Digital Scholarship in the Humanities, 2023, **38**, 1494–1507
https://doi.org/10.1093/llc/fqad032
Advance access publication 28 April 2023
Full Paper

"I would I had that corporal soundness"[a]: Pervez Rizvi's Analysis of the Word Adjacency Network Method of Authorship Attribution

Gabriel Egan[1,*], Mark Eisen[2], Alejandro Ribeiro[3], Santiago Segarra[4]

[1]School of Humanities, De Montfort University, Leicester, UK
[2]Intel, Santa Clara, California, USA
[3]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[4]Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA

*Correspondence: Gabriel Egan. E-mail: mail@gabrielegan.com

[a]A line spoken by the King of France in Shakespeare's *All's Well that Ends Well* 1.2.24.

Abstract

In his two-part article 'An Analysis of the Word Adjacency Network Method—Part 1—The evidence of its unsoundness' and 'Part 2—A true understanding of the method' Digital Scholarship in the Humanities, 38: 347-78 (2022), Pervez Rizvi attempts to replicate the Word Adjacency Network (WAN) method for authorship attribution and show that it does not produce the new knowledge that we, its inventors, claim for it. In the present essay, we will show that Rizvi misrepresents fundamental aspects of the WAN method, that his attempted replication fails not because the method is flawed but because he erred in replicating it, and that Rizvi misunderstands key aspects of the mathematics of Information Theory that the method uses.

1 What a word adjacency network captures

It has long been known that the frequencies at which the most-common words in the English language are used by any writer are peculiar to that writer. We all use the most-common word *the* about once in every sixteen words and the next most-common word *and* about once in every thirty words (If we treat as one word all the various forms of the verb *to be*—as 'was', 'am', and so on—then it takes second place ahead of *and*.). But the precise rate at which we each use these and the other most-frequent words is idiosyncratic and does not change much by genre or over time so that from a sufficiently large body of various authors' writings we may develop profiles of the differing authorial preferences regarding these words and use these profiles to attribute works of unknown or contested authorship. The most-common words in English are the function words that express syntactic relations, and for authorship attribution it is common to count the frequencies of between 50 and 100 of these function words.

Analysis of function-word frequencies has successfully determined authorship in cases as varied as the writings of the American Founding Fathers James Madison and Alexander Hamilton, the Roman statesman Cicero, the *Book of Mormon*, and the anonymized judgements of the US Supreme Court (Mosteller and Wallace, 1963; Forsyth *et al.*, 1999; Jockers *et al.*, 2008, 2019), and studies that quantify the accuracy of these authorship-attribution methods have shown function-word frequency to be objectively reliable at quantifiable levels of confidence (Hoover, 2004; Argamon, 2018).

A new refinement of this function-word frequency approach to authorship attribution called the Word Adjacency Network (WAN) was first introduced 8 years ago and has been applied by its inventors to the field of early modern drama in general and the plays of William Shakespeare in particular (Segarra *et al.*, 2015, 2016; Eisen *et al.*, 2018; Brown *et al.*, 2022). The central claim of the WAN approach is that as well as their frequencies, the patterns of clustering of the function words—their distances one from another—are distinctive of authorship. By measuring how far from one another an author places the most-common words (measured by the number of intervening words), the WAN method adds about ten to fourteen percentage points to the accuracy of authorial attributions, taking the state of the art from about 80% accuracy for the frequency-only approach to around 90–94% accuracy for the WAN approach.

Pervez Rizvi has published two critiques of the WAN method: the first in the journal *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews* (Rizvi, 2020) and the second as a two-part article in this journal (Rizvi, 2022a, 2022b). In these critiques, Rizvi attempts to show that the WAN method is incapable of establishing the authorship attributions that its inventors claim for it. Rizvi's first critique was answered by the present authors in the same journal *ANQ* (Segarra *et al.*, 2020) and the present article is their response to his subsequent two-part critique in this journal. We find that Rizvi's critique helps refine our thinking about a small point of terminology used in our method, and that his innovative experimentation adds to the mountain of evidence that function-word frequencies are indeed indicative of authorship. But the core objections to the WAN method raised by Rizvi arise from misunderstandings of how it works that vitiate his attempt to replicate it.

Part 1 of Rizvi's critique begins with a small but significant slip in the explanation of the meaning of the hypothetical data '*[and, and, 0.4]*', which we would call a 'transition', that one might find in a data structure representing a WAN. Rizvi writes that this 'tells us that when the author of the text or canon from which this profile was calculated has written *and* he is likely to follow it closely by *and* 40% of the time' (Rizvi, 2022a, p. 2). If that were true there would be an astonishingly high number of *and*s in a text, since almost every other occurrence would be followed within a few words by another occurrence of *and*, which itself would, almost half the time, be followed within a few words by yet another. In fact these numbers tell us that shortly after *and* we should expect that, if there is any occurrence at all of one of the words on the investigator's list of words-of-interest (typically a list of the 50 or 100 most-frequent words in the language), then 40% of the time this succeeding occurrence of one of those words will be an occurrence of *and* rather than one of the other words-of-interest. This qualification is essential to the method, since often there will be no such succeeding occurrence of one of the words of interest.

In attempting to replicate our work, Rizvi reports that he left out what we consider to be a vital step, which is the calculation of what are called limit probabilities. We consider these so important that we devoted almost half our explanation of our method (Brown *et al.*, 2022, pp. 325–32) to showing why they exist and how they are calculated. Without limit probabilities, the measurement of the Kullback–Leibler divergence between two WANs would not be weighted according to the different frequencies at which each of the words-of-interest appears in the texts being compared. Rizvi believes the calculation of limit probabilities to be inessential: 'For my experiments in this article, these limit probabilities are not of interest …' (Rizvi, 2022a, p. 2). As we shall see, it becomes apparent in Part 2 of Rizvi's critique that he entirely misunderstands the meaning and role of limit probabilities in our method and more generally in the mathematical analysis of Markov chains.

Before getting to the substance of his new critique, Rizvi repeats an objection he made in earlier critiques that our deducting from each relative-entropy comparison a constant—the 'background' reading, as it were—misrepresents the results we obtain. Since our authorship attributions are made solely on absolute differences between data, not on their relative proportions, this objection is spurious, as we explained at length in a prior response to Rizvi (Segarra *et al.*, 2020, p. 336). Next, Rizvi summarizes the result of what he claims is his replication of our experiments and his verdict is damning: the WAN method does not reliably distinguish authorship at all (Rizvi, 2022a, pp. 5–7). As we will show, the cause of Rizvi's failed replication is that he omitted key aspects of the method that we describe in our publications about it.

Rizvi proceeds to his major theme in this new critique, which is that we take no account of evidence for authorial habits of omission. That is, we disregard cases where an author never followed, in quick succession, a particular one of the words-of-interest with another one of the words-of-interest. To explore with Rizvi the various forms that such omissions can take, we must summarize the WAN method. A full explanation, with visual illustrations and worked examples, previously appeared in this journal (Brown *et al.*, 2022).

