chem-bio program. It is difficult to understand why the new government of Zimbabwe would not have found and publicized similar evidence about anthrax operations by the former Rhodesian government after it assumed power in 1980. As has been noted, government documents related to other chem-bio operations have emerged, so the absence of such evidence for the anthrax outbreak is unexpected if H_m were true. Therefore if E = the absence of documentation, then $\hat{\omega}(E)$ is rated “virtually certain”; this fact induces an opinion that H_m is not likely, i.e. $\hat{\omega}(H_m|E)$ = “unlikely”.

12. *Ellert's testimony - absence of evidence*

Henrick Ellert was a Rhodesian intelligence officer during the insurgency who published a book providing a fairly detailed account of a program to supply insurgents with poisoned clothing in 1977-1978, an effort which was terminated because of unintentional civilian deaths.⁴⁰ He supplies copies of classified government documents to back his claims. In addition, he describes an attempt to introduce “bacteriological cultures” (possibly cholera) into the Ruya river along the Mozambique border, and the poisoning of water tanks used by guerillas. However, nothing is mentioned about anthrax. If he was willing to reveal these operations, why not the alleged anthrax operation?

If E = the absence of Ellert testimony, then $\hat{\omega}(E)$ is rated “virtually certain”. Due to Ellert’s position within the intelligence apparatus (which ran the other operations) it is arguable that his lack of knowledge about an anthrax operation is a highly credible indication that there was no such operation. This induces an opinion that H_m is very unlikely, i.e. $\hat{\omega}(H_m|E)$ = “highly unlikely”.

13. *Parker's testimony regarding Nkai district*

In his book “Assignment: Selous Scouts” Jim Parker asserts:

“In 1979 I was tasked to travel to Wankie [district]. Planning had commenced on operations designed to destroy road and rail bridges in Zambia to disrupt Joshua Nkomo’s plans for an invasion of conventionally trained troops into Rhodesia. On the way I spent some time at the Selous Scouts’ Group Three Fort ... close to the abandoned

Khumalo airfield in Bulawayo. ... The late Doctor Sandy Kirk was based at the fort. He told me that Selous Scouts teams had recently deployed anthrax spores to infect the cattle of black tribesmen in the Nkayi [Nkai] and Lupane areas of Matabeleland. ... The purpose of the exercise, he said, was to limit the availability of food that troops could forage if a large ZIPRA force invaded the country. Outbreaks of anthrax in cattle in the Lupane and Plumtree areas had occurred in the past, so it wouldn't appear unusual if similar outbreaks happened in those adjoining areas." (Ref. 34, pages 171-172)

Of the various items of testimonial evidence proffered for the existence of a deliberate anthrax spreading operation, this is arguably the best. It provides a named source (although deceased) and places the event in the correct geographical location to be consistent with the epidemiological reports on the outbreak. Cilliers indicates that operations against the Zambian railway bridges began in the spring of 1979.³⁶ However, the Nkai outbreak began in November of 1978, so depending on when Parker visited, and how "recently" Kirk's alleged operation occurred, the consistency with the facts of the outbreak is not un-ambiguous. This story still leaves open the question of how the *B. anthracis* spores were distributed. A book written by Bob Coen and Eric Nadler and published in 2009 has a section on Parker's story.⁴⁶ On page 149 we find: "It was given to the cattle through veterinary staff, Parker told Coen." This is not elaborated on, leaving questions about precisely how the anthrax was administered, but it seems to be consistent with a scenario in which the Scouts recruited some local vets (who could very well have access to isolates of anthrax) and initiated a "one-off" effort in one district.

In assessing the reliability of this testimony and the validity of Kirk's assertion, one must take into consideration not only its consistency with the known location of the outbreak but also its relative lateness (Parker's book was published in 2006), the second-hand nature of the testimony, and the somewhat ambiguous placement of the event in time. Moreover, the rationale asserted for the operation – to deny infiltrating ZIPRA insurgents a food supply – seems questionable. Why not simply confiscate or shoot the cattle, which would be more reliable and perfectly consistent with Rhodesian government policy at the time?⁴⁴ Therefore if E = Parker's testimony regarding Nkai district, then $\hat{\omega}(E)$ is rated only "moderately likely"; if the underlying assertion made by the now-deceased Dr. Kirk were, in fact true, it arguably induces an opinion $\hat{\omega}(H_m|E) =$ "highly likely".

14. McGuinness testimony regarding Plumtree district

Ellert's book The Rhodesian Front War asserts that:

"... members of the Rhodesian Special Air Service (SAS) regiment delivered the anthrax bacteria by dropping it from an aircraft near Plumtree, on the Botswana border." (Ref. 40, page 27)

Ellert cites two 2002 interviews with Mac McGuinness, a former commander of counter-terrorism for the Rhodesian Police Special Branch for this information. However, according to Ellert it is based on information that McGuinness received second-hand, and was “surprised” to hear.

I rate the assertion itself, $E = \text{“SAS air-dropped anthrax near Plumtree,”}$ to be moderately unlikely, given the lack of any other mention of the SAS with regard to known chem or bio operations (these were generally attributed to the intelligence service or the Selous Scouts.) In addition it leaves open obvious questions such as where this anthrax was obtained and in what form the anthrax was dispensed. Therefore I have rated $\hat{\omega}(E) = \text{“moderately unlikely”}$.

If this item of evidence were true, it would still not account for the anthrax outbreak as actually experienced, because Plumtree was not the epicenter. In effect, as evidence this testimony would have only minor relevance. Thus I rate $\hat{\omega}(H_m|E) = \text{“even odds”}$.

15. Baxter testimony regarding Malvernia

In his book Rhodesia: Last Outpost of the British Empire Peter Baxter asserts:

“Anthrax was also used by the Selous Scouts to infect cattle in the Malvernia area. The spore would then be passed on to humans consuming the meat. The distribution of anthrax was carefully controlled in Gaza lest infections move to the Kruger National Park to the detriment of wildlife and South African goodwill.” (Ref. 47, page 399)

Malvernia is in Mozambique, bordering the Chiredzi district in Rhodesia. Gaza province is in Mozambique and borders Chiredzi and Beitbridge districts in Rhodesia which are about 200 km southeast of Nkai. So this is another example of a claim of anthrax dissemination that is remote from the location of the actual outbreak. In addition, there is no evidence that Malvernia experienced an outbreak of anthrax during this period (recall Nass’s argument about anthrax stopping at Rhodesia’s borders.)

I rate the assertion $E = \text{“the Selous Scouts infected cattle with anthrax in Malvernia”}$ to be “moderately unlikely”, given the lack of any evidence of anthrax outbreaks in that area.

If E were true, and this described an actual but unsuccessful Selous Scouts operation, it would clearly be of minor relevance to the proposition that the anthrax outbreak in Nkai was man-made. Thus I rate $\hat{\omega}(H_m|E) = \text{“even odds”}$.

16. Peter Stiff testimony regarding Mazoe valley

In Assignment: Selous Scouts, Jim Parker asserts:

“An anthrax deployment also designed to infect cattle was carried out by the SAS in the Mazoe Valley in Rhodesia’s far north-east. (Ref. 28, page 172, citation to “Email, Peter Stiff, June 29, 2002.”)³⁴

No additional information is given, and I have found no literature record of such an outbreak. Again, the Mazoe Valley is very remote from the epicenter of the 1978-1979 outbreak, so this evidence is also rated $\hat{\omega}(E)$ = “moderately unlikely” and $\hat{\omega}(H_m|E)$ = “even odds”. This appears to be another “barracks rumor”.

