

Marlowe and Overreaching: A Misuse of Stylometry

Abstract

In 'Christopher Marlowe: Hype and Hoax' (2018), Hartmut Ilseman implies that his application of the Rolling Delta feature of R Stylo is sufficiently robust that a century and a half of traditional scholarship should be overturned, and Marlowe stripped of the majority of his canon, including *Doctor Faustus* and *Edward II*. The article concludes that 'Marlowe is totally overrated in his influence on modern English drama' (26), the natural consequence of stripping away 5/7ths of his canon. In this response, I demonstrate that the assumptions underlying this application of the Delta method, and the application itself, are fundamentally flawed, leading to predictably erroneous conclusions. Problems with the study include a poorly designed test environment; incorrect preparation of texts; assuming that 'Marlowe's style' can be determined by a single early play; selecting and constructing Shakespeare's comparison texts in a manner likely to prejudice results; ignoring the effect upon style of a play's date and genre; failing to consider the effect of different-length comparison texts; dismissing external evidence of authorship that conflicts with the test outcomes. I argue that in the light of these issues, the results and conclusions must be dismissed. Further, the question is raised as to whether the current methods of computational stylistics, even when more rigorously applied, are equipped to challenge the attribution of the accepted Marlowe canon.

A Summary of the Study

‘Christopher Marlowe: Hype and Hoax’ (Ilsemann 2018) describes using a version of John Burrows' Delta method, with rolling variable sample sizes, to overturn Marlowe's authorship of every play that has been attributed to him except for the two Tamburlaine plays.¹ The author begins by accepting the attribution of *Tamburlaine I*, and then uses this as a comparison text to determine that *Tamburlaine II* is also by Marlowe. He uses the same method to give Marlowe sole attribution of the anonymous play *Lochrine*. Expanding the method to other plays from both the Marlowe and Shakespeare canon, he reaches the following conclusions:

- *Dido Queen of Carthage* is by Kyd, with the possible exception of Act I (by Marlowe)
- *The Massacre at Paris* is by Rowley, possibly with Lodge
- *The Jew of Malta* is by Nashe, Kyd and Shakespeare
- *Doctor Faustus* is by Nashe, Kyd, Shakespeare, Jonson, Rowley, Dekker, Chettle, and Greene
- *Edward II* is by Shakespeare and Rowley
- *Henry V* is by Shakespeare, based on an old play by Marlowe
- *3 Henry VI* is by Shakespeare and Kyd
- *King John* is ‘unambiguously a Shakespeare play’.

In each case except for Shakespeare, the comparison text used to determine an author's style was a single play. I have listed the plays in probable date order.² As can be seen, they cover a period of eighteen years, with Marlowe's play being the earliest:

- Marlowe – *Tamburlaine I* (1587)

- Kyd – *Soliman and Perseda* (1588)
- Peele – *The Battle of Alacazar* (1588)
- Greene – *Friar Bacon and Friar Bungay* (1589)
- Munday – *John A Kent and John a Cumber* (1590)
- Nashe – *Summer's Last Will and Testament* (1592)
- Chettle – *The Tragedy of Hoffman* (1603)
- Rowley – *When You See Me You Know Me* (1604)

Using these comparison texts (and what are described as ‘collated reference texts’ for Shakespeare), the author ran a number of Delta tests with sample sizes varying from 1000 words to 8000 words, in 500-word increments (i.e. 1000 word, 1500 word, 2000 word etc. windows) with each ‘window’ rolling forwards 250 words at a time. He conducted three different types of Delta test, measuring:

- frequency of single function words (which he calls MF1W)
- frequency of character bigrams (MF2C)
- frequency of character trigams (MF3C)

For those unfamiliar with these terms, in character bigrams, the word ‘corner’ would be represented by 5 separate groupings: ‘co’-‘or’-‘rn’-‘ne’-‘er’. In character trigrams, it would be represented as ‘cor’-‘orn’-‘rne’-‘ner’. Spaces between words are also included in bigrams and trigrams i.e. ‘the’ as it would be rendered in a flow of text would become ‘ t’-‘th’-‘he’-‘e ’.

Where the results conflicted (as they did in several of the tests), the author favoured the results of the character n-grams over the function word tests, and the character trigrams over the bigrams.

Not all aspects of the method are clear from the published paper and for that reason it was necessary for me to contact the author directly. In that correspondence he wrote that ‘the paper

deals with the stylistic deviations of the other plays from the *Tamburlaines*.³ If the paper merely addressed stylistic deviations of the rest of Marlowe's canon from the *Tamburlaine* plays, there would be no argument; there's no question that the Marlowe canon is diverse. But what the paper does is claim that those stylistic differences are sufficiently marked (and the tests sufficiently robust) to declare that the rest of the Marlowe canon is not, in fact, by Marlowe. The author's confidence in his test results leads him to pronounce that 'Marlowe's stylistic presence occurred mostly at the beginning of plays. He seems to have begun writing full of ideas, but then did not have the perseverance to carry things through to the end.' (14) The purpose of this paper is to show that this conclusion rests on erroneous results, which have arisen from a combination of false assumptions and a poorly-executed method.

