ESSAYS

CAN AN ALGORITHM BE DISTURBED? MACHINE LEARNING, INTRINSIC CRITICISM, AND THE DIGITAL HUMANITIES

JAMES E. DOBSON

> *Never act except in such a way that your action may be programmed.*
> —*Jacques Lacan,* The Ethics of Psychoanalysis, 1959–1960 *(1986)*

> *He who regards poems only as objects to be "processed" according to one or another method should admit to himself that the processing of leather into shoes is more useful to mankind than the processing of poems into interpretations.*
> —*Sigurd Burckhardt,* Shakespearean Meanings *(1968)*

Within literary and cultural studies there has been a new focus on the "surface" as opposed to the "depth" of a work as the proper object of study. We have seen this interest manifested through what appears to be the return of prior approaches including formalist reading practices, attention to the aesthetic dimensions of a text, and new methodologies that come from the social sciences and are interested in modes of description and observation.[1] In arguing for the adoption of these methodologies, critics have advocated

COLLEGE LITERATURE: A JOURNAL OF CRITICAL LITERARY STUDIES 42.4 Fall 2015
Print ISSN 0093-3139 E-ISSN 1542-4286

for an end to what Paul Ricoeur has termed "the hermeneutics of suspicion" and various forms of ideological critique that have been the mainstay of criticism for the past few decades.[2] While these "new" interpretations might begin with what was once repressed through prior selection criteria, they all shift our attention away from an understanding of a "repressed" or otherwise hidden object by understanding textual features less as signifier, an arrow to follow to some hidden depths, and more as an object of interest in its own right. Computer aided approaches to literary criticism, or "digital readings" (not an unproblematic term, to be sure), have been put forward as one way of making a break from the deeply habituated reading practices of the past; but their advocates risk overstating the case and, in giving up on critique, they remain blind to untheorized dimensions of the computational methods on which they rely. While digital methods enable one to examine radically larger archives than those assembled in the past, a transformation that Matthew Jockers characterizes as a shift from micro to "macroanalysis," the fundamental assumptions about texts and meaning implicit in these tools and in the criticism resulting from use of these tools belong to a much earlier period of literary analysis.

Sharon Marcus and Stephen Best's well-known essay, "Surface Reading: An Introduction," which introduced a volume of the journal *Representations* in 2009, is dedicated to the topic of "How We Read Now" and examines several variants of surface reading as an alternative to depth or "symptomatic" reading. Marcus and Best name the digital humanities and computer-assisted reading as one important and particularly hopeful methodology for the future of humanistic study. They write: "Where the heroic critic corrects the text, a nonheroic critic might aim instead to correct for her critical subjectivity, by using machines to bypass it, in the hopes that doing so will produce more accurate knowledge about texts" (2009, 17). Replacing the "heroic critic" of the symptomatic era with the heroic code, they imagine an objective world of bypassed subjectivity. Without cultural knowledge, biases, political commitments, in other words, without being situated, Best and Marcus believe that the machine and the algorithm can produce more "accurate knowledge" about the world brought into being by subjective human beings. This is to say, that digital or computer-aided readings are imagined as escaping the subjective constraints that draw us to certain passages, figures, or conclusions. An algorithm can be excluded from the hermeneutics of suspicion because it knows nothing of the concept of "hidden" depth. This leads Best and Marcus to claim that digital readings might restore a "taboo" set of goals for humanistic study: "objectivity, validity, truth" (17).

Heather Love's recent and provocative essay on surface reading, "Close but not Deep: Literary Ethics and the Descriptive Turn" (2010), takes up some of the challenges to close reading identified by Best and Marcus. Two main concerns with criticism as now practiced concern Love: first, literary critics have privileged witnessing and empathy and thus turned literary criticism into an ethical act that draws its power from the charisma of the critic rather than the text. Secondly, the hermeneutical method of close reading that is the trademark of humanistic disciplines and the methodological *lingua franca* of most of the criticism of the past fifty years—one should note, from the New Criticism to Marxism to feminism to deconstruction—has isolated the work of humanists from other disciplines within the university that do not closely attend to language. While digital humanists like Jockers and Franco Moretti have imagined the only solution to these two problems to be the rejection of reading literature as we read now, to give up on the singular text (the corpus) in favor of collections of texts (corpora), Love wants to keep reading singular objects like the novel but to rethink the activity of reading. She suggests that we renounce "depth hermeneutics" in favor of an approach she provocatively turns into a motto for our moment: reading closely, but not deeply. What keeps Love securely on the surface of her essay is her belief in what she herself terms the "normative view" of science, and in particular the social sciences. She wants to bring literature and humanistic study more broadly into the sphere of the scientific view in order to participate within the currency presently available to this discourse. This supposedly benign "normative" aspect enables Love to reconceive interpretation as description. This is to say that Love presupposes an uncontested descriptive view of literature. This so-called "normative" view of the sciences links Love's surface reading and Best and Marcus's belief that the digital humanities can deliver "objectivity, validity, truth" (2009, 17).

In this essay I will examine practices and methods of computer-aided text mining because these collectively represent what I take to be the strongest form of digital humanities. The machine learning algorithms that enable the majority of text mining efforts are widely used in other disciplines and are not a marginal and arbitrary corner of the digital humanities but central to the effort to reposition humanistic research within the bounds of current university research protocols. Matthew Jockers asserts that computational text analysis is "by all accounts the foundation of digital humanities and its deepest root" (Jockers 2013, 15). This "root" of the digital humanities has a long history, one that Jockers connects to the digital concordances of Thomas Aquinas produced by Father Roberto Busa in the 1940s and,

as I will show, is attached to the deep dreams of structuralism and its desire for a science of interpretation. I will begin by first discussing in detail several proposed methods of machine reading as possible answers to the call for surface reading. These quantitative methods come to literary studies from outside the humanities and are well understood in certain contexts, especially those within the empirical sciences; my reframing of computer-aided text mining will draw out what I take to be the theoretical assumptions implicit within these models of meaning. I'll then turn to an analysis of structuralism in order to demonstrate the degree to which the digital humanities and machine readings of text have resurrected key structuralist presuppositions. In the process, I will discuss two important critiques of structuralism from within literary studies and consider what we still have to learn from these interventions. Finally, I will argue that these critiques enable us to call into question the division between the act of interpretation and objective, scientific methodology in the rhetoric of this strong form of the digital humanities.

