

# JOURNAL OF LAW AND POLICY

---

VOLUME XXI

2013

No. 2

---

## CONTENTS

*Symposium*

### **AUTHORSHIP ATTRIBUTION WORKSHOP**

*Preface by Lawrence M. Solan, Ph.D., J.D.*..... 283

### **STYLOMETRY AND IMMIGRATION: A CASE STUDY**

*Patrick Juola, Ph.D.*..... 287

### **A SYSTEMIC FUNCTIONAL APPROACH TO AUTOMATED AUTHORSHIP ANALYSIS**

*Shlomo Argamon, Ph.D. and Moshe Koppel, Ph.D.* ..... 299

### **AUTHORSHIP ATTRIBUTION: WHAT'S EASY AND WHAT'S HARD?**

*Moshe Koppel, Ph.D., Jonathan Schler, Ph.D., and Shlomo Argamon, Ph.D.* ..... 317

### **BEST PRACTICES AND ADMISSIBILITY OF FORENSIC AUTHOR IDENTIFICATION**

*Carole Chaski, Ph.D.* ..... 333

### **ATTRIBUTION OF MUTUAL UNDERSTANDING**

*Carl Vogel, Ph.D.*..... 377

### **ON THE ROBUSTNESS OF AUTHORSHIP ATTRIBUTION BASED ON CHARACTER N-GRAM FEATURES**

*Efstathios Stamatatos, Ph.D.*..... 421

### **ON ADMISSIBLE LINGUISTIC EVIDENCE**

*Malcolm Coulthard, Ph.D.*..... 441

### **TXT 4N6: METHOD, CONSISTENCY, AND DISTINCTIVENESS IN THE ANALYSIS OF SMS TEXT MESSAGES**

*Tim Grant, Ph.D.* ..... 467

*JOURNAL OF LAW AND POLICY*

<b>TOWARDS AN INDEX OF IDIOLECTAL SIMILITUDE (OR DISTANCE) IN FORENSIC AUTHORSHIP ANALYSIS</b> <i>M. Teresa Turell, Ph.D. and Núria Gavaldà</i> .....	495
<b>LINGUISTIC CONFUSION IN COURT: EVIDENCE FROM THE FORENSIC SCIENCES</b> <i>Jonathan J. Koehler, Ph.D.</i> .....	515
<b>BEING PRAGMATIC ABOUT FORENSIC LINGUISTICS</b> <i>Edward K. Cheng, J.D.</i> .....	541
<b>INTUITION VERSUS ALGORITHM: THE CASE OF FORENSIC AUTHORSHIP ATTRIBUTION</b> <i>Lawrence M. Solan, Ph.D., J.D.</i> .....	551
 <i>Notes and Comments</i>	
<b>CODIFYING COMMON LAW: THE SELF-CRITICAL ANALYSIS PRIVILEGE AND THE NEW JERSEY PATIENT SAFETY ACT</b> <i>Adam Blander</i> .....	577
<b>IMPLIED PREEMPTION AND ITS EFFECT ON LOCAL HYDROFRACKING BANS IN NEW YORK</b> <i>David Giller</i> .....	631
<b>IS LITIGATION YOUR FINAL ANSWER? WHY THE HEALTHY WORKPLACE BILL SHOULD INCLUDE AN ADR PROVISION</b> <i>Florence Z. Mao</i> .....	679
<b>“EYE SEE YOU”: HOW CRIMINAL DEFENDANTS HAVE UTILIZED THE NERD DEFENSE TO INFLUENCE JURORS’ PERCEPTIONS</b> <i>Sarah Merry</i> .....	725

# AUTHORSHIP ATTRIBUTION WORKSHOP

## PREFACE

*Lawrence M. Solan\**

In October 2012, Brooklyn Law School's Center for the Study of Law, Language and Cognition held a two-day workshop on Forensic Authorship Attribution. This volume contains its proceedings. The workshop, sponsored by the National Science Foundation, to which Brooklyn Law School expresses its gratitude,<sup>1</sup> brought together leading scholars from around the world who approach the question of authorship attribution from disparate, and seemingly incompatible, perspectives. Represented among the articles that appear in this volume are works based on the algorithms of computer scientists and computational linguists (Argamon & Kopell; Chaski; Juola; Koppel, Schler & Argamon; Stamatatos; and Vogel), works by linguists who evaluate and compare stylistic regularity and nuance, often on a case-by-case basis (Coulthard; Grant), and research comparing the extent to which linguistic nuance results from regional differences between one speech community and another, and the extent to which it is based within the individual (Turell and Gavalda). Other prominent linguists—Ronald Butters and Edward Finegan—moderated workshop sessions.

Despite these differences in approach, what emerged from the workshop and is reflected in the published articles is a recognition that those who work algorithmically can improve their models by incorporating into them some of the insights of those who work with stylistic markers, and those whose work is less computational can develop quantitative techniques to

---

\* Don Forchelli Professor of Law and Director, Center for the Study of Law, Language and Cognition, Brooklyn Law School.

<sup>1</sup> NSF Award SES-1160828.

improve the reliability of their conclusions. In my own essay in this volume, I explore the advantages and disadvantages of the various techniques employed in this volume, commenting on how the legal system tends to react to them. While it might be too ambitious to predict that the various methods will converge, there is already reason for optimism that the insights of the various approaches will influence those of the others, creating a field with a healthy combination of cooperation and competition. Thus, in this volume, we see the expanded use of linguistic features by the computational researchers, while at the same time we see efforts by stylistic researchers to introduce statistical modeling into stylistic authorship analysis.

A quick glance at the table of contents shows a lot of contributors with Ph.D.s and very few with law degrees, an unusual array for a law journal publication. Nonetheless, the legal academic community is very much present in this volume. In addition to my own essay, comments by two prominent law professors who specialize in scientific evidence—Edward Cheng and Jonathan Koehler—consider the legal community’s likely response to the advances in authorship attribution described by the linguists.<sup>2</sup> Moreover, participating in the workshop were two statistical “consultants” (Stephen Fienberg of Carnegie Mellon University and Robert Carpenter of Columbia University).

This interaction between the scientific community and the evidence scholars was one of the workshop’s main goals. Forensic identification sciences have been under severe attack as inadequately grounded in science over the past decade—largely for good reason. Just as the scientists had a lot to learn from the reactions of the legal scholars to their work, we believed that the legal scholars could benefit from seeing in action a relatively young forensic science that takes itself seriously as science. While the evidence literature decries the absence of concern over the rate of error in one forensic science after another, an annual workshop on authorship and plagiarism identification actually requires that algorithms presented be subjected to a

---

<sup>2</sup> Also present at the workshop were D. Michael Risinger and Michael Saks, two additional prominent scientific evidence scholars.

proficiency test before a paper is accepted for presentation at the conference.<sup>3</sup>

At the same time, though, some important insights that may assist a trier of fact in a case concerning authorship, such as multiple possible authors making the same kind of grammatical error in known writings (as illustrated by Malcolm Coulthard's piece in this volume), are not yet subject to mathematical analysis, creating a dilemma for the legal system. The legal commentators observed and comment on both the promising progress of the field and the challenges it still faces.

I end on a very sad note. One of the authors, María Teresa Turell, passed away on April 24, 2013, just before this volume went to press. Maite, as everyone knew her, was a Professor of English at Universitat Pompeu Fabra in Barcelona, where she directed the Forensic Linguistics Laboratory. She was an important figure in the field, devoted to bringing quantitative rigor to stylistic insight, one of the themes of the workshop. More importantly to those of us who knew Maite, her intellectual toughness was matched with a loving and generous character that will remain with us for a long, long time to come.

---

<sup>3</sup> The conference is known as the "PAN/CLEF" conference. *See* PAN WORKSHOP & COMPETITION, <http://pan.webis.de/> (last visited May 3, 2013). As of the date of this writing (May 2, 2013), 107 teams from forty-two countries have registered for the next PAN/CLEF conference, which will be held in Valencia, Spain in September 2013.

# STYLOMETRY AND IMMIGRATION: A CASE STUDY

*Patrick Juola\**

## INTRODUCTION

This paper describes “authorship attribution” as the process of inferring authorial identity from writing style and presents several classic studies as examples. This paper further explores a case of attribution “in the wild,” so to speak, where there are a number of additional constraints and challenges. These challenges, fortunately, are not insurmountable. The background of the case, an asylum case in immigration court; responses to the challenges of the case; and the results of the analysis are discussed.

