Journal of Scientific Exploration, Vol. 31, No. 4, pp. 677–686, 2017 0892-3310/17

ESSAY

Shakespeare: The Authorship Question, A Bayesian Approach

P. A. Sturrock
sturrock@stanford.edu

Reprinted from Journal of Scientific Exploration, Volume 22, No. 4, pp. 529–537, 2008.

Abstract—Bayesian probability theory can be helpful in organizing the multiple evaluations required in analyzing complex problems that involve the comparison of several hypotheses on the basis of several datasets. The problem of deciding the authorship of the Shakespeare literary material falls under this heading. We here discuss just one aspect of this major problem: whether or not the available evidence indicates that "William Shakspere," of Stratford-upon-Avon, was a writer. We consider 24 known writers who lived in England at the same time as Shakspere. For each of these writers, and for Shakspere, we follow Price in considering whether or not there exists evidence in each of 10 categories relevant to the literary profession. We find that there is evidence conforming to at least 3 categories for each comparison author, but none for Shakspere. We evaluate the probability, based on this information, that Shakspere was a writer similar to the 24 comparison writers. According to this analysis of Price's data, we find that there is only one chance in 100,000 that Shakspere was a writer. These considerations support the heretical view that Shakspere was not the author of the Shakespeare material.

Keywords: Shakespeare—statistics—probability

Introduction

It is generally—but not universally—assumed that the plays and poems associated with the name "Shakespeare" were written by a man who was born and raised, and died and was buried, in Stratford-upon-Avon in the county of Warwickshire in the West Country of England, ninety miles northwest of London. For recent accounts of the orthodox "Stratfordian" position, and for references to supporting material, one may refer for instance to Bryson (2007) and Honan (1998). The available records refer to the "Stratford" person variously as "Shackespere," "Shackspeare," "Shaxper," "Shaxpere," and "Shexpere," as well as "Shakespeare." It is convenient to follow Price (2001) in referring to the man from Stratford as "Shakspere," reserving the

TABLE 1
24 Comparison Authors and Shakspere

Author	DOB	DOD	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Francis Beaumont	1584	1616	y						y	y		y
George Chapman	1559	1634			y	y		y	y	y	y	
Samuel Daniel	1562	1619	y	y		y	y	y	y	y		
Thomas Dekker	1572	1632		y	y			y	y	y		
Michael Drayton	1563	1631		y	y	y		y	y	y		y
William Drummond	1585	1649	y	y			y	y	y	y	y	
John Fletcher	1579	1625			y				y	y	y	
Robert Greene	1558	1592	y		y	y			y	y		y
Gabriel Harvey	1550	1630	y	y		y	y	y	y	y	y	
Thomas Heywood	1573	1641			y		y	y	y	y		y
Ben Jonson	1572	1637	y	y	y	y	y	y	y	y	y	y
Thomas Kyd	1558	1594	y	y		y				y		
Thomas Lodge	1558	1625	y	y		y			y	y	y	
John Lyly	1554	1606	y	y		y		y	y	y		
Christopher Marlowe	1564	1593	y			y				y		y
John Marston	1576	1634	y		y	y	y	y	y	y	y	
Philip Massinger	1583	1640	y	y	y	y	y	y	y	y		
Thomas Middleton	1580	1627	y		y		y	y	y	y		
Anthony Mundy	1560	1633			y	y	y	y	y	y		y
Thomas Nashe	1567	1601	y	y	y	y	y	y	y	y	y	
George Peele	1556	1596	y	y	y	y	y	y		y		
Edmund Spenser	1552	1599	y	y		y			y	y	y	y
Thomas Watson	1557	1592	y			y			y	y		y
John Webster	1578	1632			y				y	y		
William Shakspere	**1564**	**1616**										

DOB = date of birth; DOD = date of death; y = yes for confirmation of evidence for I-1 through I-10, the 10 items related to the profession of a writer.

name "Shakespeare" for the person or persons who individually or collectively produced the "Shakespeare" literary corpus.