17. Un-named intelligence officer regarding Gutu, Chilimanzi, Masvingo, and Mberengwa districts.

In Plague Wars, Tom Mangold and Jeff Goldberg assert:

“A former Rhodesian intelligence officer, now active for the new Zimbabwe Administration, has written a kind of mea culpa which contains the ring of truth. The letter, in part, reads: ‘It is true that anthrax spoor [sic] was used in an experimental role in Gutu, Chilimanzi, Masvingo, and Mberengwa areas, and the anthrax idea came from army Psyops [Psychological operations]. The use of anthrax spoor to kill off the cattle of tribesmen ... was carried out in conjunction with the psychological suggestion to the tribes people that their cattle were sick and dying because of disease introduced into Zimbabwe from Mozambique by the infiltrating guerrillas.’” (Ref. 42, page 222).

The source cited for this is not the letter itself, but rather David Martin, a left-wing journalist. It is not clear why a member of the Mugabe government would need to remain anonymous, which prevents assessing his credibility as a source. Significantly, an article written by Martin⁴⁸ contains the same quotation *minus the reference to the districts*. The districts in question lie more than 100 km to the east of the “epicenter” in Nkai and Kwekwe. Moreover, the mistaken use of the term “spoor” instead of “spore” leads one to suspect that the author of the letter did not have direct experience with *B. anthracis*, and was at best quoting others. Finally, there is no corroborating evidence that the white Rhodesian government “blamed the guerillas” for the cattle deaths or human anthrax cases at the time.

For these reasons this evidence is also rated $\hat{\omega}(E)$ = “moderately unlikely” and $\hat{\omega}(H_m|E)$ = “even odds”.

6.2 DST analysis of the evidence

The quantitative opinion vectors derived from translating the verbal probabilities in Table 3 using the corresponding upper and lower probabilities in Table 1 and the resulting product opinion calculated using the discounting rule are provided in a

table in Appendix C. The final opinion vector for each item of evidence is reproduced in Table 4.

In Table 4 the evidence items have been re-grouped according to the degree of support they supply to each of the hypotheses or the “ignorance” state. A representation of the conflict matrix for this collection of evidence is shown in Fig. 3.

Table 3. Summary of evidence ratings

#	E Item of Evidence	$\hat{\omega}(E)$ Credibility of evidence	$\hat{\omega}(H_m E)$ Induced belief in H_m if E were true
1	Outbreaks were uncommon	Moderately unlikely	Moderately likely
2	The outbreak was unusually large	Highly likely	About even odds
3	There was "jumping" from focus to focus	Unlikely	Highly likely
4	Mostly black-owned cattle were affected	Highly likely	About even odds
5	The outbreak stopped at Zimbabwe's borders	Unlikely	Likely
6	The outbreak was "in season"	Virtually certain	Moderately unlikely
7	1975 outbreak	Highly likely	About even odds
8	Other diseases increased as well	Highly likely	About even odds
9	Other CB events occurred	Highly likely	Highly likely
10	No black witnesses	Virtually certain	Unlikely
11	No government documents	Virtually certain	Unlikely
12	Ellert's testimony absence of evidence	Virtually certain	Highly unlikely
13	Parker testimony regarding Nkai	Moderately likely	Highly likely
14	McGuinness testimony (Plumtree)	Moderately unlikely	About even odds
15	Baxter testimony (Malvernia)	Moderately unlikely	About even odds
16	Peter Stiff testimony (Mazoe valley)	Moderately unlikely	About even odds
17	Unnamed intel officer (Gutu, Chilimanzi, Masvengo, Mberengwa)	Moderately unlikely	About even odds

Table 5 shows the support afforded to H_n and H_m by the complete set of evidence items when they are combined using the Dempster or Consensus rules. In addition, the total opinion derived from each sub-group of evidence items is displayed.

Evidence items 9 and 13 provide pure and relatively strong support for H_m , while items 10, 11 and 12 provide pure support for H_n . Each of these latter items arises from an absence of evidence argument and displays the form suggested in section 5. Thus, within the set of evidence relevant to the Zimbabwe outbreak there are two sets of relatively strong but conflicting evidence that can be cited – typical of controversial historical cases. Arguments about the event typically involve attempts to discredit or discount the strength of one set or the other. For example, in his paper on South Africa’s CBW program Bale⁴⁹ attempts to explain the absence of certain items of information in spite of the extensive hearings that the post-apartheid government held by asserting:

“[E]ven democratic governments are often reluctant to ‘air dirty laundry’ by revealing sensitive state secrets, even if doing so might serve to expose their domestic political enemies’ prior crimes.” (Ref. 49, page 51)

This argument could just as easily be applied to “explain away” item 11.

Table 4. Derived opinion vectors for the 17 evidence items.

	Evidence item #	H_m	H_n	$\{H_m, H_n\}$
Supports H_m	9	0.81	0	0.19
	13	0.603	0	0.397
Supports H_n	12	0	0.882	0.118
	10	0	0.657	0.343
	11	0	0.657	0.343
	6	0.098	0.657	0.245
	7	0.09	0.603	0.307
Neutral Support	8	0.09	0.603	0.307
	2	0.297	0.297	0.406
	4	0.297	0.297	0.406
High Ambiguity	1	0.067	0.01	0.923
	14	0.033	0.033	0.934
	15	0.033	0.033	0.934
	16	0.033	0.033	0.934
	17	0.033	0.033	0.934
	3	0	0	1
	5	0	0	1

Note also that items 9 and 12 are essentially based on the same body of evidence, but provide conflicting assessments of the credibility of H_m . In assessing item 9 I take the high credibility of *other* operations to induce a positive opinion about the existence of an anthrax operation. Conversely, in assessing item 12 I take the apparent absence of knowledge about an anthrax operation by someone who knew about the other operations to undermine support for the existence of an anthrax operation.

Nearly pure support for H_n also comes from items 6,7 and 8 as well – these could be said to tip the balance of the argument to favor H_n over H_m . But note that their support is not purely for H_n and they do leave some room for doubt by admitting some belief in H_m . This stems from the fact that they simply argue that certain features of the outbreak deemed “suspicious” by advocates of H_m are, in fact, highly consistent with H_n . A more detailed discussion of the way that item 6 lends weight to H_n is discussed in Appendix D. Item 7 “explains away” the large number of human cases by showing that multiple human cases also arose from a single infected animal in an outbreak acknowledged to be natural in origin. Item 8 shows that other diseases also increased without any allegations that they were deliberately spread.