The Application and Applicability of the Delta Method

As John Burrows and Hugh Craig write in 'A Collaboration About A Collaboration' (Craig and Burrows 2012), Delta allows 'any strong affinities, authorial or not, to reveal themselves... Small difference in Delta scores can sometimes yield false rankings. For that reason, it is rarely advisable to use Delta alone' (31). Burrows, the inventor of the method, uses Delta consistently as a blunt 'winnowing' tool, in order to reduce the field of possible authors a text should be tested for. In the words of Burrows and Craig, 'the Delta procedure is genre sensitive as well as author sensitive and unsupported results should not be taken either as conclusive or as purely authorial' (36). At no point does Burrows arrive at an attribution through Delta alone. Indeed, in Burrows and Craig's testing of 3 *Henry VI*, Kyd came up as an authorship frontrunner; the conclusion that Ilseman ends with. But with more targeted testing methods, Burrows and Craig decided that Kyd was not involved: 'Clearly there are affinities between *The Spanish Tragedy*

and *Soliman and Perseda* and *3 Henry VI* across the board. This is persistent finding with Delta tests, as discussed in this chapter. It is not supported by our other tests, and we discount it in our conclusions, explaining it as a general likeness in dramatic texture which does not survive more targeted authorial testing' (Burrows and Craig 2017, 202). There is no support for Delta as an accurate test of authorship of early modern plays when it is used alone.

Outlining the Delta method in 'Delta': A Measure of Stylistic Difference and Likely Authorship', Burrows states that 'with texts of a bygone era, it is usual and desirable to standardize spelling' (269). Perusing the 50-word extracts of texts given as illustrations in the appendix of the author's paper, it is clear that this has not been done. In the list of examined texts, modern spellings, and modern punctuation (relevant to n-gram analysis), are apparent in *Tamburlaine Parts 1*⁴ and 2, *The Jew of Malta*, *Doctor Faustus A and B*, *Edward II*, *Edward III*, *Henry V*, *King John*, but not *Lochrine* ('traverst', 'drave'), *Dido* ('loue', 'iuno', 'worthles', 'blowes', 'cloath', 'fild') *The Massacre at Paris* ('admirall', 'wishe', 'joyn'd', 'nuptiall', 'desolve', 'sparkes'), *3 Henry VI* ('slily'). In the reference texts, spelling appears to have been modernized in all texts except for *Friar Bacon and Friar Bungay* ('lawnds'), *Marius and Scilla* ('wher', 'waighty', 'powre') *John A Kent and John a Cumber* ('powesse') and *The Arraignment of Paris* ('raught'). In addition, there is a question-mark over the different treatment of apostrophes in these texts where the author elided an 'e' to turn what would have been pronounced as two-syllables in this era into a single syllable. In the extracts from *3 Henry VI* and *King John*, for example, we have 'cheer'd' and 'borrow'd', but in *Satiromastix* we have 'resolv'd' and 'turn'd'; in *The Spanish Tragedy*, 'imprison'd'. To my knowledge, none of the tests would count 'imprison'd' and 'imprison'd' as the same word and the character n-grams, in particular, would make a clear distinction between words ending 'd' and those ending 'ed' or 'xd'

(where 'x' is another letter); endings which are potentially numerous in plays of this era. Note also that all the tests would count *Dido's* 'loue' as distinctly different from *Soliman and Perseda's* 'love'; *Dido's* 'iove' (not in the first fifty words but assumed from 'iuno') as distinct from *David and Bathseba's* 'jove'.⁵ The spelling issues alone are sufficient to invalidate the test results.

Joseph Rudman's 2016 critique of stylometric Shakespeare studies contains numerous points equally applicable to any study of Marlowe. One of the actions that should be undertaken in preparing texts for testing is that quotations, and languages other than English, should be stripped out of texts before testing (Rudman 2016, 315). From the extracts in the author's appendix, it is clear that at least some (if not all) of the texts have not been prepared in this way: Latin is visible in the first 50 words of *Doctor Faustus* (B), *Lochrine*, and *Summer's Last Will and Testament*. Presumably, therefore, *Henry V* was left replete with its 200-plus words of French, and there will be other non-English intrusions as well as quotations in both the test and comparison texts.

The Reliability of Character N-grams

The reason for conducting character bigram and trigram tests, according to the author, was their evaluation by Grieve (2007), who called 2- and 3-character n-gram tests 'some of the most accurate techniques tested' (263). But the context in which they were established as accurate measures of style by Grieve is so different from the context of the author's testing as to make any claim to their accuracy invalid. In 'Quantitative Authorship Attribution: An Evaluation of Techniques', Grieve was careful to establish a very large yet carefully controlled corpus for each author under consideration:

It is not a trivial matter to define the variety of language in which an author writes. Most writers interact with multiple readers, at multiple times, and in multiple registers, and so one must decide which of an author's many varieties the author based corpus will represent. When attributing an anonymous text, it is unnecessary and unsound to compile an author-based corpus that attempts to represent the variety that encompasses all that author's written utterances: the anonymous text is the product of a single situation and so each author-based corpus should be composed of texts produced in the most similar register, for the most similar audience, and around the same point in time as the anonymous text. Otherwise, the investigator might get false negatives: when the anonymous text is compared to the writings of its author they may not match because of variation that is the product of differences in audience or register or time. (255)

The conditions under which 2 and 3-character n-grams were found to be able to 'distinguish between two possible authors with 94% accuracy, and ... distinguish successfully between up to ten possible authors' (263) were when the author corpora were extremely uniform, being all taken from columns of the UK-based newspaper the *Daily Telegraph* between 2000-2005. The corpora were deliberately chosen to be similar in tone, subject matter, and register, in order to ensure that the algorithms could tell them apart on style alone:

In this study, a highly representative corpus of possible authors was compiled by combining forty author-based corpora that contain texts written by authors from similar social backgrounds (middle-aged, conservative, Anglo-Saxon, upper middle-class, well-educated, British), which are written in the same register (Telegraph opinion column) and for the same audience (the readership of the Telegraph's opinion section), and which are published over the same short span of time (2000–2005). The likelihood that the topics of the texts cluster by authorship has been minimized because newspaper opinion columnists, especially when writing at the same time and in the same city, will tend to write about a similar range of subjects. (256-7)

There is no measure of how accurate the same tests could be when conducted under different conditions; where the texts are not on the same subject, or in the same register; are not by authors of similar social backgrounds; are over a span of two decades when theatrical genres were falling in and out of fashion and writing styles (in the relatively new creative area of the public playhouse) were developing rapidly. The very rigour with which Grieve compiled his author corpora indicates that he was conscious that stylometric tests distinguish texts on aspects of style that are entirely separate from their authorial attribution: the style that comes with writing in a different genre, on a different subject matter, for a different audience, and at a different time.

Whereas Grieve painstakingly compiled 'a highly representative' corpus for each author, Ilseman did not; for all of the authors except Shakespeare he chose a single play which he considered to be reliably attributed. But this is not the same as 'representative'. Indeed, it would be unreasonable to finger any single play as representative of a multi-work author. Is *King Lear* more 'representative' of Shakespeare than *As You Like It*? Is *Tamburlaine Part 1* more representative of Marlowe than *Doctor Faustus*? The point here is that the method used to strip away Marlowe's canon has not in any way been validated. If the proposal is that a single early modern play can determine an author's style across their canon using Rolling Delta, it should be demonstrated that it can, for example, correctly assign the majority of Shakespeare's canon to Shakespeare using only a single early play as the comparison text.

This leads us to the central underlying (and problematic) assumption of the paper under critique.

Defining 'Marlowe's Style'

The author of the study refers to analysing ‘the presence or absence of Marlowe’s style in contemporary plays, starting with those plays that bear the clear mark of his style’(2). What he considers to be the definition of ‘Marlowe’s style’, and his reason for defining it thus, he does not discuss, but from the text used to define Marlowe’s style it is clear that he assumes that ‘Marlowe’s style’ is defined purely by the early play *Tamburlaine Part I*, the only Marlowe text used for comparison. In correspondence, he explained that he chose this text because “There was enough empirical evidence that the *Tamburlaines* are plays by Marlowe, so they became the starting point.’ I would argue that the empirical evidence that the *Tamburlaine* plays are by Marlowe is not markedly stronger than the evidence that other plays of his canon are by him, but that is not the most critical issue.

Both *Tamburlaine* plays were published in Marlowe’s lifetime, though without any name on the title page, as was normal before 1594 (Erne 2003, 46). As with all of Marlowe’s plays, his authorship of this play was only gradually established during the 19th century, as contemporary references to the play (and Marlowe) were gathered and interpreted (Boas 1940, I: 190-197), but in the intervening time it has been established beyond reasonable doubt. However, we have substantial reasons to challenge the idea that *Tamburlaine I* is a fair representation of ‘Marlowe’s style’.

To begin with, it is not the complete text as Marlowe wrote it. The printer, Richard Jones, admits to removing the comic scenes, believing these ‘fond and frivolous gestures ... though haply they have been of some vain-conceited fondlings greatly gaped at, what time they were shewed upon the stage in their graced deformities: nevertheless now to be mixtured in print with such matter of worth, it would prove a great disgrace to so honourable and stately a history’ (Bartels and Smith 2013, 326). Stripped of the comedy that Marlowe wrote into it, we cannot

say that *Tamburlaine Part I* represents the full gamut of 'Marlowe's style'. Indeed, we can reasonably say that it does not.

Something similar has happened to *Doctor Faustus* in the hands of modern critics. Though the comic scenes in *Doctor Faustus* (particularly those in the B-text) have sometimes troubled scholars to the extent that they have wished to attribute them to someone else,⁶ they were found 'Marlovian' by Greg (1950, 63). The assumption is partly that Marlowe didn't write comedy and partly that Marlowe didn't write prose, but both of these assumptions can be challenged.⁷ The problem may not be the scenes themselves, but rather 'the attempt to impose a generically consistent tragic or moral reading of the play' according to Suzan Last (2000, 23). Last argues that by 'focusing on the comic elements as integral and important to the text overall', one can 'see the play as containing an unabashedly comic sensibility and as essentially parodic rather than primarily tragic or moralistic in tone. It can be argued that many scholars of *Doctor Faustus*, just like *Tamburlaine*'s printer, have attempted to 'normalize' Marlowe's intent and shoe-horn his play into a more traditional tragicomic genre and structure. Misunderstanding the subversive subtext and parodic nature of the comic scenes and mistaking them for trivial 'groundling fodder', they have attempted to excise them by declaring them non-Marlovian.

Yet Thomas Nashe, in what is widely recognized as a tribute to Marlowe under the name of Aretine, called Marlowe 'one of the wittiest knaves that ever God made' (Nicholl 2002, 64); a suggestion that he was more than capable of being funny. And if we accept the traditional Marlowe canon (and I contend that this study is not sufficiently rigorous for traditional scholarship to be set aside), Marlowe did write comedy. The *Jew of Malta* is another 'mixed-genre' play, which is not easily understood as a tragedy, despite the billing of early publishers to that effect. T.S. Eliot regarded it as a 'tragic farce'; Harold Bloom essentially concurred. Brian

Gibbons, defining the generic aspects of the city comedy, was tempted to classify *The Jew of Malta* as the first play of that type, being only dissuaded by its theological complexities (Scott 2010, 93). These are similar to the comic/serious components we see complicating the reception of the texts of *Tamburlaine* and *Doctor Faustus*. In *The Jew of Malta*, instead of having separate comic scenes that might easily be removed, Marlowe shows the main character Barabas in situations imbued with both serious and comic elements.