## I. BACKGROUND

### *A. Stylometry and Authorship Attribution*

Standard practice for stylometric investigations involves a detailed comparison of stylistic features culled from a training set of documents.<sup>1</sup> The questioned document is then compared

---

\* Juola & Associates, pjuola@juolaassociates.com. This material is based upon work supported by the National Science Foundation under Grant No. OCI-1032683. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

<sup>1</sup> See, e.g., Patrick Juola, *Authorship Attribution*, 1 FOUND. & TRENDS INFO. RETRIEVAL 233 (2006); Moshe Koppel & Jonathan Schler, *Computational Methods in Authorship Attribution*, 60 J. AM. SOC’Y INFO. SCI. & TECH. 9 (2009); Mathew L. Lockers & Daniel M. Witten, *A Comparative Study of Machine Learning Methods for Authorship Attribution*,

against the training set, typically using some form of classification or machine learning algorithm. Finally, an appropriate decision is reached in line with the experimental results.

A classic example of this form is the Mosteller-Wallace study of the *Federalist* papers,<sup>2</sup> a collection of eighteenth-century political documents describing and arguing for the (newly proposed) Constitution of the United States. These documents were originally published pseudonymously under the name Publius, but are now known (via traditional historical methods) to have been written by Alexander Hamilton, James Madison, and John Jay. Historians have come to consensus about the authorship of each of the eighty-five essays in the collection.

Mosteller and Wallace investigated the authorship question through the frequencies of individual words such as prepositions.<sup>3</sup> Careful analysis of known works by Hamilton and Madison, for example, show that they vary in the use of the word “by.” For instance, Hamilton tended to use it about seven times per thousand words, rarely more often than eleven times per thousand, and never (in the samples studied) more than thirteen times per thousand words.<sup>4</sup> Madison, by contrast, used the word “by” most often in the range of eleven to thirteen times per thousand words, never less than five per thousand, and as much as nineteen per thousand.<sup>5</sup> Similar studies show that Hamilton used the word “to” more often than Madison, that Madison almost never used the word “upon,” and so forth.<sup>6</sup>

We can therefore infer that a thousand-word document with seventeen tokens of “by” is more likely to be from Madison’s pen than Hamilton’s. If this document also contains relatively few “to’s” and “upon’s,” our inference is strengthened. The

---

25 LITERARY & LINGUISTIC COMPUTING 215 (2010); Efstathios Stamatatos, *A Survey of Modern Authorship Attribution Methods*, 60 J. AM. SOC’Y INFO. SCI. & TECH. 538 (2009).

<sup>2</sup> See generally FREDERICK MOSTELLER & DAVID L. WALLACE, *INFERENCE AND DISPUTED AUTHORSHIP: THE FEDERALIST* (1964).

<sup>3</sup> *Id.* at 29 tbl.2.3–3.

<sup>4</sup> *Id.* at 17 tbl.2.1–1.

<sup>5</sup> *Id.*

<sup>6</sup> *Id.*

notion of “more likely,” with respect to identifying authorship, can be formalized using statistics (particularly Bayes’ theorem)<sup>7</sup> to yield a precise odds ratio. With enough data, the odds ratio can achieve practical certainty. For example, Madison is millions of times more likely to have written Federalist Paper 51 than Hamilton.<sup>8</sup>

A similar example is the study by Binongo of the fifteenth *Oz* book, *The Royal Book of Oz*.<sup>9</sup> The original *Wonderful Wizard of Oz* was of course written by L. Frank Baum, as were the second through fourteenth books in that series. When Baum died, the publisher found another writer, Ruth Plumly Thompson, to serve as Baum’s successor, working from “notes and a fragmentary draft”<sup>10</sup> for the fifteenth book and then writing eighteen more original *Oz* books. The question is whether a substantial “draft” of the fifteenth book ever existed, or whether the *Royal Book* was also largely Thompson’s work.

Similarly to the Mosteller-Wallace study, Binongo chose to study lexical items, analyzing the relative frequency of the fifty most common words in the combined *Oz* series, a set containing words like “the,” “and,” “with,” “into,” and so forth.<sup>11</sup> Using a dimensionality reduction technique called Principal Component Analysis (“PCA”), he combined the variation among these fifty words down to two dimensions and plotted each work on a two-dimensional graph.<sup>12</sup> The results were clear and compelling; there were distinct clouds representing Baum’s and Thompson’s respective work, with a notable separation between them (in Binongo’s words, a “stylistic gulf”).<sup>13</sup> The *Royal Book* fell squarely on Thompson’s side of the fence, “reveal[ing] that the

---

<sup>7</sup> Here and elsewhere, we omit the detailed mathematical description for clarity and brevity.

<sup>8</sup> *Id.* at 211 tbl.5.5–2, 263.

<sup>9</sup> José Nilo G. Binongo, *Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution*, 16 CHANCE, no. 2, 2003 at 9.

<sup>10</sup> RAYLYN MOORE, WONDERFUL WIZARD, MARVELOUS LAND 89 (1974).

<sup>11</sup> Binongo, *supra* note 9, at 11–12.

<sup>12</sup> *Id.* at 12.

<sup>13</sup> *Id.* at 15.



writing style in the 15th Book of *Oz* is more compatible with Thompson's than Baum's."<sup>14</sup>

There are notable differences between these studies. Mosteller and Wallace studied a variety of possible features before settling on a hand-picked set of thirty words (including some rather rare words such as "direction") chosen for their discriminative abilities in this specific study.<sup>15</sup> Binongo, on the other hand, simply used the fifty most common words in the corpus.<sup>16</sup> In this volume, Stamatatos argues for the use not of words but of character sequences;<sup>17</sup> we have argued elsewhere for the use both of character sequences and word sequences.<sup>18</sup> Mosteller and Wallace used a form of Bayesian statistical analysis,<sup>19</sup> Binongo used PCA,<sup>20</sup> Stamatatos uses a third technique called "support vector machines,"<sup>21</sup> and we have argued elsewhere for similarity-based nearest neighbor methods.<sup>22</sup>

More striking than the differences, however, are the similarities in both the Mosteller-Wallace and Binongo studies:

- the set of candidate authors was limited to only a small and clearly defined group of people;
- all candidate authors had an extensive body of unquestioned work to compare;

---

<sup>14</sup> *Id.* at 16.

<sup>15</sup> MOSTELLER & WALLACE, *supra* note 2, at 67–68.

<sup>16</sup> Binongo, *supra* note 9, at 11–12.

<sup>17</sup> See generally Efsathios Stamatatos, *On the Robustness of Authorship Attribution Based on Character N-Gram Features*, 21 J.L. & POL'Y 421 (2013).

<sup>18</sup> See generally Patrick Juola & Darren Vescovi, *Analyzing Stylometric Approaches to Author Obfuscation*, in ADVANCES IN DIGITAL FORENSICS VII, at 115, 115–25 (Gilbert Peterson & Sujeet Shenoj eds., 2011).

<sup>19</sup> See generally MOSTELLER & WALLACE, *supra* note 2.

<sup>20</sup> Binongo, *supra* note 9, at 12–17.

<sup>21</sup> Stamatatos, *supra* note 17, at 431.

<sup>22</sup> John Noecker, Jr. & Patrick Juola, *Cosine Distance Nearest-Neighbor Classification for Authorship Attribution*, PROC. DIGITAL HUMAN., 2009, at 208.

- this body of work was huge (in the *Oz* study, more than a dozen novels each), large enough to provide statistical confidence; and
- the body of work was similar to the disputed document in style, topic, and genre, and thus provided a representative sample.<sup>23</sup> This is key because many of the factors that separate individuals also vary systematically between types of writing. Passive writing is very common in technical prose, for example, but uncommon in conversation or narrative.<sup>24</sup>

One might suspect that the choice of topics and works to study was in part driven by these considerations. Unfortunately, many cases of practical interest (especially in the court system) do not have these attributes, as will be seen in Part II.

### B. JGAAP

In light of the differences among possible analyses, an obvious question is “which method works best?” To address this question, the Evaluating Variations in Language Laboratory at Duquesne University has developed a modular system for the development and comparative testing of authorship attribution methods.<sup>25</sup> This system, Java Graphical Authorship Attribution Program (“JGAAP”), provides a large number of interchangeable analysis modules to handle different aspects of the analysis pipeline such as document preprocessing, feature selection, and analysis/visualization. Taking combinatorics into account, the number of different ways to analyze a set of documents ranges in the millions and can be expanded by the inventive user with a moderate knowledge of computer programming.

