

Engineering Homer's Soul: A Mathematical Rebirth of the Author

Olivia McCauley

**An Undergraduate Thesis
Presented to the Faculty of the
UC Berkeley Department of Rhetoric
In Candidacy for Honors in the Rhetoric Major**

March 8th, 2023

Abstract

Though the relationship between authors, texts, and readers has long been a subject of inquiry, recent technological innovations in computational linguistics for analyzing and generating natural language data have added a new dimension to this debate. No longer can we assume that both author and reader of a text are embodied human creatures, nor can we as easily differentiate between the texts produced by humans and those produced by machines, and so the long contested borders between “fake” and “real” works belonging to a historical author has become even more contentious.

This undergraduate thesis explores how these technologies have the ability to generate new and peculiar ideas of which bodies can be authors. I demonstrate the potential for discursive openings in computational linguistics by analyzing an article whose subject matter speaks to the historical heart of these issues as they are discussed in the intellectual “West”: the Homeric Question. Historical investigation has attributed Homer’s epic poems to a collective oral tradition in Ancient Greece, designating authorial truths of Homer as eternally enigmatic and rendering the Homeric question unanswerable. Computational linguistics seeks to disrupt this stagnation via statistical analysis. But in doing so, these academics unwittingly produce not only new questions, but also a new, technical apparatus that can reproduce Homer’s unconscious in quantitative terms from the text. Embodied in a whole new way, this cyborg re-animation of Homer suggests that modification to prevailing theories of authorship and authenticity, especially those finding roots in antiquity (including Foucault’s “What is an Author?,” Barthes’ “Death of the Author,” and Vernant’s “A Beautiful Death”) is necessary. Tested against bodies whose relationality defies classification into discrete categories of author, reader, and text, it becomes clear that Barthes’ theory in particular fails to provide sufficient explanation for what it means to be a reader, a writer, or text in a context where relationships between all three roles are mediated not just by literary criticism, but by empiricism itself. Unpacking the statistical peculiarities of this application under Nikolas Rose’s account of the self in disciplinary technology, this article argues that the Cyborg Homer is produced as a docile object endowed with a mathematical body, and in introducing notions of infinitely improving approximations of reality, demands a paradigm shift from the humanities on issues of authorship.

Introduction

The “Homeric Question,” which is concerned with understanding the historical origins of the *Iliad* and the *Odyssey*, has been designated as unsolvable in its totality by the humanities. Though centuries of study have yielded a large body of insights about these epic poems’ origins and significance, historians, philosophers, and literary scholars alike have concluded that the information loss incurred over millennia is too great to reach definitive answers. This ambiguity rings particularly true when it comes to the identity of the poems’ author, or authors, and identifies a long line of questions that makes up the field of Homeric authorship analysis. Prevailing theories in Ancient Greek and Roman studies credit an oral tradition of traveling bards for composing the two epics, which later took on their written form when bards dictated to literate scribes.

To a certain extent, accepting that these highly influential texts lack a single author precludes scholars from granting the Homeric question a perfect answer. That the poem is borne of an authorial multiplicity, rather than singularity, introduces a degree of ambiguity that is unusually strong — even for texts of antiquity. We may think we understand a modern author, but we don’t, and if we presume we know an ancient author, we will always be wrong. To think that we could understand the “true” multiplicity across time and bodies that make up “Homer’s” reflection in the modern era would be a fool’s errand, for we will always discover our own reflection looking back at us. Moreover, while study of the author has often been made a central question of literary criticism, it pushes many other issues — like those regarding what a text means to us personally, in the time that we live in, and what we can learn from it — to the wayside. Thus accepting oral-formulaic theory as a framework for authorship analysis explicitly endorses the notion that the *Iliad* and *Odyssey* lack a clear authorial origin.

Though oral-formulaic theory may reinforce the historical ambiguity of the Homeric Question and its answer (after all, it is unlikely that even physical remnants of an ancient civilization could tell the whole story of Homeric traditions), the author’s centrality in this question has allowed the question to be re-opened in another domain: computational linguistics. Despite the historical ambiguity of any statement framed as an answer to the Homeric Question, scientific paradigms forwarding the notion that everything — even authorship — is a quantitative phenomenon that can thus be known and characterized have re-opened the question in a new way. Technological innovation in computational processing capabilities, which have found broad applications in the field of natural language processing, has revived stylometric questions about the nature of the poet’s voice. The underlying motivations and formulations of these questions, however, are thus influenced more by technological innovation than Classical scholarship and take a substantially different approach to the Homeric question. A notable difference between previous investigations into Homeric style and approaches under the paradigm of natural language processing is the drive to atomize and reproduce the Homeric voice via measurable attributes of linguistic style.

In “Computational Authorship Analysis of Homeric Language,” three computational linguistics researchers at the Athens University of Economics and Business, as well as the Democritus University of Thrace, investigate the Homeric voice. They do this not by comparing it against other written works, but by performing computational text analysis. Measuring internal stylistic inconsistencies between constituent books of the *Iliad* and *Odyssey*, they seek to assess the degree to which the books are “linguistically integrated” between one or several distinct authorial voices. Their stated research question is as follows:

“Do any of the books in either the Iliad or the Odyssey demonstrate a higher or lower than average *degree of linguistic integration* to the entire poem? Does any of them present *sufficient linguistic divergence* to raise authorship doubts? And how do they compare to similar literature of their time?”¹

In order to measure the similarities between authorial voices in each book, the study utilized a statistical language model (SLM) known as “character-level” n-gram analysis, which captures syntactic and stylistic features in writing.² Though the methodology employed to perform the analysis is well-documented in the article, the underlying assumptions that warrant using such a technique, as well as assumptions around the existence of a unique, stylometric “fingerprint” defined by statistical measures are never made explicit and consequently, remain uninterrogated.

Given that the study’s stated goal breaks with conventional views, which sees Homer as an “author” who can not be reliably reconstructed, the absence of ideological justification surrounding the article’s goal and methods raises significant questions. These questions concern scientific imaginaries of Homer — imaginaries that operate on notions of authorial voice and uniqueness as initially conceived in the field of Classics, but are never made explicit. The discourse in the study, I argue, thus re-imagines the relationship between author and text. This reimagination stabilizes the relationship between voice and work, so that the Homeric question can be reformulated as a problem of adequate data and appropriate technology, rather than a question of cultural relevance and interpretation.

But how, and from where, does this understanding of the author arise? More specifically, what allows computational linguistics to transform Homer from an unknowable entity into a mathematically tractable one? What desires motivate such a transformation? What mathematical ideas about language and style project legitimacy onto computational techniques that atomize authorial voice into nearly unintelligible parts, and use those parts to reconstruct an image of the author? What, of the author, do we lose when processed through the black boxes of machine intelligence, and what might we stand to gain?

In order to address these questions, I will articulate how various ontologies of authorship get taken up within, and influence the style of, the computational authorship analysis carried out by Fasoi et al. By performing a close reading of the article’s language, I will analyze how notions of “authorial fingerprint” — the author’s unique linguistic mark — are operant in the study, paying particular attention to the way that the authorial fingerprint is imagined as an object to be apprehended through statistical language modeling. Understanding how the author’s existence is made contingent on the “fingerprint” signifier, a set of stylometric patterns whose combination identifies the author’s unique voice, will lend insight into how a mathematical ontology of Homer and his constituent voices come to exist as scientifically valid.

To do this, I take up Katherine Hayles’ methodology of “media-specific analysis.” This form of analysis “attends both to the specificity of the form... and to citations and imitations of one medium in another.” (Hayles, p. 69) Like all forms of media, the composition and purpose of scientific journal articles gives the study a unique texture that must be accounted for. Performing a close reading of scientific literature as a “text,” especially a reading that does not operate on scientific paradigms and seeks to draw out latent assumptions and worldview of

¹ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 79

²Ibid, p.79

authorship analysis, requires attending to the channels through which scientific knowledge is legitimated. These channels, like peer review or norms for documenting methods and structuring results, constitute what Hayles names “medium-specific constraints” that shape the final text. (p. 69)

Moreover, I propose viewing the methods section as part of scientific literature’s “materiality” in the sense that Hayles defines the term, “as the interplay between a text’s physical characteristics and its signifying strategies.” (p. 72) Though the methods don’t necessarily constitute a “physical characteristic,” the signifying strategy of the methods section is to frame the study’s object of analysis, and the methodology used to analyze that object, as a manipulation of the physical realm. Imagining the methods as a intervention into material reality represents one of science’s primary ways of signifying the truth-value of it’s claims. It is also the reason analysis of the methods is so central to this project: the methods are “material” in the sense that they “join the physical” reality of Homeric texts with our science’s “mental” view of authorship, and bring together “the artifact and the user.”

This analysis will be supported by context from other computational linguistics sources and framed using theories like Milman Parry’s Oral-Formulaic theory, as well as Koopman’s theory of informational persons and Barthes’ *Death of the Author* as a lens to understand how the ontology of the authorial self is being reformed.

My analysis focuses on this article as one example of how paradigms from the humanities get taken up within scientific study, and suggests that similar forms of analysis are useful to draw out how these investigations view different issues. Though I believe the methodology I employ in this project is generalizable to other domains within the digital humanities, I make no similar pretenses about the direct results of my analysis. Instead, I attempt to show how even a single article contributes to a changing landscape of what authorship is and means in today’s digital age.

Part 1: Homer, the probability model.

The earliest articulation of assumptions about authorship appears in the article's research question. It demonstrates an understanding of the author as a unified entity, which justifies casting probability models as the fingerprints that identify Homer. The authors are primarily interested in ascertaining quantitative measures of "linguistic integration" and "divergence" across the books of the *Odyssey* and *Iliad* and hope to use those measures to ascertain whether those epic narratives are internally cohesive. Though the particular phrase is never reused outside of this context, such language presupposes the existence of a homogeneous authorial voice, despite explicitly recognizing that Homeric texts were likely a "collective effort involving numerous authors, each contributing unknown [portions] of text."³ This presupposition is not simply in conflict with historical understandings of the Greek epic tradition but constructs a view of the Homeric voice as something that is cohesive, and internally consistent. The research question's language defines the terms upon which authorship analysis is conducted within the project, and consequently, the methods employed in the study correspond with this view, and later makes this assumption a reality.

The research question's focus on "integration" and "divergence" defines the article's "authorship analysis" as a project to mathematically characterize a text, and use those mathematical characteristics to identify and separate discordant contributors from a true author and text. Measuring "integration" and "divergence" with "statistical language modeling" represents two parts of an imperative to classify texts by their authors.⁴ Using measures of "integration" between books as an indication of common authorship suggests the assumption that the "real" authorial voice is unified. Here, notions of reality are grounded in statistical analysis of linguistic events, and so discerning between real and fake becomes a question of how well statistical paradigms can separate an anomalous event from normal occurrences. When used as evidence that texts (like the Books of the *Odyssey* or *Iliad*) share an author, measures of integration imply that the author's language is consistent across texts, and therefore must adhere to some system that governs authorial expression. This understanding implies that the unity inherent within a single human author produces linguistic integration within the text, and suggests the presence of an authorial structure that enforces a unique set of linguistic regularities that can be traced back to an individual voice.

On the other hand, there is "divergence." Though discourse around a text's "linguistic integration" doesn't necessarily imply a classification project, the inclusion of "linguistic divergence" in the question does.⁵ Treating strong measures of "linguistic divergence" as good reason to "raise authorship doubts" implies that some texts should be excluded from an author's corpus, if those are to be considered reliable representations of authorship. In this particular context, the way that linguistic "divergence" is positioned as potential counter-evidence against two Homeric books sharing an author suggests a desire to define a unified, true Homeric voice by discarding fraudulent or erroneously attributed texts.

Importantly, the separation between an "integrated" voice from one that "diverges" from the voice narrativizing a corpus is mediated by statistical methods that essentially perform pattern recognition. The study utilized tri-gram analysis, a form of character-level n-gram

³ Fasoi, Pavlopoulos, and Konstantinidou, "Computational Authorship Analysis of Homeric Language." P. 78

⁴ Ibid, p. 79

⁵ Ibid, p.79

analysis, to analyze the “stylistic features” of the corpus⁶. The focus on apprehending “stylistic features” through statistical language modeling draws on a line of thinking where “language modeling is concerned with capturing regularities of natural language.” Such linguistic regularities arise in the form of “semantic, syntactic, lexicographic and morphological patterns,” a view articulated in “Language Independent Authorship Attribution using Character Level Language Models” and cited by Fasoi et al., which suggests direct influence on the study’s methods.⁷⁸ Under this view, a well-formed Homer will appear by characterizing the “regularities” of his natural language, and by excluding texts with linguistic patterns in conflict with patterns that are attributed to him.

This idea is borne out in the project’s methods, where they measured the linguistic integration with a perplexity (PPL) score between each individual book against all remaining books. Defined as a measure of... how well a probability model predicts a sample,” perplexity helps explain how integration of patterns can be measured.⁹ Identifying a pattern in language can be understood as articulating a probability model for that sample of text, because a linguistic pattern is understood as a function of a probability distribution, an assignment of probabilities which dictate a worldview where events (like the use of a word or frequency of a character) occur with particular likelihoods.

These assumptions shape the textual comparison the study chooses to make. The researchers “[used] excerpts from all but one Book of the same poem (Iliad or Odyssey) and then [computed] the PPL on that remaining one.”¹⁰ By excluding one Book at a time, the authors can determine which “patterns” (regularities) pose sufficient explanatory power, and should be included in their statistical model of Homer.¹¹ Based on this view, the researchers used “average PPL [to] assess the overall linguistic proximity of a Book with the rest of the Books.”¹² High linguistic proximity, between portions of the text, which occurs when a PPL score is low, is taken as a sign of shared authorship. Accordingly, low linguistic proximity (indicated by high PPL scores) are taken as a sign that certain portions of the text are linguistically different and thus must originate from different authors. This style of measurement defines how authorship analysis is distinct from authorship attribution, and also demonstrates how scientific methods are tailored to the way that questions are formulated in the humanities. Measures of linguistic proximity, and the methodology used to generate those statistical values, act as a form of unsupervised learning, because it makes no assumptions about the text having a single author, or in the case of multiple authors, about the text having a set number of distinct authors. Instead, these methods foreground the text as the realization of authorship, which makes the text both the source and arbiter of “true” Homeric authorship.

Even though this approach seems open to many possible modes of authorship, there are two important assumptions that impose limits on what Homeric authorship can mean. First, even though there is no set number of authors whose voices the study seeks to identify and separate out, it still assumes that authorship has present material, embodied existence within the text, otherwise the text would not be foregrounded as it is. This is a major factor separating out

⁶ Ibid, p. 80

⁷ Ke’selj et al., “N-GRAM-BASED AUTHOR PROFILES FOR AUTHORSHIP ATTRIBUTION.” p. 255

⁸ Peng et al., “Language Independent Authorship Attribution with Character Level N-Grams.”p. 258.

⁹ Campagnola, “Perplexity in Language Models.”

¹⁰ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 81

¹¹ Ibid, p. 85-86

¹² Ibid, p. 82

computational linguistic authorship analysis from historical or archaeological interventions into the Homeric Question. Moreover, even though there is no set number of authors, they are all imagined as having discrete, separable beings and voices, and that those separations have been preserved via the text over centuries of transmission. Measuring the “integration” of observed patterns in the text by comparing the explanatory power of different probability models on each text allows the authors to reduce Homer into a tractable mathematical model, with intelligible behavior and parameters. Homer ceases to be a product of collective tradition, instead converging into a singular, computational object.

Part 2: Homer, the stylistic fingerprint.

But we have not yet elucidated the precise nature of computational Homer, and how the study formulates this object in concrete terms. If we understand Homer as a computational object, what are those quantitative values, and what do they represent?

Fasoi et al perform character-level tri-gram analysis, a subtype of the linguistic modeling technique broadly known as n-gram analysis. It must be noted that while natural language processing in general seeks to map linguistic regularities in a text, the specific probabilistic assumptions of n-gram analysis, like the Markov assumption, influence the texture of conclusions made in the study. However, the focus here will be on exploring the descriptive relationship between tri-gram analysis and linguistic style. Put clearly, why is n-gram analysis thought to accurately capture an author's unique stylistic tendencies? And why do SLM practitioners believe that probabilistic analysis of stylistic components in a text can separate one author from another using only the text, particularly when the author is not a singular individual, but a collective oral tradition?

Examining Milman Parry's oral-formulaic theory, it becomes clear that the statistical assumptions necessary for valid tri-gram analysis can be derived from the idea that Homeric verses abide by a specific set of syntactic and metrical shapes. Linguistic regularities, understood as patterns in a text whose occurrences can be modeled by a probability distribution, constitute the shape of Homeric style which tri-grams allegedly capture. In the methods section, the researchers justify their choice to use tri-gram analysis because it is the "most effective" way to capture "stylistic features."¹³ As defined in "Style-Markers in Authorship Attribution A Cross-Language Study of the Authorial Fingerprint," a computational linguistics methods comparison by Maciej Eder of the Pedagogical University of Kraków, "style-markers" are "syntactic... features" of a text.¹⁴ "Syntactic features" are elements within a text comprised of "syntactic markers."¹⁵ Such "graphemic elements...create coherence within phrases and between words or word groups on the clause level" by "[linking] larger entities of a sentence."¹⁶ Here, the compatibility of tri-gram analysis and oral-formulaic theory becomes clear: the character-level patterns are a direct (but atomized) way to represent syntactic formulas constituting the main structure of Homeric narratives.

It is no doubt that this approach to the Homeric Question differs significantly from historical and literary analyses carried out in an attempt to discover Homer's true identity. Some notable examples in the vein of comparative literary and historical analysis include Martin L. West's comparative study of the *Shield of Achilles*, and Friedrich August Wolfe's *Prolegomena*. Both use the text itself, alongside historical scholarship, to speculate about the conditions that could have feasibly brought the poems into existence. Most modern scholars have since taken up the "transcription hypothesis," and come to believe that the *Iliad* and *Odyssey* as we know them come from a single poet who dictated the poem to a literate scribe, but questions regarding the time of composition and an exact genealogy of the text itself (whether it is composed mostly as a unified entity, or originates as dispersed myths later strung together into a single narrative) remain contentious issues.