A WAN represents a text's patterns of word adjacency, which is the habit of following one of our words-of-interest by another with only a few intervening words separating them. The list of words-of-interest is selected by the investigator, and because habits regarding use of function words—the small and highly frequent words in English—have been shown to reveal authorship, a common choice is the top 50 or 100 most-frequently occurring words in a corpus. Likewise, the number of intervening words allowed to appear between the two occurrences (the 'window') is set by the investigator, with five and ten words being common window sizes.

To capture the authorial habits of word adjacency averaged for all occurrences of the word-of-interest across a text, a WAN takes the form of a table (in computing and mathematical terms, a matrix) in which each row has as its heading one of the words of interest and so does each column. Thus the rows might be labelled from, say, *the* (the most-common word) and *and* (the second-most-common word) down to *me* (the 50th most-common word), and each column likewise labelled from *the* and *and* up to *me* running left to right. Thus the first row holds a list of fifty adjacency values for *the*, the second row a list of fifty adjacency values for *and* and the 50th row a list of fifty adjacency values for *me*. The first cell of the matrix (in the top-left corner) holds a value representing how often, in the text the matrix was made from, the word *the* is followed within the 'window' by another occurrence of *the* (the first column heading). The next cell to the right holds a value representing how often *the* is followed by *and* (the second column heading), and the last (right-most) cell in this first row holds a value representing how often *the* is followed by *me* (the 50th column heading). And so on down the rows to the 50th row (headed *me*), which begins with a cell representing how often *me* is followed by *the* (the first column heading), with to its right a cell representing how often *me* is followed by *and* (the second column heading), and the last cell in this last row represents how often *me* is followed by an occurrence of *me* (the 50th column heading).

For authorship attribution experiments, we use a computer to construct such a WAN for a text to be attributed, representing a summary of the authorial habit of placing within a given number of words of word *x* an occurrence of word *y*. For each candidate for the authorship of this text, we also make a WAN based on the candidate's entire body of sole-authored works. Then we compare the WAN for the text to be attributed to each of the candidates' WANs, looking to see which is least different regarding these habits of word adjacency. Each WAN is a set of probabilities, one each for the probability of finding (in our example) *the* followed shortly thereafter by *the*, of finding *the* followed by *and*, of finding *the* followed by *me*, and so on up to the probability of finding *me* followed by another occurrence of *me*. In our example using 50 words, the WAN is a list of fifty lists of probabilities, one list for each of the fifty words.

Being a list of probabilities, a WAN is thus what is known as a frequency distribution. We can therefore ask the following question: how accurate is this set of probabilities derived from the text that we want to attribute at 'predicting' the adjacencies we find in the complete works of each of the candidate authors who may have written the text? This is what we measure when we compare WANs to see how alike they are. Derived from Shannon's work on information entropy, we use what is known as Kullback–Leibler Divergence as the measure of this predictive power (Kullback and Leibler, 1951). If the habits of word adjacency found in the text to be attributed are similar or identical to the authorial candidate's habits of word adjacency, the Kullback–Leibler Divergence (colloquially, the 'relative entropy') will be low or zero.

The higher the relative entropy, the less alike regarding habits of word adjacency are the text under study and the candidate author's writings. In our method, we declare that the candidate author whose works are least unlike the text to be attributed is the one most likely of this field of candidates to be the one who wrote the text. Whether this is true in practice is something we tested extensively in validation runs using around 100 early modern plays of known authorship. The method was able to 'predict' the correct author in around 90–94% of cases for which we have sufficient text to measure in the sample and in the canons of the candidate authors.

2 Acts of commission versus acts of omission

We may now return to Rizvi's critique of how our method deals with habits of omission by authors. There are three possible kinds of omission. The first is where the text to be attributed has a blank cell in its WAN because the particular transition represented by that cell is not found in the text. If the text had no examples of *the* followed within a few words by *and*, the first row's second cell—the cell for '*the*-followed-by-*and*'—would contain a zero. The second kind of omission is of the same form, but concerning cases where the habit is omitted not in the text to be attributed but in the complete works of the candidate author. The third kind of omission only applies to experiments in which we have multiple WANs for multiple authorial candidates. If we have six candidates, candidates *A*–*F*, it may be that a certain transition is found in the works of candidates *A*, *B*, *C*, *E*, and *F* but not in candidate *D*'s works. This possibility of certain candidates' works lacking certain transitions is particularly great when some candidates have small canons. At over a million words, the Shakespeare canon has at least one example of almost every possible transition that the fifty top most-frequent words might be involved in, but the much smaller canon of Christopher Marlowe lacks some of these transitions.

It may be that if Marlowe had left us more works then we would have examples of these missing transitions, but equally it may be that Marlowe disfavoured them and that even if we had twice as many of his works they would not be found. We cannot argue anything from missing evidence, yet the key to Rizvi's critique is that we ought to. In the first and second cases of omission, where a text under examination or an candidate author's known canon lacks a particular transition, our method simply takes no account of that transition: it contributes nothing to our calculation of the relative entropy between the text and a candidate author's profile.

Strictly speaking, in the first case (an omission in the text to be attributed) the mathematical equation we use inherently discounts this omission since it involves a multiplication by the cell-value from the text's WAN, which cell-value will be zero. This results in zero being added to the running total as we proceed, cell-by-cell, to tally the relative entropy between two WANs. In the second case, our algorithm programmatically discounts the omission: we test for the cell-value from the authorial WAN being zero and we simply move on to the next cell if this is the case. In the third kind of omission, where in a multi-candidate experiment one or more candidates' known works lack a particular transition, we again programmatically ignore this transition and move on. That is, for each transition, we test whether any of the candidate authors' WANs have a cell containing a zero and if it does then this transition is ignored for all candidates.

Thus we choose to ignore adjacencies that are not found in the texts under consideration. Although acts of omission—of never placing word *x* near to word *y*—might in principle be evidence of authorship, we must bear in mind that they are not evidentially equivalent to acts of commission. In authorship studies, our data are necessarily limited. We cannot possess everything an author wrote, only what happened to survive. For this reason, negative claims about what a writer never does are open to refutation by future discoveries of lost writing in which the supposedly absent habit does in fact occur. Positive claims about what a writer demonstrably has written are inherently free from this danger.

It is perhaps worth quantifying how often we ignore certain transitions. If we look for the 100 most-common words in the six dramatic canons of George Chapman, John Fletcher, Ben Jonson, Christopher Marlowe, Thomas Middleton, and William Shakespeare there are 10,000 (100 × 100) possible transitions to consider. We can ask three pertinent questions: (1) how often does one of these transitions fail to appear in a particular play?, (2) how often does one of these transitions fail to appear anywhere in any particular author's canon, and (3) how often does one of these transitions fail to appear in at least one of the six authorial canons? These three questions cover the three kinds of omission that Rizvi objects to us making.

The answers to these questions will vary from play to play and canon to canon since these vary in size. A long play or large canon has, as it were, more 'opportunity' than a short play or small canon to use any particular transition. For these six dramatists, we find in answer to question (1) that typically a play will contain 6,000 to 7,000 of the 10,000 possible transitions (so, 60–70%). In answer to question (2), we find that across the Shakespeare canon over 9,700 of the 10,000 possible transitions (97%) are present while across the smaller Marlowe canon only 8,498 of the 10,000 possible transitions (85%) are present. Regarding question (3), we find that for these six authors, 7,714 of the 10,000 possible transitions (77%) are present in all their canons, so by our method that excludes transitions for which one or more of our authors provides no evidence 23% of the possible transitions are excluded.