---

<sup>23</sup> MOSTELLER & WALLACE, *supra* note 2, at 2–3; Binongo, *supra* note 9, at 9–10.

<sup>24</sup> DOUGLAS BIBER, *VARIATION ACROSS SPEECH AND LANGUAGE* 50 (1988).

<sup>25</sup> Juola, *supra* note 1; Patrick Juola et al., *JGAAP 4.0—A Revised Authorship Attribution Tool*, *PROC. DIGITAL HUMAN.*, 2009, at 357.

## II. A CASE STUDY

To illustrate the issues and complications that can arise in “the real world,” we present the following as a case study in the application of authorship attribution in actual forensic practice. All identifying details have been changed to protect the privacy (and possible physical well-being) of the individuals involved.

### *A. Statement of the Case*

Bilbo Baggins, a native of Mordor, was facing immigration procedures that might have led to his removal from the United States. He claimed in immigration court that deportation was inappropriate and sought asylum because he was a noted and published activist against the Mordor government and he feared negative consequences if forcibly repatriated. As evidence for this claim, he offered a number of articles he had written for an Elvish-language newspaper, as well as a set of newer (antigovernment) articles he claimed to have written but that had been published anonymously while outside Mordor. Juola & Associates was asked by Baggins’ counsel to analyze these articles. The basic theory of the case was that if Baggins had, in fact, written the newer articles (the older articles were unquestioned, as they had been published under his name), and if that fact could be demonstrated, that would establish that his fears were well founded.

Superficially, this appears to be an ordinary questioned-documents case, but there are a few twists. We started by rejecting “traditional” document forensics, handwriting analysis and such, as there are no original documents to study. All documents had been submitted to newspapers and subjected to editorial review and publication; the older documents were in the form of photocopies of printed clippings, while the new documents were born-digital web pages that had no originals. All that was available was the content of the documents, suggesting a need for authorship analysis as defined above.

At the same time, there was no clearly defined set of candidate authors; either Baggins wrote the questioned documents or “someone else” did, and all we know about this

“someone else” is that they had access to the Internet. Additionally, the set of documents available was rather small: a dozen newspaper articles each in the known and questioned sets. The documents were also in Elvish, an understudied language with little computational support available.

The last point is probably the least important, as JGAAP provides a relatively language-agnostic method of analysis. Certainly, the idea of “fifty most common words” is computationally tractable in any language with a clear notion of a word (such as a language like English, German, Russian, or Spanish where spaces separate words). Furthermore, previous research has shown that there is a high cross-linguistic correlation in performance of authorship attribution methods or, in other words, that in the absence of compelling counterinformation, methods that are known to perform well in English are likely to perform well in other unstudied languages.<sup>26</sup> But structuring the problem as a verification instead of classification problem forced us to use a somewhat nonstandard approach. In a typical classification problem, there are a number of possible answers, one “correct” answer and a number of “distractor” answers. (In an authorship context, Marlowe and Kyd could be distractors for a play we believe to be written by Shakespeare; in the context of criminal investigation, all of the suspects except for the actual guilty party are *de facto* distractors.) By contrast, in a verification problem, we have only one “suspect” but need to evaluate whether the evidence is sufficient to tie him to the acts in question.

### *B. Materials and Methods*

Baggins himself supplied us with ten copies of newspaper articles published under his name approximately ten years before the date of the case; these articles comprised a set of known documents. These documents (photocopies of clippings) were hand-transcribed by Elvish-speaking typists into a machine-

---

<sup>26</sup> Patrick Juola, *Cross-Linguistic Transference of Authorship Attribution, or Why English-Only Prototypes Are Acceptable*, PROC. DIGITAL HUMAN., 2009, at 162.

readable corpus. In addition, he supplied us with eleven web page images from a recent news site, published anonymously, as the set of questioned documents.<sup>27</sup>

The JGAAP software package provided the necessary technology for this text analysis. All relevant files were preprocessed to convert them into plain text (Unicode) format. All case distinctions were neutralized, and all whitespace (interword spacing, line breaks, paragraphing, etc.) was normalized to avoid any spurious findings of dissimilarity caused by simple formatting and editing issues. (Again, JGAAP has a button for this kind of preprocessing, and in fact no manual processing was required at all for this analysis.) All documents were converted into word trigrams (phrases of three adjacent words, as in the English phrase “in the English”), a unit of processing known to give good results in authorship queries.<sup>28</sup>

To establish with reasonable certainty that Baggins had or had not written the document, it was necessary for us to create our own distractor set, which we did by gathering a collection of Elvish-language newspaper articles on political issues from another online newspaper. This corpus consisted of 160 news articles by five different named authors, none of whom were Baggins. This provided us with five separate comparison “baseline document corpora” each containing at least thirty articles known to be authored by a distractor author.

The word trigram distributions of the ten documents in the known document set were averaged to produce a central or typical example of Baggins’ writings. Each individual document in the questioned corpus as well as the five baseline corpora was individually compared against this “typical” Baggins style to determine a stylistic distance—a numerical measure of stylistic similarity. Two identical documents would be at distance zero, and, in general, the smaller the distance (the “closer” the document pair), the more likely two documents were to share

---

<sup>27</sup> Of these eleven documents, one was in English and unsuitable for study, so the actual questioned documents comprised ten web pages from which text was extracted. No typists were needed to extract text from these pages as they were in standard HTML; JGAAP will in fact do that automatically.

<sup>28</sup> See Juola, *supra* note 1, at 265–66.

authorship. These distances were averaged to produce a per-author average distance from the known documents.

1. Preliminary Results

The preliminary results can be summarized in Table 1.

Table 1: Preliminary results using cosine distance

Subcorpus	Distance to KD (Known Document Set)
BD-1 (Baseline Document Set 1)	0.9437975
BD-2	0.9517967
BD-3	0.9576155
BD-4	0.9530338
BD-5	0.9534134
QD (Questioned Document Set)	0.8840330

These results provided preliminary evidence in favor of Baggins’s claim; his style is notably closer to that of the questioned documents than it is to other, similar writers. But can we turn this preliminary observation into quantifiable probability judgments? And if so, how compelling are these probabilities? Unfortunately, standard parametric tests (such as *t*-tests) did not help. Interdocument variation (not shown here) dominated the small differences between groups, and the difference in distance was not significant, in a technical sense.

However, there is still an argument to be made here using a non-parametric framework. Assuming that the questioned documents were written by a seventh author outside the set, we have no *a priori* reason to assume that this seventh author would be particularly similar or dissimilar to Baggins. Thus, the probability of this seventh author being the closest to Baggins (as we found in this study) is one in six, approximately 16.7%. Nonparametrically, we can reject this idea (that the documents were written by a seventh author) at the *p*-value of 0.167. This confirms our intuitions that the results support his claim and provide (weak) numerical support, but enough, perhaps, to overcome a “balance of probabilities” burden of proof in a civil case.

## 2. Ensemble Methods and Mixture of Experts

We can, however, (potentially) improve upon these results using ensemble methods.<sup>29</sup> The basic idea is the one behind getting a second opinion: if two (or more) independent experts agree in their analysis, our confidence in that result is increased.<sup>30</sup> This can be formalized using probability theory: if the chance of an expert being right is  $x$ , the chance of her being wrong is therefore  $(1 - x)$ . The chance of two such experts independently being wrong is  $(1 - x)(1 - x)$  or  $(1 - x)^2$ , and in general, the chance of  $k$  experts all being wrong is  $(1 - x)^k$ . For example, if experts in general are right 90% of the time, the chance of one expert being wrong is 0.1 or 10%. The chance of two both being wrong is 0.01 or 1%, and for three experts, 0.001 or 0.1%. In this case, the chance of our analysis being wrong, from above, is 16.7%. If a similar analysis yields the same result, the chance of them both being wrong is a mere 0.167 times 0.167, one chance in thirty-six, or about 2.78%.

We therefore performed these distance comparisons twice, using two different distance formulae and hence two different analyses. The first analysis was performed using normalized dot product or cosine distance,<sup>31</sup> in which the frequency of each individual word trigram is taken into account. The second was done with Jaccard or intersection distance<sup>32</sup> between the sets of word trigrams, which does not take into account frequency but simply measures whether or not a particular author used a particular three-word phrase at any point in the samples.