What has survived of *The Massacre at Paris* also resists easy categorisation; like *The Jew of Malta* it highlights farcical elements of what was a serious historical event. The Duke of Guise is a villain so extreme that at times he appears to veer into pantomime, and the sheer relentlessness of the ever-rising death-toll is laughable (and I would argue, intentionally so) in performance. As with *The Jew of Malta*, what comic scenes have survived are inherently part of the action, as when a soldier, trying to talk himself into the murder of Mugeroun, is surprised by his victim's sudden appearance⁸ and the scene where two soldiers debate how to dispose of the body of Admiral Coligny.

We know that genre, including sub-genre, affects stylometric tests. As Rudman points out, 'It has been shown in many studies that genre trumps authorship — there is a greater stylistic difference between one author in different genres than between two authors writing in the same genre' (Rudman 2016, 318). All the most recent reliable studies published acknowledge the power of genre to skew results and the need control for it. Following the methodology evaluation of Stamatatos (2009), Luyckx and Daelemans stress that 'a good evaluation corpus should be controlled for genre and topic' (2011, 41). This is why it is a common practice in computational stylistics to choose comparison texts in the same sub-genre category: a history play will be tested against other history plays, a comedy against other comedies. The cross-genre

sensibilities of Marlowe may well complicate stylometric testing in his case. But what we can say for certain is that the text of *Tamburlaine 1*, which we know to have had its comic scenes excised, cannot be taken as typical of Marlowe's style.

Stylistic Development Negated

The removal of its comedy is not the only reason why *Tamburlaine 1* is not representative of Marlowe's style. Equally significant in the issue of its unsuitability is the fact that it is an early play. The article defends using a single play as a comparison text on the basis that 'stylistic developments are negated when the whole corpus is used for analytical purposes' (Ilseman 2018, 2). Yet stylistic developments are negated to an even greater degree when only a single text from a corpus is used as a comparison.

It is ingenuous to imply that Marlowe's style was fixed in a single play, let alone a play that 'fits in accurately at the bottom of the curve of literary development shown by his later plays' (Boas 1940, I: 197). *Tamburlaine 1* was a very early play; Marlowe's first hit on the public stage. The only play that might claim to be earlier was *Dido Queen of Carthage*, which may have been written while he was a student at Cambridge.

That a writer's style develops with experience is self-evident. As Rudman points out, '[t]here can be no doubt that for the majority of authors, style changes over time' (Rudman 2016, 319). One only need to look at the Shakespeare canon to appreciate how much that style can change even in so small a period as six years: the approximate gap between *Titus Andronicus* and *Hamlet*. Brian Vickers noted that his tests on Shakespeare's canon in the period between 1595 and 1599 proved less reliable than for later plays (Vickers 2004, 106-7) and no valid stylometric study exists that shows how Shakespeare's changes over time (Rudman 2016, 319). Nor does

such a study exist for Marlowe, yet Marlowe's style also developed noticeably in the six years or so that he was writing for the public stage. In *Doctor Faustus* and *The Jew of Malta*, interaction between the characters became more naturalistic in pacing; a more rapid to-and-fro between characters emerged, with lengthy speeches reserved for critical moments. In *Doctor Faustus* and *The Massacre At Paris*, Marlowe developed the use of soliloquy to reveal a character's internal thoughts. The heavy end-stopped blank verse style of the *Tamburlaine* plays shifted into a more fluid and flexible style, with greater enjambment and more feminine endings, in later plays. Other changes to his prosody have been noted by Marina Tarlinskaja (2014, 70-87).

That such developments would not include the aspects of style that are measurable by stylometry is unlikely and is part of the reason why multiple works from an author's corpus are generally used to create a reference text. Tests on the voluminous canon of Henry James (Hoover 2007) demonstrated that an author's style can change dramatically and still be distinguishable from that of other authors using a variation of Delta, but this result was achieved through a careful validation process that determined the method concerned correctly attributed 29 out of 30 novels. The size of the reference corpus was substantial. Had the reference text been a single early work by James (as with Marlowe in the study under scrutiny), it is most unlikely that his late works would have been attributed to him.

Andrew Gurr notes Marlowe's 'constant changes of mode in his playwriting--from classical tragedy in *Dido* to heroic tragedy in *Tamburlaine*, religious morality in *Faustus*, satire in *The Jew of Malta*, contemporary political drama in *The Massacre at Paris*, and English history in *Edward II*' (Gurr 2017, 79). Against this backdrop, the choice of a single early play as the epitome of 'Marlowe's style' shows an absence of any genuine understanding of the very thing it seeks to define. The article's declaration 'it is a truth universally acknowledged that if the

axiomatic assumptions are faulty any outcome must be erroneous' (2) must therefore be applied to itself.

Stylochronometry has been ignored in this study in a wider sense, too; not only the way that writers develop their style within their own canons, but the way that writing for theatre evolves more generally over time. The theatre of the late 1580s was very different stylistically from the theatre of the early 1600s. More careful stylometric analyses take this into account by ensuring that comparison texts are from a similar time to the text being tested. With Elizabethan and Jacobean plays this can only be a rough measure, since we are rarely able to date them with great accuracy. Nevertheless, the range of probable dates of the comparison plays spans approximately two decades, and these are being tested largely against a set of plays written in the late 1580s and early 1590s. Without this variable being eliminated or controlled for, what effect this has upon the test results can only be guessed at.