THERE IS NO SUCH THING AS THE "UNSUPERVISED"

Even though his methodology is not explicitly or even necessarily digital, Franco Moretti is a useful figure to examine the stakes of the digital intervention being proposed by critics like Best and Marcus. Like Love, Moretti has, for some time now, articulated his frustration with close reading. Unlike those critics exhausted with suspicious close reading because its moves have been appropriated by those with differing political agendas (Latour 2004, 225–30), or because criticism encourages forms of exposure that reinforce the critic's sense of knowing more than the object studied (Sedgwick 2003, 138–43), or even because critique has just become rote and boring (Felski 2009, 31), Moretti's frustration originates within the limitations of the narrow scope of close reading that complicates his ability to criticize larger forces and systems. Calling the slow, careful close reading of an individual text a "theological exercise," he accuses literary critics of giving too much attention to a small set of mostly canonical texts. Moretti wants his proposed practice of *distant reading* to enable an understanding of "the system in its entirety" (Moretti 2000, 57). In other words, it is precisely the failure of abstraction foreclosed by the specificity of the singular close reading that motivates Moretti's desire for a distanced position capable of a producing a systemic critique.

When he puts his distant reading theory into practice in *Graphs, Maps, Trees* (2005), Moretti presents an alternative approach to the digital yet still qualitative methodology imagined by Best and Marcus. What keeps

Moretti's claims to systematicity resulting from his quantitative analysis "honest" is the fact that his target is not something like a hard, empirically knowable reality, but rather the socially constructed fiction known as the market. Thus, Moretti is able to stake out a stronger position for quantitative research within the humanities:

> Quantitative research provides a type of data which is ideally independent of interpretations, I said earlier, and that is of course also its limit: it provides *data*, not interpretation. . . . Quantitative data can tell us when Britain produced one new novel per month, or week, or day, or hour for that matter, but where the significant turning points line lie along the continuum—and why—is something that must be decided on a different basis. (Moretti 2005, 9)

Moretti's turn to the scientific quantitative from the humanistic qualitative takes as its presupposition some fundamental distrust of the act of interpretation. The interpretive act of reading, in his account, is too tied up with evidence. Literary critics have what social scientists would call a selection bias that always informs the practice of close reading. Moretti, like almost all digital humanists, seeks to address this problem through the separation of his scientific methodology with its accompanying "data" from the hermeneutic act of interpretation. The substitution of what close reading would call textual evidence with quantitative data—for Moretti, the length of book titles, the number of books within specified categories sold, the number of booksellers—enables his strong claim for a quantitative approach to literature.

However, there is of course no such thing as context-less data. The concept of raw data, as Lisa Gitelman and Virginia Jackson have recently argued, is something of a misnomer, an oxymoron as they point out in the title of their jointly edited volume (Gitelman and Jackson 2012). We should doubt any attempt to claim objectivity based on the notion of bypassed subjectivity because human subjectivity lurks within all data. This is because data does not merely exist in the world: each data point is an abstraction imagined and generated by humans. Not only that, but there are always criteria informing the selection of any quantity of data. This act of selection, the drawing of boundaries that names certain objects a "data set" introduces the taint of the human and subjectivity into supposedly raw, untouched data.

Data, contra to the desires of Best and Marcus, cannot ever be said to be computed, distilled, and analyzed free of subjective intent. Best and Marcus do not elaborate on the specific digital technologies that they believe will lead to objectivity and they do not differentiate between

computer-aided and supposedly completely automated approaches. Even if data were free of subjectivity, the approaches that have been presented as completely automated, unsupervised, and human-free turn out to be even more tainted by subjectivity than the original "data" selection process. Machine learning, the set of algorithms that enable computer-aided text mining, is a relatively recent technology. It provides us with an ideal test case for examining the possibility of objective readings of text. To provide some background, we first need to situate machine learning in its place as a branch of artificial intelligence. These techniques uniquely have the ability to address an incredibly large amount of data with varying degrees of input from a researcher. They have been used to transform approaches in fields as diverse as economics and cognitive neuroscience. The "learning" element of machine learning suggests the continued repetition of an automated task that can reflexively integrate the results of past tasks. Ideally, with each repetition the algorithm improves the accuracy with which it performs the task and therefore it can be considered to be "learning."

There are two kinds of machine learning algorithms: supervised and unsupervised. Supervised machine learning can categorize data into predefined and predetermined categories; data, in the case of the application of machine learning to literature, would be groupings of words of various lengths and sizes. What makes supervised machine-learning algorithms "supervised" is the existence of a training dataset. In this form of machine learning, there are always two datasets. The human researcher parcels a set of texts or other objects into two buckets. The first bucket is the known and familiar bucket, the training dataset. Here "labels" are attached to each object that defines its membership within a category. The algorithm "trains" itself on the training dataset. After this training it extracts sets of features from this data and uses them to categorize the objects into the researcher-created categories. These features are then used on what is called the "test" dataset. The texts comprising the "test" dataset should be similar to those found within the training dataset. The algorithm then automatically sorts the data within the test dataset into the categories defined by the researcher.

