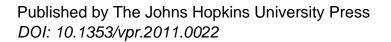
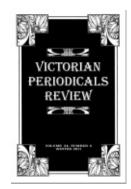


## Did Dickens write "Temperate Temperance"?: (An Attempt to Identify Authorship of an Anonymous Article in All the Year Round)

John Drew Hugh Craig

Victorian Periodicals Review, Volume 44, Number 3, Fall 2011, pp. 267-290 (Article)





For additional information about this article

http://muse.jhu.edu/journals/vpr/summary/v044/44.3.drew.html

# Did Dickens Write "Temperate Temperance"? (An Attempt to Identify Authorship of an Anonymous Article in All the Year Round)

JOHN DREW AND HUGH CRAIG

This article is the result of a collaborative exercise carried out by the Dickens Journals Online (DJO) project and the Centre for Literary and Linguistic Computing (CLLC) at the University of Newcastle, Australia.<sup>1</sup> It presents the findings of an attempt to establish authorship of a short article published anonymously on 18 April 1863 in the weekly magazine *All the Year Round* under Charles Dickens's editorship, using computational stylistics in tandem with internal clues (in themselves far from conclusive) as to author. The reporting of the results forms part of a series of new attributions to be presented by DJO following its public launch in 2012, but merits, we hope, more elaborate discussion because the findings challenge an existing form of attribution, as well as offering a further demonstration of how the so-called "Burrows method" of establishing authorship can be configured for work with Victorian periodicals.

To deal with the existing form of attribution first: a clear distinction can be made between the way in which articles in *Household Words* under Dickens's editorship (1850–59) are attributed, and the operation of this process with *All the Year Round* (1859–70; hereafter *AYR*). With the former, modern readers have to look no further than the definitive table of contents and contributor list published by Anne Lohrli in 1973, which derives its authority from the original *Household Words* "Office Book" of the journal, held in the Morris L. Parrish Collection of Victorian Novelists, at the University of Princeton Library. The Office Book was scrupulously compiled by the journal's sub-editor, W. H. Wills, and recorded, in addi-

tion to each article's title, its author(s), date/place of insertion, the price paid, date of payment, and any additional memoranda Wills deemed necessary. No equivalent ledger for AYR has survived, though it is presumed one existed. Thus, while there is a Descriptive Index and Contributor List to AYR, compiled by E. A. Oppenlander and published in 1984, it is not a transcript of a complete record, but rather an incomplete compilation of attributions based principally on external evidence, such as the subsequent republication of individual items under an author's name.<sup>3</sup> Authorship of something over two thirds of AYR articles is still unidentified in this Index. Thus, if an article is among the unascribed items in Oppenlander, the way is usually left clear for a positive attribution based on internal and/ or external evidence. However, where the suggested author is Dickens himself, there is a significant obstacle, in the guise of a rather unusual form of negative (or non-)attribution, for which an early Dickensian scholar called Frederic G. Kitton is responsible.<sup>4</sup>

Sometime before March 1900, as he reports in *The Minor Writing of Charles Dickens*, Kitton was given access to something he calls "an 'office' set" of AYR, which is described on two occasions in his book, as follows:

[B]y great good fortune, I discovered in the possession of Mr. W. H. Howe an "office set" of [the] journal, in which had been inscribed against each article, etc., the name of the author thereof—satisfactory and conclusive proof as to its origin.

I have fortunately been able to examine a complete "office" set of *All the Year Round*, in which each article has appended the name of the author, written by a member of the printing staff, so that any doubt that may have risen in respect of authorship has thus been satisfactorily disposed of.<sup>5</sup>

However satisfactory this all may have seemed to Kitton at the time, it is certainly frustrating now that he should have omitted to record and publish anything other than a listing of those 126 items (including instalments of serial fiction) against which we are to presume he found Dickens's name.<sup>6</sup> The identity of W. H. Howe is uncertain,<sup>7</sup> the whereabouts of this "office set" is unknown,<sup>8</sup> and an ideal opportunity to create a comprehensive author index for *AYR* has been lost.

Given that "Temperate Temperance" is *not* among the Dickens titles listed by Kitton, we must assume that he did not read Dickens's name appended or inscribed to or against the article by a member of the printing staff. However, to be confident in this assumption, and then to infer that as a result, Dickens did not write the article, poses a number of questions about Kitton's "methodology." And if it seems odd to apply the term to the proceedings of an amateur *aficionado* (Kitton was an artist by profes-

sion) working at the turn of the last century, it perhaps reinforces doubts one might have about how Kitton in fact approached this task, the results of which unfortunately are not now verifiable. For example, we have no way of knowing how accurate a record this "office set" may have been in itself, before Kitton accessed it. He clearly regards it as unimpeachable, but if W. H. Wills were already keeping a detailed Office Book for AYR like the Household Words ledger described above, and the office set was kept by the printing staff, it is not obvious what its primary purpose would be. We do know that thousands of copies of the journal were sold from its offices in Wellington Street,9 of which the ground floor was a shop. It is possible the office set was kept and compiled by those manning the shop, to answer enquiries from customers about authorship, but if so, its degree of accuracy in comparison with the Office Book would perhaps be secondary, particularly if based on verbal communication. It was clearly sufficiently detailed to distinguish between the authors of separate short paragraphs of jointlyauthored items such as the "Occasional Register" in AYR's first issue, but can we be sure that this degree of granularity was maintained throughout the 24 bi-annual volumes published under Dickens's editorship? If an article were based on a recent publication and loosely "crammed" by a member of the writing staff, would its compiler's name be recorded as author, or that of the author of the book being reviewed? When the writer of a serial novel was given in print at the start of each instalment as "by the author of," did the printers bother on every occasion to inscribe the authors' name against it, and if not, what other omissions may there have been? Questions such as these, more or less unanswerable, start to crowd in. It is not that Kitton's willingness to record any attributions to Dickens that he saw is in dispute: more that his materials may not have been definitive, and his method (potentially) cursory.

For example, in working on his listing of Dickens's contributions to *Household Words*, unaware of the existence of its Office Book (which did not surface until 1903<sup>10</sup>), Kitton undertook, "a most careful perusal of the entire set of nineteen volumes [which] became imperative before it was possible to draw up a list of the papers for which 'the Chief' was ostensibly responsible."<sup>11</sup> This list was then checked against similar research and "endorsed" by Dickens's eldest son, Charley, shortly before his death. While this listing was based on internal evidence only, and is indeed not wholly accurate, it seems at least to have involved a full reading of the contents of *Household Words*. In seizing upon the "Office Set" as a quick way of arriving at his *AYR* listing, based on external evidence, Kitton may not have fully assessed its quality, nor, as his account implies, actually sat down to read the articles themselves.