	Supports H_m		Supports H_n					Neutral		High Ambiguity								
Supports H_m	0.000	0.000	0.714	0.532	0.532	0.532	0.488	0.488	0.241	0.241	0.008	0.027	0.027	0.027	0.027	0.000	0.000	9
	0.000	0.000	0.532	0.396	0.396	0.396	0.364	0.364	0.179	0.179	0.006	0.020	0.020	0.020	0.020	0.000	0.000	13
Supports H_n	0.714	0.532	0.000	0.000	0.000	0.086	0.079	0.079	0.262	0.262	0.059	0.029	0.029	0.029	0.029	0.000	0.000	12
	0.532	0.396	0.000	0.000	0.000	0.064	0.059	0.059	0.195	0.195	0.044	0.022	0.022	0.022	0.022	0.000	0.000	10
	0.532	0.396	0.000	0.000	0.000	0.064	0.059	0.059	0.195	0.195	0.044	0.022	0.022	0.022	0.022	0.000	0.000	11
	0.532	0.396	0.086	0.064	0.064	0.000	0.118	0.118	0.224	0.224	0.045	0.025	0.025	0.025	0.025	0.000	0.000	6
	0.488	0.364	0.079	0.059	0.059	0.118	0.000	0.109	0.206	0.206	0.041	0.023	0.023	0.023	0.023	0.000	0.000	7
	0.488	0.364	0.079	0.059	0.059	0.118	0.109	0.000	0.206	0.206	0.041	0.023	0.023	0.023	0.023	0.000	0.000	8
Neutral	0.241	0.179	0.262	0.195	0.195	0.224	0.206	0.206	0.000	0.176	0.023	0.020	0.020	0.020	0.020	0.000	0.000	2
	0.241	0.179	0.262	0.195	0.195	0.224	0.206	0.206	0.176	0.000	0.023	0.020	0.020	0.020	0.020	0.000	0.000	4
High Ambiguity	0.008	0.006	0.059	0.044	0.044	0.045	0.041	0.041	0.023	0.023	0.000	0.003	0.003	0.003	0.003	0.000	0.000	1
	0.027	0.020	0.029	0.022	0.022	0.025	0.023	0.023	0.020	0.020	0.003	0.000	0.002	0.002	0.002	0.000	0.000	14
	0.027	0.020	0.029	0.022	0.022	0.025	0.023	0.023	0.020	0.020	0.003	0.002	0.000	0.002	0.002	0.000	0.000	15
	0.027	0.020	0.029	0.022	0.022	0.025	0.023	0.023	0.020	0.020	0.003	0.002	0.002	0.000	0.002	0.000	0.000	16
	0.027	0.020	0.029	0.022	0.022	0.025	0.023	0.023	0.020	0.020	0.003	0.002	0.002	0.002	0.000	0.000	0.000	17
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5
	9	13	12	10	11	6	7	8	2	4	1	14	15	16	17	3	5	

Figure 3. Conflict matrix for the 17 evidence items in the Zimbabwe outbreak analysis. Pink = high conflict ($\kappa > 0.3$); Yellow = moderate conflict ($0.1 < \kappa \leq 0.3$); Green = Low conflict ($\kappa < 0.1$). The individual item numbers are displayed along the bottom and right hand side.

Table 4. Summary opinion vectors for various sub-groupings of the evidence.

	# items	Dempster Rule			Consensus Rule		
		H_m	H_n	$\{H_m, H_n\}$	H_m	H_n	$\{H_m, H_n\}$
All evidence items	17	0.01	0.99	0.00	0.29	0.68	0.03
Items Supporting H_m	2	0.92	0.00	0.08	0.85	0.00	0.15
Items Supporting H_n	6	0.00	1.00	0.00	0.05	0.90	0.05
Neutral and ambiguous items	9	0.44	0.40	0.16	0.39	0.38	0.23

Most notable is the way that the evidentiary force of items 1-5, which Nass presented as providing strong support for H_m , becomes emasculated either by reliable scientific evidence that contradicts them (especially item 3) or because they are seen to be characteristics that are actually consistent with *either* hypothesis. Thus they are rendered essentially irrelevant, either because they provide identical support for both hypotheses or because all of their support is assigned to the ambiguous assertion $\{H_m, H_n\}$. Note, for example, how the “unlikely” ratings given to Nass’s “jumping” and “borders” assertions moved the opinion weight of both of these into the $i = 1$ “total ignorance” class, rather than move opinion weight from H_m to H_n .

Finally, items 14-17 represent primarily ambiguous opinions ($i > 0.9$). For these items it is important to note that the relevant question is not “were there attempts to use anthrax as part of the counterinsurgency campaign?” Instead the salient question is “did the *observed* outbreak (centered in Nkai district) start because of deliberate actions?” Because the locations of these alleged operations were so remote from the epicenter of the known outbreak, they become irrelevant as evidence.

The reader might wonder why items 14 to 17 don't argue in favor of H_m like the assertions of cholera and organophosphate operations (item 9) do. The answer is that the latter were apparently real events while the anthrax operations asserted to have occurred in 14-17 have no evidence to indicate that they are real - their credibility is affected by the perception that they represent barracks bragging, rumor, and hearsay, or have a fairly transparent political purpose.

A striking final result of the DST analysis is that the aggregate opinions resulting from both methods of combination support a belief in H_n over belief in H_m . This result is a direct consequence of the larger number of items that provide nearly pure support for H_n (6 items to 2). The Dempster rule leads to a very high level of belief in H_n , while the consensus rule leads to a more modest contrast between belief and disbelief. However, both rules indicate that the combination of all evidence items greatly lowers the “ignorance” component $\{H_m, H_n\}$ of the final opinion compared to the average uncertainty associated with the individual evidence items.

6.3 Would molecular biological evidence help resolve the controversy?

The molecular typing and sequencing methods available for analysis of any anthrax samples obtained during the Zimbabwe outbreak were much more primitive and costly than those available now. In fact, the explosion of applications of bacterial molecular genetics in epidemiology did not begin until nearly a decade after the event. Because of this there may have been little incentive to gather microbiological samples as evidence at the time. Nonetheless it seems possible that either clinical or veterinary samples from the 1978-9 outbreak were collected incidentally. Certainly if there were such samples, sequencing them now would reveal their relationship to other *B. anthracis* strains worldwide.⁵⁰ In this section we will speculate about ways such evidence might have an impact on opinions regarding H_m and H_n .

Some narratives of a man-made origin of the outbreak presume that a large volume of anthrax must have been manufactured by laboratory culture and processing in order to create enough contaminated cattle cakes, pellets or grain to account for the extensive operations where anthrax was disseminated. For example,

“Weighing all available evidence, it is suggested here that a plausible explanation for the sudden peak of anthrax in the Tribal Trust Lands beginning in November, 1978, is that one or more units attached to the Rhodesian military may have air dropped anthrax spores in these territories.” (Ref. 27, pages 203-204)

Although no data was collected on the number of infected cattle at the time, Burgess and Purkitt have proffered an estimate of the size of the outbreak:

“Debates continue today about the veracity of the claim that former Rhodesian forces, with South African involvement, planted anthrax spores in grain fed to cattle in guerrilla-held areas and caused an anthrax epidemic that struck 10,000 cattle in Zimbabwe in the early 1980s. (Ref. 41, pages 9-10)

Given the relative inefficiency of airdropping, such a scenario requires a manufacturing step. A single infected cattle cake requires at least 5×10^8 spores per cake⁵¹; since a typical culture concentration is around 5×10^8 spores per ml, the manufacture of 1000 cakes would require roughly a liter of culture. (Reliable infection of 10,000 cattle by airdrop might require about 30,000 cakes.) Nass argues that such manufacturing is well within the capabilities of the Rhodesian government:

“Technologically, production of anthrax spores is not a difficult problem. Anthrax weapons were developed and tested by at least the Japanese, British, and United States governments during the Second

World War and it is suspected that a number of other nations have developed or acquired the technologies since.

To manufacture anthrax weapons under ideal conditions, high-containment suites are employed. However, such facilities were not available to the nations that manufactured such weapons during World War II. ... Therefore, use of such suites, though desirable, is not mandatory for production of anthrax weapons.” (Ref. 27, page 204)

The discovery of one or more of the alleged cattle cakes with such a high dose of anthrax spores would certainly have provided evidence for Nass’s theory. The discovery of a manufacturing venue with extensive anthrax contamination (that could not be accounted for by environmental background levels) would have as well. Burgess and Purkitt have assessed that a candidate manufacturing site existed:

“The Rhodesian defense budget was very small, and the regime had one rudimentary chemical and biological warfare plant that received outside aid from South Africa.” (Ref. 41, page 8)

Nass also suggests:

“Soil sampling could be used to detect the presence of anthrax in soils. The extent of anthrax found in communal versus commercial farming areas would be interesting. Finding high anthrax spore counts in unusual locations, such as in places that do not support its growth, would contribute to an understanding of the epizootic.” (Ref. 27, page 207)

However, this suggestions seem less compelling in the sense that more rigorous anthrax management would probably account for less soil-accessible anthrax in the commercial farming regions, while the extensive sharing and remote butchering of anthrax infected cattle could easily account for contamination in areas with the wrong soil ecology to support endemic anthrax.