If the use of the term "supervised" by computer and information scientists suggests the presence of what Best and Marcus would call "critical subjectivity," then "supervision" must be understood as the unavoidable presence of the human subject within this area of machine learning. Supervision means that our interpretation of the results, the output from the algorithm, must take into account decisions made by the researcher to establish a set of initial conditions. These conditions might be the existence of labels that, while not providing explicit rules

or criteria for categorization, mark each dataset, each text or grouping of words, as unambiguously a member of a particular category. Thus the results of any supervised algorithm contain traces of decisions made by the researcher—precisely the "subjectivity" this work might be imagined to lack. Unsupervised algorithms would presumably be uncontaminated by any such influencing traces of the researcher. Yet this unsupervised state cannot be said to exist. The researcher must, as Gitelman and Jackson remind us, necessarily make a set of decisions in forming the original input dataset, even if it is completely unlabeled and considered disorganized. We must also choose an algorithm from the range of available options and then an implementation of this algorithm. Not only will different machine-learning algorithms give different results, but differing implementations of the same algorithm may not agree. Reproducible results will depend upon the precise replication of the software and hardware environment used. Reproducibility remains an ideal for all computational fields, but in practice is very difficult to achieve, even more so when we are searching for small yet statistically significant bits of evidence for our claims.

Supervised methods are frequently called computer-aided. A program or application using machine learning might, for example, make use of a dictionary of key terms that define topics of interest that can be used to index documents. One such method is known as sentiment analysis. One of the leading researchers in the field defines sentiment analysis as "the computational study of opinions, sentiments, and emotions expressed in text" (Lui 2010). Used mostly by social scientists and those in marketing fields, sentiment analysis takes a set of terms associated with positive and negative emotions and then automatically sorts texts or fragments of texts into these pre-defined categories. The sub-categories and key terms are hardly universal; these terms are necessarily the product of the specific period and cultural milieu in which the dictionary was assembled. A notable example can be found within the psychology dictionary that forms part of the sentiment analysis dictionary distributed with one popular commercial text mining package, Provalis Research's QDA Miner/WordStat.[3] WordStat's psychology dictionary contains a set of 3,150 total terms that align concepts and phrases into groups associated with psychoanalysis. Not just any "dialect" of psychoanalysis, however: Colin Martindale, the author of this dictionary, chose to organize his terms into areas associated with the popular practice of the 1970s, Jungian psychoanalytic analysis.[4] One slightly idiosyncratic grouping found within the dataset, "Icarian Imagery," demonstrates the limitations of the model used by this dictionary. Martindale identified these terms, and

also those terms making use of these root-words, with the sub-category of "ascension." They have been grouped together within the larger Icarian category:

ICARIAN_IM
ASCEND
ALOFT*
ARIS*
ARISEN*
AROS*
ASCEND*
ASCENS*

Of course psychological concepts from more widely accepted strands of psychoanalysis, including Freud's own theories as well as developments in the field after the influence of the cognitive and brain sciences, have no representation within the dictionary. Thus this dictionary would enable one to locate potential sources of evidence for reading Jungian imagery and associated categorizations of sentiment within a text but not, say, the terms used by the New Psychology of the 1890s that preceded psychoanalysis as the dominant discourse or those from the present that reflect an understanding of the mind derived from empirical studies of the brain.

I invoke this dictionary and the relatively new science of sentiment analysis to question some of the assumptions held by those promoting versions of machine reading and also to question the possibility of formalized and fully automated reading. This is to say, there cannot be an automated reading of a text that is free of the "taint" of subjectivity. Reading, I would claim, is always situated. Best and Marcus were wrong to imagine and hope for analysis free of subjectivity and digital humanists are wrong to insist on the separation of methodology and data from the act of interpretation. But I would hesitate to ask digital humanists to limit the application of their methods to historicist approaches that would take, as an example, this Jungian dictionary and apply it to literary works that appeared at the exact same time and in circulation with this psychoanalytic discourse. To do so would be to give up on much of the promise of digital methods and produce only a slight improvement over existing historicist readings. At the same time, we should recognize that computational science itself is always historicizable. Even though they tend to increase, data resolutions and system capacities are subject to hardware limitations. Algorithms change or "evolve" and are constantly subject to modification. Bugs are discovered and new ones introduced. Scientists

depending on complicated configurations of software known as "pipelines" or workflows are discovering this. In addition to archiving collected data, these scientists are now seeking to archive the exact versions of software used to analyze and produce the final end products of their pipelines.[5] Not just software, but also the hardware used in data analysis can produce differences that can introduce variance into the results of computation. In short, just because we are using machines to read text doesn't mean that they give the final, definitive reading.[6]

TOPIC MODELING AND CRITIQUE

There are, however, other widely used methods of digital reading making use of machine learning that do not depend upon either pre-labeled data or the assistance of a user-created or pre-supplied set of key terms. The method of topic modeling would seem to relieve us of the need for specialized dictionaries like the one distributed with WordStat. Probabilistic topic modeling, or simply topic modeling, is an emergent digital reading method that is quickly becoming popular. This method comes to the humanities from the information sciences; to what extent it might still belong to the latter is an open question. Topic models are a way to organize a large and unlabeled collection of documents into computer-generated thematic categories. Rather than supplying a list of hierarchical keywords to group documents, the algorithm "discovers" shared topics based on textual features that are used to fit documents into the discovered categories. Using single words to build our list, we might receive the following output of possible topics from Henry Adams's *The Education of Henry Adams*:

> Topic 0: adams, henry, minister, felt, john, washington, young, hay, asked, came, saw, went, point, took, wanted, long, war, century, reason, father.

If we decide to use multiple words to locate possible themes, the multiple term units popularly called "n-grams," we might receive the following output:

> Topic 0: private secretary, knew better, lord russell, diplomatic education, young adams, lord palmerston, young man, young men, free soil, earl russell, eighteenth century, english society, fayette square, fifty years, foreign affairs, francis adams, henry adams, george washington, half dozen, harvard college.