---

<sup>29</sup> See generally Patrick Juola, Authorship Attribution: What Mixture-of-Experts Says We Don't Yet Know, Conference Presentation at AACL 2008 Am. Ass'n for Corpus Linguistics (Mar. 13, 2008), available at <http://corpus.byu.edu/aac12008/ppt/115.ppt> (discussing various authorship attribution studies).

<sup>30</sup> See *id.*

<sup>31</sup> Noecker & Juola, *supra* note 22.

<sup>32</sup> Tanguy Urvoy et al., *Tracking Web Spam with Hidden Style Similarity*, PROCEEDINGS OF AIRWEB'06 (Aug. 10, 2006), available at <http://airweb.cse.lehigh.edu/2006/urvoy.pdf>.

As hoped, the results of the second experiment (Table 2) confirmed the first:

Table 2: Results using Jaccard/intersection distance

Subcorpus	Distance
BD-1	0.806731
BD-2	0.739381
BD-3	0.852844
BD-4	0.747444
BD-5	0.777530
QD	0.735449

An alert reader will see the card that has just been palmed. Our argument for ensemble methods hinges on an assumption of independence, an assumption that is almost certainly untrue. A document in another language or *a fortiori* another alphabet/writing system will share almost no words or phrases, and hence be strongly different. But within a set of documents of more limited scope—in this case, sharing language, genre, and even general topic—we can argue that a certain amount of independence can be expected. From a purely empirical standpoint, the fact that the baseline distractor authors are ordered differently in the two experiments (e.g., #2 is the closest in Jaccard distance, followed by #4; #1 is first in cosine distance) suggests that these analyses are to a large degree independent. From a theoretical standpoint, Jaccard distance is sensitive only to the distribution of rare features (word trigrams that one author does not use at all), while cosine distance is more sensitive to more common features (as they have greater frequency variance). But in light of the fact that we have no formal measure of the degree of independence, we can, strictly speaking, only say that the chance of this result occurring is no more than 16.7% and could be as small as 2.78%.

### C. Why Stop Here?

JGAAP provides many more than two possible methods. However, we provided no further analysis for this particular case. In theory, we could have used ten methods, and if they all



showed the same result, the odds of a false positive would have been approximately 0.000000165% or one in just over sixty million. However, we would also have run a risk of significantly weakening the case if the analyses did not turn out the way Baggins hoped. The additional costs and risks were, in the opinion of Baggins's counsel, not worth the marginal increase in confidence. This, of course, is a tactical and legal decision based in part on the type of case and the strength of the other evidence available.

#### CONCLUSION

Authorship analysis in the field can pose substantially different challenges than in the lab. The Baggins case presented several unusual aspects in stylistic investigations; the standard stylometric analysis paradigm selects among others rather than giving a simple yes/no answer. Using nonparametric rank order statistics and an ad-hoc set of distractor authors, we could still get an answer and validate it statistically.

Oh, and Bilbo Baggins himself? The judge permitted him to remain in the United States.

# A SYSTEMIC FUNCTIONAL APPROACH TO AUTOMATED AUTHORSHIP ANALYSIS

*Shlomo Argamon\* and Moshe Koppel\*\**

## INTRODUCTION

Attribution of anonymous texts, if not based on factors external to the text (such as paper and ink type or document provenance, as used in forensic document examination), is largely, if not entirely, based on considerations of language *style*. We will consider here the question of how to best deconstruct a text into quantitative features for purposes of stylistic discrimination. Two key considerations inform our analysis. First, such features should support accurate classification by automated methods. Second, and no less importantly, such features should enable a clear explanation of the stylistic difference between stylistic categories (read: authors) and why a disputed text appears more likely to fall into one or another category. The latter consideration is particularly important when a nonexpert, such as a judge or jury, must evaluate the results and reliability of the analysis.

We start from the intuitive notion that style is indicated in a text by those features of the text that indicate the author's choice of one mode of expression from among a set of equivalent modes for a given content. There are many ways in which such choices manifest themselves in a text. Specific words and phrases may be chosen more frequently by certain authors than others, such as the phrase "cool-headed logician" favored by the Unabomber. Some authors may habitually use certain syntactic

---

\* Linguistic Cognition Laboratory, Department of Computer Science, Illinois Institute of Technology, [argamon@iit.edu](mailto:argamon@iit.edu).

\*\* Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel, [moishk@gmail.com](mailto:moishk@gmail.com).

constructions more frequently, as in Hemingway's preference for short, simple clauses. Differences between authors will also arise at the level of the organization of the text as a whole, as some people may prefer to make reasoned arguments from evidence to conclusions, and others may prefer emotional appeals organized differently.

However, all of these "surface" linguistic phenomena have multiple potential underlying causes, not only authorship. They include the genre, register, and purpose of the text as well as the educational background, social status, and personality of the author and audience.<sup>1</sup> What all these dimensions of variation have in common, though, is independence, to a greater or lesser extent, of the "topic" of the text. Hence the traditional focus in computational authorship attribution on features such as function word usage; vocabulary richness and complexity measures; and frequencies of different syntactic structures; which are essentially nonreferential.

Early statistical attribution techniques relied on relatively small numbers of such features, while developments in machine learning and computational linguistics over the last fifteen to twenty years have enabled larger numbers of features to be generated for stylistic analysis. However, in almost no case is there strong theoretical motivation behind the input feature sets, such that the features have clear interpretations in stylistic terms.

We argue, however, that without a firm basis in a linguistic theory of meaning (not just of syntax), we are unlikely to gain any true insight into the nature of any stylistic distinction being studied. Such understanding is key to both establishing and explaining evidence for a proposed attribution. Otherwise, an attribution method is merely a black box that may appear to work for extrinsic or accidental reasons but not actually give reliable results in a given case. Furthermore, an attribution method that produces insight into the relevant language variation is more likely to be useful and accepted in a forensic context, all else being equal, as the judge and jury will be better able to understand the results.

---

<sup>1</sup> DOUGLAS BIBER & SUSAN CONRAD, REGISTER, GENRE, AND STYLE (P. Austin et al. eds., 2009).

We therefore sketch here a computationally tractable formulation of linguistically and stylistically well-motivated features we have developed that permits text classification based on specific variation in choice of nonreferential meanings. The system produces meaningful information about the stylistic distinctions being analyzed, which can be used for interpretative and forensic purposes. We will explain our methodology and then use it as a case study for what any such methodology should provide.

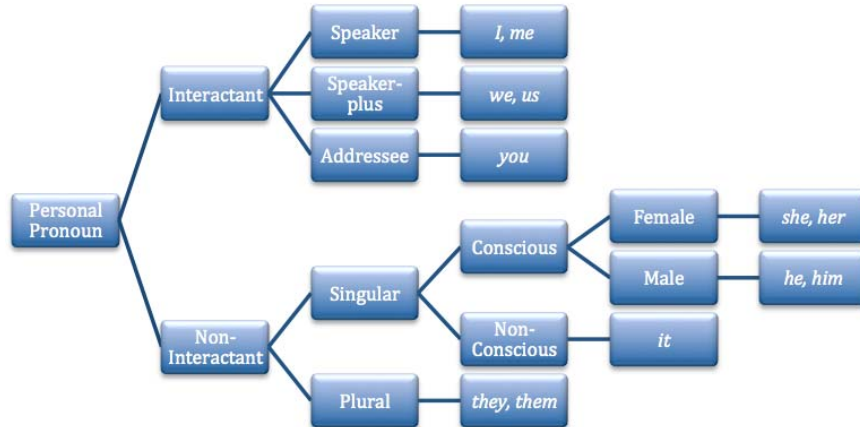
Before we begin, it is worth briefly surveying the variety of problems that fall under the umbrella of “authorship analysis.” The simplest form of the problem is where an anonymous document is potentially attributable to one of a relatively small number (two to fifty, or so) of suspects. The question is then simply which of the suspects has a writing style most like that of the anonymous document. More difficult (and much more likely in the real world) is the case where the document might not be authored by any of the suspects at all—in this case we must be able to determine that the document is not enough like any of the suspects to attribute authorship. The hardest version of this scenario is *authorship verification*, where the question is whether a single suspect did or did not author the anonymous document. All such *authorship attribution* scenarios assume a known set of suspects who are being evaluated for authorship of the questioned document. We require some quantity of texts written by each of the suspects to determine authorship. On the other hand, if, as is often the case in police investigations, specific suspects are not known, we must consider the task of *authorship profiling*, determining as much about the author as possible, based upon clues in the document. As we will discuss below, a number of personal characteristics of an author can be reliably estimated from stylistic cues in a document. But first we will consider generally how we can quantitatively characterize the style of a text for computational analysis.