As discussed in section 4, molecular biological evidence can address questions of origin as well as document the presence or absence of a pathogen at tested locations. Nass’s speculates on potential origins of the outbreak strain:

“... cultures might have been available from the American Type Culture Collection in Rockville, Maryland, or from the Centers for Disease Control. Anthrax is also easily cultured from soil found in endemic areas or from some infected animal remains.” (Ref 27, page 205)

More sinister origins are also considered possible:

“One cannot totally exclude the possibility that some biological munitions were transferred from the United States to other countries prior to their destruction. It is also not impossible to imagine that such weapons could have been produced by a nation that was not a complying party to the Biological Weapons Convention (which in any case only entered into force in 1975), or even by a renegade group, and could have found their way to Zimbabwe.” (Ref. 27, page 205)

Burgess and Purkitt suggest who the “renegade group” might be:

“Officials of the Zimbabwean veterinary service repeated the claim [that the 1978 outbreak was man-made] in 1999. They noted that the strain of bacteria responsible for the outbreak was not native to Zimbabwe and immediately alleged that these incidents could be linked to South Africa’s past CBW program.”⁵² (Ref. 41, pages 9-10)

Genetic evidence might therefore be highly relevant:

“Characterization of the genetic structure of Zimbabwe anthrax strains can be used to estimate the likelihood that the strains found originated from locally occurring southern African strains, as opposed to strains that are found elsewhere or are held in laboratories.” (Ref. 27, page 207)

Arguably, however, Rhodesian counterinsurgency forces would most likely have chosen a “domestic” strain collected from a prior outbreak in the recent past – perhaps the 1974 outbreak discussed in reference 37. This strain would not only have been “at hand”, but also would be close to what would be expected if H_n were true. Thus, the interpretation of genetic comparison data might depend on very subtle genetic differences between the outbreak strain and reference strains held by veterinary or clinical laboratories within Zimbabwe. Given the slow rate of genetic change in anthrax lineages, a close match between the outbreak strain and a strain from the recent past might easily be considered non-informative.

7. Summary and conclusions

DST and related inferential methods such as Bayesian networks are attempts to “tame” the expression of intuitive uncertainty in complex situations where information is incomplete. It is irrelevant to consider whether analyses like the one presented here are “correct”. It is always possible that another person might render different judgments for the opinion vectors, or insist upon a different correspondence between verbal probability expressions and the probability intervals they represent. I can attest that ambivalence often arises through second-guessing one’s own initial assessments as one attempts to clarify or justify them.

Nonetheless, as a tool for organizing and assessing opinion evidence, DST does provide a transparent and internally consistent method for expressing belief and uncertainty and aggregating opinion over many items of evidence.

One of the salient findings of this analysis is that most, if not all of the most reliable and probative evidence items suggest a narrative centered around a localized outbreak, rather than the highly non-local hypothesis advanced by Nass. Moreover, a surprising conclusion is that, upon reflection, many items of testimonial evidence that have been proffered over the years for the man-made hypothesis can be recognized as nearly irrelevant. These may provide evidence that the Rhodesian counterinsurgency effort included attempts to use anthrax at various times and in various locations, but they do not support the conclusion that such activities were responsible for the outbreak that actually occurred. Conversely, the absence of certain expected types of evidence plays a critical role in tipping the balance of opinion to the “natural” hypothesis when all the evidence is aggregated. The observation²⁵ that “absence of evidence can be evidence of absence” certainly comes to the fore in this analysis.

Another surprise is that combining all the evidence reduces the overall level of ignorance to negligible levels, resulting in opinions that are very close to statements of probability. While one evidence combination rule (Dempster’s) implies that H_n is near certain, the other (Josang’s) leaves much more room for doubt. Josang formulated his rule as a weighted “consensus” of the combined opinions and has asserted that it alone is consistent with Bayesian probability theory.⁹ In more complex multi-hypothesis contexts, Dempster’s rule sometimes produces paradoxical results and has been criticized for that reason.^{53,54} In our context Dempster’s rule appears to more rapidly converge on the hypothesis favored by the larger number of pure or nearly pure opinions (i.e. where b or d are near zero.) Is there a “correct” rule for combining evidence? In spite of much research on the interpretation of, and alternatives to Dempster’s rule, this aspect of DST still appears to be somewhat murky.^{9,14,15}

One of the debatable virtues of the analysis presented here is its simplicity. The implementation of DST described in this report occupies a level of complexity somewhere between more qualitative methods like the Analysis of Competing Hypotheses and more rigorous, but complex methods (including DST frameworks that involve additional hypotheses.) Van Gelder has pointed out weaknesses of ACH that are clearly meliorated by DST, for example:

“By its very nature, being based on a matrix structure, the ACH approach does not consider what is “behind” or “underneath” any given piece of evidence. From a piece of evidence, it looks “forwards” or “upwards” to its bearing on the hypotheses under consideration. However the weight of a piece of evidence with respect to an hypothesis depends on information bearing upon that piece of

evidence. *e* may be quite (in)consistent with *h*, but how seriously we take this (in)consistency depends on how seriously we take *e* itself (its plausibility or credibility.)”⁴

Moreover, Van Gelder notes that ACH cannot handle the hierarchical structure of many hypotheses. This does not enter into our discussion, but is important for analyses where more than two focal hypotheses must be considered.

Katz has applied a more complex semiquantitative version of ACH to the evidence associated with the Yellow Rain controversy.^{55,56} This structured analysis scheme does include separate assessments of evidence credibility (denoted *veritas*) and probative value (denoted strength-of-association) but simply assigns a value of strong, medium, or weak to these assessments. In place of the discounting rule used in our analysis (equations 2 and 3), the combined *veritas* and strength of association for an evidence item are mapped to a numerical ranking with a 0 – 10 scale. Ranking of hypotheses depends on statistics calculated from these ranking values over all the evidence items. This level of algorithmic complexity is very similar, at least in spirit, to the DST approach used in this report. It would be interesting to compare Katz’s method with the DST approach for some common historical incident.

Other ways of organizing and analyzing evidence^{7,8} attempt to delve more deeply into the logical interconnections between evidence items and the inferences drawn from them, and result in more complex “diagrammatic” representations of narratives or probabilistic arguments. While these methods have the cache of formal rigor, they are very often difficult to explain to the non-expert, and are not necessarily more persuasive. Extracting quantitative metrics of the likelihood of a focal hypothesis from such analyses often requires eliciting a troublingly large number of parameters. The method presented here ignores fine points about the complex inter-relationships among the evidence items and intermediate steps in the chains of inference that connect them to the focal hypothesis. It simply elicits quick intuitive opinions about the reliability of each item of evidence and its probative strength. In return, one gains a transparent, if non-rigorous measure of the total support for the focal hypothesis that can be subjected to further scrutiny if desired, and easily reassessed as new evidence becomes available.