Some of the same words are captured in this second example; "henry" and "adams" are now grouped together as they most frequently appear this way in this third-person authored autobiography. While the term "washington" appears as the name of the city when searching for single

words, when we search for two or more word phrases, the algorithm returns the proper name. We note that the term "war" no longer appears in this, the first topic grouping returned by the algorithm, although "civil war" does appear in the second list of possible topics (not displayed).[7]

However, even these unsupervised implementations of machine-learning algorithms are subject to some of the critiques outlined above. For example, all machine learning implementations (both supervised and unsupervised) capable of performing text-mining need to make some conversion and initial reduction of the string of characters that comprise a text or document. And if a text has been digitized from a print edition, then one has to make a selection of the digital edition. Potentially the machine-learning package will attempt to convert the text from one encoding to another, for example from the simple ASCII encoding to the more modern UTF-8 or vice versa. In the process, accents, diacritical marks, and other textual features may be removed or translated to equivalent marks. The workflow, or automated set of procedures, might perform what linguists refer to as lemmatization on the string of words, which is to say, the trimming of each word into its smallest meaningful components, as well as removing plurals, capitalization, punctuation, and tense. For most humanists, this process produces potentially large-scale "information" loss.[8] In addition, almost all machine learning implementations used on text include what is called an exclusion list, or stop words. Stop words are terms that are considered to be lacking in semantic content. These words are removed before running the text through the algorithm because they are considered superfluous; they are "noise" that would make that task of document classification much more difficult. MALLET ("MAchine Learning for LanguagE Toolkit"), a popular and free topic-modeling package, contains a default stop word list of 524 English language words (see McCallum 2002). While this set contains words like "you," "no," "but," "and," and "whatever," it also contains terms of potential interest to the humanities researcher like "associated," "appreciate," "sorry," and "unfortunately." And even the words from this previous list—take "whatever" as an example—might have important meanings and semantic value depending on genre and period (think 1980s valley girl). Each methodological decision in pre-processing involves some aspect of interpretation. All of this is to say that within humanistic approaches there are no words that do not signify—everything is potentially signal and nothing is noise.

As many digital humanists have argued, topic modeling and other computer-aided reading methods are just the beginning (see for example Liu 2013, 409–23). Traditional interpretive activities take over once the distilled and computed results have been generated. Unfortunately, the

division between interpretation and method enables us to forget what deconstruction has taught us to recognize as the pretext—the whole range of presuppositions for conducting the reading, including the set of initial conditions or states such as the selection of stop words, a spelling and internationalization standard, codes, algorithms, and text encoding schemes. Jacques Derrida has shown the way in which an "outwork" (*hors d'ouvre*) functions in Descartes and Hegel to preface the "main" philosophical text by stating methodological commitments (Derrida 1981, 3–22). For Derrida, the pretext of these "outworks" is that they are introductory, that they come before the text; rather, he argues, they are produced after the text and are predetermined by the text (20).[9] Likewise, in digital criticism the task of interpretation cannot be framed as isolated from methodological decision making and all the algorithmic and computational presuppositions. In a recent article examining the history of literary criticism through machine learning and topic modeling, Andrew Goldstone and Ted Underwood (2014) attempt a rigorous theorization of the digital methods they deploy. They applied the MALLET topic modeling application described above to a digitized archive of academic journals dedicated to the field of literary studies. They deployed these techniques in order to answer the question of when "criticism" and "critique" came to dominate literary studies. Goldstone and Underwood's project evinces what might come to be recognized as an important turning point in the application of computer reading techniques to literary studies: they bring a sharp and critical account of these tools while using them to produce readings and counter-narratives of their own field formation. Yet there are still some unquestioned assumptions that persist in their methodology.

When Goldstone and Underwood search through their archive of journal articles for the term "criticism" they depend upon an understanding that articles using what they term "*critic*-words" (words beginning with the root-word "critic") are doing the work of criticism and those that do not are not "critical." Perhaps recognizing this would not alter the story told by their model, as they claim that the model only adds "nuance" to an already familiar account of the "emergence and subsequent naturalization of the discourse of criticism over the whole course of the twentieth century" (Goldstone and Underwood 2014, 370). But a nuanced reading of the qualitative difference between critical practices and the specific language of criticism that self-referentially invokes "*critic*-words" is foreclosed by the rejection of reading practices resulting from theoretically informed close reading. There is a risk of creating categorical errors through a reliance on the self-evident stability of these categories.

STRUCTURALISM AND SYSTEMIC CRITICISM

Heather Love's turn to the social sciences and Goldstone and Underwood's use of topic modeling are part of a larger movement that advertently or inadvertently functions to reposition literary study as a science. Goldstone and Underwood explicitly place their work in "the recent tendency for literary studies to develop stronger connections to social science" (Goldstone and Underwood 2014, 379). This movement tends toward sharing the desire to answer to Best and Marcus's call for "objectivity, validity, truth" in literary criticism (2009, 17). There are, to be sure, dissenting voices within the digital humanities. While arguing that all criticism is algorithmic, Stephen Ramsay suggests that rather than longing for a scientific criticism "we would do better to recognize that a scientific literary criticism would cease to be criticism" (2011, 15). Yet all of these differing approaches can be understood to be part of a retrograde movement that nostalgically seeks to return literary criticism to the structuralist era, to a moment characterized by belief in systems, structure, and the transparency of language. It was well before our present concern with re-theorizing the surface of the text, prior even to the advent of "symptomatic reading," that those working within literary studies dreamed of the possibilities of a scientific criticism. Northrop Frye's *Anatomy of Criticism*, originally published in 1957, serves as one such early work. Frye made the polemical case for a systematic and scientific criticism derived from an inductive reading of literature that could encompass all of literature. He outlines the expansive scope of his approach by creating "a theory of criticism" explicitly modeled after Aristotle "whose principles [would] apply to the whole of literature and account for every valid type of critical procedure" (1971, 14). This approach would work, he argues, because like a scientific investigator, he assumes the existence of an order of nature, an order of meanings that lies behind the enterprise known as literature and exists as a coherent whole. Discovering the laws governing this order becomes the task of the critic. This understanding enables Frye to read widely across numerous literatures, to extract major modes and archetypes, and to produce a categorization of all these into a single organizing schema. Individual texts are then brought, either by Frye or a future critic, into the law of the schema and used to establish minor variations on a theme. This is what he believes makes his system scientific: each revision made by critics and scholars builds progressively on the entire body of prior humanistic research. Above all, Frye's schema works in pursuit of what he sees as a set of unalterable structural principles that can guide future criticism and reading. It is a "genuine" mode of criticism—to be differentiated from the accretion of judgments made

by literary taste makers, or what he calls "meaningless" criticism—that follows the research models provided by science and "progresses toward making the world of literature intelligible" (9).