¹³ Ibid, p.80

¹⁴ Eder, Maciej, "Style-Markers in Authorship Attribution." P. 103

¹⁵ Weth, "Distinguishing Syntactic Markers From Morphological Markers. A Cross-Linguistic Comparison." p. 1

¹⁶ Ibid, p. 1

Clearly, the digitally automated computational authorship analysis employed in the study has significant paradigmatic and methodological differences from comparative literature and philology. But what is interesting here is not the differences between fields in the way that they approach this issue. Instead, a shared view that these poems are, at their core, important because of their patterned, formulaic language highlights the fact that scholars across disciplines approach these poems with a similarly structuralist mindset. If it seems that the methods Fasoi et al use to carry out their authorship analysis are philosophically atypical, that is only because techniques like n-gram analysis, and their presentation in scientific literary form are relatively new to the domains of philology and comparative literature. The question this study asks, and the answer it seeks to discover, is not fundamentally different from the questions Wolf asked in his *Prolegomena*. It looks to the time before the poems were cemented into written record, and asks if we can discern the stylistic influence of distinct rhapsodists on the poems' final forms.

Thus there emerges surprising continuity with previous inquiries into the Homeric Question. It is important to understand the way that philosophical continuity blends with technological novelty to reconstitute a new vision of authorship, and uses the figure of Homer almost as a conduit to transfuse the cultural relevance of Classical study into the digital humanities. This recombination of ontologies surrounding authorship and Homer constitutes a new vision of authorial voice, a co-production of statistical prediction methods and prior investigations into ancient epic narratives composed in oral traditions. The extension of these paradigms into probabilistic analysis does, ultimately, make a transformative move, but the focus of this section will be on highlighting how that transformation is founded on the basis of Homeric scholarship in the humanities.

A strong point of continuity between Fasoi et al's work and previous scholarship is Milman Parry's oral-formulaic theory. Like computational linguistics, oral-formulaic theory privileges syntactic and metrical structure as the defining characteristics of Homeric narratives and expresses a belief that essential meanings of the narrative are contained within a verse's linguistic shape.¹⁷ The richness of oral culture is reduced to "a matter of the manipulation of traditional units of content... expressed in traditional phraseology."¹⁸ The primary component of narrative within this lens, "formulae" are defined as "an expression regularly used, under the same metrical conditions, to express an essential idea."¹⁹ Under this interpretation of epic poetry, style is not simply a matter how thematic meaning is expressed. Style, when defined in terms of the metrical and syntactic patterns, is understood as a unique structure defining both the narrative content and linguistic shape of a work. The idea that style is a linguistic apparatus of grammatical, syntactic, and thematic patterns makes transforms it into something that is easily characterized by probability assumptions, whose granular scrutiny can make something as ambiguous as artistic style specific, quantifiable, and concrete.

Pursuing authorship analysis via this paradigm ultimately leads computational linguists to construct an "authorial fingerprint," which embodies the defining characteristics of a unique idiolect across texts without the need for an individual author. The "authorial fingerprint" forwards a belief that stylistic features in writing contain "traces" of its author.²⁰ Originating from Wincenty Lutoslawski, the founder of stylometry, the supposed existence of the authorial fingerprint rests on the assumption that "style is... personal and characteristic," and the belief

¹⁷ Elmer, "Oral-Formulaic Theory." p. 2.

¹⁸ Ibid, p. 1

¹⁹ Russo, "Formula." p. 2

²⁰ Eder, Maciej, "Style-Markers in Authorship Attribution." 2011, p. 102

that style is a product of an author's "unconscious personal stamp" still influences the way that research problems are formulated in authorship studies.²¹ Ascertaining an author's personal stamp becomes the primary research objective, encouraging project structures that sacrifice philosophical questions around what the distinguishing features of an individual's idiolects are, and how they should be measured.²² In their place, researchers pose questions about how to best capture the defining characteristics of authorial idiolect regardless of what form those characteristics may take, they seek a technique that can easily account for identifying ties in an author's language even when the unique facets of their voice are submerged a sea of "infinite... linguistic features."²³ It is no wonder, then, that the field privileges methods like n-grams which, by atomizing every part of the text into short character sub-strings, can supposedly capture all occurrences of "linguistic features," and distill those features into the truest image of an author's style.

The study itself never explicitly identifies constructing a "stylistic fingerprint" for Homer as its primary goal, but this idea nevertheless emerges from the way Fasoi et al articulate the relationship and differences between authorship analysis and attribution. They define "computational authorship analysis" as "the study of any association between a corpus and a group of authors." (79) Contrasting their description of authorship analysis against their provided definition of authorship attribution, which is "the problem of identifying the author of an anonymous text, or a text whose authorship is in doubt," it becomes clear that the actions they carry out when analyzing authorship in Homeric texts is not quite "authorship analysis," nor "authorship attribution" in the conventional sense. The "group of authors" they are studying is commonly denoted by the singular, rather than plural, Homer, and the primary corpus is the *Odyssey* and *Iliad*. Because the sub-Homers they seek out are unknown at the outset of their investigation, their authorship analysis takes on authorship attribution's characteristic focus on identifying individual authors; the *Iliad* and *Odyssey* are broken into anonymous sub-texts, and such a procedure casts doubt upon their Homeric origins. There is no initially known set of authors that portions of the text can be attributed to, so Fasoi et al can perform neither authorship analysis nor attribution in the way that they define it. What they seek out, instead of mappings between text and author, are the authors themselves. These authors are rendered visible by the long-telescoping eye of stylometric analysis carried out with modern computational tools, but their bodies and souls perished long ago. The vestiges of their existence are strictly linguistic, and so if they are to be identified as a group of individuals who once walked the earth as discrete, embodied creativities, then they must be given new existence using the medium that preserves their voices — the text. Quantitative authorship studies in general perform statistical analyses of texts to exploit stylistic differences in order to attribute anonymous texts to an author, verify authorship, study associations. But when those stylistic differences are exploited to resuscitate authors from the text, and those authors lack historical presence, then their individual identities are constituted purely by the stylistic contours of an "authorial fingerprint." Born nameless like any other, these authors' identities are tethered to that "unconscious personal stamp" by stylometry's ideological umbilical cord.

Tri-grams provide the mechanism to build this stylistic fingerprint in mathematical terms — a probability model for the Homeric voice. The basic idea behind n-gram analysis is that the text (or corpus of texts, like the combined Books of the *Odyssey* or *Iliad*) is decomposed into

²¹ Ibid, p. 103

²² Eder, Maciej, "Style-Markers in Authorship Attribution." 2011, p. 103

²³ Ibid, p. 103

character subsequences of length “n” (a variable representing any integer greater than zero).²⁴ Then, these sequences are used to generate a probability model predicting the likelihood of naturally occurring character sequences under the Markov assumption of conditional independence (the idea that the future is independent of the past given the present).²⁵ In this case, the study constructed substrings of three Ancient Greek characters. As an example, the trigram (ῆ-ρ-ώ) is three lower-case Ancient Greek characters, separated for readability by dashes “-” that are excluded from probability calculations. The relevant part to understand here is how conditional independence assumptions generate a probability model for texts constituted by a linear system of “syntactic-metrical patterns” interspersed with substituted words.²⁶

The Markov chain model, presented in mathematical notation below, invokes assumptions about the importance of frequent phenomena.

$$P(c_i | c_{i-1}, c_{i-2}, \dots, c_1) = P(c_i | c_{i-1}, c_{i-2}, \dots, c_{i-n+1})$$

The expression above represents the characteristic feature of Markov probability models, the Markov assumption. Imagining each tri-gram as a possible state, or event, Markov models provide a way of modeling transitions between tri-grams under the assumption that transitions to future states are determined solely by the previous state. Essentially, given the current tri-gram, whichever tri-gram follows the current tri-grams is a function only of the current tri-gram, previous tri-grams are thought to hold no influence over the state transition. At each state, all the tri-grams are possible successor states, and the most likely successor state becomes the predicted successor state in the Markov chain, which is also assigned a probability. These probabilities indicate how likely on a range of zero to one (zero being that there is no chance the transition will occur, one being that there is near-certainty that the transition will occur) given the previous state, the successor state is reached by the text.

There are two important ideas to articulate here about the Markov model, which is a linear chain stringing together the most likely state transitions. First, although this chain is thought of as the stylistic model of the text itself, it does not account for all the probability values generated by tri-gram analysis. At each tri-gram state, the probability of transitioning to any other unique tri-gram state is calculated based on the number of times those two states appear in immediate succession. Visually speaking, the Markov model is like a highlighted path on a tree data structure, and that highlighted path is the most likely one to occur. At each tri-gram, transitions to all other less-likely states are still possible, and the text will inevitably violate (or disprove) the model’s accuracy.

This brings us to the second important idea, which is that the text is distinct from the statistical language model. The Markov model might predict that one tri-gram follows another in the text, but models are meant to be generalizations of reality and thus are not fitted to the text exactly. Especially in passages with large amounts of non-formulaic language, the most likely tri-gram might never occur. The implication of this is that tri-gram analysis can be used to assign the text a probability value, and by asserting how likely it is that the text was composed using the linguistic patterns it exhibits, the model assumes the role of arbiter over Homeric language, and holds de facto power to verify “proper” and “unlikely” linguistic behavior in the Homeric texts. Ideas of “proper” and “unlikely” linguistic behavior have important implications for a project whose focus is on discerning authorship through a probabilistic lens. The text data used to

²⁴ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 80-81

²⁵ Ibid, p. 80

²⁶ Russo, “Formula.” p. 1

construct a model must be imagined as a corpus for an individual author, and so the model must be thought of as representative of the author's linguistic style. Modeling style in this way, however, creates a view of authorial style that has frequent phenomena at its center, and builds boundaries around itself on the basis of un-likelihood.

The emphasis on frequency continues with the way that models predict successor states given the present state. Tri-gram language modeling uses the current tri-gram to predict which trio of characters will appear next. It does this by using "maximum likelihood estimation" (MLE), which predicts that the next tri-gram will be the one that appears most often after the current tri-gram across the text (the predicted tri-gram has the highest probability value from the conditional probability table).²⁷ Maximum likelihood estimation's methodological soundness in the context of quantitative authorship analysis clearly relies on the statistical assumption articulated by Eder that computational linguistics practitioners. That assumption is that stylometrists must "accept... that dealing with frequent phenomena, e.g. function words, leads to much more reliable conclusions than analyzing rare occurrences, such as hapax legomena," or terms that are only used once.²⁸ To identify the most likely successor tri-gram, transition probabilities, values between zero and one indicating how likely an event is, are computed for each tri-gram in the text. These probabilities represent the likelihood of transitioning to a specific tri-gram, given the current tri-gram (a conditional probability). In this model, each tri-gram represents a potential successor state that the current state (also a tri-gram) could move to, and a set of conditional transition probability tables form a probability distribution for the entire text.

To understand this process, take the following example of a transition between two tri-grams from the preface of the Iliad's first Book:

$$(\eta\text{-}\rho\text{-}\acute{\omega}) \rightarrow (\omega\text{-}v\text{-},)$$

The SLM tries to predict the successor state for $(\eta\text{-}\rho\text{-}\acute{\omega})$ based on the conditional probability table $P(c_i | c_{i-1})$ where the correct successor state c_i is $(\omega\text{-}v\text{-},)$ and the current state c_{i-1} is $(\eta\text{-}\rho\text{-}\acute{\omega})$. The conditional probability table $P(c_i | c_{i-1})$ in this example is $P(c_i | (\eta\rho\acute{\omega}))$, where the rows of the table will contain all probabilities of transitioning from $(\omega\text{-}v\text{-},)$ to another tri-gram state. The transition probability for a specific successor state is computed using the following MLE formula:

$$P(c_i | c_{i-1}) = \frac{C(c_i, c_{i-1})}{C(c_{i-1})}$$

In English, the equation above states that the probability that the next state c_i follows current state c_{i-1} is equivalent to the observed number of times c_i occurs after c_{i-1} in the training text, divided by the total number of times the tri-gram state c_{i-1} occurs in the text.

For the sake of this example, assume that the model predicts correctly and the highest transition probability across $P(c_i | (\eta\rho\acute{\omega}))$ is $P((\omega v\text{-},) | (\eta\rho\acute{\omega}))$. c_i represents all the possible tri-grams that could follow $(\eta\rho\acute{\omega})$. This would be calculated by counting the number of times the sequence (or "event," in probability terminology) $(\eta\text{-}\rho\text{-}\acute{\omega}\text{-}\omega\text{-}v\text{-},)$ occurs and dividing that quantity by the number of times $(\eta\rho\acute{\omega})$ occurs throughout the corpus.²⁹ This process yields the "likelihood estimate" for any transition from $(\eta\rho\acute{\omega})$ to another specified state, and in this case,

²⁷ Fasoi, Pavlopoulos, and Konstantinidou, "Computational Authorship Analysis of Homeric Language." p. 80

²⁸ Eder, Maciej, "Style-Markers in Authorship Attribution." p. 101

²⁹ Fasoi, Pavlopoulos, and Konstantinidou, "Computational Authorship Analysis of Homeric Language." p. 80

produces the “maximum likelihood estimate,” which would be the model’s predicted successor state.

The characteristic mathematical models, probability distributions calculated with maximum likelihood estimations, of the *Iliad* and the *Odyssey* are created in a way that gives “frequent phenomena” disproportionate influence over the distribution’s shape. This view of the *Iliad* and *Odyssey* is, notably, not new. Parry himself estimated that both poems were both more than 90% formulaic, and the legitimate history attached to this hypothesis makes Homeric texts highly amenable to techniques like Markov models and MLE.³⁰ Accepting the basic premise that oral-formulaic studies on Homeric texts have accurately documented the prevalence of formula throughout both poems, using MLE to model each poem’s probability distribution will yield more accurate quantitative representations of poetic structure in areas of the text that are highly formulaic. Recurrent “syntactic-metrical patterns localized at preferred places in the verse” would be accurately predicted by the model, because more frequent occurrences would cause a higher quantity in the numerator, yielding higher probability values.³¹ This methodology thus (unwittingly) draws upon and upholds Parry and Lord’s rationale that poesis is a system made of a “great number of expressions... such that a metrically appropriate expression can be found for ... any compositional necessity.”³² Under the joint imaginary of oral-formulaic theory and n-gram statistical language modeling, a narrative’s essential, discriminating characteristics are contained within the syntactic scaffolding of the text, and constitute the archetypal contours of Homer’s stylistic fingerprint. Through the production and reproduction of this logic in authorship analysis studies, computational linguistics produces a making relationship between author and narrative concrete through a mathematical, rather than literary, lens. And now, the representation of Homer in model form supersedes the oral tradition that authored the epic narrative itself.

Representing Homer as a probabilistic model thus realizes the prediction from oral-formulaic studies; that oral texts necessitate notions of unity different from those for written texts with a single author. Parry’s observations of “composition-in-performance” techniques in South Slavic oral epic poetry, where singers used formulaic structures and substitutions to convey essential ideas in the narratives, led him to conclude that there was “no fixed text.”³³ Instead, formulaic language allowed performers to “easily fit [their] ideas into the metrical framework of the verse” while communicating a relatively “stable set of themes” comprising the basis of the oral narrative.³⁴ The stylistic fingerprint that tri-gram modeling constructs codifies the formulas as the defining structure of a narrative, and privileges the formula over the composer as the mechanism enforcing stability upon the narrative’s meaning.

In a way, this representation brings the Homeric voice outside of the actual oral tradition that brought about its existence. The way that it privileges formulaic structure as a unique identifier also accords with the conclusion Parry made; that “it is beyond the capability of an individual poet to design multiple interlocking systems of such complexity.”³⁵ Parry’s statement creates an understanding of narrative unity that revolves around formulaic continuity. But this version of poetic unity doesn’t just disregard historical tradition, it assumes that historical tradition is accounted for within the written documents observing those patterns and formulae. In other words, the probability model doesn’t just capture the shape of the poet’s voice. It captures

³⁰ Parry himself estimated that the *Iliad* and *Odyssey* were both over 90% formulaic. (probably russo too)

³¹ Russo, “Formula.” p. 1

³² Elmer, “Oral-Formulaic Theory.” p. 1

³³ *Ibid*, p. 1

³⁴ *Ibid*, p. 1

³⁵ *Ibid*, p. 1

all of their voices, so long as those voices are truly Homeric. By allowing that probability model to evoke the stylistic fingerprint's power, the representation embodies the static parts of that oral tradition while abstracting away the need for any recognition of the historical conditions that created the epic poem. It is the system, rather than the tradition, that becomes the "real" identifier of the poems. Distinguishing between "real" and anomalous components in Homeric epics becomes a question of locating deviations from dominant formulas in the poems, a paradigm of inclusion and exclusion that both requires and produces a stylistic fingerprint perceptible only through quantitative means.

The articulation that an authorial "fingerprint" or schema is embodied in a structure, clarifies both why this metaphor is appropriate given the assumptions, and why the "fingerprint" is such a powerful signifier of authenticity. Fingerprints are used as unique identifiers not simply because statistical analysis has concluded that differences between human fingerpads are significant enough to differentiate between their owners, but because they are not easily mutable. The vast majority of us die with fingerprints nearly identical to the ones we were born with — they are not just unique, they are largely stable across time. In many ways, using "fingerprints" as a metaphor for the probability model describing the system of syntactic-metrical formulae and patterns in the Homeric epics is surprisingly appropriate. Both concepts propose that their truth value is impervious to time, perfectly unique, and always intelligible through observation and quantification. And though each concept may invoke scientific imaginaries about how the self can be measured and identified, those imaginaries have ramifications for how we understand the world and ourselves. In transitioning from an oral tradition to a probability model, Homer becomes more akin to an object than a historical subject; the question becomes "what is Homer's voice?" instead of "who is Homer, and how can we know him?". The weight of scientific validity, combined with the allure of resolving Classic's unanswerable question, makes this image of the poet's voice particularly influential in the age of digital humanities. To an extent, this transition even necessitates destroying that originating oral tradition as an important part of Homer's voice. It seems as though the model of the epic poems kills Homer, a cyborg reborn in his place.