The process of slowly reading through the contents of AYR, in order to write synoptic introductions to the range of contents and editorial issues

presented by each bi-annual volume, combined with knowledge of Dickens's working practices as an editor, leads us to suspect a greater degree of involvement by Dickens in writing for the journal than Kitton's listing, and in turn Oppenlander's *Index*, allows for. We can reliably quantify Dickens's appearance rate in *Household Words* as a solo or joint author at about 57%; the comparable figure for AYR, at present, is only 22%. <sup>12</sup> Even allowing for factors inevitably drawing Dickens's attention away from his journal during the 1860s, the AYR figure is surprisingly low. A similar situation applied to the canon of Dickens's known contributions to The Daily News during the brief period of his editorial involvement in 1846, and recent scrutiny of the paper's contents has revealed five articles, one of them approaching 4,000 words in length - which are now claimed as Dickens's own. 13 Clearly, while Kitton's listing can neither be ignored nor overturned, its unverifiable nature, uncertain manner of compilation and double-negative formula so far as Dickens is concerned ("because an AYR article is not included in the listing, it therefore can't be by Dickens") seem to make it worthwhile querying, when other factors start to point in Dickens's direction as a potential author.

To query it effectively, one would either have to produce new external evidence that demonstrated Kitton's listing to be in error, or offer more convincing forms of internal evidence than have hitherto been available, and this is where CLLC and the "Burrows method" come into play. In their 2006 article in *Victorian Periodicals Review*, Jordan, Craig and Alexis offered to demonstrate how the methodology developed in the Centre "can be used to test and verify the kind of hunches scholar, and . . . general readers . . . can develop about the authorship of nineteenth-century periodical articles." <sup>14</sup> By way of a quick resumé, we quote from the introduction to their paper, noting also various further publications since its appearance, detailing the deployment of the same methodology in supporting new attributions:

In the early days of computer analysis of texts, Professor [John] Burrows made the discovery that the incidence of the very common words of English, the "function" words, varies significantly between texts by different authors, while remaining comparatively constant within a single author's work. During the past 25 years he and his colleagues . . . have been devising and refining the statistical procedures that can most effectively isolate these distinctive usages, and they now have a suite of procedures and tests that can be used to identify the authorial "signature" embedded within a text or group of texts. / [O]ur initial application of the "Burrows method" to nineteenth-century periodical literature suggests that this technique can open a new avenue within the ongoing project of attribution. . . . Most scholars using [anonymously-published] periodical literature develop hunches about who was or was not the author of

particular pieces. . . . The "Burrows method" provides a way of testing such hypotheses, and, although unable to "prove" authorship n the way an editor's letter or a notation in a diary does, it can transform a hunch into a statistically assessable possibility. <sup>15</sup>

With the above in mind, a hunch that the author of "Temperate Temperance" sounded very much like Dickens, could be tested in considerable detail.

Before presenting the analysis, it should be stressed why the article stood out in the first place from hundreds of other anonymous articles in AYR. Firstly, at only just over 1,300 words (2½ columns), it is much shorter than the average length of piece that, as we can see from editorial correspondence and the Household Words Office Book, was commissioned from outside contributors, and paid for at so many pounds, shillings, and pence. In other words, it has the appearance of an in-house "filler." Second, it is written throughout, and emphatically so, in the editorial "We." This point, it should be said, has to be balanced carefully. Many writers other than Dickens write thus, "editorially," in Household Words and All the Year Round, both for humorous and comic effect; Dickens, equally, frequently writes in his journals in the first person singular (as in many of his "Uncommercial Traveller" papers). On the other hand, rhetorically, the editorial weight of this article falls squarely throughout, on a set of opinions about the injustice of excluding beer from dining halls for the working classes, and the absurdity of middle-class patronage of the "working man" that Dickens expresses vociferously at this period in a range of journalistic writings that are known to be his. 16 It begs the question, why would Dickens need to commission a short piece of editorial on this topic from anyone else? Space precludes republication of the entire piece here, but some samples from its 7 paragraphs will help readers familiar with Dickens's querulous and polemical journalistic style<sup>17</sup> to understand why it was felt to be worth testing thoroughly:

WE want to know, and we always have wanted to know, why the English workman is to be patronised? Why are his dwelling-place, his house-keeping arrangements, the organisation of his cellar, and his larder . . . why are all these things regarded as the business of everybody except himself? Why is his beer to be a question agitating the minds of society, more than our sherry? Why is his visit to the gallery of the theatre, a more suspicious proceeding than our visit to the stalls? Why is his perusal of his penny newspaper so aggravating to the philanthropical world, that it longs to snatch it out of his hand and substitute a number of the Band of Hope Review?

. . . Heaven knows, the working classes, and especially the lowest working classes, want a helping hand sorely enough. No one who is at all familiar with

a poor neighbourhood can doubt that. But you must help them judiciously. You must look at things with their eyes, a little; you must not always expect them to see with your eyes. . . . Your standard is too high. The transition from the Whitechapel cellar to the comfortable rooms in the model-house, is too violent; the habits which the cellar involved would have to be abandoned; a great effort would have to be made; and to abandon habits and make great efforts is hard work even for clever, good, and educated people.

. . . It is proposed just now, as everybody knows, to establish, in different poor neighbourhoods, certain great dining-halls and kitchens for the use of poor people, on the plan of those establishments which have been highly successful in Glasgow and Manchester. The plan is a good one, and we wish it every success on certain conditions. The poor man who attends one of these eating-houses must be treated as the rich man is treated who goes to a tavern. The thing must not be made a favour of. The custom of the diner-out is to be solicited as a thing on which the prosperity of the establishment depends. The officials, cooks, and all persons who are paid to be the servants of the man who dines, are to behave respectfully to him, as hired servants should; he is not to be patronised, or ordered about, or read to, or made speeches at, or in any respect used less respectfully than he would be in a beef and pudding shop, or other house of entertainment. Above all, he is to be jolly, he is to enjoy himself, he is to have his beer to drink; while, if he show any sign of being drunk or disorderly, he is to be turned out, just as I should be ejected from a club, or turned out of the Wellington or the Albion Tavern this very day, if I got drunk there.

There must be none of that Sunday-school mawkishness, which too much pervades our dealings with the lower classes; and we must get it into our heads—which seems harder to do than many people would imagine—that the working man is neither a felon, nor necessarily a drunkard, nor a very little child. . . . Encourage him to take an interest in the success of the undertaking, and, above all things, be very sure that it pays, and pays well, so that the scheme is worth going into without any philanthropic flourishes at all. He is already flourished to death, and he hates to be flourished to, or flourished about.

Who has not been outraged by observing that cheerfully patronising mode of dealing with poor people which is in vogue at our soup-kitchens and other depots of alms? There is a particular manner of looking at the soup through a gold double eye-glass, or of tasting it, and saying, "Monstrous good—monstrous good indeed; why, I should like to dine off it myself!" which is more than flesh and blood can bear.

The cause of temperance is not promoted by any intemperate measures. It is intemperate conduct to assert that fermented liquors ought not to be drunk at all, because, when taken in excess, they do harm. Wine, and beer, and spirits, have their place in the world.... There is no intrinsic harm in beer; far from it;

and so, by raving against it, we take up a line of argument from which we may be beaten quite easily by any person who has the simplest power of reasoning. The real temperance cause is injured by intemperate advocacy; and an argument which we cannot honestly sustain is injurious to the cause it is enlisted to support. Suppose you forbid the introduction of beer into one of these institutions, and you are asked your reason for doing so, what is your answer? That you are afraid of drunkenness. There is some danger in the introduction of gas into a building. You don't exclude it; but you place it under certain restrictions, and use certain precautions to prevent explosions. Why don't you do so with beer? (AYR IX, 188–9; 18 April 1863).