If instead of the 100 most-common words, we confine our attention to the fifty most-common words, we find of course that a higher proportion of the 2,500 (50 × 50) possible transitions are present in each play and each canon. This is because the fifty most-common words are much more common than the 51st to the 100th most-common words. Indeed, as we would expect from Zipf's power law, in any substantial text or corpus the most-common word occurs about 100 times more often than the 100th most-common word. Because in our multi-candidate tests, we exclude any transitions not found in all the candidates' canons, the use of the 100 most-common word instead of the fifty most-common words does not bring in proportionally more evidence: the rarer transitions are perforce eliminated by our rule that all candidates must have used the transitions that we take into account.

Rizvi accurately details our choices about excluding certain adjacencies in the central section of his critique (comprising over a third of his essay), which is headed 'The Mass Exclusion of Evidence' (Rizvi, 2022a, pp. 7–13). He writes that 'Sometimes a word adjacency does not occur in the text . . .', and '. . . it is possible that [the value for a transition] will be zero for some of them [the candidates] but not for others', and '. . . we might find a word adjacency does not occur in the scene, does not occur in the canon of one candidate author, but occurs often in the canon of the rest' (Rizvi, 2022a, p. 8). For Rizvi, 'Common sense suggests that' such cases 'should be treated as at least modest evidence' for authorship (Rizvi, 2022a, p. 8). We think, on the contrary, that common sense is misleading in this matter, as we argue here and in our previous response to Rizvi (Segarra *et al.*, 2020, pp. 333–34).

To illustrate what he thinks would be the effect of not excluding authors' habits of omission, Rizvi offers a worked example concerning the attribution of Scene 4 of Marlowe's play *Edward II* using data 'obtained from running the inventors' software' (Rizvi, 2022a, p. 9). As we will show, Rizvi did not run the inventors' software but his own that has crucial differences. Rizvi attempts to discover the effect on our results—really his results, since he is not applying our method—of including rather than excluding the acts of omission. For cases where the text to be attributed omits a particular transition the very mathematics of the relative-entropy equation—involving a multiplication by the zero probability in the text's WAN—nullifies (sets to zero) the effect of this transition upon the final verdict. Rizvi explains that 'The formula does not allow us to calculate the effect of these exclusions on the relative entropies . . .' (Rizvi, 2022a, p. 9). This phrasing betrays a misunderstanding of the notion of entropy that the formula embodies. It is not that the formula fails to take such cases into account—that we made an 'unwise choice of formula' (Rizvi, 2022a, p. 12) as Rizvi later puts it—but that there really is in such cases no relative entropy to be measured.

The calculation of relative entropy effectively measures how well the habits of word adjacency found in the text to be attributed (and represented in its WAN) would serve as predictors of the habits found in the candidate author's complete works (and hence represented in their WANs). A zero probability for a feature in the text is not the prediction of a zero in the candidate author's works but rather is the absence of any prediction at all, since the text contains no evidence about that habit. By analogy, that a particular play by Shakespeare lacks the rare word *prohibition* would not serve as an accurate predictor that none of the other Shakespeare plays will be found to contain the word *prohibition*. Rather, this rare word is unlikely to be found in any particular textual sample by any writer even if that writer uses the word elsewhere, simply because it is rare. Shakespeare uses *prohibition* precisely once across his entire canon of about a million words, in *Cymbeline*. In essence, the absence of a word is just the highest state of rareness.

If the play *Cymbeline* had not survived for us to read, then by Rizvi's logic in which the absence of evidence can be treated as evidence of absence, we could construe from a sample text's omission of this word and the corresponding omission of it from Shakespeare's canon that we had found evidence that Shakespeare wrote the sample text. This construal would be an overvaluing of the importance of an extremely small datum. For any transition that has a zero score in a sample text's WAN, we cannot tell if the reason for its absence is that the author strongly avoids it (as Rizvi's approach would assume) or merely that the transition is insufficiently common to turn up in any sample of that particular size.

Having rightly concluded that he cannot calculate the effect upon relative entropy of there being zeroes in the WAN for a text to be attributed (arising because the corresponding transition does not occur in that text), Rizvi turns his attention to cases of zeroes in the WANs for candidate authors' canons. He believes he can calculate the differing effects on the final relative entropy of the investigator either including (as he prefers) or omitting (as we prefer) this 'evidence'. Rather than running the experiment twice, one time including the evidence and one time excluding it, Rizvi calculates the effect using a spreadsheet that he helpfully includes among the supporting materials for the essay, which for the case of *Edward II* Scene 4 is a spreadsheet called 'calculation-edward-ii-scene-4-scene-to-author.xlsx'.

Rizvi reports that '. . . there are fifteen word adjacencies found in the scene [*Edward II* Scene 4] and in Shakespeare but not in Marlowe; for example, {*any*, *away*}, which is found in the scene and in Shakespeare but not in Marlowe' (Rizvi, 2022a, p. 9). For each transition, the relative entropy calculation depends on dividing the probability value in the WAN for the text to be attributed by the probability value in the WAN for the candidate author's canon, and since Marlowe's canon contains no *and*-to-*away* transitions the value in the corresponding cell (G229 in Rizvi's spreadsheet) is zero. Because there are no *and*-to-*away* transitions in the Marlowe canon, the cell in Rizvi's spreadsheet that is meant to hold the value for the relative entropy calculation in respect of this transition (cell H229), the cell in which the probabilities are divided, contains the error message '#DIV/0!'. This is the warning by which Microsoft Excel (the proprietary software used to create Rizvi's spreadsheet) alerts the user of an attempt to divide a number by zero, which is not a meaningful operation in mathematics.

To assess what difference it makes if we include transitions that are not found in a candidate author's canon, Rizvi's analysis ignores the spreadsheet's division-by-zero errors in the calculation for Marlowe but nonetheless counts the corresponding values in the calculation for Shakespeare, whose canon does include instances of the *and*-to-*away* transition. Rizvi reports the difference it would have made if we had followed him in taking this approach: '. . . the scene's relative entropy to Shakespeare would have risen by 1.9, from 57 to 58.9, overtaking the one to Marlowe, which would have remained at 57.6' (Rizvi, 2022a, p. 9). The relative entropy between the scene and Marlowe's works is, of course, unchanged because Rizvi treats his spreadsheet's 15 '#DIV/0!' errors as if they each represent the value zero when in fact the software is complaining that the result for each of these transition is, if anything, infinite. We consider such results to be neither zero nor infinite but rather non-results, which is why we exclude such transitions from our calculations.