## I. FUNCTIONAL LEXICAL FEATURES

Our methodology is based on Halliday’s Systemic Functional Grammar<sup>2</sup> (“SFG”), which we find to be particularly well-suited to the sort of computational analysis we seek. SFG explicitly recognizes and represents various aspects of nonreferential meaning as part of the general grammar, which makes it directly adaptable to stylistic classification.<sup>3</sup> We do not claim, of course, that SFG is the only, or even necessarily the best, approach but rather one that we have found convenient.

We start from the SFG idea that grammar is a set of constraints on how one may express meaning.<sup>4</sup> Grammar is thus a network of possible choices, with more general or abstract choices constraining which more specific choices are allowed. This network of choices is called a *system network*.<sup>5</sup> As a simple example, consider the (partial) system network for pronouns in English, seen below in Figure 1. This network forms a neat hierarchical taxonomy, though not all do. As an approximation we can extract a set of taxonomies (trees) from the full network.

Figure 1. System diagram for Personal Pronouns, shown as a taxonomic tree.



<sup>2</sup> See M.A.K. HALLIDAY & CHRISTIAN M.I.M. MATTHIESSEN, AN INTRODUCTION TO FUNCTIONAL GRAMMAR 37–63 (3d ed. 2004).

<sup>3</sup> *Id.* at 50–53.

<sup>4</sup> *Id.* at 1.

<sup>5</sup> *Id.* at 23.

Given this taxonomy, we may define numeric features describing the statistical “stylistics” of a text via the collection of conditional frequencies of each node in the tree given its parent. Thus, for example, we measure the frequency of “Speaker” pronouns out of all occurrences of “Interactant” pronouns, and so on. This has a straightforward interpretation of measuring the biases of how texts of a given style (e.g., by a given author) prefer certain choices of how to express more general meanings. By using such biases to analyze authorship, we seek to capture relevant *codal variation*, as contrasted with register<sup>6</sup> (variation in these probabilities due to a text’s functional context), or *dialect* (variation in how specific meanings are realized (e.g., use of “y’all” for plural “you”)).

To give a flavor of these features, here are brief descriptions of several system networks that we have found useful for stylistic classification.<sup>7</sup>

#### A. Conjunctions

How an author conjoins phrases and clauses is an indication of how the author organizes concepts and relates them to each other. Words and phrases that conjoin clauses (such as “and,” “while,” and “in other words”) are organized in SFG in the CONJUNCTION system network.<sup>8</sup> Types of conjunctions serve to link a clause with its textual context, by denoting how the given clause expands on some aspect of its preceding context. The three top-level options of CONJUNCTION are Elaboration, Extension, and Enhancement, defined as:

- Elaboration: Deepening the content in its context by exemplification or refocusing (“for example,” “in other words,” “i.e.”);

---

<sup>6</sup> Ruqaiya Hasan, *Code, Register, and Social Dialect*, in 2 CLASS, CODES AND CONTROL: APPLIED STUDIES TOWARDS A SOCIOLOGY OF LANGUAGE 224, 253–92 (Basil B. Bernstein ed., 1973).

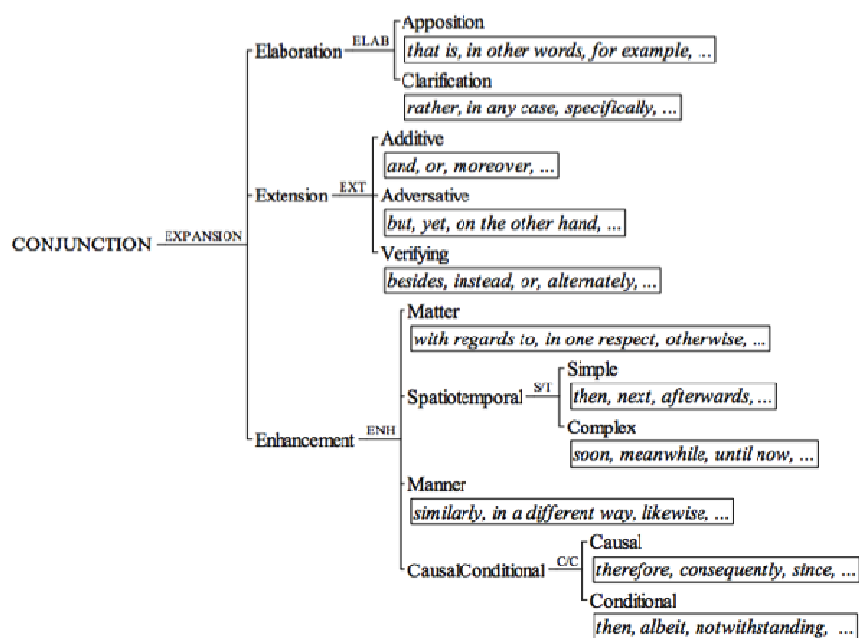
<sup>7</sup> For a more detailed discussion of these features, and the mathematical models involved, see Shlomo Argamon et al., *Stylistic Text Classification Using Functional Lexical Features*, 58 J. AM. SOC’Y INFO. SCI. & TECH. 802, 802–22 (2007).

<sup>8</sup> See HALLIDAY & MATTHIESSEN, *supra* note 2, at 538–39.

- Extension: Adding new related information, perhaps contrasting with the current information (“and,” “or,” “furthermore,” “on the other hand”);
- Enhancement: Qualifying the context by circumstance or logical connection (“and then,” “because,” “similarly”).<sup>9</sup>

Each option also has several subcategories that further subdivide the ways in which information units in a text can be linked together.

Figure 2. System diagram for Conjunction.



### B. Prepositions

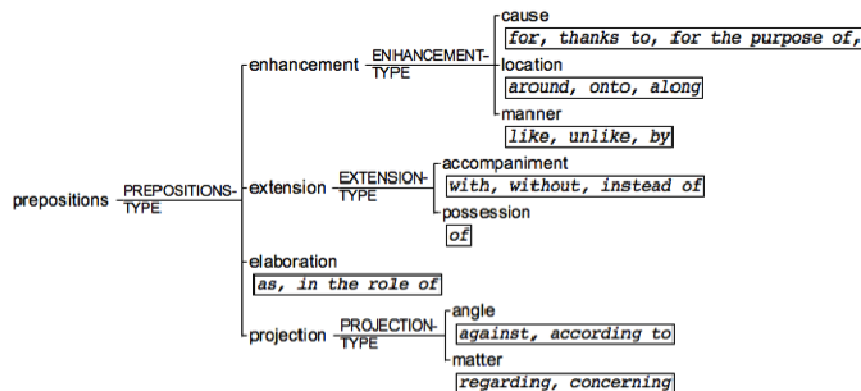
Similarly, prepositions serve to expand the meaning of a phrase or clause by connecting to it a phrase (usually a noun phrase). The high-level structure of the PREPOSITION system is thus similar to that of CONJUNCTION, with four top-level options:

- Elaboration: Exemplification (“as,” “in the role of”);

<sup>9</sup> *Id.* at 540–48.

- Enhancement: Qualifying context temporally, spatially, or causally (“around,” “thanks to,” “during”);
- Extension: Adding related information about an object or event (“of,” “without,” “besides”);
- Projection: Using an object to construe the meaning or significance of another (“against,” “regarding,” “according to”).<sup>10</sup>

Figure 3. System Diagram for Prepositions.



### C. Modality

The MODALITY system comprises four taxonomies describing choices in how to describe the level of typicality or necessity of facts and events. Syntactically, modality can be realized through modal verbs (e.g., “can,” “might,” “should,” “must”); adverbial adjuncts (e.g., “probably,” “preferably”); or projective clauses (e.g., “I think that,” “It is necessary that”). The four attributes of any modal expression are:

- Type: What kind of modality is being expressed?
  - Modalization: How “typical” is it? (“probably,” “seldom”)

<sup>10</sup> See CHRISTIAN M.I.M. MATTHIESSEN, LEXICO-GRAMMATICAL CARTOGRAPHY: ENGLISH SYSTEMS (1995).