The publication of this article between known Dickens contributions such as "The Poor Man and his Beer" (AYR I, 30 April 1859) and the signed "Uncommercial Traveller" paper later retitled "The Boiled Beef of New England" (15 August 1863; indexed first in AYR as "At a Cooking Depôt for the Working Classes") makes it easy to see as part of a series. Dickens went on to inspect one of the new London dining halls for working men in early July 1863, at the invitation of the Secretary of the Association that founded them. Significantly, on the first page of "The Poor Man and His Beer" Dickens had actually deployed the phrase "temperate temperance" to set up the narrative of his article, thus: "And on the day fixed, I, the Dreary one, accompanied by Philosewers, went down Nor'-West per railway in search of temperate temperance."18 The phrase seems, then, to be a Dickens coinage, although it is often incorrectly attributed to Mark Twain, who used it in a notebook entry of 1896. 19 However, Dickens was clearly putting the phrase in currency and making this point nearly forty years earlier. He, or possibly Wills, must have recalled it from the earlier article sufficiently to use it again as the title for this short editorial opinion piece. The curious parallel drawn at the end of the article, comparing the introduction of beer into a club to the introduction of gas into a building, with the argument that both should be encouraged, but under suitable restrictions and controls, also offers a potential connection to Dickens. The analogy is one that might have occurred more naturally to him than another writer, since his ongoing public reading tours all involved bringing a rig of gas lamps into buildings to light his performance. Malcolm Andrews, in his definitive study of Dickens's public readings, comments that these "portable lighting arrangements were potentially hazardous" and gives four examples of accidents, or near-accidents, involving them.<sup>20</sup> Dickens was clearly perfectly placed to understand how "there is some danger in the introduction of gas into a building," and to understand its benefits.

Of course, setting aside these small (but telling?) clues as inconclusive, we must concede that a "filler" article espousing in this fashion a set of

views Dickens presents in other journalistic writings could simply have been written to his instructions by another member of the *AYR* staff or inner circle of regular contributors. For this further reason, testing it using "Burrows" methods, which rely on establishing "signature" profiles for a range of potential authors of a particular anonymous text, seemed particularly appropriate. The other authors profiled in addition to Dickens were therefore selected from a group of likely suspects: Edmund Ollier (a minor poet and jobbing journalist who was to co-author an article on "Working Men's Clubs" with Dickens a few months later),<sup>21</sup> sub-editor W. H. Wills (who often wrote/compiled articles at Dickens's suggestion), Wilkie Collins (contracted to the *AYR* writing staff for a two-year period from July 1860 to July 1862), and Henry Morley (often used as a conduit for Dickens's views on social and industrial matters). An outline of the specific methodology employed, together with results, now follows.

Nine articles by Dickens ("Author B"), nine each by Wilkie Collins ("Author D") and Henry Morley ("Author C"), eight by Edmund Ollier ("A"), and eleven by William Henry Wills ("E"), were selected, proof-read in digital form, and prepared for counting. A series of one-on-one tests, which have the advantage of marking differentiation in a relatively crisp fashion, was run, opposing Dickens to each of the four authors in turn, using a single Dickens training and test set, and a single training and test set for each of the other authors. The texts were presented as by "Author A", "Author B", "Author C" etc., so that CLLC staff were unaware of their origin until after the results were analysed. Three separate tests were carried out on each of the four pairings: a rare-words test, a mid-range words test (as described below), and an additional function-words test. In ten of the twelve different test results "Temperate Temperance" clustered significantly with the Dickens segments rather than with the segments of the other author. Two examples of test results are given in Figs. 1 and 2, one showing an "Iota" test, and the other a "Zeta."

In fig. 1, the training sets have no instances of words in the other author's list, by design, so their segments appear along the two axes. The test sets have fewer of their own author's rare words, and more of the other author's rare words, but nevertheless separate into clusters. The Morley test segments all have more of the Morley words than any Dickens test segment, and "Temperate Temperance" has more of the Dickens words, and fewer of the Morley words, than any Morley test segment. It also has more Dickens words than any Dickens test segment and a score on Morley words that places it in the middle of the Dickens test segments. The test separates all the test segments into clusters which align with authorship, and indicates that the vocabulary of "Temperate Temperance" is much more like a Dickens vocabulary than a Morley one. In fig. 2, the result of a "Zeta" mid-range vocabulary test, the training sets are off the axes this time, as

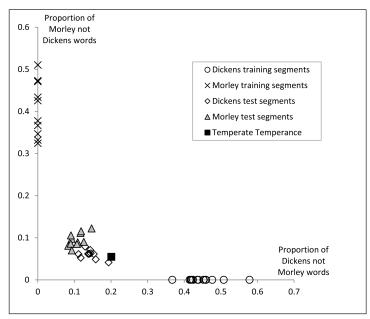


Figure 1: Iota<sup>22</sup> test: Dickens versus Morley

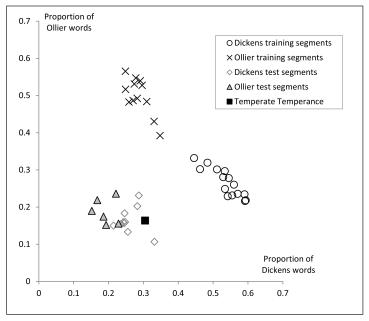


Figure 2: Zeta<sup>23</sup> test: Dickens versus Ollier

the words chosen appear relatively frequently in one set rather than in the other, as opposed to appearing in one set and never in the other. The test sets have lower scores than the training sets, as expected; the marker words were chosen precisely because they differentiated the two training sets, and the test sets do not have that advantage. There is a tendency for the two clusters of test segments to separate, but there is also some overlap. Again, "Temperate Temperance" appears clearly in the Dickens cluster.

Overall, the test segments were generally well differentiated throughout this set of twelve one-on-one trials, though in the Iota (rare-words) and PCA (function-words) experiments opposing Dickens and W. H. Wills there was some overlapping of the clusters, and therefore only fair to poor differentiation. This however, represents a level of "noise" or interference to be expected: authors do vary from themselves, and overlap occasionally, especially in shortish samples. Nothing here pointed to any evidence contrary to Dickens's authorship of "Temperate Temperance," but it was felt that if Kitton's non-attribution were to be challenged effectively, we would need to approach more closely the goal of an analysis which shows if a given sample is by a particular author or not, rather than one which shows whether it is more like one author than a single other. To this end, a series of challenging "Author B against the world" tests was carried out, in which, as there is more variation to be taken into account, clusters may overlap with greater readiness, but at the same time, any consistently positive readings form stronger positive indicators of authorship.