By treating the error '#DIV/0!' as a zero result, Rizvi finds that 'if the evidence from the Shakespeare canon had not been voluntarily excluded, the scene would have been correctly attributed to Marlowe' (Rizvi, 2022a, p. 9). This achievement emboldens Rizvi to speculate about what would be the effect if the zeroes in the scene's WAN could also have been admitted as evidence, even though this effect is, as he admits, quite impossible to calculate because it is meaningless. Like a lawyer frustrated by a judge finding inadmissible some potential evidence that she thinks would clinch her case, Rizvi is forced to concatenate negatives in the effort to imply a positive: '... it would be unjustified to suppose that the indirect exclusions have made no difference to the result' (Rizvi, 2022a, p. 9). Regarding a particular attribution to which he objects, Rizvi later repeats this rhetorical manoeuvre: '... we cannot assume that the excluded evidence here would not have overturned the attribution to Marlowe if the formula had allowed us to calculate it' (Rizvi, 2022a, p. 12).

3 Big numbers from small numbers and systemic bias

Rizvi correctly observes that if the evidence he wants to admit—transitions found in one author's canon but not another's—were deemed admissible, it would contribute to the final results numbers that are larger than the differences by which we make our authorship attributions. Thus 'The value of the excluded evidence is more than three times the margin by which the attribution was made' (Rizvi, 2022a). True, but that could be said to be as good a reason to exclude it as include the disputed evidence, it being 'noise' that swamps the 'signal' we hope to detect. He is also correct to observe that even confining attention to just the evidence we admit—where all the candidates' canons show the transitions we count—the margins by which the candidates' final tallies differ can be smaller than most of the individual data points that make up these tallies. Citing the largest numbers that figure in a particular tally, from 0.54 to 1.41, Rizvi remarks that 'When almost a thousand numbers, some as high as this, are being added up to produce the relative entropy, it is indefensible to base an attribution on a margin of only 0.6' (Rizvi, 2022a, p. 10).

This is flawed reasoning. The 2016 Brexit referendum in the UK was decided on a difference of 1.27 million votes, out of 33.5 million votes cast. It was close, but even those most bitterly opposed to Brexit do not argue that there is uncertainty about which side won. Necessarily, a referendum sums many locally organized counts involving thousands of individual batches of votes in order to produce a tally representing millions of votes. The difference in scale between the local counts and the national one was more than three orders of magnitude (1,000:1), which is the scale that Rizvi claims makes the whole process 'indefensible'. This is not so if the local counts are accurate. Indeed if Rizvi were right, democracy itself would be indefensible since we would have to reject, on principle, the summing of many small results no matter how carefully that counting and summing were undertaken.

Accuracy in these matters is not a matter of common sense or approximation: it is a measurable statistic involving quantifiable margins of error. Our validation of the WAN method using around 100 plays of known authorship generated a play-wise false-attribution rate of under 10%, which is about the state of the art in this field. Scene-wise attributions are inherently less reliable because there is less writing to go on, but Rizvi's replication was particularly unsuccessful: '... the attribution of the 234 scenes in my experiment were not much more accurate than we would get by tossing a coin' (Rizvi, 2022a, p. 10). We share Rizvi's disappointment at his results, but will show that the replication failed because of mistakes he made in following our method, not because of flaws in the method itself. Before turning to these mistakes, we will make some general points about methodologies.

Even if we accept on their own terms Rizvi's misgivings about the relative magnitudes of the admissible and inadmissible evidence and the final differences, his critique points to no systemic bias. Indeed, he admits that although 'we might be tempted to think that we can solve the problem by changing the method to include such evidence after all', he has found that this 'would make it more accurate in some tests but less accurate in others' (Rizvi, 2022a, p. 10). Indeed, that is why we think the evidence inadmissible: it is not really evidence of authorship at all; it is not 'signal' but 'noise'.

In the case of Scene 3 from Shakespeare's *Henry V*, which Rizvi's replication misattributes to Marlowe, the inclusion of the excluded evidence 'would have increased Marlowe's winning margin' (Rizvi, 2022a, p. 10). That sounds like bias towards Marlowe, but digging into why this happens Rizvi finds the opposite effect. Because Shakespeare's canon is larger than Marlowe's, the Shakespeare-canon's WAN has many non-zero transitions where the corresponding values in the Marlowe-canon WAN are zeroes, and in deference to those Marlovian zeroes we disregard these transitions. Rizvi finds that this explains 'why the method was noticeably more successful at correctly attributing the Shakespeare scenes, among the 234 that I tested, than the Marlowe scenes' (Rizvi, 2022a, p. 11). That is, our method was 'excluding more of Shakespeare's evidence, and that was almost always helping Shakespeare's case for authorship, even of the Marlowe scenes, by keeping his relative entropies low' (Rizvi, 2022a, p. 11). That sounds like bias towards Shakespeare. Rizvi does not explicitly state which way he thinks our method is biased overall. Bearing in mind his wider argument, he can do no more than hint at a pro-Shakespeare bias, since his strongest objection—the one with which he begins and ends Part 1 of his essay—is that our method wrongly attributes to Marlowe some scenes in the *Henry VI* plays that Rizvi believes are wholly by Shakespeare.

Rizvi finds it improper that in our method a positive effect from one transition may be cancelled out by the negative effect from another, as when the effect of the *the*-to-*to* transition (−0.1814) is almost exactly cancelled out by the effect of the *one*-to-*as* transition (+0.1813). For Rizvi, this is 'theoretically unjustifiable, because if some texts differ in two ways, then they differ in two ways, and I cannot see how it could make sense to allow cancelling out and to treat them as if they did not differ' (Rizvi, 2022a, pp. 11–12). We would respond that such cancelling out is precisely what a method ought to do in order to detect trends across a large swathe of text. Attributions based on small quantities of textual evidence are over-sensitive to local variations arising from subject matter, so investigators rely on large textual samples across which, they hope, the merely local effects cancel one another. For the same reason, our investigations typically track authors' habits of adjacency for around 50 to 100 words rather than fewer, in the hope that local variations involving a few words will mutually cancel one another out. To object to cancelling out is to object to the very principles on which investigations of large textual corpora are founded.

Rizvi concludes Part 1 of his critique with the assertion that 'The relative entropy formula has its uses in some disciplines' but is unsuited to analysis of 'the diversity of early modern play texts' (Rizvi, 2022a, pp. 12, 13). For this reason, he urges that Shakespearians disregard our attribution to Marlowe of parts of the *Henry VI* plays he thinks are by Shakespeare, as announced in our article in *Shakespeare Quarterly* (Segarra *et al.*, 2016). Not the least difficulty with this conclusion is that by the same method our *Shakespeare Quarterly* article also confirmed several recent re-attributions that Rizvi agrees with. Like him, we find that the first act of Shakespeare's *Titus Andronicus* is by George Peele. We agree with him that the first two acts of Shakespeare's *Pericles* are by George Wilkins. We agree in finding that *Timon of Athens* was co-written by Shakespeare and Thomas Middleton, and that *Henry VIII* was co-written by Shakespeare and John Fletcher.

If our method of authorial attribution were no better than the tossing of a coin, as Rizvi repeatedly puts it (Rizvi, 2022a, pp. 5, 7, 10, 13), the mystery to be solved is why this allegedly flawed method—which is a new approach not used by previous investigators—confirms so many previous investigators' findings of authorship in cases about which we and Rizvi agree. If our method were no better than tossing a coin, its agreements with these independently achieved conclusions would be miraculous. We propose instead that the explanation is mundane: the method works and provides independent corroboration of previous investigations.