Variable Shakespeare Comparison Text

For all writers except Shakespeare, a single play was used as the comparison text. Shakespeare has been treated differently. The test text for Shakespeare is given in most of the results tables as 'Shakespeare collated reference text'. Against *The Massacre at Paris*, the Shakespeare text was (according to the tables) collated from *1 and 2 Henry IV*; against *Henry V* and *King John*, it was collated from *Richard III, 1 & 2 Henry IV*. Against *3 Henry VI*, the comparison text is given as 'collated history reference texts' (which I understand from the author are the three texts just named). Only in one case, testing against *The Jew of Malta*, was the Shakespeare comparison text a single play: *The Taming of the Shrew*.

No explanation has been given for why particular Shakespeare texts were chosen, or why Shakespeare was treated differently to the other playwrights in this regard. Testing against *Henry V* the Shakespeare reference text ‘was merged from those parts of *Richard III* and *1* and *2 Henry IV*, where MF1W, MF2C, and MF3C were in total congruence,’ and it seems this is describing all those Shakespeare texts described as ‘collated’. In private correspondence, the author gave me further insights into this process, which was essentially to test each of those plays (*Richard III*, *1 Henry IV* and *2 Henry IV*) against the same set of comparison texts in this study e.g. *Tamburlaine 1* for Marlowe, *Soliman and Perseda* for Kyd etc. and a Shakespeare history play that had already proved to have a low Delta score against that text. He then selected only the portions of the test text that came up as ‘Shakespeare’ for all three tests (single word frequency, and 2- and 3-character bigrams).⁹

For *Richard III*, the reference text was *1 Henry IV*.

For *1 Henry IV*, the reference text was *Richard III*.

For *2 Henry IV*, the reference text was *1 Henry IV*.

In this manner, a combined Shakespeare reference text was created from portions of these English history plays by Shakespeare that had been pre-selected for what you might call ‘heightened’ Shakespearean qualities, and specifically the sub-genre ‘English history plays’. This is a very different treatment to any of the other authors, and presumably helped create the situation where all the canonical Shakespeare English history plays could be safely attributed to Shakespeare. Because of the sub-genre bias of this specially created English history-play text, it enabled the situation whereby Marlowe’s only English history play (*Edward II*) tested as primarily Shakespearean. This could occur for the simple reason that the Marlowe comparison text, *Tamburlaine I*, despite having historical elements, contains none of the formal language of

an English court, or indeed the names of noble families (Derby, Percy, Mowbray, Warwick, Kent) with which English history plays are littered. I am not speaking of the speech prefixes, which should have been stripped out of the texts, but of characters being addressed or spoken about by name e.g. ‘Wherefore is Guy of Warwick discontent?’ (*Edward II*).

Considering only the history plays mentioned in the article, ‘Warwick’ occurs throughout *Edward II* and the texts of *2H4* and *Richard III*; in both *Edward II* and Shakespeare’s plays the formulations ‘of Warwick’ and ‘Lord of Warwick’ are common. Both *Edward II* and *1H4* include many references to ‘Mortimer’ including, repeatedly, the phrases ‘of Mortimer’, ‘Lord Mortimer’, ‘name of Mortimer’ and ‘the Mortimers’. ‘Lancaster’ occurs frequently in *Edward II*, *1H4*, *2H4* and *Richard III*. ‘Pembroke’ occurs in the texts of *Edward II* and *Richard III*. Other frequent words that *Edward II* shares with the Shakespeare history plays with which it was compared include ‘Edward’, ‘Edmund’, ‘Plantagenet’, ‘Scots’, ‘King’, ‘Queen’ and ‘Lord’, and this list is hardly exhaustive. These words are not function words so wouldn’t be part of the function word tests, but would undoubtedly impact on the n-gram results which the author favoured. These words are genre markers, not authorship markers. The single-play-comparison tests formulated in ‘Hype or Hoax’ can only mark that that *Tamburlaine I* is free of such words whereas *Edward II* and Shakespeare’s history plays share these words (and the constituent n-grams) in common. It is extremely likely that the affinity discovered here between *Edward II* and the pre-selected and collated Shakespeare comparison text is one of genre, not author.

In the case of *The Jew of Malta*, the Shakespeare text with the lowest Delta values (i.e. the best match) was *The Taming of the Shrew*, a text whose provenance we do not know, as it only appeared in this form in the Folio in 1623, at least thirty years after *The Jew of Malta* was written. In the case of Shakespeare there appears to have been a range of possible texts that the

tests could match against, whereas for Marlowe and Kyd (for example) there was only one. As noted above, *The Jew of Malta* has elements of both tragedy and comedy, which might also be fairly said of *The Taming of the Shrew*, but not *Tamburlaine I*. Though it would need to be confirmed by further tests, it seems most likely that what is being successfully detected by the tests here is again genre.