Frye's *Anatomy* creates what he calls a "conceptual universe" in which all of literature can be plotted, located, and mapped. His schema are ultimately less rigid than we might expect and one particularly important feature of Frye's system is its own open-endedness: he intended that categories beyond those major labels that have made his book famous—the mythic, generic, and archetypal—would de added in order to improve his theory and even make it obsolete. Yet as Geoffrey Hartman notes in an important critique, the mythic holds a central place in Frye's system. Hartman selects this category because he believes that myth occupies a blind spot in Frye's system. Like our present moment, the possibility of a scientific criticism was deployed as a "surface" against the concept of depth. This leads Hartman to call Frye's method a "flattening out" of literature in opposition to the "depth criticism and depth psychology" of their shared historical moment. Frye's *Anatomy* is ultimately spatial. The *Anatomy* charts and maps the literary terrain and in so doing it drops what Hartman believes to be an important dimension: time. Claiming that "literature unfolds in time rather than quasi-simultaneously in space," Hartman criticizes Frye's understanding of temporality and literary history (1971, 33). The system evades the question of historical development by treating all literature as essentially co-occurring and finding little use for concepts like tradition, influence, and inheritance. This leads to Hartman's greatest concern. He worries that such "archetypal analysis can degenerate into an abstract thematics where the living pressure of mediations is lost and all connections are skeletonized" (30–31). Without the literary-historical network, the system that takes its place finds a series of dead-end nodal points.

Yet this network is precisely what myth requires and what it reworks. There are no "pure" forms of myth in Frye's system, only multiple appearances of historically situated myths. Hartman writes, "a writer does not confront a pure pattern, archetype, or convention, but a corpus of tales or principles that are far from harmonized" (Hartman 1971, 37). Hartman was right about the disappearance of history from Frye's system as a cause for concern. History is, after all, one key to humanistic inquiry. Without the nuanced understanding of the ways in which ideas and representation unfold throughout time, literary critics would be the social scientists that Frye seeks to distance himself from. Thus, Frye explicitly rejects the sociological reading advocated by critics like Heather Love and the work of contemporary digital humanists like Goldstone and Underwood:

> I understand that there is a Ph.D. thesis somewhere which displays a list of Hardy's novels in the order of the percentages of gloom they contain, but one does not feel that that sort of procedure should be encouraged. The critic may want to know something of the social sciences, but there can be no such thing as, for instance, a sociological "approach" to literature. (Frye 1971, 19)

What seems most interesting about the many contemporary digital humanities projects when compared to prior forms of scientific criticism is the deep focus on history. Indeed, these projects seem to have incorporated critiques such as the one Hartman makes of Frye, and make the temporal dimension central to their inquiry. Very large-scale archives such as Google's "Google Books" enable heretofore impossible readings across the *longue durée* of literary history. Tools like the "Ngram Viewer" make the historical tracking of word or phrase genealogies through almost all of print history a trivial task. Thus these projects could be understood as answering Hartman's main complaint of the systemic and scientific approach to literature. And yet I want to argue that the hermeneutical critique of what Hartman calls the "sweet science" remains helpful advice to the would-be scientific literary critic.

DISTURBING CRITICISM?

Frye's archetypal system shares much with structuralist criticism of the 1960s and 70s. Both approaches seek to organize all of literature into well-defined categories and take as a founding assumption the existence of an ordered world that could be illuminated through progressive critique. Like Frye, the structuralists explicitly referred to their practices as a science. This was in part because structuralism came to the humanities from the social sciences, but also due to its status as a classificatory methodology. In his well-known 1967 essay "From Science to Literature," Roland Barthes describes the structuralist commitment to taxonomy:

> Structuralism, by virtue of its method, pays special attention to classifications, orders, arrangements; its essential object is taxonomy, or the distributive model inevitably established by any human work, institution, or book, for there is no culture without classification; now discourse, or ensemble of words superior to the sentence, has its forms of organization; it too is a classification, and a signifying one. (Barthes 1989, 6)

For many, structuralism was essentially a formalism. Like Frye's system it erased history and like the New Historicism that would eventually follow structuralism, it operated synchronically rather than diachronically. It formed schema based on the presupposition of a closed world of meaning

that enabled the taxonomization of texts and the components of a text. The forms or categories, however, were not necessarily considered objective and arguments over selection and categorization prevented the production of any truly definitive readings.[10] It understood itself as an improvement upon what has become known as New Criticism primarily through the introduction of Ferdinand de Saussure's division between *langue* and *parole*. Structuralists understood each individual textual object, the closed world of the poem as theorized by the New Criticism, as an instance of enunciation, or what Saussure called *parole*, and recast the object of criticism as the system, the *langue*, which produces the grounds of possibility for the individual poem. Consequently, they desired a larger object of critique and to accompany it a common language to be used by the community of scholars. They realized that certain formalist methods that depended upon close reading would not, to use one of today's popular terms within computational fields, "scale."