4 Rizvi's software

In Rizvi's footnotes to his Part 1, we find at least part of the explanation for how he has misled himself about the WAN method. He gives the URL <http://www.shakespearestext.com/wan.zip> pointing to various online materials in support of his essay, including the software scripts he used (Rizvi, 2022a, p. 14n4). On 24 October 2022, we downloaded the scripts and other materials that Rizvi supplies and we include these time-stamped downloads as supplementary online materials for the present essay so that readers can do as we have done in examining his Python-language source code. From these materials, it emerges that Rizvi applied departures from the method as described in our publications, which he lists in a READ-ME file. He reports that he 'changed the adjacency window length from 5 to 10' and he 'enabled the feature to stop at speech boundaries'. When replicating previous studies, it is essential not to make arbitrary changes to the methods, and these two differences alone are sufficient to account for his results reported above differing from ours.

Concerning a ten-word instead of a five-word window, the extra information that is gathered about words that are between six and ten words apart does not substantially affect authorship attribution accuracy in our tests. Perhaps such distant co-occurrences do not register strongly in an author's mind. We know that many dramatists first wrote out all of a scene's speeches and afterwards added the speech prefixes in the left margin, as seen in the manuscript of the play *Sir Thomas More*. From the point of view of word adjacency, this writing practice puts the end of one speech and the beginning of the next in closer mental proximity than they seem to be after the speech prefixes are added. In any case, as with window length, we have found that stopping the window at the end of a speech does not make a substantial difference to overall attribution success.

Turning to Rizvi's source code, we find that it demonstrably does things he claims it does not, and it does not do things that he claims it does. In the body of his essay Rizvi writes that he omitted the calculations of what are called the limit probabilities of a WAN: 'For my experiments in this article, these limit probabilities are not of interest . . .' (Rizvi, 2022a, p. 2). But examination of his code, which was derived from our original, shows that it does calculate and use for authorial comparisons the WANs' limit probabilities. Rizvi supplies three scripts that match the experiments he describes in his essay and they use the limit-probability calculation that he says he eschews, as the following extracts show:

```
def relativeEntropy(anyWAN1, anyWAN2,
anyWAN1LimitProbs):
.
.
.
# Read the limit probabilities of the
first text into a 1-dimensional array
limit1 = [0 for x in range(length)]
.
.
.
# Output the relative entropy information
in a CSV format
print(sys.argv[1]+","+sys.argv[2]+","+
str(100 * relativeEntropy(wan1, wan2,
limit1)))
```

(lines 37, 75-76, and 102-103 of Rizvi's script called "entropy.py" found at <http://www.shakespearestext.com/wan.zip> on 24 October 2022 and now mirrored at <http://www.gabrielegan.com/WAN>)

```
def
relativeEntropy(textWanLimitProbs,
textWan, profile1Wan, profile2Wan):
.
.
.
# Read the limit probabilities of the
text into a 1-dimensional array
limit = [0 for x in range(length)]
.
.
.
score1 = 100 * relativeEntropy(limit,
textWan, wan1, wan2)
```

(lines 33, 56-57, and 92 of Rizvi's script "entropy_f7_two.py" found at <http://www.shakespearestext.com/wan.zip>) on 24 October 2022 and now mirrored at <http://www.gabrielegan.com/WAN>

```
def relativeEntropy(textWanLimitProbs,
textWan, profile1Wan, profile2Wan,
profile3Wan, profile4Wan, profile5Wan,
profile6Wan):
.
.
.
# Read the limit probabilities of the
text into a 1-dimensional array
limit = [0 for x in range(length)]
.
.
.
score1 = 100 * relativeEntropy(limit,
textWan, wan1, wan2, wan3, wan4, wan5,
wan6)
```

(lines 33, 60-61, and 132 of Rizvi's script "entropy_f7_six.py" found at <http://www.shakespearestext.com/wan.zip>) on 24 October 2022 and now mirrored at <http://www.gabrielegan.com/WAN>

The reader will also find that Rizvi's spreadsheets, such as the one called 'calculation-edward-ii-scene-4-scene-to-author.xlsx' that we refer to above, each have a column of data headed 'Scene Limit Prob'. (column C) and its values are invoked in each spreadsheet's formula for calculating relative entropy. We cannot account for Rizvi's false claim that for his experiments the limit probabilities are not of interest, since they are essential to the method and everything we have seen indicates that he did indeed calculate and use them.

The names of the second and third scripts above, involving the abbreviation 'f7' in their titles, refer to the distinction between, on the one hand, including in the calculation of relative entropy those transitions that are found in some but not all the authorial candidates' writings, and on the other hand excluding these transitions. As Rizvi rightly points out, this distinction is embodied in the difference between Formula 6 (which includes such transitions) and Formula 7 (which excludes them) in our essay 'Stylometric Analysis of Early Modern English plays' (Eisen *et al.*, 2018, p. 503).

The freely disseminated software published to accompany our essay 'How the Word Adjacency Network (WAN) Algorithm Works' (Brown *et al.*, 2022) does not exclude transitions that are found in some but not all the authorial candidates' writings. That essay attempted to explain how a WAN embodies the word-adjacency aspect of authorial style exhibited by a text and how two WANs can be compared to derive their relative entropy. We explicitly omitted from this explanation all aspects of the larger endeavour of applying this method in multiple-candidate authorship attribution experiments, which is where it becomes possible to choose to exclude transitions not used by all candidates.

Rizvi reports that in his experiments he adapted our illustrative software: 'I modified the software to make it use Formula 7' (Rizvi, 2022a, p. 13n2). That is, he took our software that makes no exclusions (thereby embodying our Formula 6) and changed it to exclude transitions that are found in some but not all the authorial candidates' writings (thereby embodying our Formula 7). But an examination of Rizvi's source code, as we downloaded it on 24 October 2022 and placed in the repository that accompanies the present essay, reveals that in fact it does not do this. There is nothing in the scripts 'entropy_f7_two.py', or 'entropy_f7_six.py' (where the 'f7' stands for Formula 7) that applies this step. Rather, Rizvi's code is in this regard—although not in others that matter, as mentioned above—functionally identical to the code we provided. Readers familiar with the Python programming language can download the source code and check this for themselves.

5 What is entropy?

Part 2 of Rizvi's critique begins with his brief introduction to the concept of entropy as it is used in thermodynamics and Information Theory. Rizvi's overview ends with a specific claim that becomes central to his critique of our work:

> In physics, entropy can be thought of as a measure of disorder in a system. A perfectly ordered system has an entropy of zero. Any disorder causes the entropy to take a positive value: the greater the disorder, the more positive the entropy. Neither the original entropy formula in physics nor the formula that Shannon invented for the entropy in information theory can ever give a negative value. (Rizvi, 2022b, p. 1)

In a strict and literal sense, Rizvi is right, but any statistical measure can meaningfully be made negative by taking up an alternative point of view.