Notes and References

1. Shafer G., The Mathematical Theory of Evidence (Stanford University Press, 1976).
2. Of particular interest in the context of this report is: Curley SP, "The application of Dempster-Shafer theory demonstrated with justification provided by legal evidence", *Judgment and Decision Making* 2007; 2(5): 257-276.
3. Heuer, Richards J., Jr, "Chapter 8: Analysis of Competing Hypotheses", in *Psychology of Intelligence Analysis*, Center for the Study of Intelligence, **Central Intelligence Agency** 1999.
4. Van Gelder T, "Hypothesis Testing – What’s Wrong with ACH?", imvangelder.com/2007/12/31/hypothesis-testing-whats-wrong-with-ach/; Accessed on March 30, 2015.
5. Pope S and Josang A, "Analysis of Competing Hypotheses using Subjective Logic", Report, CRC for Enterprise Distributed Technology (DSTC Pty Ltd), Level 7, General Purpose South, The University of Queensland, Brisbane 4072 Australia, June, 2005.
Available at www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA463907; Accessed on April 1, 2015.
6. Cohen MS, Schum DA, Freeling ANS, Chinnis JO, "On the Art and Science of Hedging a Conclusion: Alternative Theories of Uncertainty in Intelligence Analysis", Decision Science Consortium, Inc. Technical Report 84-6, December 1985. Available at <http://www.cog-tech.com/papers/Scanned/Inference/Hedging.pdf>; Accessed on April 14, 2013.
7. Taroni F, Aitken C, Garbolino P, and Biedermann A, Bayesian Networks and Probabilistic Inference in Forensic Science, (John Wiley and Sons, Chichester, England, 2006).
8. Anderson T, Schum D and Twining W, Analysis of Evidence 2nd Ed. (Cambridge University Press, 2005).
9. Josang A, "A Logic for Uncertain Probabilities", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2001; 9(3): 1-31.
10. Josang A, "Artificial Reasoning with Subjective Logic", in *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*, A. Nayak and M. Pagnucco (eds.), Perth, December 1997, Australian Computer Society.

11. Josang, A, "Conditional Reasoning with Subjective Logic. *Journal of Multiple-Valued Logic and Soft Computing* 2008; 15(1): 5-38.
12. Haenni R and Hartmann S, "Modeling partially reliable information sources: A general approach based on Dempster-Shafer theory", *Information Fusion* 7: (2006) 361-379.
13. R. Haenni, "Aggregating referee scores: An algebraic approach", in: *Proceedings of the 2nd International Workshop on Computational Social Choice*, U. Endriss and P. Goldberg (eds.), University of Liverpool, (2008); pp. 277-288.
14. Sentz K and Ferson S, "Combination of Evidence in Dempster-Shafer Theory" Sandia National Laboratory Report SAND 2002-0835, April 2002.
15. Florea MC, Josselme A-L, Bossé E, and Grenier D, "Robust combination rules for evidence theory", *Information Fusion* 2009; 10: 183-197.
16. Kesselman R.F., *Verbal Probability Expressions in National Intelligence Estimates: A Comprehensive Analysis of Trends from the Fifties Through Post 9/11*, Masters Thesis, Mercyhurst College, 2008.
17. Weiss C., "Communicating Uncertainty in Intelligence and Other Professions", *International Journal of Intelligence and Counterintelligence* 2007; 21(1): 57-85.
18. Wallsten TS, "Meanings of Non-numerical Probability Phrases: Final Report", ARI Research Note 90-57; July 1990.
19. Teigen KH and Brun W, "Verbal Expressions of Uncertainty and Probability" in *Thinking: Psychological Perspectives in Reasoning, Judgment, and Decision Making*, D. Hardman and L Macchi eds., (John Wiley and Sons, 2003) Chapter 7.
20. *Analysis, Causality and Proof in Safety Investigations*, Australian Transport Safety Bureau ATSB Transport Safety Research Report; Aviation Research and Analysis Report - AR-2007-053; Commonwealth of Australia, 2008.
21. Sanderson WT, Stoddard RR, Echt AS, Piacitelli CA, Kim D, Horan J, Davies MM, McCleery RE, Muller P, Schnorr TM, Ward EM, and Hales TR, "Bacillus anthracis contamination and inhalational anthrax in a mail processing and distribution center", *Journal of Applied Microbiology* 2004; 96(5): 1048-56.
22. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Keim PS, and Aarestrup FM, "Population Genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak", *mBio* 2011; 2(4):e00157-11. doi: 10.1128/mBio.00157-11.

23. Tweney RD, Doherty ME, and Kleiter GD, "The pseudo-diagnostic trap: Should participants consider alternative hypotheses?", *Thinking & Reasoning* 2010; 16(4): 332-345. Note that my use of "pseudodiagnostic" differs from its use in this paper, where it refers to information-seeking behavior that neglects determining $P(E|\bar{H})$ in favor of seeking out additional evidence E' for which $P(E'|H) \neq 0$. It seems appropriate to extend the definition to include cases where inferences are made based on only knowledge of $P(E|H)$ and the subjective intuition that $P(E|\bar{H}) < P(E|H)$.

24. Keim PS, et al., "Reply to "South Asia instead of Nepal may be the origin of the Haitian cholera outbreak strain", *mBio* 2011; 2(6): e00245-11.
doi:10.1128/mBio.00245-11.

25. Sober E, "Absence of evidence and evidence of absence: evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads", *Philosophical Studies* 2009; 143: 63–90. See also: Pedersen NJLL and Kallestrup J, "The epistemology of absence-based inference", *Synthese* 2013; 190: 2573–2593.

26. This single hypothesis actually contains a hierarchy of more specific hypotheses. For example: Rhodesian forces were responsible, or the ZIPRA guerrillas were responsible; similarly: the outbreak consisted of many independent foci, or a single focus. Each of these distinguishable hypotheses corresponds to a different narrative that could be proffered to elaborate the "man-made" hypothesis. Such sets of multiple hypotheses can be handled in the DST formalism, at the expense of added complexity.

27. Nass M, "Anthrax Epizootic in Zimbabwe, 1978-1980: Due to Deliberate Spread?" *Physicians for Social Responsibility Quarterly (PSRQ)* 1992; 2: 198-209.
http://www.whale.to/vaccine/anthrax_epizootic.html; Accessed on September 12, 2014

28. Nass M, "Zimbabwe's Anthrax Epizootic", *Covert Action Quarterly* 1992-93; 43: 12-18, 61.

29. Chikerema SM, Pfukenyi DM, Matope G, and Bhebhe E, "Temporal and spatial distribution of cattle anthrax outbreaks in Zimbabwe between 1967 and 2006", *Trop. Anim. Health Prod.* 2012; 44:63-70.

30. Stein CD, "The History and Distribution of Anthrax in Livestock in the United States", *Veterinary Medicine* 1945;; p. 340-349.

31. Data from outbreaks recorded within the United States between 1945 and 1949 are provided in Table II of Stein CD, "Anthrax in Livestock During 1949 and Incidence of the Disease from 1945 to 1949", *Veterinary Medicine* May 1950; pp

205-208. This data is reflective of a period in U.S. history when vaccination was not systematically and near-universally employed to control anthrax. If we plot this data we find that within this time period there were two very large “outlier” outbreaks with greater than 1000 animals lost. Plotting the frequency distribution of anthrax outbreaks with losses greater than a certain value indicates that the distribution has a fat tail – the likelihood of rare but large events does not fall exponentially with event size, but decreases much more slowly. Under these conditions, one’s intuition about the rarity of large events based on the “usual” or “average” observed event sizes (note that about 80% of losses per year in any given state were less than 100 cattle, according to Table II) tends to greatly underestimate the probability of an order-of-magnitude-or-more larger event. Note also that the actual number of cattle infected in the Zimbabwe outbreak was never determined, and could be much smaller than the observed number of human cases, see evidence item #7.

32. Davies JCA, “A Major Epidemic of Anthrax in Zimbabwe Part I: The spread of the epidemic in Matabeleland, Midlands, and Mashonaland provinces during the period November 1978 to October 1980”, *The Central African Journal of Medicine* 1982; 28(12): 291-298.