Price's (2001) "Chart of Literary Paper Trails" lends itself to statistical analysis. This chart compares personal and literary records left by 24 known

TABLE 2
10 Items of Evidence Related to the Profession of Writer

Item		Yes	No
1	Evidence of education	17	7
2	Record of correspondence, especially concerning literary matters	14	10
3	Evidence of having been paid to write	14	10
4	Evidence of a direct relationship with a patron	16	8
5	Extant original manuscript	10	14
6	Handwritten inscriptions, receipts, letters, etc., touching on literary matters	15	9
7	Commendatory verses, epistles, or epigrams contributed or received	21	3
8	Miscellaneous records (e.g., referred to personally as a writer)	24	0
9	Evidence of books owned, written in, or given	9	15
10	Notice at death as a writer	9	15

Columms 3 and 4 list the number of comparison authors for which there is ("yes") or there is not ("no") such evidence. A "yes" represents one or more qualifying pieces of evidence.

Elizabethan and Jacobean writers during their lifetimes, together with notice at death as a writer within 12 months of the writer's demise. This chart comprises 10 "categories of evidence."[1] For each category, Price follows historical and biographical practice (see, for instance, Altick & Fenstermaker 1993: especially p. 49, George 1909:48–49, Kendall 1985:xiii, Williams 2003: especially p. 58) in requiring that the evidence be (a) contemporaneous, (b) personal, and (c) related to the relevant profession—in this case, a literary life. Each category is reviewed for each of the 24 known authors and for Shakspere.

The comparison authors are listed in Table 1. The earliest (Gabriel Harvey) lived from 1550 to 1630, and the last (William Drummond) lived from 1585 to 1649. William Shakspere lived from 1564 to 1616. The average years of birth and death of these comparison authors were 1567 (standard deviation 10 years) and 1620 (standard deviation 18 years). As far as chronology is concerned, this seems a reasonable comparison set.

For each of these authors, and also for Shakespere, we note in Table 1 whether or not there is "paper-trail" evidence for each of the 10 categories. These categories are specified in Table 2, where we note the number of comparison authors for which such evidence has been found. We note from Table 1 that, for every comparison author, we have at least 3 items of rel-

TABLE 3
Probability of Evidence

Item	Yes	No	*P*	1 – *P*
1	17	7	0.69	0.31
2	14	10	0.58	0.42
3	14	10	0.58	0.42
4	16	8	0.65	0.35
5	10	14	0.42	0.58
6	15	9	0.62	0.38
7	21	3	0.85	0.15
8	24	0	0.96	0.04
9	9	15	0.38	0.62
10	9	15	0.38	0.62

For each of the 10 items, columns 2 and 3 list the number of comparison authors for which there is, ("yes"), or there is not, ("no"), such evidence. Columns 4 and 5 list the probability that there would or would not be such evidence for a 25th author, on the assumption that the authors all have similar habits, which result in the creation of literary paper trails with common characteristics including probability of survival. It may be noted that published evidence, such as a personal commendatory verse, has inherently greater probability of survival than does, for instance, a handwritten letter or manuscript.

evant literary evidence. For Shakspere, by comparison, we have none. We now examine this discrepancy statistically.

Analysis

For each category, we proceed as follows: We determine whether or not there is evidence relevant to that category for each of the N ($N = 24$) comparison authors. Suppose we find such evidence for K of those authors. The problem then is to determine the probability of finding comparable evidence for the "test" author, Shakspere, on the assumption that he belongs to the same group as the comparison authors. We may regard this as a sequence of $N + 1$ trails. We are given the result for the first N in this sequence (the comparison authors), and we wish to estimate the probability P of getting a positive result for the $(N + 1)$th trial (i.e. for Shakspere). We find (see Appendix) that this is given by