The Shakespeare history plays were pre-tested in order to ensure that the play to be tested would be compared against only those portions of Shakespeare plays that were not potentially authored by others. For the same reason, *Soliman and Perseda* was chosen as Kyd's reference text over *The Spanish Tragedy*, on the basis that some stylometric tests have claimed to find Shakespeare's hand in *The Spanish Tragedy*. Yet the author might have used a pre-1602 text of *The Spanish Tragedy* – the 1592, 1594 or 1599 quartos – which don't include the suspect passages. It seems that in both the choice and the preparation of comparison texts, attempts to control their purity have significantly skewed the tests. As a result of trying to ensure the Shakespeare reference text used only portions written by Shakespeare — something which can hardly be guaranteed by Delta testing alone and can possibly not be guaranteed by other stylometric methods — the author has essentially invalidated the results. Had he been able to produce a similar 'pre-processed' English history play collation for Marlowe or for Kyd, we would at least be reassured that all the comparison texts had been produced in an identical manner. The problem is, of course, that we only have one Marlowe history play, and that was the play being tested. What this points up is that we simply do not have an appropriate data set for producing the necessary controls that a truly scientific method would require.

The different treatment of Shakespeare, and the creation of a pre-selected comparison text for his history plays, as well as the practice of changing the reference text for him but not for

the other authors, is something that might fairly be described as ‘gaming the system’. The method here is very far from the rigorous sort required in any scientific experiment. When comparison texts for writers other than Shakespeare have been chosen for what might be regarded as spurious reasons, and Shakespeare’s comparison texts have been carefully pre-selected to give him his own plays as well as other plays of the period in the same sub-genre, the method has the appearance of having been (consciously or unconsciously) designed in order to create the results, which might be fairly described as the dominance of Shakespeare and the obliteration of Marlowe.

Reference Text Size and Statistical Significance

No figures have been supplied as to the sizes of comparison texts (either in the paper or in private correspondence, despite requests) and no consideration given as to the effect this might have on the results.

Nor are we given figures as to the differences in Delta scores for each segment; the author with the lowest Delta score is simply awarded that segment, but there is no indication on how low the next lowest Delta score, nor indeed whether the differences between those Delta scores are statistically significant. In this latter issue the study is not unusual; it is rare to see statistical significance discussed in any stylometric study, despite the fact that from a scientific perspective the neglect of this issue would be regarded as a serious oversight. However, by not publishing individual delta scores, but simply taking the lowest and marking that section with the author’s initial, the actual difference in scores (which may be minimal) is obscured.

External Evidence

Though it is probably unnecessary to defend traditional attributions of Marlowe's works in the light of the methodological problems already highlighted, there are two points with respect to the external evidence that are worth making, because of the manner in which they have been ignored or overridden.

The study mentions only in the appendix the attribution of *Lochrine* to Charles Tilney. This is dismissed on the basis that Charles Tilney died in 1586 'and the text has references to Edmund Spenser's *Complaints* (1591) and a poem by Thomas Lodge' (Ilsemann 2018, 28). *Lochrine* was published as 'newly set forth, overseen and corrected, By W.S.' in 1595, meaning that its being updated by someone other than the original author is hardly contentious. The attribution to Charles Tilney can nevertheless not be ignored. It was made by a future Master of the Revels, George Buc, who believed he recognized the play as one (previously called *Estrild*) written by his cousin. Buc had been involved with the original play; he had provided dumb shows for it (unspoken set pieces) which were still in his possession when he made the following note on the title page of his copy:

Char. Tilney wrot[e a]
Tragedy of this matt[r] [which]
hee named Estrild: [which]
I think is this. it was [lost?]
by his death. & now s[ome]
fellow[possibly fellow] hath published [it.]
I made du[m]be shewes for it.
w[hi]ch I yet haue. G. B.

When E.K. Chambers suggested the authors of *Lochrine* might be George Peele and Robert Greene in 1923, this note was still suspected as a forgery. However, Mark Eccles subsequently demonstrated that the handwriting was indeed that of George Buc (Eccles 1933). Regarding the authorship of *Lochrine*, Peter Berek argued that ‘inconsistencies in the narrative, in verse forms, and in the style of stage directions’ clearly delineated where Tilney’s old-fashioned play had been spliced together with newer portions that attempted ‘to imitate the swashbuckling rhetoric and self-assertive hero that won such popularity for *Tamburlaine*’; it had been deliberately fashioned in a Marlovian style (Berek 1980, 34). Will Sharpe also argues that the original author of *Lochrine* was Charles Tilney (Sharpe 2013, 659-60). It does not matter that the text has been altered or updated, nor that we have no other writing by Charles Tilney to assess against the style of *Lochrine*. No matter what has been done with the text since, contemporaneous and personal testimony by the future Master of the Revels, recognising the work of his cousin, a play he had himself contributed to, is surely enough to counteract the claim that *Lochrine* is a ‘sole authored Marlowe play’ (Ilseman 2018, 1). Perhaps it was indeed revised by Marlowe, but there’s no reason to think that the Delta tests applied in this study are sufficiently accurate to be able tell a play deliberately rewritten in the style of *Tamburlaine* from a play written by the author of *Tamburlaine*, since little if any allowance has been made for the fact that genre trumps authorship in such tests, if proper controls have not been set up.

The second point worth making with regards to external evidence relates to *Edward II*. The author claims that the ‘Rolling Delta attribution scheme ... makes it very clear that Shakespeare and Rowley wrote this play’ (14). *Edward II* was registered on 6 July 1593 and the text published in 1594. There is no evidence that Samuel Rowley was writing plays in 1593/4. His first appearance in Henslowe’s diary (as an actor, not a writer) is dated 14 December 1594, a

year and a half after the play was registered, and ten years before his comparison play, *When You See Me You Know Me* would be published. The first record of him being paid as a writer dates from 1602, when he and William Byrd were paid for additions to *Doctor Faustus* (Cerasano 2010). The 1594 quarto of *Edward II* declared on its title page that it was played by the Earl of Pembroke's Men, a company with which Samuel Rowley has not been associated. It seems most unlikely that Rowley was involved in writing plays at this time, and for this company, let alone that he was capable of writing a play of this level of sophistication. The match for Shakespeare can be explained by genre bias, as discussed. The match with Rowley's text cannot be considered as significant. An unsound method is bound to produce some arbitrary results.