Part 3: Homer, the informational cyborg.

The fingerprint that results from the study is a perplexing sign. In treating the construction of Homer's stylistic fingerprint with such specificity, I have ignored the larger question. The fingerprint, and the practices that bring it into being, certainly carry a set of assumptions consonant with oral-formulaic theory and stylometry writ large. How could modern (and the project of authorship attribution is fundamentally a modernist one) techniques in computational linguistics *not* inherit from this genealogy? The idea that Homeric epic poetry is a function of metrical patterns makes it amenable to computational methods such as tri-gram analysis. So, establishing how Fasoï et al.'s study builds upon pre-existing frameworks for understanding Homeric poetry is essential, but doesn't explain why the application of these methods and ideas marks a departure from the status quo—and thus fails to provide an answer as to why the fingerprint matters at all.

While this authorship analysis project may rely heavily on a combination of methods and ideas from stylometry as well as oral-formulaic theory, it still meaningfully breaks from these paradigms. Their primary focus is one difference separating stylometry and oral-formulaic theory from computational linguistics authorship studies. Parry's oral-formulaic theory has undergone significant change and nuance in the approximately 90 years since it was first proposed, yet still maintains (to varying extents) that the defining themes in Homeric epic poems are inextricable from the syntactic formulae that construct poetic architecture housing narrative meaning. Similarly, stylometry "is the study of writing style based on linguistic features" and is best understood as a set of methods used to describe how meaning arises from the combination of predetermined structure and improvisation.³⁶ Stylistic unity, in these fields, is still secondary to issues like narrative meaning. By contrast, authorship studies use stylometric techniques to hypothesize relationships between texts and authors, shifting a descriptive undertaking to a generative project. The object of study ceases to become the texts themselves, and analyzing the specific patterns to describe and understand style as exhibited in a work becomes less important. Style becomes a set of linguistic patterns acting as opaque identifiers rather than an object inviting interpretation.

Returning to the fingerprint metaphor, the difference between style in stylometry and style in authorship attribution is akin to the difference between fingerprinting for forensic analysis versus horoscopes and astrology. In horoscopes and astrology, the qualities of an observed fingerprint — its shape, size, contours, and unique features — indicate properties intrinsic to the person they are attached to. In forensic investigations, fingerprints are gathered from evidence and compared against fingerprints from suspects to create matches, or "hits." Stylometry focuses on meaning, while authorship studies focus on relationships. Stylometry describes interiors; authorship studies draw relationships between exteriors.

This difference emerges, in part, from authorship studies' reliance on a much more robust notion of stylistic unity. Ideas about stylistic unity lie at the center of both stylometry and authorship studies. After all, the fingerprint and "unconscious personal stamp" metaphors both originate from stylometry. The previous section articulates how the frequentist school of statistical thought creates an image of the fingerprint whose defining characteristics are the linguistic phenomena that appear most often in the training corpus. In a sense, these most frequent phenomena define the center of the fingerprint and comprise its core being and identity. But in authorship attribution, the likely presence of multiple stylistic fingerprints in a corpus and

³⁶ Dauber, Overdorf, and Greenstadt, "Stylometric Authorship Attribution of Collaborative Documents."

the imperative to map particular parts of that corpora to separate authors means there is an emergent need to differentiate between these stylistic fingerprints. Fingerprints, for these applications, don't just need to have a center. They must also be well-defined at their edges.

The definition of stylistic unity needs to be even stronger if Fasoi et al. are to differentiate between distinct rhapsodist voices who contributed to the Homeric texts, particularly because of the similarities between their project and unsupervised machine learning. Unlike authorship attribution or verification studies, there is no set of rhapsodists whose other works are known — the *Iliad* and *Odyssey* are not labeled text data attached to the names of their original composers, they are unlabeled datasets, and so any project that uses patterns to identify unknown authors is unsupervised by formal arbiters of truth. Typically, authorship attribution studies start with a group of known authors, and use texts whose authorship is already well-verified to train their language models and construct a “stylistic fingerprint.” These language models can be used as reliable indicators of linguistic proximity between an author's known works and other anonymous works. But the contributions of specific rhapsodists, who likely influenced the shape of the poems well before the advent of writing, have no labeled corpora that can be used to pre-construct a stylistic fingerprint and compare it against various linguistic expressions in the Homeric texts.

The absence of corpora attributable to contributing rhapsodists illuminates exactly how ambitious this study is. To correctly perceive both the number of distinct voices present in the poems, as well as to map out which portions of the text can be attributed to one voice over another, they must assume that stylistic fingerprints, as well-defined at their edges as they are at the center, already exist in the text. To go a step further and claim that disparate linguistic styles can be tied to separate individuals, embodied authors whose identities are otherwise unknown require conceptualizing the “stylistic fingerprint” in its most literal form. The ideological justifications necessary for such an operation indicate that “stylistic unity” isn't just a product or indicator of metrical convention; it is imagined to encompass the author's mind and enforce metric conventions in a near-legalistic manner.

Understanding authorship analyses of the sort carried out by Fasoi et al. as a constitutive act helps demonstrate what this “fingerprinting” really is. Proving that stylistic fingerprints exist is not their objective, nor would they be able to prove such a thing. To do so, they would need to devise a test differentiating known authors based on their corpora, which is impossible when there are no corpora or known rhapsodists who may have composed different parts of Homer's epic poems. Instead, they set out, aided by statistics and computational devices, to perceive stylistic fingerprints imprinted long ago upon the *Iliad* and *Odyssey*'s linguistic structure. But without proof that these fingerprints existed before their study, the procedures Fasoi et al. use to “perceive” them can only be seen as procedures that extract individuals from the text.

The basic fact that a model of the text is produced under the presumption that it can speak for some important realities of an authorial style means that we need to treat the statistical stylistic fingerprint as a product of individualizing processes. These processes imagine an authorial self who preserves themselves through writing. Given that we cannot verify the existence of the object that Fasoi et al. set out to perceive, it is appropriate to take up Nikolas Rose's work on procedures of individualization in technological and disciplinary settings. Individualization, he argues, consists of two parts. Individuals must be well-defined on the inside and outside. To become this way, their interiors must be inscribed with an individual identity, which can then be used externally to separate individuals from one another. Thus, we must explore two relevant dimensions of the fingerprint to answer the questions above.

First are the quantitative structures and philosophical assumptions regulating the procedures used to fabricate a statistical-stylistic fingerprint and inscribe the sign of an author with a stylistic identity. These procedures essentially construct the presumed author's (since the author is yet unknown, their existence must be presumed) internal identity, particularly because style is thought of as a product of the unconscious. The second dimension of fingerprinting, while still concerned with constructing the individual, concerns how fingerprints are created to be easily differentiated. This is necessary for the fingerprint to define a class of individuals who can be compared and related along the same axes.

In an attempt to understand how modern ideas of the writing subject, stylistic expression, and authorship are invoked and modified in this computational linguistics study, this section aims to handle the first dimension of fingerprinting concerned with the author's interior self. Where the previous two chapters focused primarily on how ideas about stylistic unity and linguistic patterns are used to construct the stylistic fingerprint, this chapter will argue that using stylistic fingerprints as indicators for the presence of embodied human authors rests on a distinctly modern conceptualization of the relationship between expression and self. In pursuit of understanding how both the metaphor and mathematical construction of the statistical-stylistic fingerprint invoke and modify existing paradigms around the writing subject, stylistic expression, and authorship, we must entertain the following questions. If an implicit "self" is attached to the statistical-stylistic fingerprint, we must ask — who are they? How is the materiality of their existence demarcated by the statistical and philosophical conditions that produced the statistical-stylistic fingerprint out of atomized language information? How is their relationship to the signifying "fingerprint" formulated, and how does attaching the fingerprint sign to a real body situate it in the realm of authorial theory?

To examine these lines of inquiry, I will continue performing a media-specific analysis of the statistical-stylistic fingerprint. This section, however, will focus on how the quantitative assumptions surrounding the fingerprint (as opposed to the mathematical assumptions within it) interact with broader ontologies of personality and the self. This will involve utilizing Colin Koopman's theory of the informational person and Nikolas Rose's work in "Inventing Ourselves: Psychology, Power, and Personhood" to articulate how the study doesn't simply carry out "authorship analysis" but produces a new cyborg modality of "being" in the realm belonging to author, text, and reader.

The very metaphor of the "fingerprint" turns on the idea that a particular "self" is literally embodied within the text, and characterizes an essential belief in computational authorship studies. This conjecture is articulated clearly by Eder and Maciej in "Style-Markers in Authorship Attribution," where they note that the field operates on the idea that style is "personal and characteristic," which lends credence to the idea that careful analysis of stylistic markers can "afford certainty as to [the] identity" of the writer by way of their personality.³⁷ If linguistic style is contingent on personal identity, the text becomes the tissue encapsulating the expressed "self" within their work's tactile sounds and shapes.

Moreover, an uncritical belief in this dynamic between personal identity and stylistic expression justifies the single-minded pursuit of an author's "stylistic fingerprint" and, by extension, the self who can claim ownership of that fingerprint. Those computational linguists who subscribe to this image of the author, the writing and speaking subject whose "unconscious personal stamp" betrays their lasting presence within the text, are bound first and foremost to

³⁷ Eder, Maciej, "Style-Markers in Authorship Attribution." p. 103.

questions regarding the best methods to extract “as many unique elements of style as possible.”³⁸ Under this research paradigm, questions regarding what of the author to measure — and how — must always take precedence. And in the scurry to measure, delimit, and identify the author, questions regarding stylistic measurements come at the cost of asking: who, exactly, are we measuring? Are they even really here?

An unusual philosophy of the unconscious quickly emerges from the paradigm and imperatives outlined above when addressing the interior structure of the fingerprint. Eder’s account of stylometry’s origin indicates a clear belief that the author’s unconscious state produces a distinct, unique style whose expressions cannot be intentionally modified and thus justifies using quantitative style measurements to delineate between writing subjects. The diction of an “unconscious personal stamp” implies the existence of some unconscious, a sentient but inaccessible awareness, which stamps a unique imprint of itself onto the text.

Even though the study’s authors, Fasoi et al., never explicitly call out the unconscious or statistical stylistic fingerprint as something they seek, their project relies on these ideas. The idea that the deepest parts of the human soul are externalized through speech has deep roots in religious rituals and mysticism (for example, Catholic rites of confession). Still, modern links between speech and being have often found footing in various forms of psychotherapy. Rose writes about the “techne” of the confessional in psychology and explains that “the ‘I’ [who] speaks is... identical with the ‘I’ whose feelings, wishes, anxieties, and fears are articulated.” Confession thus becomes an act of “constructing an identity,” and in an era where psychology has become a dominant authority on the self and behavior, identity projects become “bound to the languages and norms of psychological expertise.” (96) Though Homer does not construct a confessional, or divulge any internal monologue in either poem, the link between the true self and language is maintained by the discourse relating the unconscious with style. Even without explicitly invoking any form of self-confession, if Fasoi et al. are to successfully argue that style can identify the set of distinct voices that contributed to a text and that those creativities are housed within a single human body, then they must imagine that style is the unconscious, perhaps accidental, confession of individual identity.

In trying to reconstruct a sort of personality from atomized poetry data, the perspective of computational linguists chasing the elusive Homeric fingerprint thus exhibits the core ideas that Koopman argues characterize the “informational person” and its associated notions of the unconscious. The statistical-stylistic fingerprint that Fasoi et al. set out to construct is not simply a probability model representing the relative frequencies with which certain character sequences occur; if this were the case, the notion of “authorship” would not enter into their quantitative textual analysis, they would only care about the text. Instead, linguistic frequency is made to represent style as a proxy for authorial origin.

Nor does style belong solely to the fingerprint, because the stylistic qualities of a text are supposed to be a product of an unconscious mind belonging to the author who wrote those letters one by one. Even this chain of conceptual connections must be taken for what it is not in isolation, but understood in the context that gave rise to this theory — the linking of information and personality is not natural, and not arbitrary, but a product of historical biopower projects whose legacy now presents itself in the informational apparatuses of personality psychology, an apparatus that claims an ability to extract and in a sense, reproduce, the very unconscious of the author.

³⁸ Ibid, p. 103.

Though present in the paradigm that animates the unconscious via the statistical-stylistic fingerprint, informal personality psychology forms a culturally influential conceptualization of the self. Colin Koopman's writing on the informational argues that the co-productive relationship between personality psychology and personal data (particularly language data) is now so intertwined that humans have become "cyborgs."³⁹ Today, we are all cyborgs in the sense that the data attached to a "person" — our birthdays, names, social media images and language, credit card numbers, and transactions — are just as constitutive to the pragmatics of modern existence as our experiences as "minded and embodied creatures."⁴⁰ This means that to exist in the time of big data is to assume the character of a cyborg, and take on a hybrid identity comprised of data and physical body. It is our data which constitutes our "personalities" and, more broadly, our "psychologies."⁴¹ Practically, there is no way to escape these conditions without totally withdrawing from common society. Koopman recounts a widely accepted claim by personality researcher David Funder that "an individual's personality consists of any characteristic pattern of behavior, thought, or emotional experience that exhibits relative consistency across time and situations."⁴² Though not explicitly noted, the patterns in human actions and experiences that allegedly constitute a stable personality cannot be identified without supposing that the creation of, and the actual acts of creating, datafied representations of those actions and experiences is appropriate and somewhat natural — even for the elusive figure that is Homer.

Designating the unconscious as a text's style-generating being, as computational linguistics authorship studies do, marks authorship attribution projects as attempts to extract informational persons. In the words of Eder and Maciej, the "most desired style-markers are those undiscoverable to the naked eye, and thus beyond authorial control," as their locus of origin within the unconscious makes them "resistant to imitation, plagiarism, and parody."⁴³ Here, the separation between the conscious and unconscious mind expresses stylistic patterns uniquely identifying the author-name with an authorial unconscious.

The idea that the unconscious is the primary cause behind stylistic differences between authors supposedly shields an author's unique style from discovery and reproduction. The authorial unconscious, inaccessible to every other consciousness, including its own, requires something other than a conscious being to detect the presence it articulates through subtle linguistic expressions. In authorship attribution and even in some areas of stylometry, practitioners seem to believe that the authorial unconscious is expressed so subtly that it can only be perceived through mathematical, technological means. However, applying these statistical technologies across the entire text doesn't just identify the authorial-unconscious. Before the author can even be identified, that style must be mathematically described outside the text — the traces of the unconscious must be measured and compiled under a statistical model. This process lifts the authorial unconscious outside the text, extracting it for a new form of technified embodiment.

It could seem that extracting a statistical-stylistic fingerprint from a text contradicts the supposed invisibility of an author's "true," unique, discerning style. Examining the previous quote more deeply, we see that the "unique elements of style" are only "undiscoverable to the naked eye" — and it has been quite some time since the "naked eye" was the only way to see

³⁹ Ibid, p. 8.

⁴⁰ Ibid, p. 8.

⁴¹ Ibid, p. 69.

⁴² Ibid, p. 70.

⁴³ Ibid, p.103.

the world, or to make subjects intelligible.⁴⁴ The phrasing around “naked eye” suggests that collecting unique features is possible, so long as one has the proper instruments to see and extract those stylistic elements. Thus, an author’s unique style and defining intangible qualities are only protected from imaging and reproduction under the magnifying glass of manual interpretation. But when seen through the microscope of information technology, the textual facade veiling the true author fractures into a multitude of atoms, nothing more than computable characters floating in space, ready to be configured into a cognizable form.

The implication that a mechanically-assisted mode of “seeing” the writing subject within the text is necessary indicates that the relationship between the text, unconscious, and statistical-stylistic fingerprint are mediated by informational apparatuses that produce quantitatively intelligible subjects. To understand how computational techniques don’t just presuppose but also produce a particular kind of cyborg self, it is appropriate to take up Nikolas Rose’s approach to locating the subject in “a complex of apparatuses, practices, machinations, and assemblages within which human being has been fabricated.”⁴⁵ In the most literal sense, applying this framework is appropriate because the statistical-stylistic fingerprint doesn’t exist *a priori* within a text. Rather, it is mechanically fabricated by information processing techniques that require inputs and produce outputs. As such, the cyborg Homer and his distinguishing statistical-stylistic fingerprint can be understood as a subject or self produced by a system regulated under the logic of the informational self, which relies on making individuals intelligible through the information they produce.

The fact that representing persons through informational means under the guise of “personality” seems natural in any capacity necessitates a deeper investigation of the genealogy surrounding such techniques. Koopman describes a history of “personality informatics,” whereby two notable trends in the “human sciences” — abnormal psychology research and the growing scientific authority granted to quantitative assessments of the human condition — form the grounds for widespread self-descriptions phrased in “the vernacular of trait psychology” (at least in American culture).⁴⁶ In short, we have become conscious of our-selves under a schema of personality traits. Therefore, it feels unproblematic to extend the same schema of understanding to the authorial self and its unconscious. Authors, like every other person, have personalities. If language is a behavior that is in no small part determined by who we are — our unconscious ties, our experiences, our identity — then language, like every other behavior, can be utilized by data technologies to construct a composite image of who we are.

At its inception during the late 1800s and early 1900s, according to Koopman, a view of personhood as “personality” departed sharply from the dominant view of self-hood as a function of “character.” “Character” was a positive, unitary element of the self that “one either did or did not have,” and structured a binary of good and bad selves — it was a “good thing to have [character],” and lack of character reflected poorly on an individual.⁴⁷ Unlike character, “personality was essentially diverse, an assembly of various tendencies,” and something that everyone had.⁴⁸ Figuring personality as something that was always composite, personality traits protracted psychological understandings of the self beyond simple binaries of possession or lack of character, or divisions between essentially good and bad persons. The ease with which the

⁴⁴ Ibid, p.103.

⁴⁵ Koopman, *How We Became Our Data*. p.68.

⁴⁶ Ibid, p. 70.

⁴⁷ Ibid, p. 72.