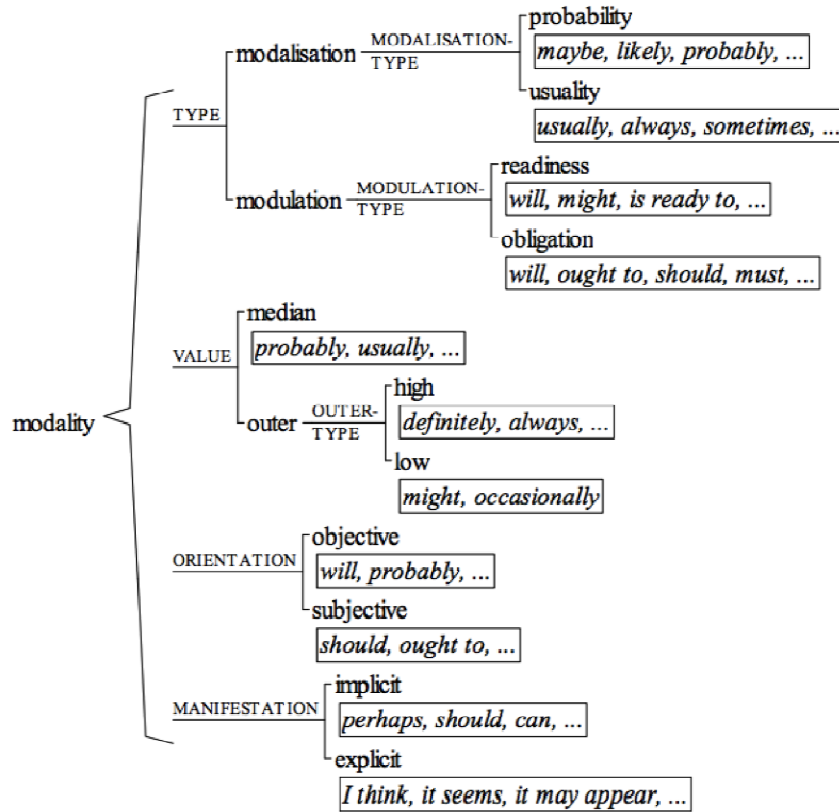


- Modulation: How “necessary” is it? (“ought to,” “allowable”)
- Value: What degree of the relevant modality scale is being averred?
  - Median: The “normal” amount (“likely,” “usually”)
  - Outer: An extreme (either high or low) amount (“never,” “maybe,” “must”)
- Orientation: What is the relation to the speaker/writer of the modality expressed?
  - Objective: Modality expressed irrespective of the speaker/writer (“maybe,” “always”)
  - Subjective: Modality expressed relative to the speaker/writer (“We think,” “I need”)
- Manifestation: How is the modal assessment related to the event being assessed?
  - Implicit: Modality realized “in-line” by an adjunct or modal auxiliary (“preferably,” “maybe”)
  - Explicit: Modality realized by a projective verb, with the nested clause being assessed (“It is better to,” “It is possible to”)<sup>11</sup>

---

<sup>11</sup> See HALLIDAY & MATTHIESSEN, *supra* note 2, at 612–25.

Figure 4. System diagram for Modality. Note the four parallel taxonomies.



## II. EXPERIMENTS IN AUTHORSHIP PROFILING

The uses of these features can be seen both in authorship attribution and in *authorship profiling*, where we seek to determine characteristics of a text’s author (such as sex, age, or personality), even in the absence of any specific candidate authors. We describe here some experiments we have done on authorship profiling for author sex, age, native language, and personality.<sup>12</sup>

In these experiments, we compared the use of functional lexical features as above with content-based features, namely,

---

<sup>12</sup> See Shlomo Argamon et al., *Automatically Profiling the Author of an Anonymous Text*, COMM. ACM, Feb. 2009, at 119.

individual words. In order to keep the number of features reasonably small, we consider just the 1,000 words that appear sufficiently frequently in the corpus and that discriminate best between the classes of interest (determined by “information-gain” on a holdout set).

We note that the use of content-based features for authorship studies can be problematic. One must be even more wary of content markers potentially being artifacts of a particular writing situation or experimental setup and thus producing overly optimistic results that will not be borne out in real-life applications. For example, were we to seek to identify Arthur Conan Doyle’s writing by the high frequency of the words “Sherlock,” “Holmes,” and “Watson,” we would misattribute any works not part of that detective series. We will therefore be careful to distinguish results that exploit content-based features from those that do not.

Whatever features are used in a particular experiment, we represent a document as a numerical vector  $X$ . Once labeled training documents have been represented in this way, we can apply machine-learning algorithms to learn classifiers that assign new documents to categories. Generally speaking, the most effective multiclass (i.e., more than two classes) classifiers for authorship studies all share the same structure: we learn a weight vector  $W_j$  for each category  $c_j$  and then assign a document,  $X$ , to the class for which the inner product  $W_j * X$  is maximal. The weight vector is learned based on a *training set* of data points, each labeled with its correct classification. There are a number of effective algorithms for learning such weight vectors; we use here Bayesian Multinomial Regression (“BMR”),<sup>13</sup> which we have found to be both efficient and accurate. BMR is a probabilistically well-founded multivariate variant of logistic regression, which tends to work well for problems with large numbers of variables (as here).<sup>14</sup> BMR has

---

<sup>13</sup> See Alexander Genkin et al., *Large-Scale Bayesian Logistic Regression for Text Categorization*, 49 *TECHNOMETRICS* 291, 291–304 (2007).

<sup>14</sup> When seeking to construct predictive models from data with a very large number of variables, it is possible that a model can easily be found to fit the known data accidentally, just because there are many parameters in the model that can be adjusted. Such a model will then not classify new data

also been shown specifically to be effective for text classification and related problems.<sup>15</sup> Other learning methods such as support vector machines<sup>16</sup> generally work just as well.

### A. Test Data

In the experiments described below, we sought to profile documents by four common author characteristics: sex, age, native language, and personality type. The first three of these have obvious application in the investigative and forensic contexts. Personality type is more useful for investigations but can also provide corroborative evidence for identification when personality information about a suspect is known. We first describe in this section the data sets, comprising labeled collections of texts, that we used to learn and test our classification models. In the following section, we will describe the experimental procedure and results.

**Sex and Age.** Our corpus<sup>17</sup> for both author sex and age consists of the full set of postings of 19,320 blog authors (each text is the full set of posts by a given author) writing in English. The (self-reported) age and gender of each author is known and for each age interval the corpus includes an equal number of male and female authors. The texts range in length from several hundreds to tens of thousands of words, with a mean length of 7,250 words per author. Based on each blogger's reported age, we label each blog in our corpus as belonging to one of three

---

well. This problem is known as *overfitting*. See Tom Dietterich, *Overfitting and Undercomputing in Machine Learning*, ACM COMPUTING SURVS., Sept. 1995, at 326–27. BMR, and other modern learning algorithms, seek to minimize this problem by various mathematical methods.

<sup>15</sup> See Genkin et al., *supra* note 13; see also Moshe Koppel et al., *Automatically Classifying Documents by Ideological and Organizational Affiliation*, PROC. 2009 IEEE INT'L CONF. ON INTELLIGENCE & SECURITY INFORMATICS, at 176.

<sup>16</sup> See NELLO CRISTIANINI & JOHN SHAWE-TAYLOR, AN INTRODUCTION TO SUPPORT VECTOR MACHINES AND OTHER KERNEL-BASED LEARNING METHODS 7 (2000).

<sup>17</sup> First described in Jonathan Schler et al., *Effects of Age and Gender on Blogging*, AAAI SPRING SYMPOSIUM: COMPUTATIONAL APPROACHES TO ANALYZING WEBLOGS, 2006, at 199.

age groups: thirteen to seventeen (42.7%), twenty-three to twenty-seven (41.9%) and thirty-three to forty-seven (15.5%). Intermediate age groups were removed to avoid ambiguity since many of the blogs were written over a period of several years. Our objective is to identify to which of these three age intervals an anonymous author belongs.

**Native Language.** We used the International Corpus of Learner English (“ICLE”),<sup>18</sup> which was assembled for the precise purpose of studying the English writing of nonnative English speakers from a variety of countries. All the writers in the corpus are university students (mostly in their third or fourth year) studying English as a second language. All are roughly the same age (in their twenties) and are assigned to the same proficiency level in English. All texts are short student essays on a similar set of topics, so they are in the same genre. We consider five subcorpora from Russia, the Czech Republic, Bulgaria, France, and Spain. To balance the corpus, we took 258 authors from each subcorpus (randomly discarding any surplus). All texts in the resulting corpus are between 579 and 846 words long. Our objective is to determine which of the five languages is the native tongue of an anonymous author writing in English.