33. Davies JCA, “A Major Epidemic of Anthrax in Zimbabwe Part II: The Distribution of Cutaneous Lesions”, *The Central African Journal of Medicine* 1983; 29(1): 8-12.

34. Parker J, Assignment Selous Scouts, Inside story of a Rhodesian Special Branch Officer (Galago Publishing, Alberton, South Africa, 2006)

35. Pugh AO and Davies JCA, “Human Anthrax in Zimbabwe”, *Salisbury Medical Bulletin* 1990; 68(Supp): 32-33.

36. Cilliers JK, Counter-Insurgency in Rhodesia (Croom Helm Ltd., London, 1985).

37. Roberts CJ and Chambers PG, “An Outbreak of Anthrax in the Mondoro Tribal Trust Lands”, *The Central African Journal of Medicine* 1975; 21(4): 73-76.

38. Lawrence JA, Foggin CM, Norval RAI, “The effects of war on the control of diseases of livestock in Rhodesia (Zimbabwe)” *Veterinary Record* 1980; 107: 82-85.

39. Brickhill J, “Zimbabwe’s Poisoned Legacy: Secret War in Southern Africa”, *Covert Action Quarterly* 1992-3; 43: 7-10.

40. Henrik Ellert, The Rhodesian Front War: Counter-Insurgency and Guerrilla War in Rhodesia, 1962–1980 (Gweru, Zimbabwe: Mambo Press, 1989), pp. 109–12.

41. Burgess SF and Purkitt HE, “The Rollback of South Africa’s Chemical and Biological Warfare Program”, USAF Counterproliferation Center, Air War College,

April 2001.

42. Tom Mangold and Jeff Goldberg, *Plague Wars: The Terrifying Reality of Biological Warfare* (MacMillan Publishers, Great Britain, 1999).

43. Le Roux P, Stubbs AJ, and Donnelly PH, "Problems and prospects of increasing beef production in the tribal trust lands", *Zambia* 1978; VI(i): 37-48.

44. Alexander J, McGregor J, and Ranger T, *Violence and Memory: One Hundred Years in the 'Dark Forests' of Matabeleland*, (Weaver Press, Harare, 2000).

45. Mark Wheelis, "A Short History of Biological Warfare and Weapons" in *The Implementation of Legally Binding Measures to Strengthen the Biological and Toxin Weapons Convention*, Marie Isabelle Chevrier, Krzysztof Chomiczewski, Henri Garrigue, Gyorgy Granstztoi, Malcolm R. Dando, and Graham S. Pearson, eds., Joint proceedings volume based on the NATO Advanced Research Workshop held in Warsaw in 2000 and the NATO Advanced Study Institute held in Budapest in 2001; NATO Science Series; Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

46. Coen B and Nadler E, *Dead Silence: Fear and Terror on the Anthrax Trail*, (Transformer Films, 2009). Kindle version.

47. Baxter P, *Rhodesia: Last Outpost of the British Empire 1890-1980* (Galago Publishers, Republic of South Africa, 2010)

48. Martin D, "Human Anthrax Scares were a Major Reality in Southern Africa" *Southern Africa News Features*, 16 November 2001; <http://webcache.googleusercontent.com/search?q=cache:enVrH1CZriIj:www.sardc.net/editorial/sanf/2001/iss21/specialreport.html+&cd=4&hl=en&ct=clnk&gl=us>; Accessed on October 24, 2014.

49. Bale JM, "South Africa's Project Coast: 'Death Squads', Covert State-Sponsored Poisonings, and the Dangers of CBW Proliferation" *Democracy and Security* 2006; 2: 27-59.

50. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, Zanecki SR, Pearson T, Simonson TS, U'Ren JM, *et al.*: "Global genetic population structure of *Bacillus anthracis*", *PLoS ONE* 2007; 2(5): e461.

51. Turnbull PCB and Shadomy SV, "Anthrax from 5000 BC to AD 2010, Chapter 1 in *Bacillus Anthracis and Anthrax*, Nicholas H. Bergman, ed. (John Wiley & Sons, Hoboken N.J., 2011).

52. I did not include this allegation as an evidence item because the South African program was completely exposed in extensive public hearings and no connection with the Rhodesian outbreak was ever found. As Burgess and Purkitt make clear, it is more likely that the Rhodesian security services influenced the South African program, which began *after* the fall of the Rhodesian government.

53. Zadeh LA, "On the validity of Dempster's rule of combination of evidence", Technical Report 79/24, University of California, Berkely, 1979.

54. Haenni R, "Shedding New Light on Zadeh's Criticism of Dempster's Rule of Combination", FUSION'04, 7th International Conference on Information Fusion, Stockholm, Sweden, July 1st, 2004.

55. Katz RL, *YELLOW RAIN REVISITED: LESSONS LEARNED FOR THE INVESTIGATION OF CHEMICAL AND BIOLOGICAL WEAPONS ALLEGATIONS*; PhD Thesis, Princeton University, 2005.

56. Katz R and Singer B "Can an attribution assessment be made for Yellow Rain? Systematic reanalysis in a chemical-and-biological-weapons use investigation" *Politics and the Life Sciences* 2007; 26(1): 24-42.

Appendix A. Opinions in two dimensions

The opinion vectors (b,d,i) are nominally three dimensional objects, but the normalization relation $b+d+i=1$ means that they only carry two dimensions of information. In these notes I explain the meaning of the opinion vector in two dimensions, as drawn from Shafer, Josang and other authors.

The basic idea is that an opinion represents an “investment” of probability in a given assertion that represents an inference from evidence. When we feel that (a) we have a sufficient amount of evidence relative an assertion, (b) we are very confident in the truth of that evidence, and (c) we are certain about inferring the truth or falsity of the assertion from that evidence, we are willing to “invest” a large fraction of our subjective probability in the belief or disbelief of that assertion. To the extent that we are uncertain about the evidence or the inference, we reserve some of that probability and assign it to “ignorance” i , rather than to b or d . In other words, we “hold back” deciding between belief and disbelief to some extent.

So, when we quantify an opinion we can think of starting with a “full measure” of probability ($= 1$), and then partitioning it among b , d and i such that $b+d+i = 1$.

We can take the amount of probability we are willing to invest in b and d as a measure of our confidence in our opinion. Formally, we can define a measure of confidence χ by:

$$\chi = 1 - i = b + d \quad (\text{A1a})$$

Given the fraction of probability we have invested in $b+d$ we can also define a measure of the fraction invested in b alone, which we will call the “likeliness” of the assertion:

$$\mathcal{L} = \frac{b}{b+d} \quad (\text{A1b})$$

Note that we can define b , d , and i in terms of just χ and \mathcal{L} :

$$b = \mathcal{L}\chi \quad (\text{A2a})$$

$$d = (1 - \mathcal{L})\chi \quad (\text{A2b})$$

$$i = 1 - \chi \quad (\text{A2c})$$

Equation (A2a) expresses the idea that belief = likeliness x confidence.

Equation (A2b) expresses the idea that disbelief = unlikeliness x confidence

Equation (A2c) expresses the idea that the degree of ignorance = the lack of confidence in our evidence and/or our inference. Thus, every opinion vector is equivalent to a 2-dimensional likeliness-confidence vector (\mathcal{L}, χ) .