$$P = \frac{K+1}{N+2} \quad , \tag{1}$$

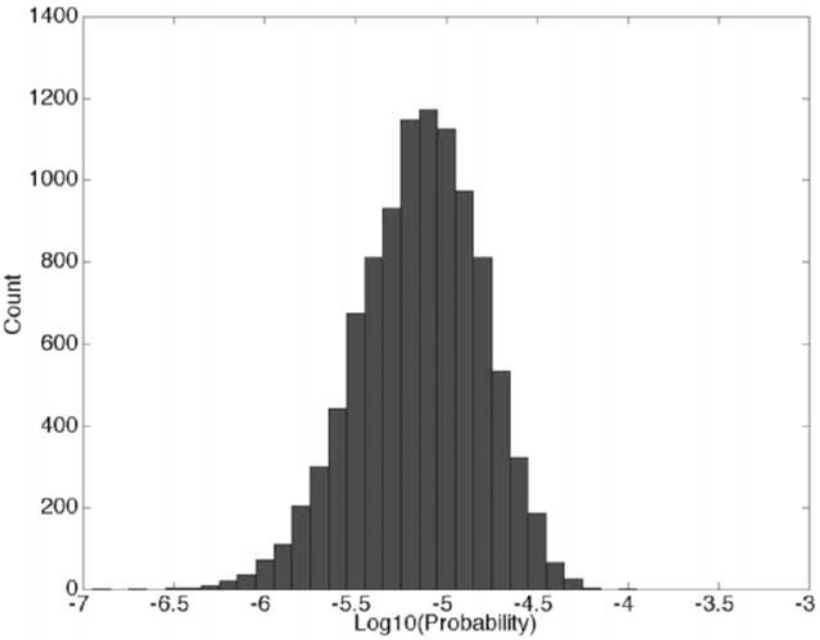

Figure 1. Histogram of $\log_{10}$(Probability) found from Rule of Succession estimates derived from 10,000 Monte Carlo simulations of the dataset shown in Table 3.

which is known as the Laplace "rule of succession" (Howson & Urbach 1989, Jaynes 2003). These estimates are listed in Table 3.

As we see from Table 1, the Shakspere entries are remarkable in that they show no evidence for any of the 10 categories that we have found to be characteristic of most of the comparison authors. The question is whether or not this discrepancy is significant. The usual procedure, in statistical analyses, is to compute the probability that a certain result may have occurred by chance. We see from Table 1 that the probability that there might by chance be no evidence for item 1 is 0.31; for category 2 it is 0.42, etc. The probability that all 10 estimates are unrepresented purely by chance is the product of these 10 estimates, which is found to be 10^{-3}. If we expect evidence relevant to Shakspere to conform to the evidence we find for the 24 comparison authors, there is only one chance in 100,000 that the results concerning Shakspere would have occurred by chance.

One may be justifiably concerned that the above estimate may depend critically on the precise selection of comparison authors. The best response for this concern would be for other Shakespeare scholars to develop independent "charts of literary paper trails." However, one can to some extent judge the dependence of the result on the selection process by using the

bootstrap procedure (Efron & Tibshirani 1993). For each bootstrap simulation, and for each cell in Table 1, we enter a value selected randomly (with replacement) from the column to which the cell belongs. We have carried out 10,000 such bootstrap simulations, and we show in Figure 1 a histogram of the resulting estimates of the logarithm (base 10) of *P*. We see that our estimated value 10^{-5} is very close to the peak of the histogram, indicating that our probability estimate is not sensitive to the precise selection of comparison authors.

The probability of obtaining the results of an experiment or test on the basis of an assumed "null hypothesis" (in this case, the hypothesis that Shakspere was in most respects similar to the comparison authors) is known as a "*P*-Value." It is generally recognized that this quantity should not be interpreted as the probability that the null hypothesis is false. (See, for instance, Sturrock 1997, Utts 1996). However, the latter quantity may be estimated by means of Bayesian procedures (Sturrock 1973, 1994). To use the Bayesian approach, one must consider a complete set of hypotheses, such that one and only one of the hypotheses must be true. For the present problem, clearly one hypothesis could be

H1. Shakspere wrote the Shakespeare material (plays and poems),

and one may complete the set by adopting

H2. Shakspere did not write the Shakespeare material.

We consider the two statements:

S.1 Shakspere was a writer, and

S.2 Shakspere was not a writer.

We need to evaluate the probability of each of these statements on the basis of the evidence and on the basis of each hypothesis.

We have found that $P(S1|E) = 0.00001$.

Since one of *S1.S2* must be true, $P(SE|E) = 0.99999$.

If *H1* is correct, he must have been a writer, and one might therefore expect that he would have the same "paper trail" as the comparison authors. This proves to be a key decision. If we make this assumption, then $P(S1|H1) = 1$.

If Shakspere is assumed not to be the author of the Shakespeare material, we can be noncommital and assume that it is even odds whether or not he was a writer, i.e. $P(S1|H2) = 0.5$. Then, since *S1* and *S2* are mutually exclusive, $P(S2|H1) = 0$ and $P(S2|H2) = 0.05$.

We can now calculate the post-probabilities of *H1* and *H2* from

$$P(H_i|E) = \left[\sum_k \frac{P(S_k|H_i)P(S_k|E)}{\sum_j P(S_k|H_j)P(H_j|-)} \right] P(H_i|-), \quad (2)$$

where $P(H_i|-)$, $i = 1,2$, are the prior probabilities.