**Personality.** We used essays written by psychology undergraduates at the University of Texas at Austin collected by James W. Pennebaker.<sup>19</sup> Students were instructed to write a short “stream of consciousness” essay wherein they tracked their thoughts and feelings over a twenty minute free-writing period. The essays range in length from 251 to 1,951 words. Each writer also filled out a questionnaire testing for the “Big Five” personality dimensions: neuroticism, extraversion, openness, conscientiousness, and agreeableness. We consider here just the dimension of neuroticism (roughly, tendency to worry or be anxious), as methods and results for other personality factors are qualitatively similar. We defined “positive” examples to be the

---

<sup>18</sup> *International Corpus of Learner English*, UNIVERSITE CATHOLIQUE DE LOUVAIN, <http://www.uclouvain.be/en-cecl-icle.html> (last visited Mar. 2, 2013).

<sup>19</sup> Shlomo Argamon et al., *Lexical Predictors of Personality Type*, PROC. JOINT ANN. MEETING INTERFACE & CLASSIFICATION SOC’Y N. AM., 2005.

participants with neuroticism scores in the upper third of the authors, and ‘negative’ examples to be those with scores in the lowest third. The rest of the data were ignored, and the final corpus consists of 198 examples.

### *B. Procedure and Results*

Accuracy results for the above profiling tasks are given in Table 1 for different combinations of features. Recall that a training set is required for the system to learn a classification model for any given task. The accuracy of the system must be evaluated on data separate from the training data, since even perfect performance on the training data is easy to achieve and meaningless in terms of the real-world potential accuracy of the system. Hence each dataset needs to be divided into disjoint training and test sets for evaluation. To maximize use of limited data, a standard technique, called *ten-fold cross-validation*, is used to divide the data randomly into ten equal parts, then to perform ten train-test runs, each run training on nine-tenths of the data and testing on the remaining tenth. The average accuracy over these ten runs is a good estimate of the actual performance of the system on new data.

Accuracy is measured simply as the percentage of text examples that the system classified correctly. In any given classification problem, there is a baseline performance, given by the percentage of the data falling into the majority class. This percentage indicates the performance of the trivial classifier that just classifies every example as that majority class. If the accuracy of our classification system is significantly higher than this baseline performance, the system can be said to work; the higher the accuracy, the better it works.

Consider now the results for authorship profiling given in Table 1. We first note that while in most cases (other than neuroticism) content words help, style features often give good results on their own. More informative are the highest weighted features for each output class, given in Table 2. For sex, the style features that prove to be most useful for gender discrimination are determiners and certain prepositions (markers of male writing) and pronouns (markers of female writing),

which is consistent with other studies. For age, we see a preference for more formal writing in the older bloggers (prepositions and determiners), though the content features in this case give more insight, in terms of the usual concerns of people in different age groups. For native language, we see some interesting stylistic patterns, in that native speakers of Slavic languages have clear preferences for personal pronouns, particularly first person, while Romance language speakers have distinctive (and different) patterns of verb auxiliary use. The content features in this case, while more dispositive, are clearly not useful in any context where deception would come into play, as they can be easily planted by a deceptive writer.

Finally, we see that neurotics tend to refer more often to themselves, use pronouns as subjects rather than as objects in a clause, and consider explicitly who benefits from some action (through prepositional phrases involving, e.g., “for” and “in order to”); nonneurotics, on the other hand, tend to use less precise specification of objects or events (determiners and adjectives such as “a” or “little”) and show more concern with how things are or should be done (via prepositions such as “by” or “with” and modals such as “ought to” or “should”).

In other experiments we have done using features of lexicogrammar indicative of writers’ attitudes, we found (unsurprisingly) texts by neurotic individuals to be characterized more by focus on, e.g., negative orientation and affect, whereas texts by nonneurotics focused more on positive orientation and appreciation.<sup>20</sup> That is, neurotics evaluated objects and propositions more negatively and more in terms of feelings, while nonneurotics did so more positively and more in terms of objective characteristics.

---

<sup>20</sup> *See id.*

*Table 1. Classification accuracy (10-fold cross-validation) for authorship profiling using different feature sets.*

	<b>Baseline</b>	<b>Style</b>	<b>Content</b>	<b>Style+ Content</b>
<b>Gender (2 classes)</b>	50.0	72.0	75.1	<b>76.1</b>
<b>Age (3 classes)</b>	42.7	66.9	75.5	<b>77.7</b>
<b>Language (5 classes)</b>	20.0	65.1	<b>82.3</b>	79.3
<b>Neuroticism (2 classes)</b>	50.0	<b>65.7</b>	53.0	63.1



*Table 2. Most important Style and Content features (by information gain) for each class of texts in each profiling problem.*

<b>Class</b>	<b>Style Features</b>	<b>Content Features</b>
Female	<b>personal pronoun, I, me, him, my</b>	<i>cute, love, boyfriend, mom, feel</i>
Male	<b>determiner, the, of, preposition-matter, as</b>	<i>system, software, game, based, site</i>
Teens	<b>im, so, thats, dont, cant</b>	<i>haha, school, lol, wanna, bored</i>
Twenties	<b>preposition, determiner, of, the, in</b>	<i>apartment, office, work, job, bar</i>
Thirties+	<b>preposition, the, determiner, of, in</b>	<i>years, wife, husband, daughter, children</i>
Bulgarian	<b>conjunction-extension, pronoun-interactant, however, pronoun-conscious, and</b>	<i>bulgaria, university, imagination, bulgarian, theoretical</i>
Czech	<b>personal pronoun, usually, did, not, very</b>	<i>czech, republic, able, care, started</i>
French	<b>indeed, conjunction-elaboration, will, auxverb-future, auxverb-probability</b>	<i>identity, europe, european, nation, gap</i>
Russian	<i>can't, i, can, over, every</i>	<i>russia, russian, crimes, moscow, crime</i>
Spanish	<b>determiner-specific, this, going_to, because, although</b>	<i>spain, restoration, comedy, related, hardcastle</i>
Neurotic	<b>myself, subject pronoun, reflexive pronoun, preposition-behalf, pronoun-speaker</b>	<i>put, feel, worry, says, hurt</i>
Nonneurotic	<b>little, auxverbs-obligation, nonspecific determiner, up, preposition-agent</b>	<i>reading, next, cool, tired, bed</i>

## III. DISCUSSION

We have sketched here a framework for addressing authorship attribution as a question of evaluating codal variation by estimating the probabilities of different grammatical choices by different authors or kinds of authors. These features perform as well or better in our empirical tests as other sorts of features and (often) have the advantage of giving meaningful insight into the underlying stylistic differences between authors.

As we have argued above and elsewhere,<sup>21</sup> such insight should be considered a key criterion for authorship attribution methods, along with accuracy and reliability. Without such understanding, it is extremely difficult, or impossible, to have real confidence that results in any specific instance are reliable, due to the large number and variety of possible confounding factors (dialect and register variation and the like). Results that can be meaningfully interpreted, however, also make the task of conveying their import to nonexperts, including judges and juries, much easier.

It also seems likely that an operationalization of idiolect as a systematic skewing of probabilities in system taxonomies, as developed above, helps to put the problem of author analysis into a larger theoretical context. This context recognizes language variation due to code, as in authorial differences, as well as variation due to register and genre. By identifying author analysis as one aspect of a continuum of similar kinds of variation, we may hope to disentangle the omnipresent effects of register and genre variation when analyzing authorship.

---

<sup>21</sup> See, e.g., Shlomo Argamon & Moshe Koppel, *The Rest of the Story: Finding Meaning in Stylistic Variation*, in *THE STRUCTURE OF STYLE: ALGORITHMIC APPROACHES TO UNDERSTANDING MANNER AND MEANING* 79 (Shlomo Argamon et al. eds., 2010); see also Argamon et al., *supra* note 7.

# AUTHORSHIP ATTRIBUTION: WHAT'S EASY AND WHAT'S HARD?

*Moshe Koppel,\* Jonathan Schler,<sup>†</sup> and Shlomo Argamon\*\**

## INTRODUCTION

The simplest kind of authorship attribution problem—and the one that has received the most attention—is the one in which we are given a small, closed set of candidate authors and are asked to attribute an anonymous text to one of them. Usually, it is assumed that we have copious quantities of text by each candidate author and that the anonymous text is reasonably long. A number of recent survey papers<sup>1</sup> amply cover the variety of methods used for solving this problem.