Suppose that something causes us to become more confident in the evidence and/or the inferential path that leads from the evidence to the assertion in question. To represent this stronger opinion, we transfer some portion of the probability that was reserved in i to b . The maximum amount of additional probability that we could transfer to b is the amount that was initially reserved to i . We can express this as an upper bound to the probability that could be assigned to belief in the assertion in question:

$$P_{upper} = b + i = 1 - d \quad (\text{A3a})$$

On the other hand, it might turn out that the additional evidence leads us to transfer the probability from i to d instead. Thus, the lower bound on the probability is simply b :

$$P_{lower} = b \quad (\text{A3b})$$

Thus, we can also express an opinion as a two dimensional vector (P_{upper}, P_{lower}) . This provides a basis for assigning belief vectors from verbal probability expressions, which correspond to ranges with upper and lower bounds. For completeness we have:

$$b = P_{lower} \quad (\text{A4a})$$

$$d = 1 - P_{upper} \quad (\text{A4b})$$

$$i = P_{upper} - P_{lower} \quad (\text{A4c})$$

$$\chi = 1 - (P_{upper} - P_{lower}) \quad (\text{A5a})$$

$$\mathcal{L} = \frac{P_{lower}}{1 - (P_{upper} - P_{lower})} \quad (\text{A5b})$$

$$P_{lower} = \chi \mathcal{L} \quad (\text{A6a})$$

$$P_{upper} = 1 - \chi(1 - \mathcal{L}) \quad (\text{A6b})$$

Appendix B: DST and probability

DST can be thought of as a formal system for expressing evidential weight, in contrast to likelihood (Curley, reference 1b). However, there is a formal connection between DST and likelihood descriptions of uncertainty that can be useful in many cases. For example, it provides a means for translating opinion conclusions that result from DST calculations directly into Bayesian-like probabilities. However, this mapping is not unique, as will be discussed in this appendix.

Suppose that we have a piece of evidence E , say an assertion from a witness that has a bearing on a hypothesis H we are considering. Let E = “the assertion is true” and \bar{E} = “the assertion is false”. Based on what we know about the witness, we may (subjectively) assign a likelihood $P(E)$ to the truth of the assertion, and hence a likelihood

$$P(\bar{E}) = 1 - P(E) \quad (\text{B1})$$

that the assertion is false. Next, consider how the truth of E impacts the truth of our hypothesis. $P(H|E)$ is the likelihood of H given the truth of E . In many cases, a piece of evidence can have a very simple relationship to the hypothesis based on logic, either $P(H|E) = 1$, or $P(H|E) = 0$. For example if E = “the suspect was dead at the time”, and H = “the suspect committed the murder” then $P(H|E) = 0$. In some cases, of course, E may only imply that H is more or less likely than not, and it would be necessary to provide a subjective estimate of a value for $P(H|E)$.

Note that the quantities $P(H|E)$ and $P(E)$ respectively correspond closely to intuitive notions of the “likeliness” of a hypothesis given the evidence, and the “degree of confidence” we have in the evidence itself. Haenni has pointed out that our belief in H that is induced by E is naturally represented by the product $P(H|E) \cdot P(E)$. Thus:

$$b = P(H|E)P(E) \quad (\text{B2a})$$

$$d = P(\bar{H}|E)P(E) = (1 - P(H|E))P(E) \quad (\text{B2b})$$

$$i = P(\bar{E}) = 1 - P(E) \quad (\text{B2c})$$

Equation (B2c) is not transparently obvious, but is implied by (B2a) and (B2b) combined with the relation $b+d+i = 1$.

If we interpret $P(H|E)$ as a conditional probability, then (B2a) implies that our belief in H induced by E is simply the joint likelihood: $b = P(H, E)$. Similarly, our degree of disbelief is $d = P(\bar{H}, E)$. The degree of belief we withhold from either hypothesis is simply the likelihood that the evidence itself is false.

From the perspective of probability theory, equations (B2) do not provide complete information for inferring the probability of H . We also need the joint probability

$P(H, \bar{E})$, because $P(H) = P(H, E) + P(H, \bar{E})$. There are two choices for a complete theory of evidence. The first is that we somehow acquire a subjective estimate for the conditional probability $P(H|\bar{E})$, then use the equation

$$P(H, \bar{E}) = P(H|\bar{E})P(\bar{E}) \tag{B3}$$

to obtain the joint probability we need. Equation (B3) could be said to define a “counter-evidence belief” parameter for H – the belief in H induced by knowing that E is false. If we are able to do this, we simply arrive at Bayesian inference for H. The practical problem with this is that often subjective estimates $P(H|\bar{E})$ are very difficult, even though assigning a value to $P(H|E)$ is simple. To see this, refer back to the “dead suspect” case cited above. Because of this difficulty, evidence theories based on DST choose a different method to estimate $P(H|\bar{E})$.

Table B1. Correspondence between joint probabilities and DST variables.

$P(H, E)$ b	$P(H, \bar{E})$ αi	$P(H) = b + \alpha i$
$P(\bar{H}, E)$ d	$P(\bar{H}, \bar{E})$ $(1 - \alpha)i$	$P(\bar{H}) = d + (1 - \alpha)i$
$P(E)$ $1 - i$	$P(\bar{E})$ i	1

The basic idea is to transfer some part of the ignorance i into the unknown joint probability. Table B1 illustrates this process by showing each joint probability of interest and the corresponding DST variable, along with the column and row sums in the bottom row and rightmost column respectively. The variable α represents the fraction of the ignorance i that is transferred to the joint probability $P(H, \bar{E})$. Note that $\alpha = P(H|\bar{E})$ by definition. Therefore, in cases where H is logically incompatible with \bar{E} , we must set $\alpha = 0$. Conversely, if \bar{H} is logically incompatible with \bar{E} , we must set $\alpha = 1$. Given the definitions of plausibility in main text equations (5) and the expressions for $P(H)$ and $P(\bar{H})$ in table B1 we see that

$$b \leq P(H) \leq Pl(H) \tag{B4a}$$

$$d \leq P(\bar{H}) \leq Pl(\bar{H}) \tag{B4b}$$

In the absence of a logical argument regarding the value of $P(H|\bar{E})$ the choice of α depends not on a subjective probability estimate, but rather on a subjective choice between rules that have different consequences when calculating $P(H)$. In addition to the two rules based on logic discussed above, three additional rule choices discussed by Haenni [x] are summarized in Table B2.

When we have no reason to think that \bar{E} favors H over \bar{H} , then the natural heuristic is to set $\alpha = 1/2$. This results in a rule for mapping opinion to probability that Haenni (reference 8) has labeled “pignistic”. In the belief-weighting scheme we simply

apportion i in proportion to the belief accorded to H and \bar{H} . Finally, plausibility weighting, while less intuitive than the other two rules, has two advantages. First, it guarantees that if we first combine evidence according to the Dempster rule (2), then calculate $P(H)$, we get the same result as if we first calculated $P(H)$ for each evidence item separately then combined the results as if $(b, d, i) = (P(H), P(\bar{H}), 0)$. A consequence of this is that the logarithms of the odds ratios $\frac{P(H)}{P(\bar{H})}$ of each separate evidence item add up to the logarithm of the odds ratio calculated for the combined evidence. Thus, the odds ratios calculated this way behave like likelihood ratios, which are arguably the best way to represent evidence strength.

The second advantageous property of the plausibility weighting rule is that it is more conservative than either the “equal weighting” or “belief weighting” rules in the sense that odds ratios calculated from $P(H)$ using the plausibility weighting rule are always smaller. Thus, it always assesses a lower “weight of evidence” than the other two rules.

Table B2. Five potential choices for α .