If we adopt the noncommittal values $P(H_i|-) = 0.5$, we find that the post-probabilities are $P(H1|E) = 7 \times 10^{-6}$ and $P(H2|E) = 1 - (7 \times 10^{-6})$.

Note that, in the present context, the evidence E comprises Price's (2001) "Chart of Literary Paper Trails," nothing more and nothing less. To get an "absolute" post-probability, one would need to convolve the probability of the hypotheses on Price's evidence with the probability of Price's evidence on some (nonexistent) "absolute" database.

Further Hypotheses and Discussion

This very strong result hinges on the key assumption that $P(S1|H1) = 1$, which rests implicitly on the assumption that if Shakspere was a writer, he would have footprints similar to those of the comparison authors. The assumption needs careful consideration. If one can plausibly argue that, for instance, Shakspere had some strong incentive to hide the fact that he was a writer, we can no longer conclude from our statistical analysis that he was not a writer.[2] This then leaves open the possibility that he might have been the author of the Shakespeare works. One way to cope with this possibility is to divide *H1* into two sub-hypotheses:

H1,1. Shakspere was a writer and did not hide the fact.

and

H1,2. Shakspere was a writer but hid the fact (as best he could).

Shakspere would not have been able to suppress all the items of evidence listed in Table 2, but he might (for unknown reasons) have taken steps to suppress those that he could. This seems an unlikely prospect for Shakspere, which could be reflected in a low prior probability for *H1.2*.

If we conclude that Shakspere was not the author of the Shakespeare material, we are left wondering who was. We could then proceed to allow for more options, such as:

H1. Shakspere was the sole author of the Shakespeare material.

H2. Shakspere produced the Shakespeare material in collaboration with another commoner, or with more than one commoner.

H3. Shakspere produced the Shakespeare material in collaboration with at least one member of the upper class,[3] and possibly one or more other commoners.

One must also consider the possibility that Shakspere had no part in the writing of the Shakespeare material. This may be broken down into the following possibilities:

H4. The Shakespeare material was written by one or more commoners, excluding Shakspere.

H5. The authorship of the Shakespeare material involved at least one member of the upper classes, possibly in collaboration with others, but excluding Shakspere.

Various specific proposals have been made which are special cases of the above hypotheses. The basic Stratfordian position corresponds to *H1*. The proposals for Ben Jonson and Christopher Marlowe are special cases of *H4*. The proposals for Francis Bacon, Edward de Vere Earl of Oxford, and Mary Sidney, Countess of Pembroke, are special cases of *H5*. Hypotheses *H2, H3, H4,* and *H5* allow for possible collaborations.[4]

In pursuing this topic, the first requirement would be to assess or supplement Price's (2001) "Chart of Literary Paper Trails," on which our estimates have been based. An ideal procedure would be for several scholars to agree on a list of comparison authors, and a list of categories of evidence, and then for each scholar to make his or her own assessment of whether or not the evidence for each category exists for each comparison author and for Shakspere. One would then obtain a list of *P*-Values, one for each scholar, which could be converted into post-probabilities for the proposed hypotheses. These post-probabilities could if necessary be combined using Bayesian procedures.

It would be most desirable to evaluate other relevant evidence, which we refer to as "items," such as (a) chronological analysis, comparing the known history of Shakspere and the dates of first mention of the plays and poems; (b) content analysis (as indicative of knowledge of other languages and other countries, and of the interests and pastimes of commoners and of the nobility, etc.); and (c) textual analysis, comparing samples of the Shakespeare material and the writing of specified candidates. We could evaluate each hypothesis on the basis of each "item," and then combine the judgments using procedures described elsewhere (Sturrock 1973, 1994).

Notes

[1] Price (2008) explains that her list of just 10 categories represents a convenient packaging of diverse pieces of evidence. For instance, Price can list over 30 pieces of evidence for Drayton and more than 20 for Chapman, down to 5 each for Fletcher and Kyd. Hence a checkmark for evidence in a particular category for a particular author may represent a number of

separate pieces of evidence. For instance, Price collapses more than 20 records concerning Marlowe's presence and education at Cambridge into a checkmark for one category of evidence (Evidence of education). For recent updates, see http://www.shakespeare-authorship.com/resources/errata.asp.

[2] If Christopher Marlowe's murder was merely a staged event to save him from the not-so-tender mercies of the Court of Star Chamber, he would have had the best reason in the world to keep a very low profile.

[3] The term "upper class" is used to connote a member of a noble family, or any person with a title (such as Bacon, who was knighted in 1603).