To confirm the reliability of the test procedures we needed an estimate of how well they worked with freshly introduced anonymous samples, so we used a "nine-fold validation" scheme.<sup>24</sup> This means withdrawing some of the articles each time and re-introducing them as if they were anonymous. We did this nine times, with-holding a different Dickens article and a different four or five from the non-Dickens set at each stage. For each iteration, there was thus a different "training" set of samples used to form a Dickens and a non-Dickens profile, and a different "test" set of samples used to estimate how well the test performs. By the end, each Dickens article and each article by the four other writers had served once in the test set, and only once.

"Temperate Temperance" is short, as we have said: 1,336 words in total. Classification tends be less reliable with shorter samples, so to give a comfortable margin of error we divided all the articles into 1,000-word segments, adding any residue to the last segment. An "Iota" test was first made, separating all the words that occurred in the Dickens training set and not in the non-Dickens one, and then all the words that occurred in the non-Dickens training set and not in the Dickens one. We then calculated the percentage that these words formed of all the different words in all the segments. Fig. 3 shows the results for the first round of the nine valida-

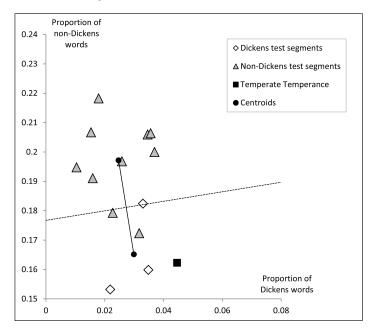


Figure 3: Iota test with 3 Dickens test segments, 10 non-Dickens test segments, and "Temperate Temperance"

tions, using as test articles "The Refreshment of Travellers" (see Appendix) and five from the non-Dickens collection in the corpus.

The horizontal axis arrays the segments according to the proportion of different words formed by the Dickens list in each case. The vertical axis does the same for the non-Dickens words. We averaged the counts for the two sets, the Dickens test set and the non-Dickens one, and these are marked as "centroids." Remembering that the two lists of words have been chosen with no reference to the test sets, we are interested to see if the Dickens test segments (the grey triangles) are generally higher on Dickens words and lower on non-Dickens words than the non-Dickens test segments (the diamonds). The centroids indicate that this is so: the Dickens centroid is to the right and lower than the non-Dickens one.

We made a division between a "Dickens" and "non-Dickens" area of the chart by joining the two centroids (the solid black line) and then finding the perpendicular bisector of this line (the dotted line). Above the dotted line we regard as "non-Dickens" territory, below it as a "Dickens" area. The division leaves one Dickens segment above the line and two non-Dickens segments below it. The system we have set up has thus misclassified these segments. This gives us an idea how often the system could be expected

to fail in the case of a newly introduced segment of comparable text type. Overall, ten of the thirteen segments are correctly assigned by authorship. Clearly there are broad differences between Dickens's vocabulary and that of his four peers, measured in these terms, and they make it possible for the test to do a reasonably good job in distinguishing freshly-introduced, smallish sections of prose as by Dickens or not. But we can expect mistakes to be made in the odd case, unexceptionably, given the size of the samples, and the variation within them, which can mean that Dickens sometimes overlaps with his peers in style, and vice versa. Finally, we can find the proportion of Dickens words and non-Dickens work in "Temperate Temperance" and plot the outcome, shown by the asterisk symbol. In this case it is firmly in Dickens territory.

The results for all nine validation rounds are tabulated in Table 1, below. The number of segments in the test sets varies because the articles are divided into 1,000-word segments and the articles are uneven in length. The centroid for the Dickens test group is always higher on Dickens words and lower on non-Dickens words than the centroid for the non-Dickens test group, which is a broad indication that, even though the system is "blind" to the provenance of the test segments, it can tell the difference between a typical "Dickens" vocabulary and a typical "non-Dickens" one. The success rates are another way of estimating this. They vary between the rounds, with a particularly poor result from round 8, where one of the three Dickens test segments and four of the eight non-Dickens test segments were mis-classified. At the other end of the scale all the test segments in Round 6 were correctly classified. The rest of the rounds fell somewhere in between. Overall, 21 out of 25 Dickens test segments and 70 out of 85 non-Dickens segments were correctly classified, a success rate of 84% and 82% respectively.

"Temperate Temperance" itself was always placed in the Dickens area of the chart. It is thus rated as, if anything, more characteristically Dickens than the general run of Dickens test segments. The results for the latter lead us to expect that a genuine Dickens segment will fall on the other, wrong, side of the line once in eight or nine trials; "Temperate Temperance" did better than this, being assessed as Dickens nine times out of nine. It comes through the Iota test as an unexceptionably Dickensian sample (though care must be taken, as ever, in applying this Protean epithet, even in a narrowly stylistic sense).

As a check on this result we can make a second round of "Zeta" test classifications with the same samples in the same sets, but this time excluding all the words used for the Iota test. From the rest, the less rare words, we choose the 500 that appear more regularly in the Dickens training segments than in the non-Dickens training segments, and the 500 that appear more regularly in the non-Dickens training segments than in the Dickens

Table 1. Nine-fold validation: lota test results

Validation round	Dickens test centroid higher on Dickens words	Dickens test centroid lower on non-Dickens words	Dickens test N	Dickens test correct	Non-Dickens test N	Non-Dickens test correct	"Temperate Temperance" classified as
П 2	Yes Yes	Yes Yes	$\omega$ $\omega$	3 2	0I 6	8 2	Dickens Dickens
€ 4	Yes Yes	Yes Yes	4 4	7 7	5 15	4 12	Dickens Dickens
6 5	Yes Yes	Yes Yes	€ 4	<b>4</b>	9 14	7 14	Dickens Dickens
<b>⊳</b> ∞	Yes Yes	Yes Yes	н «С	1 2	01	o 4	Dickens Dickens
9 Total	Yes	Yes	4 4 5	4 21	8 8 8 5	7 70	Dickens
Percentage	ıtage			84%		82%	

ones. As before we then find the proportion of the total of different words in each segment formed by these word sets and plot all these proportions in a chart. Fig. 4 shows results for the first round of validation, with the segments divided between test and training sets as in fig. 1.

The Dickens centroid is higher on Dickens words and lower on non-Dickens words than the non-Dickens centroid. All the non-Dickens test segments this time are on the correct side of the dotted line. One of the Dickens test segments (as it happens, the same one as in the Iota test, the third segment of "Refreshment of Travellers") is just across the line into the non-Dickens area. The "Temperate Temperance" datapoint is on the Dickens side, if only by a whisker.

Table 2 shows the Zeta Variant results for all nine rounds.

This time there is one round, Round 7, where the Dickens test centroid is lower on Dickens words than the non-Dickens test centroid, indicating a failure in this case to detect a "Dickens signal" in the test segments. Overall, though, the success rate for the classification of individual segments is very similar to that achieved in the Iota test: the same percentage as for the Dickens test segments (84%), and fractionally lower for the non-Dickens test segments (81%). "Temperate Temperance" is classified as Dickens seven times out of nine, just below the expectation for a regular Dickens segment this time. In this separate sampling of the vocabulary of the various segments, the system again has a good general success rate in discriminating Dickens from non-Dickens segments, even though they are presented to the analysis as anonymous, and "Temperate Temperance" once again is clearly assigned to Dickens, though in a majority of rounds this time rather than every time (7 of 9).