	α	$1 - \alpha$	Reasoning
Logical	1	0	H is incompatible with \bar{E}
	0	1	\bar{H} is incompatible with \bar{E}
Equal weighting	$\frac{1}{2}$	$\frac{1}{2}$	No reason to favor/disfavor $P(H, \bar{E})$ versus $P(\bar{H}, \bar{E})$
Belief weighting	$\frac{b}{b + d}$	$\frac{d}{b + d}$	Distribute $P(\bar{E})$ in proportion to $P(H, E)$ and $P(\bar{H}, E)$
Plausibility weighting	$\frac{1 - b}{1 + i}$	$\frac{1 - d}{1 + i}$	Mapping between (b, d, i) and $(P(H), P(\bar{H}), 0)$ is homomorphic wrt the Dempster combination rule

It is not always necessary, of course, to calculate probabilities when DST is used to analyze evidence. If the only evidence at hand is of the sort where subjective assessments of confidence $\hat{\omega}(E)$ and likeliness $\hat{\omega}(H|E)$ are necessary, then the DST formalism can be used to calculate belief and plausibility values alone and Bayesian probabilities need not enter the picture.

Appendix C: Opinion vectors for the 17 evidence items.

#	Evidence	Credibility of evidence (Antecedent likelihood)	Induced belief in Hm if true (Implication likelihood)	Antecedent			Implication			Result		
				b(E)	d(E)	i(E)	b(H E)	d(H E)	i(H E)	b(H,E)	d(H,E)	i(H,E)
1	Outbreaks were uncommon	Moderately unlikely	Moderately likely	0.10	0.67	0.23	0.67	0.10	0.23	0.067	0.01	0.923
2	The outbreak was unusually large	Highly likely	About even odds	0.90	0.00	0.10	0.33	0.33	0.34	0.297	0.297	0.406
3	There was "jumping" from focus to focus	Unlikely	Highly likely	0.00	0.67	0.33	0.90	0.00	0.10	0	0	1
4	Mostly black-owned cattle were affected	Highly likely	About even odds	0.90	0.00	0.10	0.33	0.33	0.34	0.297	0.297	0.406
5	The outbreak stopped at Zimbabwe's borders	Unlikely	Likely	0.00	0.67	0.33	0.67	0.00	0.33	0	0	1
6	The outbreak was "in season"	virtually certain	Moderately unlikely	0.98	0.00	0.02	0.10	0.67	0.23	0.098	0.6566	0.2454
7	1975 outbreak 5:1 human to cattle	Highly likely	Moderately unlikely	0.90	0.00	0.10	0.10	0.67	0.23	0.09	0.603	0.307
8	Other diseases increased as well	Highly likely	Moderately unlikely	0.90	0.00	0.10	0.10	0.67	0.23	0.09	0.603	0.307
9	Other CB events occurred	Highly likely	Highly likely	0.90	0.00	0.10	0.90	0.00	0.10	0.81	0	0.19
10	No black witnesses	virtually certain	Unlikely	0.98	0.00	0.02	0.00	0.67	0.33	0	0.6566	0.3434
11	No government documents	virtually certain	Unlikely	0.98	0.00	0.02	0.00	0.67	0.33	0	0.6566	0.3434
12	Ellert's testimony absence of evidence	virtually certain	Highly unlikely	0.98	0.00	0.02	0.00	0.90	0.10	0	0.882	0.118
13	Parker testimony regarding Nkai	moderately likely	highly likely	0.67	0.10	0.23	0.90	0.00	0.10	0.603	0	0.397
14	McGuinness testimony (Plumtree)	moderately unlikely	About even odds	0.10	0.67	0.23	0.33	0.33	0.34	0.033	0.033	0.934
15	Baxter testimony (Malvernia)	moderately unlikely	About even odds	0.10	0.67	0.23	0.33	0.33	0.34	0.033	0.033	0.934
16	Peter Stiff testimony (Mazoe valley)	moderately unlikely	About even odds	0.10	0.67	0.23	0.33	0.33	0.34	0.033	0.033	0.934
17	Unnamed intel officer (Gutu, Chlimanzi, Masvengo, Mberenga)	moderately unlikely	About even odds	0.10	0.67	0.23	0.33	0.33	0.34	0.033	0.033	0.934

Appendix D. The “In-Season” evidence from another point of view

Among the evidence items from the Zimbabwe outbreak was the observation that the outbreak occurred “in-season” for natural outbreaks (evidence item 6.) Chikerema reported that there was about a 3x higher probability for anthrax outbreaks to occur in the September to November “season” in Rhodesia relative to other months. The actual 1978-79 outbreak began in November 1978 – in other words it was “in-season”. The question is, does this fact provide support for the hypothesis that the outbreak was natural H_n , and not man-made H_m ?

Let S = “the outbreak occurs in-season”

Then \bar{S} = “the outbreak occurs out-of-season”

According to Chikerema

$$P(S|H_n) = 3P(\bar{S}|H_n)$$

Since $P(S|H_n) + P(\bar{S}|H_n) = 1$ we have:

$$P(S|H_n) = 3/4$$

$$P(\bar{S}|H_n) = 1/4.$$

We want to know:

$$P(H_n|S) = \frac{P(S|H_n)P(H_n)}{P(S|H_n)P(H_n) + P(S|H_m)P(H_m)}$$

What can we say about $P(S|H_m)$?

We can argue that the Selous Scouts (say) were so clever that they were sure to schedule their anthrax spreading operation to be “in-season”, in which case

$$P(S|H_m) = 1 \text{ (the “clever Scouts” assumption)}$$

In this case, the evidentiary weight of S , expressed as the likelihood ratio is:

$$\frac{P(S|H_n)}{P(S|H_m)} = 3/4.$$

Therefore the observation of an in-season outbreak is very weak evidence, slightly favoring the man-made hypothesis H_m .

At the other extreme we could argue that the Scouts might have scheduled their operation based on other considerations, not feeling constrained to disguise it as a natural event (especially if they were going to spread the rumor that the outbreak was caused by the ZIPRA insurgents.) In this case there is an equal probability that the operation might have occurred any time during the year. The length of the in-season is about 3 months, with the remaining 9 months “out-of-season”, so in this case

$$\frac{P(S|H_m)}{P(\bar{S}|H_m)} = 1/3 \text{ (The “any-time is a good time for anthrax” assumption)}$$

Again, using $P(S|H_m) + P(\bar{S}|H_m) = 1$ we have:

$$\begin{aligned} P(S|H_m) &= 1/4 \\ P(\bar{S}|H_m) &= 3/4. \end{aligned}$$

Therefore, under this assumption the evidentiary value of S is:

$$\frac{P(S|H_n)}{P(S|H_m)} = 3,$$

and we would conclude that S provides modest support for the natural origin hypothesis.

Note, however, that the “clever Scouts” hypothesis assumes that the veterinarians supposedly co-opted into the operation felt strongly enough about the seasonality of natural anthrax outbreaks that they would have argued for an in-season operation, and that this consideration would have dominated the scheduling decision. Doubt about this underlying assumption arguably weakens the credibility of an assertion that the Scouts were sure to choose an in-season operational schedule, and implies that $P(S|H_m) < 1$. This reduces an already small interval over which S supports H_m rather than H_n .

Given these various arguments about the values of the likelihoods, the posterior probability of H_n can be written:

$$P(H_n|S) = \frac{0.75P(H_n)}{0.75P(H_n) + \alpha P(H_m)}$$

where $\frac{1}{4} \leq \alpha < 1$.

If we assumed equal priors $P(H_n) = P(H_m) = 1/2$, then

$$0.43 < P(H_n|S) \leq 0.75$$

Using the conservative plausibility weighting scheme (Appendix A) to derive probabilities from the opinion vector for item 6, we find that $P(H_n/S) = 0.72$. This suggests that the opinion derived for this evidence item was somehow equivalent to assuming either the “any-time” hypothesis was true or that the prior probability of H_n was larger than H_m .