[4] Ms. Price advises me that Shakespeare editors and scholars have long known that other hands were responsible for parts of *Pericles*, *Henry VII*, and parts of the *Henry VI* trilogy, to name the best-known Shakespeare "collaborations" (in quotes, since the nature of such collaborations remains elusive). More recent scholarship has succeeded in identifying or confirming specific collaborators in those and other plays in the canon, as well as finding Shakespeare's hand in plays attributed to others or published anonymously. The process of analyzing texts has been facilitated by the Chadwick-Healy database Literature Online, which provides scholars with tools to compare and quantify vocabulary, function words, syntax, prosody, stylometry, parallel passages, and other linguistic features. These techniques have been discussed in book form by Vickers (2002) and by Jackson (2003). Both books provide the interested student with helpful bibliographies. Recent journal articles include "Shakespeare and the Quarrel Scene in Arden of Feversham" by Jackson (2006) and Vickers (2007) in which Vickers makes the case for the presence of the hand of Thomas Nashe.

Acknowledgments

The author is indebted to Diana Price for generously reviewing and commenting on earlier versions of this article, and to Henry Bauer, Elliot Bloom, Jeffrey Scargle, and Sarah Webster Goodwin for helpful discussions.

References Cited

Altick, R. D., & Fenstermaker, J. J. (1993). *The Art of Literary Research*. Norton. p. 49.

Bryson, B. (2007). *Shakespeare: The World as Stage*. Harper Collins.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

George, H. B. (1909). *Historical Evidence*. Clarendon Press. pp. 48–49.

Honan, P. (1998). *Shakespeare: A Life*. Oxford University Press.

Howson, C., & Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Jackson, M. P. (2003). *Defining Shakespeare: 'Pericles' as Test Case*. Oxford University Press.

Jackson, M. P. (2006). Shakespeare and the quarrel scene in *Arden of Faversham*. *Shakespeare Quarterly*, *57*:249–293.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science* edited by G. L. Bretthorst. CUP.
Kendall, P. M. (1985). *The Art of Biography*. Norton. p. xiii.
Price, D. (2001). *Shakespeare's Unorthodox Biography: New Evidence of an Authorship Problem*. Westport, CT: Greenwood Press.
Price, D. (2008). Personal communication.
Sturrock, P. A. (1973). Evaluation of astrophysical hypotheses. *Astrophysical Journal*, *182*:569–580.
Sturrock, P. A. (1994). Applied scientific inference. *Journal of Scientific Exploration*, *8*:491–508.
Sturrock, P. A. (1997). A Bayesian maximum-entropy approach to hypothesis testing, for application to RNG and similar experiments. *Journal of Scientific Exploration*, *11*:181–192.
Utts, J. (1996). *Seeing Through Statistics*. Pacific Grove, CA: Duxbury Press.
Vickers, B. (2002). S*hakespeare, Co-Author: A Historical Study of Five Collaborative Plays*. OUP.
Vickers, B. (2007). Incomplete Shakespeare: Or, denying coauthorship in *1 Henry VI*. *Shakespeare Quarterly*, *58*:311–352.
Williams, R. C. (2003). *The Historian's Toolbox: A Student's Guide to the Theory and Craft of History*. New York: Sharpe. (especially p. 58)

APPENDIX

Suppose that we begin with a competely open mind, and assume initially that *P* can have any value between zero and unity, with a uniform distribution in that range. We denote the probability that *P* is in the range *p* to *p* + *dp* by *Q*(*p Z*)*dp*, where the symbol "*Z*" indicates that we have zero relevant information. The likelihood of getting a positive result (a "*Y*") for any trial is *p*, and the likelihood of getting a negative result (an "*N*") is *1 – p*. Hence the likelihood of getting *K* positive results and (*N – K*) negative results is

$$L(K|N,p) = p^K(1-p)^{N-K} \quad . \tag{A.1}$$

By Bayes Theorem, we may obtain the "post-probabiity distribution function" for *p,* given *N* and *K*, as follows:

$$Q(p|N,K) = \frac{L(K|N,p)}{\int_0^1 dp' L(K|N,p')} \tag{A.2}$$

Based on this information, the probability of getting a positive result for the (*N* + 1)*th* trial is

$$P = \frac{\int_0^1 dp Q(p|N,K)\, p}{\int_0^1 dp Q(p|N,K)} \tag{A.3}$$

This is found to have the value

$$P = \frac{K+1}{N+2} \quad . \tag{A.4}$$