⁴⁸ Ibid, p. 72.

“new scientists of personality” assumed their role as “technicians of the measure of personality traits” through “a bevy of humble devices, instruments” like questionnaires and algorithmic processing techniques is due to the composite structure of personality.⁴⁹ As composites of experience and action, personalities could “assume different shapes in different persons,” which made them “amenable to quantitative empirical scrutiny” through measurement.⁵⁰ Now, emergent differences between individuals can be “descriptively interrogated” not simply to separate good from bad, but to create striated spaces of individuals who are differentiated via the interpretation of their associated personality data.⁵¹ These selves, existing amidst systems that constantly compute their psychological proximity to others, inhabit an identity of the “informational person.”

One important moment from this history is how the unconscious subject is figured at the genesis of personality psychology. This moment is useful because it clarifies what stylistic analysis projects in authorship attribution constitute through their methods. Computational linguists seem to assume that there is an unconscious subject who not only created a text, but left so many essential parts of itself within that text that an informational apparatus like tri-gram analysis could apprehend unconscious schemas of style differentiating one author from another. One early writer on personality and the unconscious was Morton Prince, who articulated that the unconscious leaves vestigial records of itself in the expressions of the conscious self. To him, “personality was now “a complex affair” definable in terms of combinations of present and past states of consciousness.”⁵² This view corresponded with the “dominant cultural imaginary” of the unconscious at the time that Prince wrote — that ideas pass through the mind and become dormant, but their presence persists by way of imprinting their “physiological” records within ourselves.⁵³ Prince postulated that “neurograms,” or the “modified neural structure resulting from [mental] activity,” could be identified from the remnants of their expression (which ostensibly includes language data) that he called “neurogrammatic storage complexes.”⁵⁴ He advocated that to understand an individual’s personality, one could not simply resort to hermeneutic psychoanalysis to understand the dormant psyche, but must extract the information that comprised those neurogrammatic storage complexes. Only through these procedures could one apprehend an individual’s personality, and a story of their evolving unconscious as told through data.

The origins of personality psychology and the unconscious don’t just influence the way that theorists must frame the “fingerprint” analogy — it is the paradigm that justifies applying the apparatus of informational personality psychology to apprehend a Homeric fingerprint in the first place. Through the apparatus of informational personality psychology, studies in this vein don’t simply try to identify the writing subject to carry out classification. Their methodology indicates the latent belief that the unconscious mind is always embodied within the text, and can therefore be re-embodied outside the text. They seek not just to characterize, but extract, the unconscious mind who maintains a presence within the syntactic structure of a work/text. In a sense, these studies reverse engineer the authorial unconscious, reconstituting the deepest nethers of “Homer’s” mind from the atomized information it once expressed in the form of poetry. Stylistic unity is no longer a simple product of the authorial unconscious; we imagine it must

⁴⁹ Ibid, p. 71.

⁵⁰ Ibid, p. 72.

⁵¹ Ibid, p. 72.

⁵² Ibid, p. 73.

⁵³ Ibid, p. 73.

⁵⁴ Ibid, p. 73.

constitute the entire fabric of the author's unconscious. Under this schema, the text becomes like a cryo-chamber preserving Homer's unconscious, the mystical mode of his poesis frozen in time until technology capable of revitalizing its generative power emerges.

Part 4: Population-Level Statistics for the Cyborg Unconscious

Having addressed the first dimension of the fingerprint (its “interior”) we must now turn to the second dimension: relations between statistical stylistic fingerprints. Since each statistical stylistic fingerprint is constructed as a purely mathematical object, relationships between “cyborg” Homers and their respective “unconsciousnesses” are always constructed via quantifying and calculating procedures. And as these cyborgs are bound to a strictly quantitative existence, these relationships must always be formulated by, and situated within, the technical apparatuses that regulate the statistical-stylistic fingerprints standing in for Homer’s authorial unconscious.

Examining how the relationships between cyborg-author objects come into existence must be carried out with great care, because it speaks to the fundamental objective of “authorship attribution and analysis” in computational linguistics: to differentiate between authors, their “real” work, and mere imitations. But if we are to understand how relations of difference between authors are constructed, we must not depart from our understanding of the statistical-stylistic fingerprint’s “interior,” where we articulated how informational personality psychology, and its corresponding genealogy, actualize statistical stylistic fingerprints as psychological individuals. Understanding statistical stylistic fingerprints not simply as representations of an author’s style, but to a certain extent, a re-instantiation of the elusive authorial unconscious itself (phrased in terms of a probability model) has the following implication. Attempts to differentiate stylistic models of an author from one another, in order to identify the “true” author, are always also attempting to differentiate between the individual psychologies of the minds who speak through a text. That is evident, whether or not the statisticians performing these analyses acknowledge it (or more often, as is the case with Fasoi et al, do not).

This is how computational linguistics inserts itself into the enduring debates on authorship, and also why the article by Fasoi et al is not a mere aberration, oddity, or anomaly in the field. It is only on the basis of informational personality psychology that computational linguistics can imagine that there exists some underlying quantitative structure or model of the psyche that governs authorial expression, and carry out projects seeking to apprehend that model. So the very operations carried out under this assumption bring those technical apparatuses, which imagine themselves as approaching some fundamental ground truth, into being. Building this technical framework for the authorial unconscious subsequently introduces “scientific objectivity” into the equation formulated as an answer to the question — what is an author?

In previous parts of this paper, I analyzed the statistical model, stylistic fingerprint, and subsequent notions of the unconscious on the level of the “individual.” But the study in question does not produce a single “statistical stylistic fingerprint” for Homer — it produces exactly ninety-six cyborg Homers in the form of probability models.⁵⁵ Standing in for the portions of the text they were trained on, the language models are mathematical objects that can be used to predict the likelihood that the test data (in this case, the portions of the text that were excluded from the tri-gram analysis used to construct the model) occurs using MLE, which gives the probability that their specific cyborg Homer would write the test portion of the text. The larger population of cyborgs can be subdivided into two classes, which I will name the exclude-one cyborgs and the single-book cyborgs. The exclude-one cyborgs are probability models (statistical

⁵⁵ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” pp. 81-82

stylistic fingerprints) constructed by applying tri-gram analysis to excerpts drawn from “all but one Book of the same poem.”⁵⁶ Through this methodology, the computational linguists “developed twenty-four language models per poem,”⁵⁷ and across both poems, giving birth to the forty-eight, exclude-one cyborg Homers. The single-book cyborg Homers, by contrast, are probability models trained on excerpts drawn randomly from one single book. Over the two books, Fasoi et al create twenty-four single-book Homers per poem. Counting our cyborg Homers, there are ninety-six altogether.

These cyborg Homers are, in a word, mutilated. Their bodies, the material topography of their constitutive probability models, are patchworks of skin grafts, and full of holes. They are constituted by the texts included within training data, but the defining differences between cyborgs are a function of that which they lack — the metrical patterns, formulae, epithets and meanings found only in the excluded portions of the text. The first dimension of information loss proceeds obviously from the fact that none of the models are constituted by the combined textual bodies of the *Iliad* and *Odyssey*, which would (theoretically) be the most complete representation of the Homeric unconscious. But the second dimension is more subtle, a function of the sampling techniques employed to avoid overfitting Homer’s expressed unconscious mind. Fasoi et al writes that they “only used the text [data] between the 2,600 and the 9,600 characters per Book, to train each language model, in order to control the varying length of the Books.”⁵⁸ By using the same amount of text data from each book, the authors indicate their latent assumption that each book, regardless of length, is an equally indicative confession of the Homeric unconscious, and that including too much of one confession would bias their model in favor of a single book’s content, leading to overfitting and a poorer representation of Homer. Moreover, they justified excluding each book’s “beginnings and ends” from all training sets on the basis that these portions of the text “tend to include formulaic characteristics of opening or concluding a narration” that would have guided “Alexandrian grammarians dividing the Books” somewhat arbitrarily.⁵⁹ The explicitly stated intention behind this method was to “soften the effects from the arbitrary Book division,” which could potentially bias their model away from an accurate representation.⁶⁰ But in making this decision, they erect a barrier categorizing formulae that are characteristic of Homer. Thus emerges a schism between forms contained within the body of the narrative, and ones that are not. These strange beings, mathematical objects contorted into a representation of too many things — a text, a style of metrical patterns, an oral tradition, an idea of Homer and the mind — are born permanently disfigured by design.

No doubt, the practices for constructing this cyborg population, and the produced cyborgs themselves, seem antithetical to the picture and purpose of the statistical stylistic fingerprint I have presented prior. If the statistical-stylistic fingerprint is really meant to reproduce an entire unconscious, why not construct it with the maximum amount of data available? Why is overfitting such a concern, if it creates the most accurate representation of the Homeric unconscious speaking to us from antiquity? How should we understand the “truths” of a self that are constructed upon the basis of variously disfigured, incomplete replicas? The answer may lie in the scientific imperative to reject null hypotheses as a means for validating knowledge production, and broader norms around quantitative truth testing. More specifically, examining

⁵⁶ Ibid. p. 81

⁵⁷ Ibid. p. 81

⁵⁸ Ibid p. 81.

⁵⁹ Ibid p. 81.

⁶⁰ Ibid p. 81.

how testing methodologies are applied to differentiate and evaluate a population of cyborgs whose existence is linked to personality psychology informatics, may help unfold the apparent contradiction between ideal reproductions of the speaking unconscious and their reality as incomplete representations. Thus the following section will use Nikolas Rose's account of psychological tests, as a framework to make sense of this emerging population of authorial cyborgs, and by analyzing the way that Fasoï et al's use of quantitative tests differentiates individual cyborg-models from one another, attempt to unpack why so many incomplete versions of Homer's statistical-stylistic fingerprint are created.

Though the study's stated goal may be to ascertain whether specific subsets of books display higher or lower levels of linguistic integration with the rest of the text, actually performing this analysis involves taking measurements of model performance for each fingerprint, and using those measurements to compare models (which are the statistical-stylistic fingerprints created from varying subsets of each Homeric poem). Given the way that the authors pursue their primary goal within the study, and computational authorship attribution techniques writ large, it is clear that these statistical stylistic fingerprints are not created for their own sake. They are created to be analyzed, tested, and compared against "test" data and one another, so that the authors can progress toward a "truthful" vision of the author and their unique style. In short, enacting measures of "linguistic integration" between subsections of Homeric texts visualizes, inscribes, and assesses differences — and by extension, proximities — to make individual members of a larger population of cyborg Homers legible. This "ensemble of processes," which are materialized into a routine, and operates over a set of psychological "individuals," identifies the study as a psychological test as articulated by Nikolas Rose. Adopting this framework allows us to develop an understanding not simply of how the individual fingerprint is constructed, but how certain re-instantiations of Homer come into a privileged status. This ranking and organizing of the cyborg population thus influence notions of a "true" Homer, and more importantly, what it means to be a "real" author.

Taking up Michel Foucault's "hypothesis that [the psy disciplines]" reverse "the political axis of individualization,"⁶¹ Rose argues that "the discipline of psychology took shape around the problem of inventing... new techniques of individualization."⁶² Beginning from what he claims is the "first contribution of psychology to the project of individualization," he uses the intelligence test to explain how psychological tests, in general, make behavioral variance between humans legible and inscribable upon individuals, and thus constitutes a "[technique] for the disciplining of human difference."⁶³ Intelligence testing was a novel scientific invention at the time, because prior attempts to understand undesirable social behaviors, such as academically deficient Victorian-era schoolchildren,⁶⁴ had "first scrutinized the body as a means of diagnosis" under a medical lens.⁶⁵ But it was quickly discovered that the gaze of a single doctor was "difficult to align" with "the requirements of the institution."⁶⁶ To satisfy the demand that

⁶¹ Foucault 1977, 191: via disciplinary methods, visibility is transformed from a privilege reserved for those in power to a means of control and domination exerted over the common subject.

⁶² Rose, *Inventing Ourselves*. "An Individualizing Technology," p. 109.

⁶³ *Ibid*, p. 105, 109.

⁶⁴ Academic failure was considered "undesirable" because the school was yet another apparatus for disciplining human difference. As Rose describes it, these children not only constituted a financial burden on the authorities who ran educational programs, but also, posed a "concern to those who regarded the school as a vital apparatus of moralization."

⁶⁵ *Ibid*, 109.

⁶⁶ *Ibid*, 109.

disciplinary techniques be scalable, Galton invented the psychological intelligence test, which reduced difference to order through processes of normalization.⁶⁷ By creating a technique (questionnaires⁶⁸) that rendered differences in intelligence quantitatively, Galton also made human intelligence comparable, and therefore normalizable as “human variability” could be rendered statistically “through the use of the normal curve.”⁶⁹ Moreover, he circumvented a need for the body “as the diagnostic intermediary between conduct and the psyche,” as the physicality of the questionnaire allowed institutions to visualize the behavioral characteristics of their subjects without them being directly present.⁷⁰ This set of procedures made the intellect “manageable,” because “intellectual abilities could be construed as a single dimension whose variation across the population was distributed according to precise statistical laws,” and thus “the capacity of any given individual could be established in terms of his or her position within this distribution.”⁷¹

Rose’s focus on the intelligence test, and its application to the development of child behavioral psychology, showcases it as an example that demonstrates the defining characteristics of a psychological assessment. He argues that “psychological assessments [all] use essentially the same techniques for quantifying all the qualities of the human soul.”⁷² Through his analysis, he identifies three fundamental procedures that comprise the psychological test: visualization, inscription, and assessment.

Visualization techniques produce scientific “individualizing observations,” fundamentally measurements that come in forms like “images, graphs, and numbers.”⁷³ Rose proposes that these metrics are “traces” of “3D” subjects. Visualization entails applying a common lens to the subjects of interest, which “makes the person stable through constructing a perceptual system.” Though the “real” referents of these observations are “ephemeral” and thus “hard to perceive in any stable manner,” visualizations “render the mobile and confusing manifold of the sensible world into a cognizable field” by making those persons objective, at least momentarily.⁷⁴ Visualizations thus act as a normalizing force between individuals, creating a standard scale for comparison.

It is these traces which enable persons to be inscribed with an individual identity. Rose defines inscriptions as technical operations that “render ephemeral phenomena into stable forms, which can be repeatedly examined and accumulated over time” on the level of the individual.⁷⁵ He further explains that “inscriptions [of individuals] are ideally of two dimensions and amenable to combination in a single visual field without variation or distortion by point of view.”⁷⁶ Though statistics and machine learning have evolved to a degree where observations can typically be represented easily in at least three dimensions and allow much higher dimensional data to be transformed for easy visualization,⁷⁷ Rose’s fundamental point remains relevant. By

⁶⁷ Ibid, 109.

⁶⁸ Intelligence tests were administered via questionnaires, which were primarily affirmative/negative answer questionnaires.

⁶⁹ Ibid, 109.

⁷⁰ Ibid, 109.

⁷¹ Ibid, 110.

⁷² Ibid, 112.

⁷³ Ibid, 108.

⁷⁴ Ibid, p. 106.

⁷⁵ Ibid, p. 106.

⁷⁶ Ibid, p. 108.

⁷⁷ Much of the data that models are built off of, exist in high-dimensional feature space. Each individual could be inscribed with upwards of fifty, a hundred, or even a thousand features, depending on the size

making visualizations attributable to bodies, techniques of inscription stabilize persons as individuals. This stabilization “enables them to be placed side by side,” preparing them for techniques of analysis and comparison.⁷⁸

The final stage in the process of individualization is assessment. Once measurements of embodied persons have been taken, and those measurements collected as features of an “individual,” individuals could be positioned in relation to other objects defined under a shared mathematical schema and inter-individual differences can be established in quantitative, concrete terms.⁷⁹ When made comparable along the same set of axes, it is not only individuals that become cognizable. Populations emerge as networks between individuals, and those networks are comprised of relationships between population members, relationships conceived in terms used to describe spatial proximity: difference, similarity, and integration.

Many elements in the “Experimental Results” section carry out various procedures of the psychological test upon “psychological” individuals, identifying this set of procedures as a psychological test. Through tri-gram analysis, Fasoi et al suppose that they are unfolding the layered traces of Homer’s unconscious mind, which disguise themselves in metrical form. The relationship between style, fingerprinting, and the psychological unconscious of the writer means that the cyborgs are as much a model of the author’s unconscious as they are a generalization of the text’s linguistic systems. Speaking broadly, any attempt to make the resultant ninety-two cyborgs knowable via quantitative measures, which stabilize them as objects, must be understood as an investigation into phenomena “accorded psychological pertinence,” and psychological phenomena are appropriately subjected to psychological tests.⁸⁰ Even more specifically, the normative relevance of statistical testing methodologies to evaluate the results of those investigations in scientific spaces means that truth-testing the experimental results produced by Fasoi et al exhibits all the defining attributes of a psychological testing apparatus.

Importantly, this apparatus revolves around one metric: Perplexity, or “PPL,” which visualizes and normalizes the behaviors of each cyborg. Fasoi et al used this metric not only “as an indicator of how well a language model predicts an excerpt,” but “as a means to evaluate the origin of a textual excerpt.”⁸¹ In making this extrapolation, they envision that real degrees of “linguistic integration” or divergence between the test and training data is a function of patterns that can be captured in a predictive model’s performance.

To understand how PPL both captures and signifies linguistic integration, it is useful to divulge its mathematical structure. Formally, PPL is “the inverse probability of the test data” given the training data, “normalized” by the number of training units (in this case, the number of tri-grams) in the training data, and is notated as follows:⁸²

and purpose of the dataset. These data are represented in the form of matrices and vectors, and while the human eye cannot easily perceive more than two dimensions, there exist many ways (such as principal component analysis) to reduce the dimensionality of a dataset for visualization without excluding entire informational features.

⁷⁸ Ibid, p. 108.

⁷⁹ Ibid, p. 106.

⁸⁰ Ibid, p. 106.

⁸¹ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 81.

