

# A reply to Pervez Rizvi's letter

Gabriel Egan<sup>1,\*</sup>, Mark Eisen<sup>2</sup>, Alejandro Ribeiro<sup>3</sup>, Santiago Segarra<sup>4</sup>

<sup>1</sup>School of Humanities, De Montfort University, Leicester, UK

<sup>2</sup>Intel, Santa Clara, California, USA

<sup>3</sup>Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA

<sup>4</sup>Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA

\*Corresponding author. Gabriel Egan. E-mail: [mail@gabrielegan.com](mailto:mail@gabrielegan.com)

We begin with an admission of error and an apology to Pervez Rizvi. Looking again at his software, we concede that, contrary to what we asserted in our essay, his software does indeed implement the distinctive feature of our ‘formula 7’, which is that of processing only those ‘transitions’ that are non-zero in all the candidates’ profiles when testing multiple candidates for authorship. We overlooked the bit of his code that does this because, unlike our code that he builds upon, Rizvi’s code is undocumented and it adds this functionality using a logical construction that computer scientists are taught to avoid. Specifically, Rizvi uses a 200-character long ‘if’ statement governing seven distinct Boolean expressions. Much virtue in ‘if’. But it works, and Rizvi is right that without it his replications would not even come close to producing meaningful results. We apologize to Rizvi for our error arising from overlooking that line of his code. We reject the rest of Rizvi’s complaints.

Rizvi’s point about the computed entropies from ‘formula 7’ becoming negative has already been addressed in our latest essay, where we wrote that ‘Rizvi is on to something here’ and agreed in future to call them ‘modified relative entropies’ (Egan *et al.*, 2023: 1503–1504). The idea of negative entropy is not completely new, and we quoted the celebrated physicist Erwin Schrödinger endorsing it. Rizvi is entitled to insist that, from first principles, entropy can never be negative, but we doubt that he will win many adherents to this view merely by pointing his readers to the Wikipedia page for Gibbs’s Inequality without further explanation.

Rizvi objects that in two of our essays (Eisen *et al.*, 2018 and Brown *et al.*, 2022), we used different settings for the size of the text window and whether speech boundaries are respected when creating a Word Adjacency Network (WAN). This difference was intentional, and we discussed it in our latest essay (Egan

2023: 7–8). Our illustrative software allows the user to make her own choices regarding these details, and its internal documentation indicates just where to uncomment-out a block of program code and where to alter a variable assignment in order to change this behaviour. It is the user’s responsibility to ensure that when replicating someone else’s experiments, she uses the same settings as the original investigators did.

We believe that the calculation of the limit probabilities of a Markov Chain is essential to its proper use in this kind of research. In his essay about our method, Rizvi wrote that ‘For my experiments in this article, these limit probabilities are not of interest, and we do not need to go into them here’ (Rizvi 2022: 348). We think he does need to go into them, and Rizvi’s letter suggests that he now realizes this: ‘At no point do I suggest that I do not include them in my calculations’. We agree that his calculations do use limit probabilities since they are built into the code that he got from us, and in our latest essay we pointed out that despite his dismissal of their importance, he does calculate them (Egan *et al.*, 2023: 1501). We maintain that if limit probabilities are important enough to use in one’s calculations, then they are important enough to explain to one’s readers.

Because many different methods have confirmed the fact, it is uncontroversial to believe that Shakespeare’s play *Titus Andronicus* was written partly by George Peele, his *Pericles* partly by George Wilkins, and his *Timon of Athens* partly by Thomas Middleton. Our WAN tests corroborate these co-authorship attributions, and in our latest essay we challenged Rizvi to explain how this could be if, as he believes, our tests are worthless (Egan 2023: 7). Rizvi ends his letter by objecting that we should not presume to even know what he thinks about these plays’ co-authorship since

he has ‘never given’ his views on the matter, ‘nor ... ever tested for’ co-authorship in these plays.

In fact Rizvi’s views on the co-authorship of these plays are a matter of record, because he did test them and he published his results. On his website ‘Shakespeare’s Text: Collocations and N-Grams’ at <<https://shakespearestext.com/CAN>>, Rizvi described his tests for word collocations shared by multiple plays (which is a method of authorship attribution) and presented his counts in spreadsheets that were named to reflect the attributions he was testing. Thus he provided a file of numerical data called ‘summary-collocations-titus\_andronicus\_excl\_act\_1\_2\_1\_2\_2\_3\_2\_4\_1.csv’ (for *Titus Andronicus* excluding Act 1, and Scenes 2.1, 2.2, 3.2, and 4.1, the bits by Peele), another called ‘summary-collocations-pericles\_acts\_1\_to\_2.csv’ (for the bits of *Pericles* by Wilkins), and another called ‘summary-collocations-timon\_of\_athens\_middleton.csv’ (for the bits of *Timon of Athens* by Middleton). In October 2017, when Rizvi published these spreadsheets, we downloaded them, and we have now added them to the online materials in support of our essay at <<https://gabrielegan.com/WAN>> so that readers can see that Rizvi once had a definite view on this matter, did the necessary tests, and published his results.

Since October 2017, Rizvi has substantially revised his website, including renaming it from <<https://shakespearestext.com/CAN>>, from where we downloaded these spreadsheets and which is now defunct, to <<https://shakespearestext.com/can>>. (The WayBack Machine of the Internet Archive contains a snapshot

of the ‘.../CAN’ version of Rizvi’s website from October 2018.) Perhaps since 2017 Rizvi has also changed his mind about *Titus Andronicus*, *Pericles*, and *Timon of Athens* and now believes that they are solely by Shakespeare. This would be a lonely position to adopt, but it would at least enable him to evade our challenge. But if he still thinks that these plays are co-authored, then our challenge stands and it is incumbent upon him to explain why our supposedly worthless tests confirm his published views on these plays. Are our tests valid only when they agree with his views on certain plays and invalid when they do not? Claiming that he never tested these attributions nor gave his views on them is untenable in the light of the evidence from Rizvi’s published spreadsheets about the matter.

## References

- Brown, P. et al (2022) ‘How the Word Adjacency Network (WAN) Algorithm Works’, *Digital Scholarship in the Humanities*, 37: 321–35.
- Egan, G. et al (2023) “‘I Would I had that Corporal Soundness’: Pervez Rizvi’s Analysis of the Word Adjacency Network Method of Authorship Attribution”, *Digital Scholarship in the Humanities*, 38: 1494–1507.
- Eisen, M. et al (2018) ‘Stylometric Analysis of Early Modern English Plays’, *Digital Scholarship in the Humanities*, 33: 500–28.
- Rizvi, P. (2022) ‘An Analysis of the Word Adjacency Network Method—Part 1—The Evidence of its Unsoundness’, *Digital Scholarship in the Humanities*, 38: 347–60.