A useful analogy to illustrate this is heat, which is the kinetic energy of atoms and molecules, or more colloquially their 'jiggling and bouncing' as the theoretical physicist Richard Feynman put it (Feynman, 2011, p. 5). Heat energy cannot be negative since at their coldest the atoms and molecules have their minimal possible jiggle and bounce. This lowest state is absolute zero on the Kelvin scale, but this provides a rather inconvenient starting point for the heat values we typically encounter in everyday life, since water remains frozen from 0 to 273 Kelvins. For convenience, we keep the Kelvin units but recalibrate our zero as the freezing point of water, giving us the Celsius scale in which the outdoor temperatures on especially cold days are negative numbers. Rizvi's objection to our use of negative numbers for entropy is as absurd as objecting to negative temperatures on the grounds that zero is the minimum value for heat.

We are not alone in using an arbitrary zero for entropy. In his classic popular-science work *What is Life?*, first published in 1944, Erwin Schrödinger wrote of a living organism 'attracting, as it were, a stream of negative entropy, to compensate the entropy it produces by living and thus to maintain itself on a stationary and fairly low entropy level' (Schrödinger, 2021, p. 73). This concept of negative entropy was not a passing fancy in Schrödinger's book but rather the essence of his answer to the question posed in the book's title. Addressing objections from physicists to his use of this concept, Schrödinger traced the idea of 'entropy with a negative sign' back to the foundations of thermodynamics, it being 'precisely the thing on which [Ludwig] Boltzmann's argument turned' (Schrödinger, 2021, p. 74). Within Information Theory negative entropy is usefully defined as the Kullback–Leibler Divergence between a given frequency distribution and a Gaussian (i.e. a 'normal') frequency distribution with the same mean and variance. We labour this point because Rizvi makes much of it: he characterizes our use of the notion of negative entropy as a foundational error that vitiates our entire method.

In support of his claim that 'One of the mathematically proven properties of the Kullback-Leibler relative entropy is that, like Shannon entropy, it can never be negative: it is always either zero or positive' (Rizvi, 2022b, p. 2), Rizvi cites the Wikipedia page for the mathematical statement known as Gibb's Inequality, named after Josiah Willard Gibbs (born 1839, died 1903). Rizvi offers the reader no conceptual bridge between Gibbs's foundational work on thermodynamics and its application to Information Theory half a century later, nor does he indicate the relevance of Gibbs's Inequality to his claim that entropy is always positive. We are happy to concede that in a strict sense Rizvi is right about quantities such as Shannon entropy and Kullback–Leibler divergence only ever being positive, and will return to this point shortly in relation to his critique of our use of only a subsection of a frequency distribution. But just as scientists performing experiments at sub-zero temperatures know that their negative temperatures do not indicate actual negative heat energy (an impossible notion), we likewise use negative numbers without misunderstanding what Shannon entropy and Kullback–Leibler divergence really are, as Rizvi thinks we do.

Rizvi expands upon his point about entropy always being positive:

> The inventors [that is, Ribeiro, Segarra, Eisen & Egan] varied the textbook Kullback-Leibler formula to insert the limit probabilities into it (Eisen *et al.*, 2018, p. 503, Formula 7). This meant that the mathematical proof that the Kullback-Leibler relative entropy is always non-negative was not applicable to their work and negative relative entropies became possible. With the exclusion of evidence that the method was then forced to perform, they became unavoidable. (Rizvi, 2022b, p. 2)

We did not 'insert the limit probabilities' into anything: they are present (as the symbol π) in the relative-entropy equation that we adopted, without modification, from the standard mathematics and presented as our Formula 4 (Eisen *et al.*, 2018, p. 502). The limit probabilities cannot make 'negative relative entropies ... possible', since they are always positive and are multipliers used to scale the logarithmic probability calculation. In the last sentence above it becomes clear why Rizvi is making these bizarre statements about this branch of mathematics: he mistakenly believes that limit probabilities are the means by which we exclude certain zero-value transitions from our calculation. Rizvi misunderstands what limit probabilities are, although the matter is explained at some length in our 2022 account of our method written in layman's terms (Brown *et al.*, 2022, pp. 325–32). This misunderstanding of what 'limit probabilities' are might also account for Rizvi's statement (Rizvi, 2022a, p. 2) that he does not use limit probabilities in his replication when, as we saw above, he demonstrably does because they are present in our code that Rizvi uses without modification.

6 What is a Markov Chain?

Rizvi rightly points out that when a Markov Chain holds probability distributions, the weights on the edges emerging from each node 'must add up to 1' (Rizvi, 2022b, p. 3). When the Markov Chain is represented as a matrix, as it is with our WANs, this rule requires that the sum of each row is also 1. But what if we find that in a text under examination, one of our words of interest is never followed (within our window of interest) by an occurrence of one of our other words of interest? This would produce a row of zeroes in the WAN matrix, seemingly in violation of the rule of summing to 1. We have explained before how we address this point in our method: '... in the event of a row being all zeroes we fill each cell with 1 divided by the number of words of interest we are using, in order to represent the absence of a preference' (Brown *et al.*, 2022, p. 328). Since there will be as many cells in the row as there are words of interest, this ensures that the row sums to 1.

Although the whole of a WAN thus meets the requirement for being a Markov Chain, Rizvi detects a new problem arising from our practice in multi-author comparisons of omitting from consideration those cells (and thus those transitions) for which one or more of the candidate authors' WANs contains a zero. Rizvi is right that the subset of a WAN row that is just the cells for which all the candidate authors give us some data is not itself a complete probability distribution. Once we have omitted from our consideration a cell in, say, Marlowe's WAN on the grounds that the corresponding cell in Jonson's WAN contains a zero—because Jonson never uses the transition that this cell represents—the values that we do consider from that row in Marlowe's WAN will no longer sum to 1.

Rizvi summarizes his objection on this point with two successive sentences, the first of which is true and the second false:

> The remaining small subsets of probabilities can in no meaningful sense be called probability distributions. It follows that in no meaningful sense can WANs be called Markov chains. (Rizvi, 2022b, p. 328)

These two sentences refer to two different things and what is true of the first is not true of the second. A set of cells that remain after certain cells in a row are ignored is not a complete probability distribution, so the first sentence is strictly true. But the full set of cells from one row in a WAN is a complete probability distribution for the word that this row represents and the whole WAN itself is a complete probability distribution for the entire set of words of interest that was used to make it. The first sentence being true does not make the second sentence true, as Rizvi seems to think. The important question is, does using a subset of the probability distribution introduce an invalidity to the method?

Rizvi is on to something here, but it is more of a semantic point than a mathematical one. Our process of focussing on only some of the cells in a WAN means that we are straying away from the computation of relative entropies. In particular, this modification may lead to negative entropies even before we move the zero point as described above, although it rarely does in practice. For this reason, we should perhaps at this stage in our procedure (but not before) stop calling the results of our calculations 'relative entropies' and instead call them something like 'modified relative entropies' to alert the reader that our interest is now in only a subsection of the full probability distribution. But it is nonetheless mathematically valid to compare subsections of probability distributions. For instance, having produced a probability distribution for the possible totals resulting from our rolling of two dice, we are entitled to look at the part of the frequency distribution concerning only rolls in which both dice show an odd number or only frequencies above a certain threshold. Using a selected part of a frequency distribution does not of itself break any rules that apply to frequency distributions, as Rizvi claims it does.