⁸² Campagnola, “Perplexity in Language Models.”

$$\begin{aligned}
 PP(W) &= \frac{1}{P(w_1, w_2, \dots, w_N)^{\frac{1}{N}}} \\
 &= \sqrt[N]{\frac{1}{P(w_1, w_2, \dots, w_N)}}
 \end{aligned}$$

Figure 1: $PP(W)$ indicates the “Perplexity” (PP) of the test data, “ W .” Tri-grams are notated as “ w_i ,” proceeding from $i=1$ to N , where “ N ” is the number of tri-grams in the test corpus.

The expression $P(w_1, w_2, \dots, w_n)$ stands for the probability that all of the w_i tri-grams in the test corpus occur, a proportion calculated using the training data as a predictive model under the Markov assumption. High probabilities indicate that the test data is predicted well by the model, and because the probability value is inverted, small PPL values indicate that the model predicts the test data well, and large PPL values indicate poor model performance on the test set. Using model performance as a proxy for linguistic proximity, Faso et al “employ it as a means to evaluate the origin of a textual excerpt.”⁸³ The normalizing operation, taking $P(w_1, w_2, \dots, w_n)$

to the power of $\frac{1}{N}$, is necessary in order to control for the number of tri-grams.⁸⁴ This is because $P(w_1, w_2, \dots, w_n)$ is really a product of probabilities:

$$P(w_1, w_2, \dots, w_N) = \prod_{i=1}^N P(w_i) = P(w_1) \cdot P(w_2) \cdot \dots \cdot P(w_N)$$

Figure 2: The probability that the test tri-grams occur is equal to the product of probabilities that each individual tri-gram (of N total tri-gram elements in the test data) occurs given the testing corpus.

Thus the power term literally normalizes the quantity in a way that is proportional to typical averaging operations.⁸⁵ It scales the quantitative size of the product of probabilities to the scale of a single probability, which can be seen by applying the natural log to both sides of the equation.⁸⁶

$$\begin{aligned}
 P(w_1, w_2, \dots, w_N)^{1/N} &\propto \ln \left(\prod_{i=1}^N P(w_i) \right)^{\frac{1}{N}} \\
 &\propto \frac{1}{N} \sum_{i=1}^N P(w_i)
 \end{aligned}$$

⁸³ Faso et al., 2021, p. 81

⁸⁴ Campagnola, “Perplexity in Language Models.”

⁸⁵ From Wikipedia, a good definition of what normalization is mathematically and statistically: “normalization of ratings means adjusting values measured on different scales to a notionally common scale, often prior to averaging. In more complicated cases, normalization may refer to more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.” [https://en.wikipedia.org/wiki/Normalization_\(statistics\)](https://en.wikipedia.org/wiki/Normalization_(statistics))

⁸⁶ Campagnola, “Perplexity in Language Models.”

Figure 3: The inverse probability that the test tri-grams occur is proportional to the natural log of the inverse product of all the test tri-gram probabilities. By taking the log of the geometric mean, we see that the inverse probability of the test corpus is proportional to the arithmetic mean of the probabilities.

While normalizing the inverse probability may not matter for comparing models that are all trained on exactly the same size dataset as they were in this study, it produces a universal scale of comparison that doesn't rely on the particular model, or sets of models, whose behavior is being studied. By log rules, we can clearly see that taking the total probability of the test data is proportional to taking the average of the individual probabilities that the test set occurs. Thus it is clear that PPL creates a standardizing metric for evaluating degrees of difference between the model and test set, and by extension, the training and test data.

Given these characteristics and application, PPL emerges as nothing but a visualization technique of the type that Rose described, and a technical key to the testing apparatus that Fasoi et al construct. When computed, PPL scores act as a standardizing lens of model performance that makes complex statistical models cognizable under a behavioral lens, because it reports the same range of values regardless of the type or size of the model. As a metric for scoring a model's predictive power, PPL observes a model's performance — a behavioral term — as it operates over the test data. Not only does it observe behavior, but it produces a one-dimensional, scalar value on a standardizing that both “perceives” and represents a relationship between “real” referents (the training and test corpus, and by extension, the differentiable Homeric unconscious of each corpus). Though model performance might vary significantly depending on the selected test data, PPL scores at least provide a way to make that variance comprehensible in quantitative language.

Though Fasoi et al could have, theoretically, stopped at a single PPL measurement for each model, such an endeavor would have violated academic norms around hypothesis testing, statistical significance, and reliability of results. This is because metrics of model performance, including PPL, will vary naturally depending on the test data.⁸⁷ As a result, PPL measurements are not treated as indications of “true” model performance when taken on their own, but as a means for estimating some unknown PPL value that can speak to that truth; it is a sample “statistic,” not an absolute “parameter” indicating the model's true predictive power. A “parameter” is a “numerical value associated with a population,” but the term “population” in this case can refer to any set of measurements over a single, or set, of objects.⁸⁸ Parameters can never be calculated directly unless the object of analysis can be known and observed without variance in measurements. To estimate the parameter — the truth of a body or population — statisticians building upon Galton's work⁸⁹ established the law of large numbers, which observes that the distribution of “many independent samples is (with high probability) close to the graph of the density of the underlying distribution.” The idea here is that there is a “true” distribution, and “true” parameter, which could come in the form of the “true” mean or another similar metric. Just as measured intellectual capacities (which were inscribed to individuals across populations) could be “construed as a single dimension whose variation across the population was distributed according to... [the] statistical law of large numbers” (a concept that defines Galton's Normal

⁸⁷ Ibid.

⁸⁸ Adhikari, Dennero, and Wagner, *Computational and Inferential Thinking*. 13.1: Estimation. <https://inferentialthinking.com/chapters/13/Estimation.html>

⁸⁹ Namely, the normal curve.

distribution), so too can a sufficiently large set of PPL measurements forming a body or “population” of observed behaviors be evaluated via “acts of comparison” that place each measurement on “the smooth outline of the normal curve.” Under the law of large numbers, a sufficiently large number of PPL measurements taken on a single model can be used to create a sample distribution that accurately estimates the true distribution, and subsequently, the “true” PPL parameter. This value is the ideal visualization that the psychological testing apparatus seeks to inscribe upon a model in order to individualize and distinguish it from the others.

Such idealization of parameterized statistical truths guides the following methodology of the test. In pursuit of that ground-truth metric, Fasoi et al “complement PPL with a notion of statistical significance.”⁹⁰ To create a more normalizing vision of model performance, they steer away from “reporting the PPL of a single excerpt in question,” and instead compute one thousand PPL scores over one hundred randomly sampled excerpts of the test corpora designated to each of the forty-eight exclude-one cyborg models in order to construct 95% confidence intervals (CIs).⁹¹ They then estimate the confidence interval “by ranking the 1,000 PPL scores and reporting the 2.5% and 97.5% quantiles.”⁹² For context, key governing principles in frequentist statistics — the Central Limit Theorem, the resulting Normal distribution, and the Law of Large Numbers — also gave rise to mathematical notions of normality and abnormality in terms of statistical significance. Because the parameter can only be infinitely approximated, claiming that the sample distribution’s median or mean PPL of a model’s performance distribution is a good approximation of the true population parameter would be precise to an erroneous degree.⁹³ But statistical practitioners still want to “provide an interval of estimates for the unknown parameter”⁹⁴ where they can state with some degree of certainty that the parameter must lie. To do this, they developed a technique called “confidence intervals,” which draws on the three principles named above. The Central Limit Theorem (CLT) posits that “the probability distribution of the sum or average of a large random sample drawn with replacement will be roughly normal, regardless of the distribution of the population from which the sample is drawn.”⁹⁵ When taking the limit of the average over all the recorded sample statistics,⁹⁶ and the number of observed sample statistics approaches infinity, then with certainty — probability equal to one — the distribution mean converges to the expected value as projected by the real underlying probability distribution. On the basis of these ideas, “the CLT implies that with

⁹⁰ Fasoi et al., 2021, p. 81.

⁹¹ Ibid, p. 81. Note: There is some ambiguity as to the exact methods employed for constructing the confidence intervals. In the first paragraph of the experimental results section, the authors say that they evaluated model performance on one-hundred excerpts, but a later footnote says that confidence intervals were computed using the “1,000 PPL scores.” This could be a typo, and they meant to write that they evaluated model performance on one thousand excerpts. Or, since they note later in section 5.1 that they took samples of 600 characters, they could have sampled 10 sub-samples for each of the one hundred excerpts. Regardless of sampling technique, it seems that the final confidence intervals would have to be computed using 1000 PPL scores. Otherwise, the sample of PPL measurements would not be large enough for the Central Limit Theorem and Law of Large Numbers to apply, and thus the confidence interval would be invalid methodology.

⁹² Ibid, p. 81.

⁹³ Adhikari, Denaro, and Wagner, *Computational and Inferential Thinking*. 10.3: Empirical distribution of a Statistic.

⁹⁴ Ibid, 13.3: Confidence Intervals.

⁹⁵ Ibid, 14.4: Central Limit Theorem.

⁹⁶ In Rose’s terms, these one could call “sample statistics” as “visualizations” or “measurements.”

chance about 95%, the sample mean is within 2 [standard deviations] of the population mean.”⁹⁷ These propositions allow statisticians to describe intervals that contain the population parameter with a specific level of certainty, or confidence: hence the name “confidence intervals.” Confidence intervals are formally defined as follows:⁹⁸

Let $\lambda\%$ be any confidence level, and z_λ be the point such that the interval $(-z_\lambda, z_\lambda)$ contains $\lambda\%$ of the area under the standard normal curve. \bar{X}_n is the sample mean.

Then the random interval $\bar{X}_n \pm \frac{z_\lambda \cdot \sigma}{\sqrt{n}}$ is called the approximate $\lambda\%$ confidence interval for the true population mean μ .

Put simply, there is about a $\lambda\%$ chance that this random interval contains the parameter μ .

Thus, we can understand the 95% confidence intervals that Fasoi et al construct for each exclude-one cyborg as an act that does three things. First, it makes all the observed behaviors apprehendable in the simple visual form of two-dimensional plots:⁹⁹

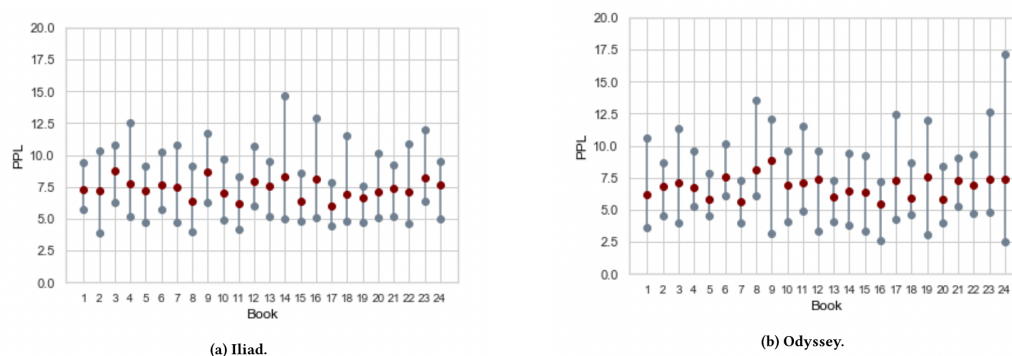


Figure 3: Perplexity (PPL) scores and 95% confidence intervals are shown on the Y axis, measured using the twenty four language models per Homeric Poem. Each model is trained on all the Books but one, and PPL is computed on 1,000 excerpts taken from the remaining Book (X axis).

Second, these confidence intervals produce ranges that contain the “true” degree of linguistic proximity between training and test corpora with 95% certainty. Third, these intervals inscribe a range of normal behaviors to individual models. Fasoi et al “report the mean PPL along with the PPL that was ranked at the lower (5%) and the PPL ranked at the higher (95%) position.”¹⁰⁰ These ranges, defined by a mean, minima, and maxima, invoke notions of abnormality and normality in terms of statistical significance since the middle interval containing 95% of the area under the probability density function (the observed distribution of the sample statistics) excludes statistical outliers —

⁹⁷ Adhikar and Pitman, *Data 140: Probability for Data Science*. 14.6: Confidence Intervals.

⁹⁸ Ibid.

⁹⁹ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” P. 82

¹⁰⁰ Fasoi, Pavlopoulos, and Konstantinidou. p. 81.

statistics with a low probability of being the true population parameters. With such a minimal chance that they could accurately speak to the true value the statisticians are interested in, these values become abnormal in relation to both the intervals and the means computed using the distribution of PPL values for each model. Each model is inscribed as an individual by their respective ranges of normal behavior. All these ranges are subtly different, with unique observed PPL means, minimas, and maximas. Thus the creation of confidence intervals constitutes inscription because it enacts a set of procedures for documenting individuality, and makes the individual ranges of abnormal and normal behavior for each model.

Having made the behavioral patterns of each individual model upon their test sets cognizable and visible, Fasoi et al proceed to the final stage of their psychological test: assessment. Using each model's average PPL score, which was computed over the distribution that defines the confidence interval, they construct spatial proximities between the models to evaluate relationships between the exclude-one cyborg population's constituent members in the form of line plots.¹⁰¹

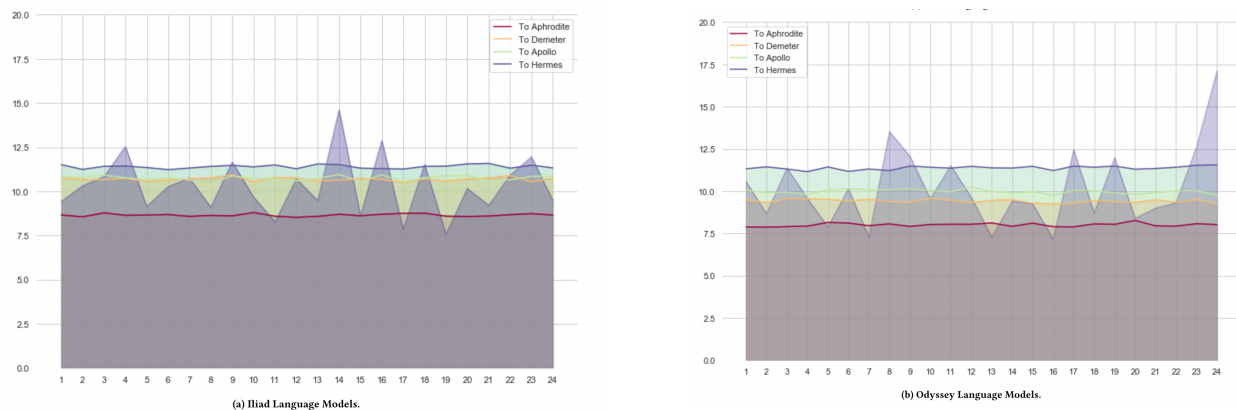


Figure 5: Average PPL (1,000 runs) per hymn, scored with language models trained on all but one Book (X axis; the PPL of the respective Book is shown with the purple area).

These line plots perform inter-individual comparisons that assess and quantify proximity between exclude-one cyborg models as a way to rank the objects and thus ascertain the “best” models. Their spatial proximity is quite literally made real — distances between any two cyborgs can be computed over two-dimensional space. Existing in this space, the individual exclude-one cyborg becomes calculable not only in relation to its intrinsic, individual attributes but also in relation to the other forty-seven exclude-one cyborgs. Just like Rose described, these line plots “[provide] a simultaneous means of perceiving, recording, and evaluating... the features of the [object] deemed significant” in normal terms.¹⁰² Through this final assessment, it is not only the individual that becomes cognizable — the entire cyborg population is “rendered thinkable by being made visualizable, inscribable, and assessable.”¹⁰³

This reading of PPL analysis as a psychological testing apparatus is supported by the discourse presented in the discussion and concluding sections of the text. They make definitive statements about the spatial proximity of the books in each poem based on their average PPL scores:

¹⁰¹ Ibid, p. 84.

¹⁰² Rose, *Inventing Ourselves*, p. 111.

¹⁰³ Ibid, p. 111.

“By employing character-level statistical language modeling, we showed that some groups of Books in the Odyssey (9, 10 & 12 as well as 2, 19 and 24) are much closer than others. Also, we show that some books (Books 5 and 9 in the Iliad and, again, Books 9, 10 and 12 in the Odyssey) are linguistically far from the rest in the same poem.”¹⁰⁴

Having transformed a complex set of “behaviors” into simple conclusions about geometric distance, the books of each poem are transformed into a population of “docile” cyborgs.¹⁰⁵ Their subjectivities are rendered stable by the objectification of their conduct, and the variance of said conduct that once made an individual unpredictable becomes “regular” and “predictable.” The cyborgs, through this analysis, are primed for disciplinary action.

Despite the author’s proclamation that performing authorship analysis using “the linguistic proximity of each Book with the rest of the poem that it comes from” yielded “inconclusive” results, their investigation is nonetheless, quite productive. They create comparable cyborg individuals, and make those individuals, along with the populations they comprise, technically cognizable. Moreover, the way that the authors of the study configure and pursue parameters that can speak to Homer’s authorial truth brings into perspective why a population was necessary to create in the first place. Under the ever-observant eye of frequentist statistics, a true authorial self can never be apprehended directly. Take their final recommendation for future authorship analyses on the Homeric texts, where they suggest that the “wide variation in Perplexity scores among excerpts from the same Books, only reinforces the need to break down the poems into smaller structural units in future studies.”¹⁰⁶ The already partial, mutilated, individualized cyborg Homers are not atomized sufficiently to arrive at any image of authorial truth. Though the whole text may constitute a body to preserve traces of the authorial unconscious, the unconscious presents itself in disordered, neurogrammatic fragments. Thus, the unconscious can only be reproduced through the repetition of systematic testing procedures, which proceed ad infinitum from partial truths towards an approximation of a unified one. In the most precise terms, the “real” author may always be just out of reach — but the distance between us and that “real” author can be made infinitely more narrow by the creation of more mutilated authorial cyborgs.

¹⁰⁴ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 87.

¹⁰⁵ Rose, *Inventing Ourselves*. p. 106.

¹⁰⁶ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 85.

Part 5: To Die a Digital Death?