We conclude therefore that the vocabulary of "Temperate Temperance" fits the Dickens profile established by these methods. In the Iota tests, the targeted article tends to have a similar share of the unusual Dickens words and of the unusual non-Dickens words to the genuine Dickens samples treated in the same way. In the Zeta test, likewise, we find that it has a Dickens-like proportion of favourite Dickens words and of the words Dickens tends to avoid. The system certainly does not overcome all the variation in the samples. There are times when a known Dickens segment can seem more like the non-Dickens group generally than like the Dickens group, and vice versa. Nevertheless, there is no doubt that the system has found broadly consistent patterns that distinguish Dickens's writing from that of the other four authors in our corpus, and that on the basis of these patterns it generally classifies "Temperate Temperance" with the Dickens samples. "Temperate Temperance" is as Dickensian in style as "The Poor Man and his Beer" or "The Boiled Beef of New England" on these measures, and as if not more so than "Refreshments for Travellers." Similar methods, relying on word frequency data, have been used previously for

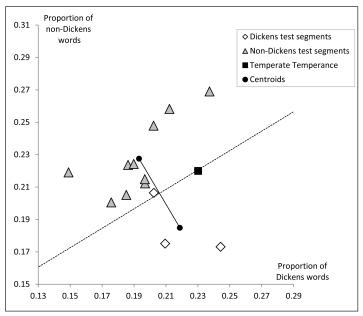


Figure 4: Zeta test with 3 Dickens test segments, 10 non-Dickens test segments, and "Temperate Temperance"

widely-accepted attributions in poetry, drama and prose, the latter both fictional and not.<sup>25</sup> Aside from this success elsewhere, the best guarantee that the present results are reliable in relation to "Temperate Temperance" is in the validation by closely comparable test segments and the fact that two different approaches, one using rare words and one using commoner words, give broadly the same result. A degree of error in the test results remains, and thus a degree of caution is appropriate, but there is no doubt that the methods detect a markedly Dickensian style and that "Temperate Temperance" fits within it.

The results of this exercise open up interesting avenues of enquiry, whether the impartial observer chooses to accept the attribution or not. Clearly, even with the strong positive index of the nine-fold validation supporting the case for Dickens, this is ultimately only a relative measure, and it remains possible, if quite unlikely, that another author could be found whose "signature" after profiling, is still more like that of the author of "Temperate Temperance" than is Dickens's. So, on the one hand, if we entertain the view (which Kitton's non-attribution, such as it is, supports) that Dickens was not the author of this piece, we must conclude that *All the Year Round* had therefore found or trained a writer who could write so like its illustrious Conductor as to be more or less indistinguishable. It

Table 2. Nine-fold variation: Zeta Variant results

	"Temperate Temperance" classified as	Dickens	Dickens	Dickens	Dickens	non-Dickens	Dickens	non-Dickens	Dickens	Dickens		
	Non-Dickens test correct	10	9	3	12	9	13	10	4	5	69	81%
	Non-Dickens test N	OI	9	2	15	6	14	OI	~	8	85	
	Dickens test correct	7	3	1	7	3	3	1	7	4	21	84%
	Dickens test N	<i>c</i>	$\kappa$	4	7	8	4	I	Е	4	2.5	
icta variant iesuits	Dickens test centroid lower on non-Dickens words	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
radic 2. ivinction variation: zeta variant results	Dickens test centroid higher on Dickens words	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes		age
14DIC 2. 14III	Validation round	I	4	Е	4	5	9	_	8	6	Total	Percentage

was a common complaint, after all, made by both readers of, and contributors to *AYR* and its predecessor *Household Words*, that the house style was "Dickensy" (the word is Gaskell's<sup>26</sup>), and our findings would therefore illustrate to a remarkable extent how well Dickens succeeded in imposing his voice on those of his staff, and they, in turn, in ventriloquising their "Chief."

Yet, those who complained that Dickens both as an editor and author of fiction was raising up a tribe of slavish imitators to the detriment of the nation's literature, often stressed that their specific complaint related to the paleness of the imitation—in other words, with its failure, as much as its success. James Fitzjames Stephen, perhaps Dickens's most virulent detractor in reviewing circles in the 1850s and 60s,<sup>27</sup> justified repeated attacks in the *Saturday Review* on the basis that "Dickens's influence over some departments of literature has been so marked, and his imitators are so numerous" (25 December 1858), but concedes that while Dickens's imitators attempt to copy his style and mannerisms,

the difference is merely that his observation is much truer and juster, and his fancy is original and endless. The wording is the least thing with him, but it is everything with them . . . (23 February 1861)

#### and that

when Mr Dickens writes at his best, it is surprising how very unlike him are all his imitators, and how subtle and numerous are the touches by which he maintains his superiority. (12 December 1863)<sup>28</sup>

As early as 1837, the nickname of "The Inimitable" had been bestowed on Dickens, who occasionally adopted this as a signature-substitute in letters,29 so this question of stylistic (in)imitability which our investigation throws into such clear relief, can be seen as central both to Dickens studies and the culture of mid-Victorian periodicals. Such matters are further nuanced by the common practice of collaborative authorship, of which Dickens and his weekly journals, as Harry Stone has shown, were pioneering exemplars.<sup>30</sup> Here, it is not simply a question of one writer attempting, successfully or otherwise, to mimic another, but of writers approximating their natural styles so as to compose joint works, seamlessly joining material by two or more contributors. That this could result in an unequal partnership between the collaborators, whose contributions might encode the tensions between them, is something Lillian Nayder has thoroughly investigated, but in terms of the resulting composition, the process of stylistic approximation between individuals is still relatively unexamined.<sup>31</sup> In mooting ideas for founding a weekly journal in late 1849, Dickens had

wanted to establish the persona of "a certain SHADOW" to "bind" the contents together, "and to get a character established as it were which any of the writers may maintain without difficulty," a desirable goal for an editor who had discovered, exhausted, a few days before launching *Master Humphrey's Clock* as a weekly miscellany, that it "has [become] quite clear . . . that I must write it *all*" (*Letters*, 2: 46). The production of composite bi-authored or multiply authored articles was, Stone suggests, a salient feature of the writing economy of Dickens's journals, and one which, if "Temperate Temperance" were the result of it, would naturally challenge the configuration of the "Burrows method" as we have deployed it (pp. 43–45).