In the particular case of authorship attribution, there is actually good reason to select a part of a frequency distribution. If there is a transition between words that appears in a text to be attributed but never in a candidate-author profile, our method, as summarized by Formula 7 (Eisen *et al.*, 2018, p. 503), ignores the contribution of this term to the relative entropy. If we stick to the definition of relative entropy in Formula 4 (Eisen *et al.*, 2018, p. 502), we should in fact add an infinite contribution because of this occurrence. However strange, this is, in a mathematically strict sense, reasonable. If we interpret the profile as a true and perfectly accurate description of an author's style and a transition appears in the text but does not appear in the profile it means that it is impossible that the text and profile come from the same author. Of course, profiles are not perfect and another possible explanation is that we are observing a rare transition. It is therefore reasonable to ignore this transition as we do in our method. In the end, experimental evidence should dictate the choice. We have tried both and we have seen that skipping null terms as we do in Formula 7 yields better attribution accuracy. We are happy to see that Rizvi has rediscovered this conclusion and thank him for this independent validation.

7 Word adjacencies versus word frequencies

The next eleven pages of Rizvi's Part 2 attempt to show 'What the Method Really Does' (Rizvi, 2022b, pp. 4–16), using his attempted replication of our method described in Part 1 of his critique. As we show above, Rizvi's replication does not do exactly what our method does, for which reason it produces different results. Naturally, we take no responsibility for Rizvi's disappointing results. As part of his exploration of our work, however, Rizvi undertakes a fresh investigation that we consider to be genuinely valuable. If our method relies on the word adjacencies in the texts it attributes then what, Rizvi asks, will happen if we jumble the order of the words in the texts so that each word appears just as often as in the original text but in an entirely random order?

Rizvi reasons that randomly reordering the words of the texts ought to remove the evidence on which our method relies, making it incapable of attributing authorship correctly. We agree that the loss of this information should harm our method's accuracy, but not that it should make the method entirely useless since there is also valuable authorship information in the raw frequencies of the words we count. And this is exactly what Rizvi finds. In a well-designed experiment, he applies this reordering to early modern plays of known authorship to see what our method then makes of them. As Rizvi reports, his replication finds that even with the words in a random order 'The method attributes sixty-eight out of eighty-six plays correctly, an accuracy of 79%' (Rizvi, 2022b, p. 4). Compared to his previous success rate of 89.5% Rizvi considers this not much of a diminution, since it 'is not statistically significant at the conventional level of 5%' (Rizvi, 2022b, p. 4).

We disagree and interpret Rizvi's results as showing that taking into account the specific adjacencies of the words, as opposed to their mere presence, raises the accuracy from the quite mundane—lots of previously tried methods can achieve around 80% accuracy—to close to the state of the art (Our own results, with subtleties of application that Rizvi does not replicate, get to 90–94% accuracy, depending on the sizes of the texts.). We are grateful to Rizvi for, in our view, independently corroborating our claim that word adjacencies are a distinctive marker of authorial style.

Rizvi interprets his results as proof that much of the success of our WAN method comes not from its capturing of word adjacencies but from its capturing of the mere frequencies of the words. We agree with him in the sense that in counting word adjacencies we are necessarily also counting their frequencies. There can be no reckoning of the frequency with which the occurrences of the word *the* are followed shortly thereafter by occurrences of the word *and* that is not, in that process, also a counting of the frequencies of *the* and *and*. The word adjacencies in a text are predicated on the words in question being present in the text.

Rizvi performs a further experiment to detect if the strongest word adjacencies found by our method—actually, his method, since his replication is imperfect—correlate with the overall frequencies of the words. Using the Shakespeare canon and the single word-of-interest *a*, he asks whether the word that is most often found following *a* is simply the most-common word, and the word that is next most often found following *a* is simply the next most-common word, and so on down the rankings of word adjacencies and word frequencies. His Fig. 1 shows that indeed as we move down the list of adjacencies in descending order of strength we are also, on the whole, moving down the rank of word frequencies in the canon.

This caveat of 'on the whole' is important because the line plotting the correspondence between rank-order of word adjacency and rank-order of word frequency in Rizvi's Fig. 1 is far from straight. Some words' adjacencies to *a* are much more highly ranked than we would expect from their frequency, and some are much more lowly ranked. The spikes and dips are visible in Rizvi's Fig. 1 and we can see which words are causing them by turning to his Table 1. The word *little*, for instance, is the 13th most likely of all the words-of-interest to be found shortly after an occurrence of *a* but is only the 70th most frequent word in Shakespeare. To the right of *little* is another spike for *most*: 18th highest in the rank order of adjacency to *a* but only 48th in the rank order of frequency. We can follow dips in the line in the same way. To show what we would expect if there were a perfect correlation between word-adjacency rank order and word-frequency rank order, Rizvi overlays a straight trend line on the uneven data line in his Fig. 1.

Rizvi explains his actual data's departures from the trend line: 'We cannot expect a perfect linear relationship because the probabilities are calculated by the method using formulae that involve exponentiation' (Rizvi, 2022b, p. 8). We are unclear why he thinks exponentiation is the cause, and he offers no elaboration. We would be interested to see Rizvi's calculation of the correlation between the two rankings, since it appears to us that the departures are most plausibly understood not as random fluctuations but as the very authorial preferences our method tracks. That is, it looks as if the data show that Shakespeare really does put *little* and *most* shortly after *a* more often than we would expect given these words' frequencies in his works, and puts other words there less often.

Rizvi's concludes that our method is 'just a proxy for information we could get from old-fashioned word-counting, the word adjacencies being a flourish' (Rizvi, 2022b, p. 8). If we can agree that our flourish has boosted the attribution success rate of around 80% achieved with mere word-frequency counting to a success rate in excess of 90% then we consider our efforts well rewarded. Indeed, we will happily accept Rizvi's compliment in the sense in which Shakespeare uses the word 'flourish', as a fanfare to accompany an important arrival.

For his final experiment, Rizvi constructs a new method, seemingly half in jest, for counting the frequencies of words in eighty-six plays in order to test whether mere word counting is sufficient for authorship attribution. He reports 89.5% accuracy when testing whole plays and 68% when testing individual scenes, the latter being only one percentage point worse than his success at attributing scenes using his supposed replication of our method (Rizvi, 2022b, p. 11). Obviously, a comparison of one relatively poor performance in Rizvi's experiments with another relatively poor performance—his 68% scene-wise score for the word-counting method versus his 69% scene-wise score when attempting to replicate our method—tells us nothing if his replication of our method is imperfect.