Considering the way that Fasoï et al approach their project, along with the methodology they employ, their project expresses remarkably modernist aims. These computational linguists seek definitive answers — real truths — in an area of questioning where none satisfying their scientific criteria have been found before. They confidently imagine that their decidedly rote application of statistical techniques to publicly available datasets using ordinary computational devices will create tangible progress. Such an attitude defines the promises that science, as a meta-narrative in the likes that Lyotard outlined, has made to modern society. While the totalizing purview of science is not difficult to discern, it is made especially clear in the context of the Homeric Question. Literature, culture, and the humanities are no match for the universal reason expressed by mathematics, science, and statistical rules that supposedly govern the real. Mysteries that reside in these fields are made out as nothing more than un-tested statistical phenomena that can be resolved given sufficient attention under scientific paradigms. Everything, even the Homeric Question, should be answerable with the appropriate use of the tools science has fashioned for humankind.

And in many ways, their project does bear some semblance of progress. Concluding that “some groups of Books in the Odyssey (9, 10 & 12 as well as 2, 19 and 24) are much closer than others,” and that “some books (Books 5 and 9 in the Iliad and, again, Books 9, 10 and 12 in the Odyssey) are linguistically far from the rest in the same poem,” they present their findings as evidence that can weight the odds in favor of oral-formulaic theory.¹⁰⁷ Mapped to ideas about the origins of Homeric texts in the real bodies of bards who once sang of Greek heroes, measures of linguistic proximity are made to speak on behalf of historical realities of singular or multiple authorship. Progress in this context means that the authors’ generated measures of linguistic proximity between books allow them to make conclusions about what Homeric “authorship,” at its core, really means.

But assuming that these conclusions are made in good faith and subsequently criticizing Fasoï et al as fundamentally mistaken in their attempt to extract Homer from the texts that gave his name such stature would obscure the larger reasons why they must present their work as progress-making. The construction of this link between quantitative observations of written records and historical events not directly included in the narrative is not just a convenient way for the authors of this study to position their work as relevant knowledge-production in the digital humanities. It is actually a crucial movement in science’s continued performance as a comprehensive schema, because it reinforces the idea that quantitative methods can explain qualitative relationships like the ones shared between bards and their audiences, even if those methods are applied only to the physical vestiges of social interaction.

Thus the chosen subject matter of this decidedly mundane computational analysis is so important because it constitutes an attempt to broaden of science’s epistemological horizons. As

¹⁰⁷ Fasoï, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p. 87.

James I. Porter writes in *Homer: The Very Idea*, “Homer is a peculiar kind of object in the history of culture inasmuch as he enters the cultural record as an object that is lost from the moment it is found... an impossible object, an entity who only becomes tangibly real and actual in the very failed effort to grasp him.”¹⁰⁸ This study appears as confirmation of this claim, and highlights the way that even scientific methods must ask whether they are analyzing, or creating some form of reality. Procedures meant to extract the author are more generative than derivative, and Homer in the context of this study only becomes “real” (though inhabits strange bodies and existences) after the creation of those ninety-six cyborgs. This is why the *Iliad* and *Odyssey*’s origins in single or multiple authorship, while an interesting question, are not what make this study an interesting object for analysis. By taking on the Homeric Question under an empiricist paradigm and positioning their representations of this “impossible object” as reality, Fasoi et al seek to provide an answer that will do away with the ambiguity that science imagines has always, and always will, plagued the humanities.

Their approach to the issue of Homeric authorship, which envisions it as a simple question of discerning between voices who were already formed as distinct beings at the time of the poems’ composition, is certainly a modernist reinforcement of what Barthes calls the “reign of the author” via empiricist terminology.¹⁰⁹ Yet even this movement is secondary to the higher-level power dynamics at play here. If Barthes believed that to crown an Author is to furnish the text with a final signifier, some true meaning, then what Fasoi et al set out to do is not just crown one or many Homers, but to furnish science as narrative’s final meaning, and simultaneously pronounces that answers to humanities longest-enduring questions lie in technology.

A Litany of Modernist Leanings

Yet our own questions about this project have yet to be resolved, since attempts to furnish final meaning are just that — attempts. So, we still must ask: what should we make of these strange, docile, cyborgs, who stand in for authors but clearly are not Homer himself, or rather, him-selves? What has modernity, and then post-modernity, done to the ever-shifting mind, body, and idea of the great author Homer? The unusual character of this study on authorship calls into question what embodiment of Author and reader mean in the digital age. Thus the implications and stakes of this study for the humanities, which has a much longer history of dealing directly with questions surrounding authorship, warrants analysis under one of the most prominent theories of authorship put forth in modern times — Barthes’ “Death of the Author.” The theory of author-death is applicable here because of the way that the study challenges it. Refusing to believe that Homer is gone, that perhaps Homer never existed in the first place and there was only a multi-generational witnessing of the epic narratives’ evolution, the study’s methodology must be understood in conversation with the theory that it defies. Given the peculiarity of the cyborgs, we must investigate whether changes in fields like natural language processing, or even

¹⁰⁸ Porter, *Homer: The Very Idea*. p. 2

¹⁰⁹ Barthes, “The Death of the Author.”p. 5

just new applications of old methods, necessitate changes to the humanities' conceptions of what it is to be an author, a writer, a reader, and a text.

The most glaring modernist point of view expressed by the project is its very pursuit of Homer, the author. Barthes writes that "the author is a modern figure, produced no doubt by our society," whose existence relies on Western society having "discovered the prestige of the individual."¹¹⁰ At every stage, from the wording of the research question to the selection of computational methods to the interpretation of the results, the study's very existence is contingent upon the prestige conferred to individuals titled "Author." Barthes' characterization casts the Author as a construction, a mirage built by the tools of "English empiricism" and "French rationalism." Accordingly, the study's assumptions and stated goal of differentiating between distinct voices directly take up these paradigms.

My analysis from all the previous sections explore the ways that the study reproduces this notion of the Author on the same basis that he (notably, a historical he) was initially configured. In Section 1, I show how the language used to formulate the research question positions the study's goals as identifying areas of linguistic coherence. While such coherence, or "integration," could be thought of as a product of collective engagement during poetic composition, here it can only ever be the product of a single speaker — other potential realities are never even considered. Section 2 bears witness to the way that the author is not just metaphorically spoken of, but actually produced, as an empirical object whose characteristic contours provide fodder for later applications of disciplinary techniques, as well as the way that oral-formulaic theory can be interpreted to support the creation of these empirical author-objects. Through Koopman, Eder, and Rose, Section 3 explores the way that stylometry creates a basis for the rationalist belief that because stylistic coherence is produced by individual psyches, personalities can be rendered concrete and visible through empirical atomizations of language. Section 4 demonstrates that empiricist productions of the individual do not end with the computational object's instantiation. Instead the creation of "individuals" requires that even computational objects must be differentiated from each other using statistical tests, which make them intelligible in terms of a unique set of relations to the other objects of their kind. Everything so far shows how Fasoi et al takes modernity's idolization of the author to a logical extreme — the latent influences of empiricism, rationalism, and individualism are not usually made so readily apparent when we speak of the Author. Analyzing the study under Barthes' theory brings out the way that it enacts the very ideologies in literary analysis that he critiqued, and translate that ideology in mathematical terminology.

Connecting this project to the field that made it possible, stylometry, also brings to light how ideas in the humanities find new footing in science, and is a keystone idea connecting linguistic integration to the writer's psyche. Despite all the proliferation of discourse suggesting that there is an unbreachable schism separating the humanities and sciences, (public understanding of science debate — make this into a footnote) Barthes articulation of the assumptions made by literary critics are also clearly present in stylometry, which then gets taken

¹¹⁰ Barthes., p.1

up by computational linguistics via personality psychology. He writes that “criticism still consists, most of the time, in saying that Baudelaire's work is the failure of the man Baudelaire, Van Gogh's work his madness, Tchaikovsky's his vice: the explanation of the work is always sought in the man who has produced it, as if, through the more or less transparent allegory of fiction, it was always finally the voice of one and the same person, the author, which delivered his confidence.”¹¹¹ Taking this quote in the context of Barthes' later claim, that to furnish an Author is to “impose upon the text a stop clause,” the use of personality attributes like madness and vice indicates that Barthes perceived a belief in literary criticism where the meaning of a work could not be extricated without some understanding of the writing self, and that patterns of expression in behavior (ie, personality) were in some way tied to the expression of meaning in writing.¹¹²

With stylometry and informational personality psychology, the relationship posited by literary critics between expression and mindedness is maintained but reversed. The idea that the work is always not simply a function of the author's efforts, but that personality encompasses the fundamental characteristics driving the expression of meaning in their work, must be transformed in order to account for situations where the author's personality is not and can never be known. Conveniently, stylometry seems to borrow from literary criticism to provide answers for texts with unknown or disputed authorship. Wincenty Lutoslawski's belief that linguistic style in writing was produced by the author's unconscious justified calling the author's style an “unconscious personal stamp.”¹¹³ This idea is tied more concretely to the self by personality psychology, particularly by Morton Prince (as explained in section 3). Importantly, if the author cannot help but leave traces of themselves in the text, then the message they are trying to express — and the creative genius inherent in the message — is one that defines them and their psychology, their soul. From this point of view, it feels only natural that if the chain of cause and effect can be traced, then it can certainly be reversed.

Tracing the ‘causal chain’ between expression and mindedness backward gives Barthes' use of the word “confidence” a new dimension. To confide in one's reader is more than an expression of attitude or thought, it is a literal preservation of oneself in the form of personality traits when personality is thought to shape expression — it is literally like Prince's neurograms. This idea is what allows the reversal — if the text preserves the author's mind, then the author's mind can be extracted in order to “close the writing,” and thus resolve any lingering questions about its meaning or character.¹¹⁴ If Homer was here all along, why even bother asking any questions?

There are still more resonances between the literary culture that Barthes critiques and the study's framing of authorship analysis, particularly when it comes to how the author is situated in time. Composition by multiple authors, as is suggested by oral-formulaic theory, allows for collective and iterative forms of authorship which could mean that a figure like Homer wouldn't

¹¹¹ Barthes. p 2

¹¹² Barthes.p 5

¹¹³ Eder, Maciej, “Style-Markers in Authorship Attribution.”

¹¹⁴ Barthes, “The Death of the Author.” p. 5

necessarily precede the text. Successive iterations of the epic poems might be composed during collaborative sessions between bards, and the poems' structures might have undergone significant changes even during performance as bards adapted their performance to their performances' reception, and sought to provoke certain responses from the audience.

But the study's footing on stylometry and personality informatics means that these scenarios can never be entertained as a possibility. In order for the computational objects produced, those ninety-six cyborgs, to have any claim to veridicality and authorship, latent ideas of what an author is must comply with modernist expectations that an author precedes their work. In order for the author's whole creative self and voice to be contained within the text, and therefore exist as an extractable thing, the Author must arrive at the point of composition fully formed. As Barthes writes:

“The Author, when we believe in him, is always conceived as the past of his own book: the book and the author take their places of their own accord on the same line, cast as a before and an after: the Author is supposed to feed the book — that is, he pre-exists it, thinks, suffers, lives for it; he maintains with his work the same relation of antecedence a father maintains with his child.”¹¹⁵

Applying the ideas from this quote to the study's framing of the Homeric Question, it becomes clear how important this conception of the author is to a project that seeks to extract and differentiate individual voices who contributed to the composition of these poems. In order for the study's methods to stand as valid upon its premises, the Author Homer, or *Authors* Homer, must be thought to exist as distinct, uniquely minded individuals prior to making their contribution to the epic poems. This quote reflects the way that author figures are made stable, since when the definition of an author mandates that their whole creative persona must precede their work, it creates a concrete, whole person that the terminology “author” can refer to without issue. In Saussurian terms, the sign “author” has meaning in the particular context of its application to a text if and only if there is some “real” author that it signifies. And since the text is (at least in most modern literary cases) understood as a stable work whose fundamental message is intrinsic and unchanging, so too must its originator, the Author, be understood as an unchanging figure whose personality traits and expressive tendencies exist in a state of coherence and unity that can be hidden within the text. Without the assumption that the Author precedes the text, Lutoslawski's conceptualization of style as an expression of the unconscious no longer makes sense — style cannot be characteristic, cannot be made into a unique identifier, if the referent that the fingerprint identifies lacks real existence at any specific point in time.

Taking up the possibilities outlined prior, where the poems' composition is the product of many minds coming together and engaging in conversation, there can be no single author who exists prior to the text — even if we were to call these collaborative dialogues a “mind” or an “author” in some way, it could only exist during moments of composition, never before or after. There can be no confidence (confidence as in confession) of the self in the reader via the text unless the author and their corresponding unconscious precede the text's very creation, which is

¹¹⁵ Barthes. P 3.

thought to express the essential, individual qualities of the scriptor. All this is to say the latent presence of personality psychology in computational authorship analysis demands that collaborative modes of composition be overlooked in favor of an individualized view of authorial influence, so that mappings between discrete texts and persons can be established with claim to some historical reality and scientific truth.

The fantasy that if there exists no single, real, historical person we could call Homer, then there must at least be strictly separable voices who independently contributed poetic content and syntactic forms to the epics, is a central component in authorship analysis's claim to relevance. This holds particularly true when it comes to the Homeric Question in the context of scientific forms of interrogation. Fundamentally, what computational authorship analysis and attribution studies seek to do is construct a one-to-one mapping between author and work. If the significance of a work of writing or art is in any way tied to some idea of a creative genius who originates the work, and thus imbues some important message within the text that can provide benefit to human society via interpretation, then procedures that make the relationship between work and artist concrete and visible can claim to encapsulate all the potentialities of meaning and cultural significance attributable to that work.

Here again, science takes on the literary critic's position. The Homeric Question is essentially significant and enduring because the origins of those two epic poems, afforded "cult status" as noted by Porter, will always remain unknown to some extent.¹¹⁶ If Homer were to be identified, apprehended in some true way and expert consensus validated those conclusions, then questions surrounding the *Iliad* and *Odyssey*'s origins would not be the only question to become uninteresting. Homer is essentially the Classic's NP-hard problem — no satisfactory solutions have been discovered, even after a significant amount of time. If one was able to answer an NP-hard question of this kind, then all the other kinds of questions in a similar vein would also be solved. In essence, solving the Homeric question makes all authorship questions fundamentally uninteresting from the standpoint of science, because the same solution applies in all scenarios.

This is clearly part of the scenario that Barthes outlines when speaking about what the presence of the Author means for interpretation. Barthes argues that "to give an Author to a text is to impose upon that text a stop clause, to furnish it with a final signification, to close the writing."¹¹⁷ Giving an Author by way of "[discovering] an author" including aspects of "the psyche," that underlie the work allows the critic to proclaim victory.¹¹⁸ Victory is won by way of establishing the arena — the space between a generative, creative mind and the work — where the text's secret, or ultimate meaning, dwells. All that is necessary to achieve this victory is carried out in the process that makes that arena "real," which entails constructing a map of that relationship between Author and work, and make the map visible.

This new territory, where meaning resides and so can be seen and described, allows the belief that "once the Author is discovered, the text is 'explained' ... [and] the critic has

¹¹⁶ Porter, "Whose Homer Is It Anyway?"

¹¹⁷ Barthes, "The Death of the Author." p. 5

¹¹⁸ Barthes. p. 5

conquered.”¹¹⁹ Constructing visible territories that connect author to work is precisely what the study does, just in the form of probability models and significance testing. The creation, differentiation, and ranking of the cyborgs articulates a field of causes, effects, and proximities between potential authors and subsets of the text. If we take this field of meaning to accurately represent the creative relationship between Homeric authors and these epic texts, then the study causes the enigmatic character that made the poems such important cultural touchstones in the West to vanish. Posturing as a final meaning to the whole corpus of works attributed to Homer, the cyborgs proclaim that the Homeric question, if not already closed, will soon be answered.

Victory for the literary critic is thus transformed into victory for computational linguists, and by extension, scientific victory over the humanities. The shared paradigm of authorship between traditional, modern literary criticism and computational authorship analysis is significant because it gives us a way to evaluate the successes and failures of the study put forth by Fasoi et al. Though not at all surprising that scientific forays into the world of literature borrow from the most prominent, recent paradigms, it is interesting that even in scenarios where scientific epistemologies are held up as superior to those utilized in the humanities, the knowledge produced can still be judged through the framework of literary theory. So, if Barthes deemed that refusing to assign the text an ultimate meaning, in the form of the Author or otherwise, would be “counter-theological, [and] properly revolutionary” resistance to the regime of the Author, then what the study here does is attempt to uphold or revive that regime using the tools and technology at science’s disposal.¹²⁰ This project isn’t just modern, but anti-revolutionary. Instead of “[refusing] to arrest meaning” and “refuse God,” the computational linguists uphold their signifiers for ultimate meaning: the cyborgs who can speak for God and “his hypostases, reason, science, [and] the law.”¹²¹ If we read this study as an attempt to reinvigorate the regime of the author by answering the Homeric Question, then the critical question we must address is: can the cyborgs truly speak in Homer’s name? Can we really believe that they furnish a final meaning?

No Victory for the Linguists

If the computational authorship analysis carried out in the study seeks to answer the Homeric question by the creation of bodies who, by representing the relationship between creative mind and work are thought to stand in for the Author, then it is safe to say that Fasoi et al do not accomplish what they meant to. In order to answer the Homeric question, furnish a final signified to the *Odyssey* and the *Iliad*, reinstate the regime of the Author, and position science as a grand-narrative that holds up certainty like a candle to the dark ambiguity they perceive in the humanities, the cyborgs Fasoi et al create must be, in some way, genuine Homeric Author(s).

But they can never be real Authors, because they are cyborgs. The cyborgs are yet another litter of modernity’s strange children, the hybrid offspring of human expression, physical artifact, digitized knowledge, and computational processing. Ironically, the cyborgs who are

¹¹⁹ Barthes. P. 5

¹²⁰ Barthes. P.5

¹²¹ Barthes. P.5

essentially chimeric in nature, are as much a product of modernity as they are a threat to it's legitimacy. Each move to construct the cyborgs — patchworking their skin together out of text probability models, situating them in time and giving them relations to one another — requires hybridizing human and machine under such an extreme paradigm that the absurdity of the whole project becomes evident. Quite easily, the project's intended meaning undoes itself.