If, on the other hand, one concludes, as the present authors do, that "Temperate Temperance" is in fact a perfectly genuine piece of vintage Dickens editorial, the implications are equally interesting. To begin with, Kitton's longstanding negative attribution can no longer be regarded as definitive. While it cannot, as earlier stressed, be wholly overturned, it can now be shown to present a probable margin of error. No floodgate of new Dickens attributions is to be expected; part of the point of the journals was that they were "Conducted," and not wholly composed, by their illustrious editor. But some, perhaps one or two papers in each bi-annual volume, might realistically be claimed as Dickens's work, if the systems we have developed give a strong positive reading, in tandem with suitable internal clues. Certainly, a programme of collaborative work proposed between CLLC and DJO looks likely, if the funding climate mellows, to turn up some other interesting discoveries. Not all, by any means, will involve additions to the canon of Dickens known journalism, because a further, and equally stimulating, implication of our investigation is that the "Burrows method" can be satisfactorily configured for general work with the varied contents of All the Year Round. The way is open for testing a whole range of possible authors, currently canonical or otherwise, to whom the Kitton caveats do not apply. Work by Wilkie Collins, Henry Chorley, John Hollingshead, Charles Allston Collins, and others, is currently scheduled for consideration, if funding can be found. It is hoped that an active forum for suggesting, discussing, selecting, testing and reporting on new attributions will be established on the DJO website, so that specialists and nonspecialists alike can participate in this fascinating collaborative enterprise of "authoring" All the Year Round.

University of Buckingham University of Newcastle

### Appendix: List of Sample and Control Texts

- 1. Author\_A(EO)\_Sampleo1 = Edmund Ollier, "Faces" *Household Words* X (16 Sept 1854), pp. 97-101 [2,558 words].
- 2. Author\_A(EO)\_Sampleo2 = Edmund Ollier, "Timber-Bending" *Household Words* XIV (30 August 1856) pp. 154–56 [2,153 words].
- 3. Author\_A(EO)\_Sampleo3 = Edmund Ollier, "A Scientific Figment" *House-hold Words* X (23 December 1854) p. 453–56 [2,247 words].
- 4. Author\_A(EO)\_Sampleo4 = Edmund Ollier, "Vampyres" *Household Words* XI (10 Feb 1855), pp. 39–43 [4,352 words].
- 5. Author\_A(EO)\_Sampleo5 = Edmund Ollier, "A Cornish Hug" *Household Words* XVIII (11 Sept 1858) pp. 296–99 [2,544 words].
- 6. Author\_A(EO)\_Controlo1 = Edmund Ollier, "Comets and their Tails of Prophets" *Household Words* XV (23 May 1857) pp. 481–84 [2,796 words].
- 7. Author\_A(EO)\_Controlo2 = Edmund Ollier, "Left Behind" *Household Words* IX, (22 July 1854) pp. 543–46 [2,456 words].
- 8. Author\_A(EO)\_Controlo3 = Edmund Ollier, "Thor and The Giants." *House-hold Words* XVI, (19 Sept. 1857) pp. 282–85 [2,530 words].
- 9. Author\_B(CD)\_Sampleo1 = Charles Dickens, "The Poor Man And His Beer" *All the Year Round* I (30 April 1859), pp. 13–16 [3,145 words].
- 10. Author\_B(CD)\_Sample02 = Charles Dickens, "The Young Man From The Country" *All the Year Round* VI (1 March 1862) pp. 540–42 [2,372 words].
- II. Author\_B(CD)\_Sampleo3 = Charles Dickens, "The Uncommercial Traveller" (repr. as "The Boiled Beef of New England" in *The Uncommercial Traveller*) All the Year Round IX (15 August 1863), pp. 588-91 [3,329 words].
- 12. Author\_B(CD)\_Sampleo4 = Charles Dickens, "Bill-Sticking" *Household Words* II (22 March 1851) pp. 601–06 [4,986 words].
- 13. Author\_B(CD)\_Sampleo5 = Charles Dickens, "New Uncommercial Samples. By Charles Dickens. A Plea For Total Abstinence" *All the Year Round* NS II (5 June 1869), pp. 13–15 [1,636 words].
- 14. Author\_B(CD)\_Controlo1 = Charles Dickens, "The Uncommercial Traveller" (repr. as "Wapping Workhouse" in *The Uncommercial Traveller*) *All the Year Round* II (18 February 1860) pp. 392–96 [4,350 words].
- 15. Author\_B(CD)\_Controlo2 = Charles Dickens, "The Uncommercial Traveller" (repr. as "Refreshments for Travellers" in *The Uncommercial Traveller*) All the Year Round II (24 March 1860) pp. 512–16 [3,635 words].
- 16. Author\_B(CD)\_Controlo3 = Charles Dickens, "Our [English] Watering-Place" *Household Words* III (2 August 1851) pp. 433–36 [3,618 words].
- 17. Author\_C(HM)\_Sampleo1 = Henry Morley, intro. to *Anecdotes of the Late Samuel Johnson* by H. L. Piozzi, ed. H. Morley (London: Longmans, 1856) [1,761 words].
- 18. Author\_C(HM)\_Sampleo2 = Henry Morley, intro. to *The Battle Of The Books and Other Short Pieces* by Jonathan Swift, ed. H. Morley (London: Cassell, 1886) [1,893 words].

- 19. Author\_C(HM)\_Sampleo3 = Henry Morley, intro. to *Playful Poems*, ed. H. Morley (London: G. Routledge, 1891) [3,946 words].
- 20. Author\_C(HM)\_Sample04 = Henry Morley, "My Wonderful Adventures In Skitzland" {excerpt} *Household Words* I (1 June 1850) pp. 225–229 [2,394 words].
- 21. Author\_C(HM)\_Sampleo5 = Henry Morley, "Letter From A Highly Respectable Old Lady" {excerpt} *Household Words* I (18 May 1850) pp. 186–7/"Jack and the Union Jack" {excerpt} *Household Words* IX (25 Feb 1854) pp. 32–33 [together 2,277 words].
- 22. Author\_C(HM)\_Sampleo5 = Henry Morley, "A Call Upon Sophy" *House-hold Words* IX (8 April 1854) pp. 174–76 [1,909 words].
- 23. Author\_C(HM)\_Controlo1 = Henry Morley, "A Russian Cauldron in Full Boil" *Household Words* IX (25 March 1854) pp. 118–22 [3,996 words].
- 24. Author\_C(HM)\_Controlo2 = Henry Morley, "Goblin Life" *Household Words* IX (8 April 1854) pp. 170–4 [4,659 words].
- 25. Author\_C(HM)\_Controlo3 = Henry Morley, "Doctor Pablo" *Household Words* IX (6 May 1854) pp. 280–4 [4,379 words].
- 26. Author\_D(WC)\_Sampleo1 = Wilkie Collins, intro. to *Rambles Beyond Railways*  $2^{nd}$  ed. (London: Richard Bentley, 1852) [2,453 words].
- 27. Author\_D(WC)\_Sample02 = Wilkie Collins, chap. vii, *Rambles Beyond Railways* 2<sup>nd</sup> ed. (London: Richard Bentley, 1852) [3,057 words].
- 28. Author\_D(WC)\_Sampleo3 = Wilkie Collins, Prologue to *After Dark* 2 vols (London: Smith, Elder & Co., 1856) [2,672 words].
- 29. Author\_D(WC)\_Sampleo4 = Wilkie Collins, "Laid up in Two Lodgings" (Part I) *Household Words* XIII (14 June 1856) pp. 517–523 [5,730 words].
- 30. Author\_D(WC)\_Sampleo5 = Wilkie Collins, "Strike" *Household Words* XVII (6 February 1858) pp. 169–172 [3,435 words].
- 31. Author\_D(WC)\_Sampleo6 = Wilkie Collins, "A Breach of British Privilege" *Household Words* XIX (19 March 1859) pp. 361–364 [2,790 words].
- 32. Author\_D(WC)\_Controlo1 = Wilkie Collins, "My Spinsters" *Household Words* XIV (23 August 1856) pp. 121–126 [3,922 words].
- 33. Author\_D(WC)\_Controlo2 = Wilkie Collins, "My Black Mirror" *Household Words* XIV (6 September 1856) pp. 169–175 [6,012 words].
- 34. Author\_D(WC)\_Controlo3 = Wilkie Collins, "To Think, Or Be Thought For?" *Household Words* XIV (13 September 1856) pp. 193–198 [4,946 words].
- 35. Author\_E(WHW)\_Sampleo1 = W. H. Wills, "Baptismal Rituals" *Household Words* I (27 April 1850) pp. 106–8 [1,724 words].
- 36. Author\_E(WHW)\_Sample02 = W. H. Wills, "The Lambeth-Marsh Mulcibers" *Household Words* XV (10 January 1857) pp. 25–27 [2,382 words].
- 37. Author\_E(WHW)\_Sampleo3 = W. H. Wills, ""A Good Plain Cook"" *Household Words* I (4 May 1850) pp. 139–41 [1,859 words].
- 38. Author\_E(WHW)\_Sample04 = W. H. Wills, "The Appetite for News" *Household Words* I (1 June 1850) pp. 238–40 [1,784 words].
- 39. Author\_E(WHW)\_Sampleo5 = W. H. Wills, "The Golden City" *Household Words* I (29 June 1850) pp. 313–7, [1,997 words].