Summary

The Delta method has become established as a reliable way of sifting from a large field of possible authors a smaller pool of authors who might be considered for more detailed stylistic testing, but with strong caveats, and only with certain controls in place. In particular, because of its sensitivity to picking up any kind of affinity in the text (and not those purely determined by authorship) it is not deemed a method which is sufficient (on its own) to determine attribution of a text. Similarly, character bigrams and trigrams, though they have proved highly reliable measures of authorship under strictly controlled conditions, cannot be deemed reliable indicators of authorship when conditions are less strictly controlled i.e. when genre, tone, audience, author's background, date of authorship and other variables have not been carefully accounted for. In all stylometric testing, the selection of a comparison corpus is critical, and the same methods must be used for all the authors being tested, in order that the results for each author can be regarded as comparable.

That this study was not conducted under such controlled conditions might partly be excused by the fact that it is almost impossible to create such conditions when testing the playtexts of the early modern era. Rudman has already identified the many challenges of conducting stylometric tests on the Shakespeare canon, but in the case of Marlowe, these issues are significantly exacerbated by the playwright's experiments in genre. Marlowe's cross-genre tendencies within plays, and the fact that every play in his corpus (if we count the *Tamburlaines* as two parts of a whole) belongs to a distinct sub-genre, makes stylometric testing of his corpus immensely difficult. In an environment where 'genre trumps authorship' one would have to test parts of a single play against other parts of the same play (which would run into the issue of reduced sample size) or find some marker of authorship that could reliably be shown to be unaffected by genre in plays of this period. In order to validate such a method, one would need to test, for example, some of Shakespeare's canon against other parts of his canon written in a different genre.

For example, in order to conduct the kind of solo-play comparison that the article proposes, one might test the authorship of *As You Like It* with *King Lear* as the Shakespeare comparison text, against some comedies by other authors. If *As You Like It* came back as Shakespeare's in such conditions, we might (if the experiment were properly controlled in other aspects) be able to say that genre was not trumping authorship. Presuming such a test would also work for numerous other out-of-genre pairings, we might apply this test to paired plays in the Marlowe canon. But computational stylistics does not yet appear to have yielded such a test. All the reasons that Rudman identifies which make Shakespeare so challenging to test stylometrically apply to Marlowe, and the sparseness and genre-leaping qualities of the latter's

acknowledged canon add another layer of unsuitability to this dataset for non-traditional attribution methods.

I agree with the author in one key regard: that ‘one might be tempted to think that Rolling Delta mirrors all sorts of things, but bears no relation to acknowledged facts and to the authorship attributions of secondary literature’ (Ilseman 2018, 16). That is certainly what Rolling Delta has revealed in this study, largely because of the manner in which it has been applied. The article states that in order to guard against this possibility ‘the two *Tamburlaines* and *Lochrine* were dealt with first to show how easily and clearly a real Marlowe text can be recognized’ (ibid).¹⁰ But the thorny question of what constitutes a ‘real’ Marlowe text is neither discussed nor answered; it is simply assumed to be any play with sufficient stylistic matches to the first successful play of the Marlowe canon without room for the playwright to either experiment or develop.

The attractions of computational stylistics as a seemingly ‘scientific’ method of attribution are obvious, but humanities scholars need to be sure that they are actually applying this method scientifically, or at least as scientifically as problematic datasets (like early modern plays) allow. They must beware of becoming so enamoured of a method that they are unable to see its inherent weaknesses and pitfalls. Where a study throws up results which appear to overturn 150 years of traditional scholarship, it is essential to ask ourselves whether these tests have genuinely been conducted under what a scientist would recognize as properly controlled conditions.

In a discipline where new or revised attributions might make one’s reputation, it can be tempting to cast doubt on the authorship of plays rather than the reliability of one’s methods. But humanities scholars undertaking or assessing stylometric studies are doing the discipline a

considerable disservice if they do not take great care over a method's application and control. It is probable that Marlowe's canon is too small, variable and fractured to be suitable for non-traditional attribution studies. Perhaps suitable stylometric methods can be developed that can be shown to be reliable for this kind of data, but they would need extensive testing on other cross-genre texts of the period. There are certainly some interesting questions to ask in relation to Marlowe's attributed canon, but the stylometric tests in 'Christopher Marlowe: Hype and Hoax' have neither asked the right questions nor provided valid answers.

Notes

¹ Readers not familiar with the Delta method need only know that it compares a sample of a (test) text against (comparison) texts by various authors, measuring the frequency of an array of certain words or character combinations. It generates a single numeric variable said to indicate the stylistic similarity of the test text to the comparison text. Those authors for whom the Delta value is lowest are considered to be the ones whose work is stylistically the most similar to the text in question. A more detailed explanation of Delta, and indeed Rolling Delta, is given by the method's inventor, in Burrows, J. (2002). "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." Literary and Linguistic Computing **17**(3): 267-287.

² The probable dates are taken from either the known date or the 'best guess' date in volumes II-V of Wiggins, M. and C. T. Richardson (2015). British Drama, 1533-1642: A Catalogue. Oxford, OUP.