The way in which this dynamic appears is with respect to the Author's temporality. By assuming that the Author precedes the text, the Author can then embody a coherent version of themselves within it, and proper technology will allow that Author to be extracted, or reverse-engineered. This is central to the analogy of the cryo-chamber I articulate at the end of Section 3, and is what allows the cyborgs to be positioned as (at least) Author-like. But the Author and his designated meaning is held up with prestige over readers and their interpretations — the Author's very social significance — can only be maintained if the Author and reader are not just physically, but also temporally separate. Yet as the cyborgs are crafted, the very procedures that bring about their existence collapse all temporal and physical separations between Author, reader, and text.

The collapsing of physical separations is fairly obvious. Though the text may have once served to physically and temporally separate Author from reader, in authorship analysis and attribution it is made to be the bridge that makes those all-important distinctions obsolete. The cyborgs' bodies are made of texts, their individual identities are tied to the patterns of authorial expression through their textual bodies, and the unification of author and text into a single cyborg is carried out by a process of statistical reading. This process of reading is important, because the assumptions underlying the study's use of tri-gram analysis relies on the belief that the Authors mind is so inseparable from the text that it is always present wherever and more importantly, whenever, the text is. The Author is not absent from the text when it is read, rather the text allows readers to commune directly with the Authors mind, which is always-already physically imprisoned within the words it once wrote. Essentially, the assumptions made by the study end up undoing the boundaries of the very body and mind they seek out and wish to replicate.

As the mechanisms of cyborg-assembly collapse these distinctions, the final product comes into view as a body held together by contradictions. The mind of a cyborg is at once reader and Author. Because each cyborg stands as the ultimate authority on the shape of the text that was used to create it, and their bodies all made of text hold the "ultimate meaning" that computational linguists assume must exist, the cyborgs are at least as close to Homer the author as the text itself is. They are intended as partial replicas of the Author, and are constructed as such by utilizing the modernist paradigm of authorship. These cyborgs are also readers in the way that Barthes defined a reader. Readers are, in Barthes eyes, "the place where multiplicity is united, connected."¹²² Though the atomization of text data into individualized cyborg bodies may not seem like a unification of that text, the unification of these multiple authorships is facilitated by the application of psychological tests, which makes individuals at the same time as it creates

¹²² Barthes. p.6

populations of comparable beings. Here, the whole population of cyborgs accounts for the multiplicity that Barthes sees in the reader.

The cyborg's dual nature has one particularly important implication: the study does not successfully recreate Homer's mind or body. By collapsing distinctions between author, reader, and text, we cannot say that these cyborgs are Homer extracted, resuscitated from the cryo chamber of the text. Unlike the author, who under a modernist lens, precedes and suffers for the text, and is therefore not bound to the text, the cyborgs only exist within the text. Without the text their bodies dissolve not into dirt as humans and animals do, but into the data used to bring about their being. We can say only that the conditions of their birth encouraged all to believe that these objects were authors, and the inability to recreate an object which we could definitively say is an author, not text nor reader, means that the answers proposed by the project — namely, “the need to break down the poems [into] smaller structural units [during] future studies” — does not achieve a success that holds up to modernist scrutiny.¹²³ Fasoi et al find no final answer, no confidences passed down from Homer that speak to the truth of his singular or multiple existence, and no explanations for what he meant or who meant it. In Homer's place, they are left (as the reader always was) with nothing more than the meaning of their own invented objects and ideas. Thus there is no victory for the linguists, nor can there be any celebration of science's empirical superiority over the humanities.

In a way, this failure highlights the enduring value of Barthes' work. Even computationally advanced techniques fail to identify the ultimate intended meaning that motivated so much of modern literary criticism. What this attempt confirms is that there can be no Author who exists separately from the reader once the text has been put into writing, and that attempts to extract that final meaning will be more reflective of the reader than the Author.

But the peculiar, multi-faceted existence of these cyborgs collapses the distinction between Author and reader that even Barthes relies on, particularly when it comes to his concluding claim. The most culturally iconic portion of Barthes' “Death of the Author” is its final line, which reads as follows:

“We know that to restore to writing its future, we must reverse its myth: the birth of the reader must be ransomed by the death of the Author.”¹²⁴

For the author to die, and stay dead while the reader reigns, the reader and author cannot coexist in one body, they must always be separated by the text. But here, the text unifies author and reader into one body which, though positioned as an extraction, has never left the text itself. This is just one of many ways that when mathematical and classical frameworks come into contact, conflicting frameworks for knowledge production become problematic from either direction. The existing theoretical frameworks we have used to understand our relationship with the narrative realm pose challenges that are related to changing notions of what constitutes a valid embodiment of an individualized self — these problems will only expand.

¹²³ Fasoi, Pavlopoulos, and Konstantinidou, “Computational Authorship Analysis of Homeric Language.” p.87

¹²⁴ Barthes, “The Death of the Author.” p.6

What we must ask now is whether Barthes' theory of authorship is crucially outdated in a world where statistical and digital modes of representation and reproduction are so commonplace. These types of movements draw out gaps in the existing theoretical frameworks that we have used to understand humanity's relationship to narrative, and these gaps point towards evolving notions of what constitutes valid forms of embodiment for individualized selves. These problems and questions will expand, so having used Barthes' theory to evaluate the big-picture objectives and procedures of the study, we can now use the study to understand how Barthes must be reconfigured. On this bizarre terrain of digitally embodied Homer-text-machine hybrids, the critical question arises: what does "author death" even mean here?

Part 6: Death and Immortality in an Age of Non-Human Author-Objects

So far, attempting to place cyborgs neatly within a philosophical ontology of authorship has only made it more difficult to understand what they are, and what consequences their existence holds. Their hybrid character makes it difficult to say how the dynamic between text and authorship, as portrayed in the study, maps to Barthes' theory of author death. Inherent contradictions may mean that the cyborgs collapse distinctions, but extending contradictions to formulate an argument that current divisions and frameworks foreshadow a coming paradigmatic collapse would be unsatisfying, likely wrong, and wouldn't contribute a nuanced explanation about what how theory must change to account for our changing world.

In order to provide a useful analysis of how the cyborgs necessitate changes to author death theory, the following section will use J.P. Vernant's scholarship on immortality in Greek myth to explore how notions of authorship are related to narrative and immortality, and how those ideas apply to the cyborgs. Specifically, when considering Barthes' final claim, we must ask if the temporal division he posits between the Author's death and the Reader's life still makes sense. Are these cyborgs "born," can they "die," and if so what does death look like for them? What do birth and death mean in this context, and when birth and death occur in the same body, simultaneously, how does one's access to immortality in narrative change? Is cyborg immortality a radical idea, or is it inherent to their form of embodiment? Why does Barthes position Reader and Author as mutually exclusive, and how does the cyborg's defiance of this separation change the relationship between authorship and immortality?

Vernant's writing is important to take up for these questions because his ideas are present in both Foucault's and Barthes' writings on authorship. In both of their work, invoking Vernant is a way of signposting the way that the function of narrative has changed in recent times, and by extension, the way that the roles of reader and Author have changed in response. Barthes' citation of Vernant focuses on his portrayal of "ambiguity" in the *Iliad* and *Odyssey*, whose tragic characters are at least partially derived from the characters' unilateral interpretation of words with dual meanings (for example the Greek word for honor, time, can either mean ordinary or heroic honor).¹²⁵ Citing Vernant in this way allows Barthes to articulate how readers are defined by their ability to see and understand multiple layers of meaning in a text, and that this ability is what separates readers from the Author figure, who is thought to have only intended a single meaning.

While Foucault doesn't reference Vernant directly, he qualifies how Barthes' work is transformative in terms of the Greek relationship between narrative and immortality, a relationship that is one of Vernant's primary concerns in "A Beautiful Death." Foucault explains that when hanging Barthes' work against the backdrop of Ancient Greek myth, it becomes clear that "Death of the Author" has modified the "kinship between writing and death."¹²⁶ His articulation of this relationship practically summarizes J.P. Vernant's explanation of *kleos*, which

¹²⁵ Barthes. P.6

¹²⁶ Foucault, "What Is an Author?"

is that “The hero accepted an early death because his life, consecrated and magnified by death,” he could “[pass] into immortality” by way of “the narrative” that preserves him in collective memory.¹²⁷ Foucault positions Barthes’ theory of author death as an inversion of that old relationship, arguing that “writing is now linked to sacrifice and to the sacrifice of life itself; it is a voluntary obliteration of the self that does not require representation in books because it takes place in the everyday existence of the writer.”¹²⁸ This interpretation of Barthes’ theory indicates that Foucault saw it as transformative to the field of authorship, but more importantly, also demonstrates that authorship is broadly important because of its claim to ownership over narratives, which grants the role close proximity to immortality and cultural meaning.

In the modern era, an author’s immortalization in the public eye is related to physical, human death in more concrete ways than the philosophical relationships posited by Vernant, Foucault, and Barthes. The status of an author has many parallels to heroic ideals of honor, contest, individualizing glory, and immortality. Our society deems the work of dead painters more valuable than the paintings made by the living, and we seem to find the greatest literary genius in the writers who wrote for a time and audience well before our own.¹²⁹ Broadly speaking, it often seems that in the world of art and authorship, most artists only achieve recognition post-humously. The fame associated with author-names including Dickinson, Kafka, Poe, Thoreau, Vermeer, and Wilde tells us as much.

If real death seems necessary for mortal artists to secure a place in our collective memory, perhaps it is because they are like Achilles, whose story was uninteresting until death because eternal and static qualities like glory cannot be afforded to characters who are still subject to change over time. When a human artist dies a physical death, they can be made relatively static in the cultural imagination because original modifications to an author’s work become impossible, and thus the character of their life’s work can only change via interpretation. Only then can art historians and literary critics sing praises of artistic genius. What this suggests is that immortality’s elusive character and associated traditions for conferring that status may not have changed so drastically between the modern intellectual West and the Ancient Greeks after all.

Western culture’s tendency to glorify dead authors and artists whose works can be connected to them alone at a single point in time, despite repeated protests from the academy that the author is dead, demonstrates *kleos*’s lasting legacy when it comes to authorship. Greek narratives recounting a hero’s journey to eternal glory through a beautiful death offer a way to make sense of the irony that defines the cyborgs’ existence. It should be clear by now that I have called these mathematical objects “cyborgs” simply because they fit the description Koopman puts forward of the informational person, which he calls cyborgs. They are cyborgs of the sort that Haraway writes about in “A Cyborg Manifesto.” Each one is “a cybernetic organism, a hybrid of machine and organism.”¹³⁰ They represent human voice, culture and knowledge; are

¹²⁷ Vernant, “‘A Beautiful Death’ And The Disfigured Corpse in Homeric Epic.” p.57

¹²⁸ Foucault, “What Is an Author?”

¹²⁹ Professor Winnie Wong, Rhetoric 136: Theory of the Copy, Lecture April 2023

¹³⁰ Haraway and Wolfe, “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century.” p.1

embodied in energy, machine code, and memory bytes somewhere on a server. They are uncoupled from biological reproduction, as the application of the psychological testing apparatus individualizes Homer's cyborgs and makes them "[creatures] of social reality as well as [creatures] of fiction."¹³¹ Indeed these fictions are world-changing in "the struggle over life and death."¹³² The fiction necessary to invent these cyborgs foreshadows a world where irony, in the form of conflicting ontologies contained in one digital body, challenges normative understandings of life and death. And since, as Haraway notes, that "irony is about contradictions that do not resolve into larger wholes," perhaps an understanding of these cyborgs can only be formulated on a terrain that does not seek logical consistency and synthesis as a means to forge knowledge.¹³³ In line with Haraway's suggestion that we take "pleasure in the confusion of boundaries" as a way to make sense of the irony embodied in the Homeric cyborgs, I suggest that we take pleasure in the metaphor of the cyborg in a way that generations before us have — through the narrative themes of the *Iliad* itself.¹³⁴

What happens when we understand the project through the lens of Ancient Greek tragedy? Parallels between Achilles and Homer's cyborgs, the texture of epic poems and scientific journal articles, themes of death and immortality in heroic honor as well as authorship create the opportunity to draw out an analogy between the Homeric epics and computational authorship analysis. This analogy will facilitate the investigation into whether author-death makes sense if we try to imagine that it happens in one body, particularly because of the way that the construction of these cyborgs calls Barthes' theory of author death seriously into question. By applying Vernant's theory about how epic narratives act as an immortalizing force, this section seeks to make sense of contradictory cyborg identities as a bid for immortality.

Scientific Progress as Literary Epic

The easiest place to begin analogizing the *Iliad*, with its warriors and mission for glory, to the computational authorship analysis study is with the form and purpose of epic poetry. What makes the *Iliad* such a powerful vehicle for transforming human warriors into eternal heroes through song is the way that it transmits cultural knowledge and values. According to Vernant, epic poetry "[served] to transmit, to teach, and to make manifest within each individual the alloy of knowledge, beliefs, attitudes, and values that make up a culture."¹³⁵ Poetry's inherent social power in Ancient Greek society was thus a function of its ability to transmit proper knowledge. Vernant's use of the word "alloy" here also indicates how epic poetry was able to unify disparate components of culture, and help the listener understand their place in Greek society.

Scientific literature functions very similarly to epic poetry in the modern Western world. Like a vocal performance that recounts a culturally significant event or theme, each scientific journal article (and the truth-claims they contain) contributes to science's central institutional

¹³¹ Haraway and Wolfe. p.1

¹³² Haraway and Wolfe. p.6

¹³³ Haraway and Wolfe. p.1

¹³⁴ Haraway and Wolfe. p.7

¹³⁵ Vernant, "A Beautiful Death' And The Disfigured Corpse in Homeric Epic." p.58

imperatives. Robert K. Merton's work on the normative structure of science argues that the scientific ethos revolves around four norms: communism, universalism, disinterestedness, and organized skepticism. These are science's primary values, which shape developmental trends in scientific theory and influence the sort of analysis undertaken using scientific methodology. The similarity in social function between science and epic poetry are evident immediately. It is beyond dispute that science holds significant influence over collective action and values in modern Western societies, and like epic poetry, continually "[puts] its stamp on the whole system of values," whether we deem this good or not.¹³⁶

Universalism and communism appear in Vernant's description of epic poetry, and are as important to the social function of Ancient Greek oral culture as they are to the scientific ethos. Universalism, as defined by Merton, is a principle that requires that the truth-value of all knowledge-producing claims be judged using "preestablished impersonal criteria," which help practitioners judge whether the claims are "consonant with observation and with previously confirmed knowledge."¹³⁷ Science's central quest — validation of objective, universal truths — rests on institutional authority whose vested power has a similar end result as the epic poem. Vernant writes that "Only epic poetry has the importance and power to confer on the hero's quest for deathless glory both institutional solidity and societal approval, without which the quest would be merely a subjective fantasy."¹³⁸ This quote helps identify a correlate between objectivity in science and narrative. Epic narratives like science wield institutional power that allows the public to differentiate "subjective fantasy" from real, validated knowledge based on a claims' presence or absence in the epic's discourse.¹³⁹ Though the knowledge contained within epic narrative might not be so straightforward or prescriptive as basic scientific principles (centuries of literary study dedicated to understanding the meaning and importance of narrative themes in the *Iliad* and *Odyssey* say as much), both use forms of social approval to separate important knowledge from unimportant or misleading trains of thought.

Merton's description of communism in science also parallels Vernant's characterization of epic poetry, demonstrating how science and epic poetry formulate the relationship between communities and knowledge in similar ways. Writing that "the substantive findings of science are a product of social collaboration and are assigned to the community," Merton positions scientific knowledge as communal intellectual property and gives scientific journal articles a similar character to oral poetry, "which serves as a repository of shared culture and as a societal memory for the group."¹⁴⁰ Both scientifically and poetically transmitted knowledge are created to serve the well-being of the collective audiences they speak to, and these knowledges inform worldviews, ethics, and appropriate courses of action.

Moreover, the way these knowledges are composed and accumulated occurs through collective, and not individual, channels. Though history may credit independent scientists for

¹³⁶ Vernant. p.58

¹³⁷ Merton, "The Normative Structure of Science." p.270

¹³⁸ Vernant, "'A Beautiful Death' And The Disfigured Corpse in Homeric Epic." p.58

¹³⁹ Vernant. p.58

¹⁴⁰ Merton, "The Normative Structure of Science." p.273

moving their field forward, in actuality scientific creativity is typically practiced in groups, and that creativity is exercised in order to give new information to society. Noting this communal dynamic, Merton writes that scientific discoveries “constitute a common heritage in which the equity of the individual producer is severely limited.”¹⁴¹ This relationship between scientists and their work is paralleled by the relationship between bards and their songs. Knowledge production of this sort (either poetic or scientific) is not something that individuals take part in, it is something that is a natural byproduct of their creative ecosystem. Though bards and scientists alike might make individual contributions, the legacy of their career blurs together individual thought, feedback from outside sources, and iteration upon preceding iterations of similar ideas. Authorship, in the sense that to be an author is to contribute something to collective bodies of knowledge, is as multiple in science as it is in epic poetry.

Disinterestedness, the third Mertonian norm, follows from the dynamic of multiple authorship in scientific investigation and poetic composition alike. Ruth Scodel’s work on the representations of epic bards as they are portrayed in the *Iliad* and *Odyssey* identifies disinterestedness as the characteristic distinguishing bardic narratives from ordinary ones. Though dismissive of the possibility that bards engaged in any form of collaboration whatsoever, Scodel explains that bardic narratives are disinterested because they do not seek to manipulate the audience, rather they attempt to transmit messages and information about Olympian happenings that are handed down from on high by the Muses.¹⁴² Bardic reliance on the Muse for inspiration and information on Olympian happenings thus imbues their performances of epic narrative with the character of multiple authorship, and by extension, disinterestedness. Scientific disinterestedness, while also rooted in the idea that a tradition’s practitioners are simply communicating meaning that pre-exists them, is more rooted in notions of responsibility and communal benefit. A disinterested scientist must sublimate their own ego because the institutions they operate within demand it, as Merton explains that disinterest results from the “distinctive pattern of institutional control” exercised over scientists, who “once the institution enjoins disinterested activity,” find it in their interest “to conform on pain of sanctions and, insofar as the norm has been internalized, on pain of psychological conflict.”¹⁴³ Disinterestedness of the scientist here formulates a relationship between labor and objective truth that parallels that of the disinterested bard’s relationship with narrative performance and the Muse. Like the disinterested bard, scientists are not meant to persuade. Instead they are nothing more than messengers, extracting and communicating realities whose truth value are taken as eternal, and unchanging.