Barthes's turn from structuralist to poststructuralist hinges on his discovery that there is no "neutral state of language" that would allow literary criticism to become a scientific enterprise; structuralism cannot "call into question the very language by which it knows language" (Barthes 1989, 7). Contra the structuralists and the New Critics, there were no closed worlds and no "common language." Barthes argues that the descriptive language of scientific discourse is not a metalanguage, a "superior code," but merely one code co-existing and layered among many others. Deconstruction, the most prominent mode of poststructuralist thought, called into question the stability of the spatial features that enabled Frye's charts and maps by drawing attention to what could be thought of as the continental drift active underneath the surface. Deconstruction questioned the oppositional forms that enable structuralism to establish categories. In pushing aside this fundamental insight from deconstruction, as well as the various forms of political critique that remain linked with this project, the digital humanities work described in this essay repeats the categorical errors of structuralism.[11]

While literary critics were still engaged in forms of critique influenced by Frye and the structuralists, Sigurd Burckhardt produced a strong critique of the mechanical tendencies found in these methodologies. Burckhardt's "Notes on the Theory of Intrinsic Interpretation" appeared as an appendix to his *Shakespearean Meanings* (1968). This essay sought to revitalize literary criticism primarily through the division of intellectual labor into two categories: explanation and interpretation. His own categorizations enable him to make an unusual defense of hermeneutics by arguing against the understanding of interpretation as the description

of the way in which a work of literature "works." Interpretation is not the accounting for why a text follows certain mythic laws or archetypes but a mode of discovery that takes as its primary object the text itself. At the same time, if the surface reading and digital readings of the present reject the conception of "depth hermeneutics," so too does Burckhardt's theory. For what calls his reading practice into action is not the deeply buried symptom, the sense of deep meaning to be revealed by the critic, but something on the surface that troubles our ability to give a structuralist account of the text.

Burckhardt argues that insofar as it has a methodology capable of supporting a theory, science is intrinsic. By this Burckhardt means, like Northrop Frye, that science understands the universe as ordered and organized by a set of discoverable laws. Like the religious belief that science has made obsolete, the entire world postulated by empiricism is subject to intrinsic analysis. Everything must have a place and meaning. Interpretation, according to his account, "would mean the attempt to know the law of a poem *solely from the poem itself,* on the necessary assumptions of the infallibility of the poem. Explanation, on the other hand, would mean the attempt to demonstrate how parts of a poem obey an already known, established principle" (1968, 298).

Interpretation and explanation map onto, respectively, intrinsic and extrinsic analysis. Frye's conception of the "order of words" necessitates an intrinsic approach and this shares some assumptions with those of Burckhardt. Burckhardt, however, places his emphasis on the hermeneutical act called into being by the intrinsic method. While the residual New Critical focus on poetry and the single poem draws Burckhardt toward the poem as his unit of interpretation, there is no reason why this procedure should be limited to a single poem, to poetry, or even to a single novel. Indeed it seems entirely likely that Burckhardt's hermeneutical approach is exactly what we need for the large archives studied with digital approaches. One does not have to necessarily follow Burckhardt in his belief that each textual object is a "unit" or world with knowable rules in order to understand the force of his critique of certain strains of structuralist thinking. What I mean to say is that Burckhardt's conception can revitalize the digital humanities-cum-structuralist reading practices that we find at the present moment.

Returning to Goldstone and Underwood, we might want to think about ways in which they invoke the concept of the hermeneutical circle as it relates to digital reading. They invoke a soft hermeneutics when they write "in the end we must always close the hermeneutic circle with human interpretation" (2014, 10). I agree with them that the digital humanities

cannot function without "human interpretation," but I resist the division between interpretation and method that renders method totally free from interpretation. In their version of the circle, the human interpreter comes after; interpretation, according to their logic, proceeds from algorithmic output. They want to extend "human interpretation" over very large archives, over collections of documents and texts that would be too large for a human interpreter. Yet in an important way the circle remains incomplete. When Hans-Georg Gadamer produced a definition of the hermeneutic circle he made the point that the concept "is based on a polarity of familiarity and strangeness" (1989, 262). The poles in Gadamer's circle are less rigidly defined than allowed for by much work in the digital humanities and "strangeness" extends all the way through the project of criticism, not just to an examination of the results.

Yet there are ways to turn the major presupposition of digital reading techniques such as machine learning back against itself. Classification, whether machine or human derived, fits observed data or objects into distinct categories. The important difference between human and machine classification, however, is what draws us to categorize data into categories and our doubt about this categorization. The algorithm assumes that all data will "fit." Within machine learning there are concepts to label the degree to which data fits into categories: we call any potential uncertainty within classification confusion or simply "error." The outlier, that peculiar object not belonging to the domain of one law or another, might present some difficulty to categorize for the algorithm, but it is of high interest to the human interpreter because it represents a problem. Burckhardt draws our attention to the way in which when we are reading we encounter something that he calls a "stumbling block" (1968, 289) that becomes the occasion for analysis:

> What occurs, then, when I really do interpret? Something which in principle is very simple. I read a poem and the poem "speaks to me." At the same time, however, or perhaps only after several readings, I get the impression that I have not yet grasped its true significance. Something "disturbs" me. What it is that will "disturb" me is never predictable. It may be a "discrepancy" (a contradiction, sometimes purely factual, which seems to reside in the poem itself); it may be an apparent whim of the poet or a seemingly inappropriate word; it may be configuration whose meaning is obscure; or it may be (as with Hölderlin's late hymns) that the coherence of the whole completely escapes me. Finally any conception of the poem which contradicts my own may also disturb me in this sense. (Burckhardt 1968, 301)

Burckhardt's "stumbling block" functions much like the effect of the *punctum* in Roland Barthes's account of the photographic image. For

Barthes, what disturbs the interpreter is the appearance of a "sting, speck, cut, little hole" within the field of the image, what he calls the *studium* (1981, 26–27). The *punctum* produces the occasion for interpretation: "the photographer's *punctum* is that accident which pricks me (but also bruises me, is poignant to me)" (27). Burckhardt faults the dominant contemporary theory of his time, structuralism, for not paying enough attention to those objects that do not fit within preexisting strategies. This critical science has pushed such difficult to categorize elements aside in favor of generalizations and secure categorizations.