³ Document dated March 28 2018.

⁴ I would contend that *Tamburlaine Part I* is not examined but is used as a reference text.

⁵ All the comparison texts have had their capitalisation removed, thus 'iuno' and 'jove' not 'Iuno' and 'Jove'.

⁶ For example, not atypically, they are assumed not be by Marlowe in Marlowe, C., E. Rasmussen, et al. (1993). Doctor Faustus : A- and B-texts (1604, 1616). Manchester, Manchester University Press..

⁷ There are prose scenes in *Edward II* and *The Massacre at Paris*, and a substantial amount of prose in *The Jew of Malta*.

⁸The fuller and more comedic version in the surviving fragment known as the Collier leaf gives a sense of how much more comedy we might have if a more complete version of the playtext had survived.

⁹ Harmut Ilsemann's pre-test tables for *Richard III*, *1H4* and *2H4* are available at <http://www.shak-stat.engsem.uni-hannover.de/eauthorrichiii.html>, <http://www.shak-stat.engsem.uni-hannover.de/eauthor1h4.html> and <http://www.shak-stat.engsem.uni-hannover.de/eauthor2h4.html>. They are created by the same simple rolling Delta method and thus subject to many of the same problems as the data in 'Hype or Hoax'.

¹⁰ In fact *Tamburlaine I* was never 'dealt with'. It was selected as the single arbiter of Marlowe's style.

References

- Bartels, E. C. and E. J. Smith (2013). Christopher Marlowe in Context. Cambridge, CUP.
- Berek, P. (1980). "Lochrine Revised, Selimus, and Early Responses to Tamburlaine." Research Opportunities in Renaissance Drama XXIII: 33-54.
- Boas, F. S. (1940). Christopher Marlowe : a biographical and critical study. Oxford, Clarendon Press.
- Burrows, J. (2002). "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." Literary and Linguistic Computing 17(3): 267-287.
- Burrows, J. and H. Craig (2017). The Joker in the Pack? Marlowe, Kyd, and the Co-authorship of Henry VI, Part 3. The New Oxford Shakespeare Authorship Companion. G. Taylor and G. Egan. Oxford, OUP: 194-217.
- Cerasano, S. P. (2010). Rowley, Samuel (d. 1624). Oxford Dictionary of National Biography. H. C. G. Matthew and B. Harrison, Oxford: OUP. online ed., ed. Lawrence Goldman, January 2008.
- Craig, H. and J. Burrows (2012). A Collaboration About a Collaboration: The Authorship of King Henry VI, Part Three. Collaborative Research in the Digital Humanities. M. Deegan and W. McCarty, Ashgate: 27-65.
- Eccles, M. (1933). Sir George Buc, Master of the Revels. Thomas Lodge and Other Elizabethans. C. J. Sisson. Cambridge, MA, Harvard UP: 409-506.
- Erne, L. (2003). Shakespeare as literary dramatist. Cambridge, UK ; New York, Cambridge University Press.
- Greg, W. W., Ed. (1950). Marlowe's Doctor Faustus 1604-1616: Parallel Texts. Oxford, Clarendon.

-
- Grieve, J. (2007). "Quantitative Authorship Attribution: An Evaluation of Techniques." Literary and Linguistic Computing **22**(3): 251-270.
- Gurr, A. (2017). Shakespeare's Workplace: Essays on Shakespearean Theatre. Cambridge, CUP.
- Hoover, D. L. (2007). "Corpus stylistics, stylometry, and the style of Henry James." Style: 174-203.
- Ilseman, H. (2018). "Christopher Marlowe: Hype and Hoax." Digital Scholarship in the Humanities: fqy001-fqy001.
- Last, S. (2000). "Marlowe's Literary Double Agency: Doctor Faustus as a Subversive Comedy of Error." Renaissance and Reformation / Renaissance et Réforme **XXIV**(1).
- Luyckx, K. and W. Daelemans (2011). "The effect of author set size and data size in authorship attribution." Literary and Linguistic Computing **26**(1): 35-55.
- Marlowe, C., E. Rasmussen, et al. (1993). Doctor Faustus : A- and B-texts (1604, 1616). Manchester, Manchester University Press.
- Nicholl, C. (2002). The Reckoning : the Murder of Christopher Marlowe. London, Vintage.
- Rudman, J. (2016). "Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats." Journal of Early Modern Studies **5**: 307-328.
- Scott, S. K. (2010). The Jew of Malta and the Development of City Comedy: 'The Mean Passage of History'. Christopher Marlowe the Craftsman: Lives, Stage and Page. S. K. Scott and M. L. Stapleton. Farnham, Ashgate: 91-108.
- Sharpe, W. (2013). Authorship and Attribution. William Shakespeare and Others: Collaborative Plays. J. Bate and E. Rasmussen. Basingstoke, Palgrave Macmillan: 643-747.
- Stamatatos, E. (2009). "A survey of modern authorship attribution methods." Journal of the American Society of Information Science and Technology **60**(3): 538-56.

Tarlinskaja, M. (2014). Shakespeare and the Versification of English Drama, 1561-1642.

Farnham, Ashgate.

Vickers, B. (2004). The Troublesome Raigne, George Peele, and the Date of King John. Words

That Count: Essays on Early Modern Authorship in Honor of MacDonald P. Jackson. B.

Boyd. Newark, University of Delaware Press: 78-116.

Wiggins, M. and C. T. Richardson (2015). British Drama, 1533-1642: A Catalogue. Oxford,

OUP.