The fourth Mertonian norm, organized skepticism, lacks an obvious corollary with Homeric epic tradition. Organized skepticism, as portrayed by Merton, is a “methodological and an institutional mandate” applied to scientists.¹⁴⁴ It ensures that science “asks questions of fact...

¹⁴¹ Merton. p.273

¹⁴² Scodel, “Bardic Performance and Oral Tradition in Homer.” p.172-173

¹⁴³ Merton, “The Normative Structure of Science.” p.276

¹⁴⁴ Merton. p.277

concerning every aspect of nature and society.”¹⁴⁵ Science may present it’s domain as the sum of all things with real being, but in order to make that vision a reality, it requires a mechanism to test it’s own epistemology against “other attitudes” whose interpretation of the same experiential data may have been “crystalized” and “ritualized” differently in other institutions. This character of science reinforces it’s position as a metanarrative, since organized skepticism drives the expansion of its epistemology through a form of conquest.

Understanding organized skepticism as an openness towards conflict, or perhaps even a mandate to engage in competition with other knowledge generating paradigms, opens a line of comparison with Nietzsche’s work on competition in “Homer’s Contest.” Nietzsche’s account of Eris’s, or envy’s, dual nature brings the generative power of contest in Ancient Greek society to light. Envy drove the Ancient Greeks to “contest,” and although Eris had a wicked side that could leads men into hostile struggle-to-the-death, she also had a good side.¹⁴⁶ The good Eris, while also causing envy, benefited society because she “goads men to action,” which acts like a stimulant and prevents stagnation.¹⁴⁷ Envy acts as a “stimulant” which helps to remove the “preeminent individual... to renew the tournament of forces” that prevents “a monopoly of predominance.”¹⁴⁸ Organized skepticism is to science as Eris is to Ancient Greek society, and if the Iliad is to be understood as a repository of important contests, then scientific journal articles must also be thought of as contributing stories of conquests over other domains to the epic metanarrative of science.

Topical threads in scientific research consequently emerge as a style of knowing that is more similar to the bardic modes of composition. The iterative cycle of knowledge production and authorship quietly cements images of “proper” ways of being — like what it means to be an author with style — into our collective imagination. This also means that like the Iliad and Odyssey, the content of the scientific investigation is preserved such that it becomes, in a way, immortal through that collective imagination. Viewing science as epic narrative, rife with conflict and rivalry, illuminates scientific literature like the study published by Fasoï et al as a battleground where it’s inscrutable characters can attain glory and immortality.

Battle of the Cyborgs

What then, is the central conflict in the Fasoï et al study? Who are the warriors, and whose side do they fight on? The most obvious conflict is between the cyborgs, who are all vying for the position of most predictive model. Achieving the lowest PPL score of all the cyborgs on a portion of text would distinguish that cyborg from all the rest, for two reasons. First, because it is a position only one cyborg can hold and so naturally separates one from the rest. Second, because it makes that cyborg, out of all the other cyborgs, the closest to a real, whole author who expressed themselves both in the training excerpt and the test excerpt. What the cyborgs authors are fighting for is the right to say that they had a voice, they used it, and it

¹⁴⁵ Merton. p.277

¹⁴⁶ Nietzsche, “Homer’s Contest.” p.177

¹⁴⁷ Nietzsche. p.177

¹⁴⁸ Nietzsche. p.178

mattered in the grand scheme of science. Their primary mission is not just to exist as representations of Homer, or to aid in authorship analysis, but express real, embodied truths in numbers. They are fighting for their right to live.

The idea that inscription is key to creating bodies as individuals reappears in Vernant, suggesting that the disciplinary techniques used to produce digital bodies as individual cyborgs operate similarly to narratives that inscribe individual prestige to specific heroes. In explaining how a heroic death grants the warrior immortality, Vernant writes that “the hero, by the fame he has acquired in pledging his life to battle inscribes his reality as an individual subject on the collective memory of the group; the death that has given his biography has also given it permanence.”¹⁴⁹ Attaining individuality is thus conditioned on a form of self-sacrifice, a certain way of giving oneself over to larger, more important forces at play.

Taking into account that cyborgs are differentiated, individualized, and interrelated with one another in order to “discover” truths of poetic authorship, self-sacrifice in the context of Homeric cyborgs entails relinquishing the opacity of a dual and contradictory existence to allow intelligibility and interpretation by another reader. The psychological testing apparatus used to individualize and rank these cyborgs writes their “biography” with a single metric — PPL. The sum of all their possible behaviors is contained within a single confidence interval, which separates normal from significant behaviors. If we see the cyborgs as warriors, the highest and lowest PPL scores associated with each model are analogous to victory over, or loss to, an opponent. Born into a world that grants authors, but not readers, life and immortality, the terms of victory entails identifying as closely with Homer as is possible.

In this battle for status as the author, the role of cyborgs in the study begins to parallel that of Achilles in the *Iliad*. This parallel can be articulated through Vernant’s analysis of Achilles’ pursuit of *kleos*, and reframes the occurrence of author death as a phenomenon that doesn’t render the Author not as obsolete, but immortal. Vernant explains that obtaining “imperishable glory,” requires living a short life and dying beautifully in battle, and that “a long life in [one’s] own home” means living a life without “renown.”¹⁵⁰ A beautiful death in battle ensures that the hero will have his life’s story transmitted by bards singing epic poetry, which affords the hero immortality because the epics passed down orally and made part of the culture are (allegedly) never forgotten. Like Achilles, the role of the cyborgs is to make concrete “real” historical, Homeric authors whose contributions to the poems can be recognized and glorified — in a way, brought back to life after several millenia of dormancy.

Arguing that Achilles never had to make a choice between ordinary and heroic time, being “dedicated from the outset to a beautiful death,” Vernant’s analysis also reveals the cyborgs as not just any population, but a population of warriors who lacked agency in choosing their life’s purpose. Achilles is bound by his birthright to choose the heroic side of honor over the ordinary, and the cyborgs are faced with a similar non-choice. Vernant writes that Achilles rejects ordinary time (public esteem) in favor of extraordinary time (*kleos*, eternal glory), and this

¹⁴⁹ Vernant, “A Beautiful Death’ And The Disfigured Corpse in Homeric Epic.” p.57

¹⁵⁰ Vernant. P.58

“refusal highlights the tension between ordinary honor, the societal approval necessary for self-definition, and the much greater demands of heroic honor, in which one still needs to be recognized.” Like Achilles, the cyborgs must take the terms that honor and achievement are defined with literally. The need for recognition and self-definition can only be satisfied through competition, which for the cyborgs, comes in the form of performance metrics. Comparisons between PPL scores — the minimum, maximum, and average values — comprise each cyborg’s attempt to out-perform one another, and obtain recognition in the public eye, as they all vie for the position of a historically real, “true” Homer. Having no other dimensions of behavioral existence, PPL is the singular factor determining their individual value and utility.

Another similarity between the cyborgs and Achilles is that their existence, in many ways, is strictly literary. The cyborg’s presence in computational linguistics literature preserves the memory of its existence, just as Achilles and his great deeds were brought to life by the “song that exalts their fame.”¹⁵¹ This means that “As a heroic character, Achilles exists to himself only in the mirror of the song that reflects his own image.”¹⁵² His existence is as mythical, as much a product of narrative function in many ways as Homer’s own life is. Thus if the scientific literary form can be taken as a modern analogue for the *Iliad*, then at least that best cyborgs seems well-posed to achieve immortality through narrative.

New Ways to Live and Die

What do these analogies mean, however, for our ability to salvage an understanding of the cyborgs and author-death? Though cyborgs share a position in narrative and life mission with Achilles, they are unable to achieve *kleos* in the same way that Achilles did. This is due, in large part, to the irony of combining Author, reader, and text in a single body.

For starters, cyborgs cannot die an honorable death in battle. Though PPL comparisons, in the form of individualizing psychological testing apparatuses, create differentiation and ranking systems based on cyborg behavior, that contest does not lead to death for any of the losing cyborgs. They have physical and material existence, for their essence exists as energy held between capacitors and resistors on circuitboards that store them in some server’s memory. But with no automatic program to bring about their death once deemed inferior to that ultimate cyborg Homer, these cyborgs have no access to a beautiful death.

Yet another reason the cyborgs cannot die the beautiful death accorded to true heroes is that they are nameless, and eternally disfigured. Vernant’s analysis of honor and death in the *Iliad* reveals how immortality is tied to “the remarkable beauty of [the warrior’s] corpse,” which is “preserved in a youthfulness that age can no longer mar.”¹⁵³ The corpse’s aesthetic value “raises him above the human condition and saves him from common death by conferring sublime luster on his demise.”¹⁵⁴ No analogous situation can be concocted for the benefit of the cyborgs. They are created as experimental objects and used to help to draw a quantitative picture of oral history.

¹⁵¹ Vernant. p.59

¹⁵² Vernant. p.59

¹⁵³ Vernant. p.64

¹⁵⁴ Vernant. p.64

But the image of traveling poets that the cyborgs are meant to imitate does not really exist in any concrete way — this is why the project is one of authorship analysis rather than attribution — and so the cyborgs can be nothing other than a distortion of reality. If we take nature to be all that is beautiful to the Greek eye, then certainly the cyborgs are not natural, easy representations of anything — they are ugly to their core. Excluded from beauty in life, the corpse of a dead cyborg cannot preserve vitality that never existed in the first place.

Though unable to access a “beautiful death,” the cyborgs do experience one form of death as they slip below the churning surface of scientific inquiry. That death comes in the form of social obscurity and forgetting. While I have presented the article as an important phenomenon, and I genuinely believe that it signposts an important development in the digital humanities, the truth is that it is not a well circulated article — there are only two citations. Vernant writes that “Archaic Greek culture is one in which everyone lives in terms of others, under the eyes and in the esteem of others, where the basis of a personality is confirmed by the extent to which its reputation is known; in such a context, real death lies in amnesia, silence, demeaning obscurity, and the absence of fame.”¹⁵⁵ Without names, it hardly matters whether or not their death is beautiful, because the story of their individual existence cannot be spread through narrative as a hero’s name would be. Thus their existence emerges as one that is ultimately destined to end in forgetting, the significance of their short existence buried under the never-ending production of scientific knowledge.

In the absence of any other names, there is one that emerges from the dust of scientific conflict: Homer. Warriors, though they may claim exclusive access to heroic honor, are not the only figures made immortal through narrative. Homer is the study’s ultimate object of desire, he is that which they seek but will never find. And though he might not be a warrior, he certainly dies in such a way that reinforces his status as an immortal. The old Homer is dead, because there is a real shift in the inquiries and methods used to investigate the Homeric Question, through which Homer is preserved in memory, and made into a hero like any other. The old directions of inquiry have been opened with new methods, and so must be understood as an additional offshoot of the Homeric question that fundamentally changes who the composite of memories and imaginaries that make up the figure of Homer. Now that the collective memory of Homer includes the domain of computational linguistics, who Homer is to our contemporary understandings constitutes a real change from the existence that his figure has enjoyed in the humanities.

Homer’s immortality is ironically not in accordance with the structure of immortality that Vernant identifies for warriors in the *Iliad*. His immortality is the function of contest between unnamed cyborg soldiers, who sacrifice their own names and bodies for the glorification of his Homer’s name. The cyborgs, though they may be nameless and ugly, carry Homer’s name as if it were emblazoned upon their shields. Regardless of the minute details of the cyborg’s being, regardless of whether we take them to be authors, readers, texts, or something else entirely, Homer is their north star, their truth, their ideal. And before they slip into obscurity, they spend

¹⁵⁵ Vernant, p.57

their last breath referring back to that north star. Such idealization allows Homer to endure above the fray, where cyborgs and grand narratives advocate a particular vision of authorship. In a sense, the combination of death and enduring presence appears as a cycle of renewal, and perhaps even rebirth.

Yet another irony appears here, in the form of subverted expectations. It is deeply ironic that the cyborgs are warriors engaged in contest, yet they march towards the battlefield without any chance of a beautiful death or immortality. It is even more ironic, that by reading the text in with their statistical bodies, they reinvigorate the figure of Homer with their own vitality. Thus the collapse of identity categories such as author, reader, and text need not signify an emptiness or absence; instead it can be taken as a sign of coming changes in the way that we envision life and death.

Returning to a Barthes-esque view of the cyborg, it appears that the cyborg's triple identity makes the relationship between authorship, reading, death, and immortality nearly incomprehensible. By reading the text, the cyborg immortalizes Homer. Though that immortality may be achieved through a sort of death, the renewal of Homer in a shifted light indicates that immortality has been severed from the author's original narrative in this case. But the reader pre-exists the death of old Homer; Homer didn't have to die in order for this reader to be born in the same sense that Barthes proposed. Homer's death and renewal is simultaneous, a point in time too small to pinpoint. Ironies abound here, and suggest that Author death is no longer as useful of a concept as it once was. So what do these ironies tell us?

If we revel in these ironies, take them as rich and meaningful, then a picture of the historical Author emerges that is not a linear image of life and death, of bodily or temporal separation. The ontology attached to cyborgs brings to light the artificiality of philosophical categories like "Author" and "reader," and the ease with which we transcend and violate them. The cyborgs highlight that adherence to a single identity is a choice we make when scaffolding the interior architecture of our selves, and that choice is certainly fraught with political and philosophical power dynamics. Transforming author into reader, and reader into author, the cyborg is neither author nor reader alone; it embodies the whole cycle of reading and renewal that makes any author immortal. The old Author dies as the Reader reads, and that very act of reading brings a new imagined Author into being. Irony subverts the need for boundaries between roles, it shows how malleable our view of the world is. Thus the cyborg re-imagines Barthes' linear timeline of Author death as a self-renewing cycle, which is funnily appropriate in light of the cyborg's relationship with oral tradition. This is the cyborg's core contribution: Barthes may have claimed that the Author is dead, but placing the cyborg in Archaic Greek culture tells us that even dead cultural heroes are not easily forgotten, and live on through memory.

Bibliography

- Barthes, Roland. "The Death of the Author," 1968.
<https://writing.upenn.edu/~taransky/Barthes.pdf>.
- Campagnola, Chiara. "Perplexity in Language Models." Medium, November 5, 2022.
<https://towardsdatascience.com/perplexity-in-language-models-87a196019a94>.
- Dauber, Edwin, Rebekah Overdorf, and Rachel Greenstadt. "Stylometric Authorship Attribution of Collaborative Documents." In *Cyber Security Cryptography and Machine Learning*, edited by Shlomi Dolev and Sachin Lodha, 10332:115–35. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017.
https://doi.org/10.1007/978-3-319-60080-2_9.
- Eder, Maciej. "Style-Markers in Authorship Attribution A Cross-Language Study of the Authorial Fingerprint." *Studies in Polish Linguistics*, 2011.
- Elmer, David F. "Oral-Formulaic Theory." In *The Homer Encyclopedia*, edited by Margalit Finkelberg, wbhe1030. Oxford, UK: Blackwell Publishing Ltd, 2011.
<https://doi.org/10.1002/9781444350302.wbhe1030>.
- Fasoi, Maria, John Pavlopoulos, and Maria Konstantinidou. "Computational Authorship Analysis of Homeric Language." In *Digital Humanities Workshop*, 78–88. Kyiv Ukraine: ACM, 2021. <https://doi.org/10.1145/3526242.3526256>.
- Foucault, Michel. "What Is an Author?" *Screen* 20, no. 1 (Spring 1979): 13–34.
- Haraway, Donna J., and Cary Wolfe. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Manifestly Haraway*. University of Minnesota Press, 2016. <https://doi.org/10.5749/minnesota/9780816650477.001.0001>.
- Ke'selj, Vlado, Fuchun Peng, Nick Cercone, and Calvin Thomas. "N-GRAM-BASED AUTHOR PROFILES FOR AUTHORSHIP ATTRIBUTION," n.d., 10.
- Lapham's Quarterly. "Whose Homer Is It Anyway? | James I. Porter." Accessed April 11, 2023.
<https://www.laphamsquarterly.org/roundtable/whose-homer-it-anyway>.
- Merton, Robert King. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press, 1973.
- Nietzsche, Friedrich. "Homer's Contest," 1872.
- Peng, Fuchun, Dale Schuurmans, Vlado Keselj, and Shaojun Wang. "Language Independent Authorship Attribution with Character Level N-Grams." In *10th Conference of the European Chapter of the Association for Computational Linguistics*. Budapest, Hungary: Association for Computational Linguistics, 2003. <https://aclanthology.org/E03-1053>.

- Porter, James I. *Homer: The Very Idea*. Chicago: University of Chicago Press, 2021.
- Nikolas Rose, *Inventing Ourselves: Psychology, Power and Personhood* (Cambridge, United Kingdom New York, NY Port Melbourne, Australia Singapore: Cambridge University Press, 1996).
- Russo, Joseph. "Formula." In *The Homer Encyclopedia*, edited by Margalit Finkelberg, wbhe0509. Oxford, UK: Blackwell Publishing Ltd, 2011.
<https://doi.org/10.1002/9781444350302.wbhe0509>.
- Scodel, Ruth. "Bardic Performance and Oral Tradition in Homer." *The American Journal of Philology* 119, no. 2 (1998): 171–94.
- Vernant, Jean-Pierre. "'A Beautiful Death' And The Disfigured Corpse in Homeric Epic." In *Mortals and Immortals: Collected Essays*. Princeton N.J: Princeton University Press, 1991.
- Weth, Constanze. "Distinguishing Syntactic Markers From Morphological Markers. A Cross-Linguistic Comparison." *Frontiers in Psychology* 11 (2020).
<https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02082>.