- 40. Author\_E(WHW)\_Sample06 = W. H. Wills, "The Manchester School of Art" *Household Words* XVI (10 Oct 1857) pp. 349–52 [2,562 words].
- 41. Author\_E(WHW)\_Controlo1 = W. H. Wills, "A Coroner"s Inquest" *Household Words* I (27 April 1850) 109–113 [3,765 words].
- 42. Author\_E(WHW)\_Controlo2 = W. H. Wills, "A Legal Fiction" *Household Words* XI (21 July 1855) pp. 598–99. . [1,726 words].
- 43. Author\_E(WHW)\_Controlo3 = W. H. Wills, "My Annular Eclipse" {excerpts} *Household Words* XVII (24 April 1858) pp. 433–36 [1,435 words].
- 44. Author\_E(WHW)\_Controlo4 = W. H. Wills, "The Monster Promenade Concerts" *Household Words* II (19 Oct. 1850) pp. 95–6 [1,582 words].
- 45. Author\_E(WHW)\_Controlo5 = W. H. Wills, "Paris Improved" {excerpt} Household Words X (17 November 1855) pp. 361–63 [2,722 words].

#### **NOTES**

- www.djo.org.uk; www.newcastle.edu.au/school/hss/research/groups/cllc/.
   Accessed March 14 2011.
- 2. Anne Lohrli, comp., *Household Words*, . . . *Table of Contents*, *List of Contributors and Their Contributions &c.* (Toronto: University of Toronto Press, 1973).
- 3. E. A. Oppenlander, *Dickens' All the Year Round: Descriptive Index and Contributor List* (Troy, NY: Whitston Publishing, 1984), 63.
- 4. 1856–1904 (DNB); artist and author, principally of books about Dickens.
- 5. Kitton, Frederic G., *The Minor Writing of Charles Dickens* (London: Elliot Stock, 1900), vii, 138.
- 6. 6 items "written in collaboration with other writers" are also included in Kitton's listing; 29 contributions to the AYR Extra Numbers for Christmas (1859–67) should also be included (figs. from D. Thomas, Dickens and the Short Story [Philadelphia: University of Pennsylvania Press, 1982], 147–52). This works out at 161 interventions in the journal, over a total of 24 biannual volumes of an average of 25 weekly numbers that appeared under his editorship i.e. an appearance in 22% of issues. The comparable figure for the 19 bi-annual volumes of Household Words is 271 i.e. an appearance in just over 57% of issues. These are not exact figures, as they do not take account of the small number of issues in which more than one item appears, but the significant discrepancy between them cannot be in doubt.
- 7. Probably Walter Henry Howe, b. 1849?, [son of George Howe, also a printer] printer and publisher; resident at 8 Bisham Gdns. N. Highgate London between 1895 and 1901 (Census returns), during which period Kitton consulted his "office set"; author of *Methods Of Voting: as They Are and as They May Be: A Electoral Revolution*, and numerous popular compilations of games, proverbs, jokes etc. published under the title *Everybody's Book*&c. We are grateful to Dr Ellen Jordan for information in this note.

- 8. Efforts to trace it have so far proved fruitless. In noting, in a review of Oppenlander's *Index*, that "a set of *All the Year Round* once in W. H. Howe's collection (*most of which is in the Berg Collection*) has dropped from view" (our italics; *Studies in English Literature* 25.4 [1984], 913), Donald H. Reiman incorrectly identifies Kitton's W.H. Howe with William Thomas Hildrup Howe (1874–1939), the former President of the American Book Company, whose private collection was purchased by Albert Berg in 1940. The latter did own a set of *AYR*, now in the New York Public Library, but it is not the annotated one that Kitton consulted. Partially *ex. inf.* Anne Garner, to whom we are most grateful, e-mail message to John Drew, February 9, 2010.
- 9. Oppenlander, *Descriptive Index*, 49; 11 Wellington Street North, re-numbered 26 Wellington Street in April 1860: see J. Drew, *Dickens the Journalist* (Basingstoke: Palgrave, 2003), 140).
- 10. Lohrli, *Household Words*, 35, explains how the Office Book, after being presented by Wills's niece, who had inherited it, to R. C. Lehmann, became available to B. W. Matz, editor of the National Edition of Dickens's works (Chapman & Hall, 1906–08), some time between 1903 and 1908.
- 11. Kitton, Minor, vi.
- 12. Figures arrived at by taking the number of known contributions by Dickens divided by the total number of issues of each journal under his editorship. In a small minority of cases, Dickens contributed more than one item to a single issue, so the figure is an appearance rate rather than a strict percentage of magazines in which his work appeared.
- 13. See John Drew, "Dickens on 'Poor Hood': A New Article," *The Dickensian* 104.2 (2008), 110–22, and John Drew and Michael Slater, "What's in *The Daily News*? A Re-evaluation," *The Dickensian* 106.3 & 107.1 (2010, 2011), 197–206, ??–?? [forthcoming].
- 14. Ellen Jordan, Hugh Craig, and Alexis Antonia, "The Brontë Sisters and the *Christian Remembrancer*: A Pilot Study in the Use of the 'Burrows Method' to Identify the Authorship of Unsigned Articles in the Nineteenth-Century Periodical Press," *Victorian Periodicals Review* 39.1 (2006), 21–45; 37.
- 15. Jordan, ibid., 22. For a recent detailed outline of CLLC's methods, see chapter 2 of Hugh Craig and Arthur F. Kinney, *Shakespeare*, *Computers*, and the Mystery of Authorship (Cambridge: Cambridge University Press, 2009). For an example of the use of "PCA" analyses of the type employed in the present article, see also Hugh Craig, "Contrast and Change in the Idiolects of Ben Jonson Characters," *Computers and the Humanities* 33.3 (1999), 221–240.
- 16. From The Examiner: "Demoralisation and Total Abstinence," (27 Oct. 1849), repr. in M. Slater, ed. The Amusements of the People and Other Papers: Reports, Essays and Reviews, 1834–51 "Dent Uniform Edition of Dickens' Journalism" vol. 2 (London: J. M. Dent, 1996), 159–69;