When the texts are smaller than whole plays we achieve rather better results. Across nearly 100 plays by six authors broken into acts, we get 93.4% accuracy, and when there are only two prime candidates to choose between, as not infrequently happens in authorship debates, our method achieves 91.5% accuracy even with as little to work on as individual scenes (Segarra *et al.*, 2016, pp. 243–44). That Rizvi is unable to get his replication to perform well is, as we have shown, a consequence of his not following our method. It cannot help that he misunderstands the role of limit probabilities and the reasons and mechanisms for our exclusion of transitions for which one of the candidate authors' canons shows no instances. As we have seen above, he acknowledges using a larger 'window'—the number of words across which a transition is allowed to count for our purposes—than we do (ten instead of five words) and he does not allow transitions to occur across speech boundaries whereas we do.

Rizvi's claimed 89.5% accuracy rate when testing plays with his new word-counting method is impressive. We applaud his result and would like to examine the software that achieved it. Unfortunately, unlike the Part 1 section of Rizvi's critique, discussed above, the online support materials at <http://www.shakespearestext.com/wan.zip> do not contain the software that achieved this result.

8 The provenance of texts

Rizvi ends with a brief discussion of the important matter of 'The Choice of Texts' (Rizvi, 2022b, pp. 16–17). He quotes the present author Egan's objection to Rizvi's decision to source all his non-Shakespearian plays from the dataset of the EarlyPrint project at Washington University in St Louis while using the Folger Shakespeare editions for all his Shakespeare plays. He points out that by getting all our texts from the Literature Online (LION, now One Literature) database, the present authors also introduce non-homogeneity into their dataset, since LION is 'a collection of transcripts from many non-homogenous primary sources, the quarto and Folio editions of early modern plays' (Rizvi, 2022b, p. 16).

This is quite true, but the important difference is that our non-homogeneity is random. Just what kind of early edition LION used for each play we took from it is mere happenstance. In contrast, by choosing to get all his non-Shakespeare plays from a source in which the modernization of spelling was done by computer (the EarlyPrint transcripts) and all his Shakespeare editions from a source in which the modernization of spelling was done by humans (the professional editors of the Folger Shakespeare series), Rizvi introduces systemic bias in provenance. And it is a bias that runs along exactly the authorial lines that he wants his attribution software to find for itself. We have not shown that this systematic non-homogeneity in Rizvi's dataset affects his results, and perhaps it does not, but careful investigators avoid introducing such unnecessary bias into their primary data.

9 Conclusion

We do not think that Rizvi's critiques of the WAN methodology have been without value. They have required us to clarify aspects of our methodology and rethink how we explain them, and to provide additional details to justify choices we have made. We believe that through critique, response, and counter-response, the field of authorship attribution builds upon each new advance, and abandons approaches that turn out to be fruitless. We are grateful to Rizvi for pointing out that once we exclude certain transitions in our comparisons of WANs we are no longer dealing with full probability distributions and hence should not call the results of these comparisons 'relative entropies'. For that reason, we will in future refer to these as 'modified relative entropies' instead. We insist, however, that it is mathematically valid to compare subsections of probability distributions in this way.

We are especially grateful to Rizvi for his innovative experiment in which he jumbled the word order of texts so that although the frequency of each word remained the same, the proximities of one word to another that our WAN method measures are lost. He reports that using only the frequencies of words he was able to achieve 79% accuracy of attribution at the level of whole plays, compared to 89% accuracy when the word order (and hence the information gathered by the WAN method) is preserved (Rizvi, 2022b, p. 4). At these levels of success, a gain of ten percentage points is worth celebrating.

We are pleased to agree with Rizvi that '... authors use some words more than they use others, and these preferences are not the same for everyone' and that we should be 'treating small differences between a text's and a candidate author's usage of words as evidence for his authorship and treating large differences as evidence against' (Rizvi, 2022b, p. 15). We restate this agreement with Rizvi because some influential scholars of early modern drama disagree. Brian Vickers recently asserted that scholars of authorship attribution should not assume that '... words chosen by a dramatist to create and differentiate *characters* can identify their *authors*' and claimed that scholars who make this assumption are committing 'a serious category error that has made all their authorship attributions unreliable' (Vickers, 2022, p. 211). Vickers stands on one side of a chasm that divides early modern authorship attribution scholarship and we are glad to stand with Rizvi on the other side.

Author Contributions

Gabriel Egan (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing), Mark Eisen (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing), Alejandro Ribeiro (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing), and Santiago Segarra (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing).

References

Argamon, S. E. (2018). Computational forensic authorship analysis: promises and pitfalls. *Language and Law/Linguagem e Direito*, 5(2): 7–37.

Brown, P., Eisen, M., Segarra, S., Ribeiro, A., and Egan, G. (2022). How the Word Adjacency Network (WAN) algorithm works. *Digital Scholarship in the Humanities*, **37**: 321–35.

Eisen, M., Ribeiro, A., Segarra, S., and Egan, G. (2018). Stylometric analysis of early modern English plays. *Digital Scholarship in the Humanities*, **33**: 500–28.

Feynman, R. P. (2011). *Six Easy Pieces*. London: Penguin.

Forsyth, R. S., Holmes, D. I., and Tse, E. K. (1999). Cicero, Sigonio, and Burrows: investigating the authenticity of the *Consolatio*. *Literary and Linguistic Computing*, **14**: 375–400.

Hoover, D. L. (2004). Delta prime? *Literary and Linguistic Computing*, **19**: 477–95.

Jockers, M. L., Witten, D. M., and Criddle, C. S. (2008). Reassessing authorship of the Book of Mormon using Delta and nearest shrunken centroid classification. *Literary and Linguistic Computing*, **23**: 465–91.

Jockers, M., Nascimento, F., and Taylor, G. H. (2019). Judging style: the case of *Bush Versus Gore*. *Digital Scholarship in the Humanities*, **35**: 319–27.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**(1): 79–86.

Mosteller, F. and Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association*, **58**: 275–309.

Rizvi, P. (2020). Authorship attribution for early modern plays using function word adjacency networks: a critical view. *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews*, **33**: 328–31.

Rizvi, P. (2022a). An analysis of the Word Adjacency Network method—part 1—The evidence of its unsoundness. *Digital Scholarship in the Humanities*, **38**: 347–60.

Rizvi, P. (2022b). An analysis of the Word Adjacency Network method—part 2—A true understanding of the method. *Digital Scholarship in the Humanities*, **38**: 361–78.

Schrödinger, E. (2021). *'What is Life?' With 'Mind and Matter' and 'Autobiographical Sketches'*. Cambridge: Cambridge University Press.

Segarra, S., Eisen, M., and Ribeiro, A. (2015). Authorship attribution through function Word Adjacency Networks. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Signal Processing*, **62**(20): 5464–78.

Segarra, S., Eisen, M., Egan, G., and Ribeiro, A. (2016). Attributing the authorship of the *Henry VI* plays by word adjacency. *Shakespeare Quarterly*, **67**: 232–56.

Segarra, S., Eisen, M., Egan, G., and Ribeiro, A. (2020) A response to Pervez Rizvi's critique of the word adjacency method for authorship attribution. *ANQ: A Quarterly Journal of Short Articles, Notes and Reviews*, **33**: 332–7.

Vickers, B. (2022). The limitations of stylometry: idiolect and the authorship of *Titus Andronicus*. *Notes and Queries*, **267**: 207–11.