To return to my titular question: can an algorithm be disturbed? In the case of computer-aided text mining and machine-learning algorithms, the ever-present risk is that they cannot. Digital readings resist and reduce disturbance—it is only when they fail to be properly iterative that they might be said to be "disturbed." Algorithms, of all kinds, are recipes for success. They are a description, an ordering of operations, which can be iteratively executed to produce a "correct" result.[12] Failure, as opposed to algorithmic success, might be the special providence of humanists. It is in another essay found in *Beyond Formalism* that Geoffrey Hartmann describes interpretation as requiring either the location of a space in between the text or the opening of that space by the critic: "Interpretation is like a football game. You spot a 'hole' and you go through. But first you may have to induce that opening. The Rabbis used the technical word *patach*, 'he opened,' for interpretation" (Hartman 1971, 255). For Hartmann, literature is special because it has the capacity to sustain the hole. Interpretation exists within a space that might be thought of as in between the "bits" of language. When we allow our algorithms to overly familiarize that which is fundamentally ambiguous, we risk turning our work, the project of literary criticism, into what Burckhardt would call explanation. This activity of explanation risks too quickly closing down the disturbing possibility of texts. In privileging explanation over interpretation digital humanists might be tempted to exploit the cunning of empiricism to ideologically suppress interpretive moves and in the process marginalize a certain kind of questioning of the critical "pretext." Perhaps we can use machine learning and other computer-aided reading techniques to open holes by deploying the algorithm against itself, but ultimately interpretation is an interesting and compelling narrative of how one deals with being "pricked" by a text, by being disturbed.

Throughout this essay I have argued that the present movement in criticism that seeks to reposition literary studies as a social science is resurrecting the project of structuralism. The computer-aided text mining practice of the digital humanities provides us with an important

case study through which we can examine the stakes of this swerve away from much of the contemporary critical discourse. This discourse remains, as Rita Felski shows us, quite suspicious (2009, 28–30). But these suspicious interpretive practices have enabled a whole range of important political projects that have made what was once invisible visible and have moved what was once on the margins to the center. An entire generation of critics disturbed by absences and tightly constricted categories that reinforced ideological thinking about difference has rightfully questioned the self-assurance with which prior critics deployed what they conceived of as politically neutral methodologies. A criticism that, once again, seeks to authorize itself through an appeal to the social sciences cannot ignore the insights of these political projects nor can it so easily push aside the deconstructive critique of the first literary science, structuralism.

NOTES

This essay was greatly improved through the critiques and contributions of my anonymous readers. Thanks also to Graham MacPhee for several clarifying comments and questions. I brought an early draft of this project to my seminar at the Futures of American Studies Institute in 2014 and would like to thank three participants: Kristie Schlauraff, Dan Sinykin, and especially Moacir P. de Sá Pereira. I must also acknowledge Donald E. Pease for his invaluable suggestions and support. Special thanks to Louis A. Renza.

[1] As I work within American literary studies, many of my references will be the local application of what I describe as larger movements within the humanities. On description as method, see Heather Love, "Close But Not Deep: Literary Ethics and the Descriptive Turn," (2010). An example of the renewed interest in literary aesthetics can be found in Looby and Weinstein 2012, 1–19. For an example of the new formalism, see Otter 2008, 116–25.

[2] Rita Felski has made this argument in several locations. She offers some suggestions of what might come after suspicion, after critique: "Critique needs to be supplemented by generosity, pessimism by hope, negative aesthetics by a sustained reckoning with the communicative, expressive, and world-disclosing aspects of art" (2009, 33).

[3] See the documentation provided at www.provalisresearch.com/products/content-analysis-software/wordstat-dictionary/sentiment-dictionaries.

[4] Martindale 1975. The digitized version of Martindale's dictionary can be found at: www.provalisresearch.com/products/content-analysis-software/wordstat-dictionary/regressive-imagery-dictionary-by-colin-martindale-free.

[5] The problem of discovering and sharing data provenance has become a pressing issue for many computational fields. The following has been offered as a vision for an "ideal world": "users would be able to reproduce their results by replaying previous computations, understand why two seemingly identical runs with the

same inputs produce different results, and find out which data sets, algorithms, or services were involved in the derivation of their results" (Moreau et al. 2008, 54).

[6] Key references on this aspect of the digital humanities include Ramsay 2011, 7–9; Jockers 2013, 26–30.

[7] Both examples are from running Henry Adams's *The Education of Henry Adams* through the CountVectorizer topic-modeling algorithm provided with the Python-based package called "sci-kit learn": www.scikit-learn.org. The complete ASCII text of *The Education of Henry Adams* used for these examples was produced by Richard Fane and distributed by Project Gutenberg, and is available at: www.gutenberg.org/cache/epub/2044/pg2044.txt

[8] One of the most popular stemming algorithms is the Porter Stemming algorithm. This is incorporated within the workflows as a preprocessing step by many packages including Provalis Research's WordStat. See Porter 1980, 130–37.

[9] Another Derridean concept, the "*parergon*" or frame that we take to exist outside the space of the work of art, might be useful in understanding the stakes involved in suggesting the existence of non-signifying elements of a text (Derrida 1987, 97–98).

[10] For a general theoretical and historical background on structuralism, its main currents of thought, and adoption within the American academy, see Scholes 1974. Scholes describes the assumption of an *a priori* order of the world: "The perception of order or structure where only undifferentiated phenomena had seemed to exist before is the distinguishing characteristic of structuralist thought" (Scholes 1974, 41). See also, Culler 1975, 37–63.

[11] Geoffrey Hartman destabilizes Frye's spatialization of literature by introducing the problem of temporality to Frye's understanding of literature as unfolding "quasi-simultaneously in space" (Hartman 1971, 32–33). The classic deconstructive critique of structuralism based on a "decentering" of the structure is Derrida's "Structure, Sign, and Play in the Discourse of the Human Sciences" (Derrida 1978, 278–93).