- from *Household Words*: "The Great Baby" XII (4 Aug. 1855), 1–4; from *AYR*: "The Poor Man and His Beer" *AYR* I (30 April 1859), 13–16; "The Uncommercial Traveller" (repr. as "The Boiled Beef of New England" in *The Uncommercial Traveller*) *AYR* IX (15 August 1863), 588–91; "Working Men's Clubs" (with Edmund Ollier) *AYR* XI (26 Mar. 1864), 149–54; "New Uncommercial Samples. By Charles Dickens. A Plea For Total Abstinence" *AYR* NS II (5 June 1869), 13–15.
- 17. The best reference point, semantically, is Dickens's article "Why?" *House-hold Words* XIII (1 Mar. 1856) pp. 145–48, composed, like the opening para. of "Temperate Temperance", entirely of rhetorical questions.
- 18. 13; repr. in M. Slater and J. Drew, eds, *The Uncommercial Traveller and Other Papers*, 1859–70 "Dent Uniform Edition of Dickens' Journalism" vol. 4 (London: J. M. Dent, 2000), 3–11; 4.
- 19. The internet is awash with Twain's so-called "famous quotation": "Temperate temperance is best; intemperate temperance injures the cause of temperance."
- 20. Andrews, *Charles Dickens and his Performing Selves* (Oxford: Oxford University Press, 2006), 144–46.
- 21. 1827-86 (DNB).
- 22. The "Iota" test is based on a search for rare words. A list was made of the words that appeared in one author's training segments but not in the other's, and vice versa. These are "Author X-not-Y" and "Author Y-not-X" words, odd words that happen to be used at least once in one author's set of "training" articles but never appear in the other author's. We then found the number of different words used in each segment and found the proportion of these that is formed first by one list, then by the other, in the training sets, the test sets, and in the mystery article, "Temperate Temperance." Each segment thus had two scores, the proportion of its words that are "Dickens-not-Another's" ones and the proportion that are "Another's-not-Dickens" ones.
- 23. The "Zeta" test trawls mid-range vocabulary: words that are commoner than the rare words used in Iota testing, but not as common as the very frequent grammatical or "function" words. The process identifies words that appear regularly in one set and rarely in the other. We find the proportion of segments by one author in which a word appears and compare this with the proportion of segments by a second author in which the word does not appear. All the words in the sets can be ranked by this index. We take the 400 highest-ranking words, the ones most characteristic of the first author's segments, and then find the number of the words from the list which appear in each segment, training, test, and mystery, and divide this by the total number of different words in the segment. We then do the same in reverse, this time making a list of words appearing more regularly in the second author's segments than in the first author's and finding the proportion that these form of each segment's total range of different words. Each segment

- then has two scores, for one author's characteristic words and for the other author's characteristic words. The zeta test, developed by John Burrows, is explained more fully in Craig and Kinney, *Shakesepeare*, *Computers* &c., ch. 2.
- 24. See Prabir Burman, "A Comparative Study of Ordinary Cross-validation, *v*-fold Cross-validation and the Repeated Learning-testing Methods" *Biometrika* 76.3 (1989), 503–14.
- 25. For examples see John Burrows and Hugh Craig, "Lucy Hutchinson and the authorship of two seventeenth century poems: a computational approach", *The Seventeenth Century* 16.2 (2001), 259–82; Hugh Craig, "The 1602 additions to *The Spanish Tragedy*," in Craig and Kinney (2009), pp. 162–80; Peter Anstey and John Burrows, "John Locke, Thomas Sydenham, and the Authorship of Two Medical Documents," *Electronic British Library Journal*, 2009, Article 3 (http://www.bl.uk/eblj/2009articles/article3.html Accessed March 14 2011); John Burrows and Anthony Hassall, "Sarah and Henry Fielding and the Authorship of *The History of Ophelia*," *Script and Print* 30 (2006), pp. 69–102.
- 26. See Gaskell, *Letters* ed. Chapple and Pollard (Cambridge MA: Harvard University Press, 1967), No. 418, *To* C.E. Norton.
- 27. 1829–94 (*DNB*); see Christopher C. Dahl, "Fitzjames Stephen, Charles Dickens, & Double Reviewing," *Victorian Periodicals Review* 14.2 (1981), pp. 51–58.
- 28. Extracts from the *Saturday Review* VI, 644; XI, 194–6; and XVI, 759–60, all excerpted in Philip A. W. Collins, ed., *Charles Dickens: the Critical Heritage* (London: Routledge, 1971), 405, 412, 413 respectively.
- 29. John Forster recounts how a former schoolmaster sent Dickens a snuffbox engraved with this nickname part-way through the serialisation of *Pickwick Papers* (*The Life of Charles Dickens*. [3 vols. 1872–74] Ed. J. W. T. Ley. [London: Cecil Palmer, 1928], I.i), and the epithet stuck. For Dickens's use of it as a signature substitute, see M. House and G. Storey, eds, *Letters* "The 'Pilgirm' Edition" (Oxford: Oxford University Press, 1965), xxviii.
- 30. Harry Stone, ed., *Charles Dickens' Uncollected Writings from Household Words*, 1850–59. 2 vols. (Bloomington & Indiana University Press, 1968). 3–69). Stone's introduction (3–69) sets out in detail the methods Dickens and his co-writers used for drafting and revising these "composite" writings: "[i]deally, at least, such articles were to be indivisible, the work, apparently, of a single author" (43).
- 31. Lillian Nayder, Unequal Partners: Charles Dickens, Wilkie Collins, and Victorian Authorship (Ithaca. NY: Cornell University Press, 2002); see esp. chaps. 1, 2, and the Conclusion, "'This Unclean Spirit of Imitation': Dickens and the 'Problem' of Collins's Influence," 198–202. For a more straightforwardly biographical emphasis on editor-writer collaboration, see P. D. Edwards, Dickens's "Young Men": George Augustus Sala, Edmund Yates and the World of Victorian Journalism (Aldershot UK: Ashgate, 1997), chaps. 3, 4.