[12] In this essay I am addressing the use of a subset of computer algorithms, machine-learning algorithms. These algorithms, by necessity, fit all supplied data into a set of categories. To my larger point about algorithms, I have invoked Burkhardt's sense of the hermeneutical "stumbling block" that appears during reading to call into question our reliance on the separation between methodology—here we should understanding reading, both computer and close as the methodology—and interpretation, and the way in which the algorithmic thinking denies the possibility of being disturbed and the situated or idiosyncratic reading. In his introductory text to algorithms, Thomas Cormen defines an algorithm as "a set of steps to accomplish a task that is described precisely enough that a computer can run it." He continues to refine this concept through the addition of iterability: "Computer algorithms solve computational problems. We want two things from a computer algorithm: given an input to

a problem, it should always produce a correct solution to the problem, and it should use computational resources efficiently while doing so" (2013, 1–2). See also the explanation of error handling algorithms in Cormen's co-authored *Introduction to Algorithms*: "An algorithm is said to be *correct* if, for every input instance, it halts with the correct output. We say that a correct algorithm *solves* the given computational problem. An incorrect algorithm might not halt at all on some input instances, or it might halt with an incorrect answer. Contrary to what you might expect, incorrect algorithms can sometimes be useful, if we can control their error rate" (Cormen, Leiserson, Rivest, and Stein 2009, 6).

WORKS CITED

Barthes, Roland. 1981. *Camera Lucida: Reflections on Photography.* Translated by Richard Howard. New York: Hill and Wang.

———. 1989. *The Rustle of Language.* Translated by Richard Howard. Berkeley: University of California Press.

Best, Stephen, and Sharon Marcus. 2009. "Surface Reading: An Introduction." *Representations* 108: 1–21.

Burckhardt, Sigurd. 1968. "Notes on the Theory of Intrinsic Interpretation." In *Shakespearean Meanings.* Princeton, NJ: Princeton University Press.

Cormen, Thomas H. 2013. *Algorithms Unlocked.* Cambridge, MA: MIT Press.

Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, Clifford Stein. 2009. *Introduction to Algorithms.* Cambridge, MA: MIT Press.

Culler, Jonathan. 1975. *Structuralist Poetics.* Ithaca, NY: Cornell University Press.

Derrida, Jacques. 1978. *Writing and Difference.* Translated by Alan Bass. Chicago: University of Chicago Press.

———. 1981. *Dissemination.* Translated by Barbara Johnson. Chicago: University of Chicago Press.

———. 1987. *The Truth in Painting.* Translated by Geoffrey Bennington and Ian McLeod. Chicago: University of Chicago Press.

Frye, Northrop. 1971. *Anatomy of Criticism: Four Essays.* Princeton, NJ: Princeton University Press.

Felski, Rita. 2009. "After Suspicion," *Profession* 35: 28–35.

Gadamer, Hans-Georg. 1989. *Truth and Method.* Translated by Joel Weinsheimer and Donald Marshall. New York: Crossroad.

Gitelman, Lisa, and Virginia Jackson. 2012. *"Raw Data" Is an Oxymoron.* Cambridge, MA: MIT Press.

Goldstone, Andrew, and Ted Underwood. 2014. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Teach Us." *New Literary History* 45.3: 359–84.

Hartman, Geoffrey. 1971. *Beyond Formalism: Literary Essays, 1958–1970.* New Haven, CT: Yale University Press.

Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History.* Urbana: University of Illinois Press.

Latour, Bruno. 2004. "Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry* 30.2: 225–48.
Liu, Alan. 2013. "The Meaning of the Digital Humanities." *PMLA* 128.2: 409–23.
Looby, Christopher, and Cindy Weinstein. 2012. "Introduction." *American Literature's Aesthetic Dimensions*. New York: Columbia University Press.
Love, Heather. 2010. "Close But Not Deep: Literary Ethics and the Descriptive Turn," *New Literary History* 41.2: 371–91.
Lui, Bing. (2010). "Sentiment Analysis and Subjectivity." In *Handbook of Natural Language Processing*, edited by N. Indurkhya and F. J. Damerau, 627–66. Boca Raton, FL: CRC Press.
Martindale, Colin. 1975. *Romantic Progression: The Psychology of Literary History*. Washington, DC: Hemisphere.
McCallum, Andrew Kachites. 2002. "MALLET Homepage." *MALLET: Machine Learning for Language Toolkit*. www.mallet.cs.umass.edu.
Moreau, Luc, Paul Groth, Simon Miles, Javier Vazquez-Salceda, John Ibbotson, Sheng Jiang, Steve Munroe, Omer Rana, Andreas Schreiber, Victor Tan, and Laszlo Varga. 2008. "The Provenance of Electronic Data." *Communications of the ACM* 51.4: 52–58.
Moretti, Franco. 2000. "Conjectures on World Literature," *New Left Review* 1: 54–68.
———. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary Theory*. New York: Verso.
Otter, Samuel. 2008. "Aesthetics In All Things." *Representations* 104.1: 116–25.
Porter, M. F. 1980. "An Algorithm for Suffix Stripping." *Program* 14.3: 130-37.
Ramsay, Stephen. 2011. *Reading Machines: Toward an Algorithmic Criticism*. Urbana: University of Illinois Press.
Sedgwick, Eve Kosofsky. 2003. *Touching Feeling: Affect, Pedagogy, Performativity*. Durham, NC: Duke University Press.
Scholes, Robert. 1974. *Structuralism In Literature: An Introduction*. New Haven, CT: Yale University Press.

JAMES E. DOBSON is Lecturer in English at Dartmouth College. He has published essays on Mark Twain, Lucy Larcom, and Ambrose Bierce. He is presently working on two book-length projects: a critical account of the digital humanities and scientism in the humanities and a study of *fin-de-siècle* American autobiography titled *The Awkward Age of Autobiography*.