Unfortunately, the kinds of authorship attribution problems we typically encounter in forensic contexts are more difficult than this simple version in a number of ways. First, the number of suspected writers might be very large, possibly numbering in the many thousands. Second, there is often no guarantee that the true author of an anonymous text is among the known suspects. Finally, the amount of writing we have by each candidate might be very limited and the anonymous text itself might be short.

---

\* Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel, moishk@gmail.com (Corresponding Author).

† Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel, schler@gmail.com.

\*\* Department of Computer Science, Illinois Institute of Technology, argamon@iit.edu.

<sup>1</sup> Patrick Juola, *Authorship Attribution*, 1 FOUND. & TRENDS IN INFO. RETRIEVAL 233, 238–39 (2006); Moshe Koppel et al., *Computational Methods in Authorship Attribution*, 60 J. AM. SOC'Y FOR INFO. SCI. & TECH. 9, 9 (2009); Efstathios Stamatatos, *A Survey of Modern Authorship Attribution Methods*, 60 J. AM. SOC'Y FOR INFO. SCI. & TECH. 538, 539 (2009).

This paper considers four versions of the attribution problem that are typically encountered in the forensic context and offers algorithmic solutions for each. Part I describes the *simple authorship attribution problem* described above. Part II considers the *long-text verification problem*, in which we are asked if two *long* texts are by the same author. Part III discusses the *many-candidates problem*, in which we are asked which among thousands of candidate authors is the author of a given text. Finally, Part IV considers the *fundamental problem* of authorship attribution, in which we are asked if two *short* texts are by the same author. Although other researchers have considered these problems, here we offer our own solutions to each problem and indicate the degree of accuracy that can be expected in each case under specified conditions.

#### I. SIMPLE AUTHORSHIP ATTRIBUTION

The simplest problems arise when, as mentioned above, we have a closed set of candidate authors as well as an abundance of training text<sup>2</sup> for each author. Our objective is to assign an anonymous text to one of the candidate authors. For this purpose, we wish to design automated techniques that use the available training text to assign a text to the most likely candidate author. As a rule, such automated techniques can be divided into two main types: *similarity-based* methods and *machine-learning* methods.<sup>3</sup>

In similarity-based methods, a metric is used to computationally measure the similarity between two documents, and the anonymous document is attributed to that author whose known writing (considered collectively as a single document) is most similar. Research in the similarity-based paradigm has focused on the choice of features for document representation—such as the frequency of particular words or other lexical or

---

<sup>2</sup> Training text is simply a collection of writing samples by a given author that can be used to characterize the author's writing style for purposes of attribution.

<sup>3</sup> Stamatatos, *supra* note 1, at 551.

syntactic features in the document—and on the choice of distance metric.<sup>4</sup>

In machine-learning methods, the known writings of each candidate author (considered as a set of distinct training documents) are used to construct a classifier that can then be used to categorize anonymous documents. The idea is to formally represent each of a set of training documents as a numerical vector and then use a learning algorithm to find a formal rule, known as a classifier, that assigns each such training vector to its known author. This same classifier can then be used to assign anonymous documents to (what one hopes is) the right author. Research in the machine-learning paradigm has focused on the choice of features for document representation and on the choice of learning algorithm.<sup>5</sup>

This section of the paper focuses on machine-learning methods. Here we consider and compare a variety of learning algorithms and feature sets for three authorship attribution problems that are representative of the range of classical attribution problems. The three problems are as follows:

1. A large set of emails between two correspondents (M. Koppel and J. Schler, co-authors of this paper), covering the year 2005. The set consisted of 246 emails from Koppel and 242 emails from Schler, each stripped of headers, named greetings,

---

<sup>4</sup> See generally Ahmed Abbasi & Hsinchun Chen, *Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace*, 26 ACM TRANSACTIONS ON INFO. SYS. 7:1 (2008); Shlomo Argamon, *Interpreting Burrows's Delta: Geometric and Probabilistic Foundations*, 23 LITERARY & LINGUISTIC COMPUTING 131 (2007); John Burrows, *'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship*, 17 LITERARY & LINGUISTIC COMPUTING 267 (2002); Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 INT'L J. SPEECH LANGUAGE & L. 1 (2001); David L. Hoover, *Multivariate Analysis and the Study of Style Variation*, 18 LITERARY & LINGUISTIC COMPUTING 341 (2003).

<sup>5</sup> Abbasi & Chen, *supra* note 4, at 7:10; Koppel et al., *supra* note 1, at 11–12; Ying Zhao & Justin Zobel, *Effective and Scalable Authorship Attribution Using Function Words*, 3689 INFO. RETRIEVAL TECH. 174, 176 (2005); Rong Zheng et al., *A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques*, 57 J. AM. SOC'Y FOR INFO. SCI. & TECH. 378, 380 (2006).

signatures, and quotes from previous posts in the thread. Some of the texts were as short as a single word. Messages sent prior to July 1 were used as training data. The task is to classify messages sent after July 1 as having been written by either Schler or Koppel.

2. Two books by each of nine late nineteenth- and early twentieth-century authors of American and English literature (Hawthorne, Melville, Cooper, Shaw, Wilde, C. Bronte, A. Bronte, Thoreau, and Emerson). One book by each author was used for training. The task is to determine the author of each 500-word passage from the other books.

3. The full set of posts of twenty prolific bloggers, harvested in August 2004. The number of posts of the individual bloggers ranged from 217 to 745 with an average of just over 250 words per post. All but the last thirty posts of each blogger were used for training. The task is to determine the author of each of the 600 (20 authors \* 30 posts) remaining blog posts.

These corpora differ along a variety of dimensions, including most prominently the size of the candidate sets (2, 9, 20) and the nature of the material (emails, novels, blogs).

For each corpus, we ran experiments comparing the effectiveness of various combinations of feature types—measurable properties of a text, such as frequencies of various words, that can be used to characterize the text—and machine-learning methods. The feature types and machine-learning methods that we used are listed in Table 1. Each document in each corpus was processed to produce a numerical vector, each of whose elements represents the relative frequency of some feature in the selected feature set. Models learned on the training sets were then applied to the corresponding test sets to estimate generalization accuracy. Table 2 shows the results for each combination of features and learning method for the email corpus. Table 3 shows the results for the literature corpus. Table 4 shows the results for the blog corpus.

As can be seen, a feature set consisting of common words and character  $n$ -grams (sequences of  $n$  characters), used in conjunction with either Bayesian logistic regression or support vector machines (SVM) as a learning algorithm, yields accuracy near or above 80% for each problem. More broadly, the results

suggest that large sets of very simple features are more accurate than small sets of sophisticated features for this purpose. Many other experiments on more straightforward problems indicate that for two-author problems and ample training text, accuracy is very close to 100%.

## II. LONG-TEXT AUTHORSHIP VERIFICATION

Next, we consider the authorship verification problem for long, book-length texts. Specifically, we seek to determine whether two specific books,  $A$  and  $X$ , were written by the same author. The “unmasking” method (described below) can be used to answer this question.<sup>6</sup> Broadly speaking, unmasking is a technique for measuring the depth of the differences between two documents.

A naïve starting point might be to apply the methods described above to learn a model for  $A$  vs.  $X$  and assess the extent of the difference between  $A$  and  $X$  by evaluating generalization accuracy through cross-validation. (That is, we use part of the available data for training and test on the rest, repeating this process according to a specific protocol, the details of which we omit here.) This intuitive model asserts that if cross-validation accuracy is high, one should conclude that the author of  $A$  did not write  $X$ ; however, if cross-validation accuracy is low (i.e., we fail to correctly classify test examples better than chance), one should conclude that the author of  $A$  did write  $X$ . This intuitive method does not actually work well at all.

Examining a real world example helps us consider exactly why the last method fails. Suppose we are given known works by Herman Melville, James Fenimore Cooper, and Nathaniel Hawthorne. For each of the three authors, we are asked if that author was or was not also the author of *The House of the Seven Gables*.<sup>7</sup> Using the method described and using a feature set consisting of the 250 most frequently used words in *Gables* and

---

<sup>6</sup> See generally Moshe Koppel et al., *Measuring Differentiability: Unmasking Pseudonymous Authors*, 8 J. MACH. LEARNING RES. 1261 (2007).

<sup>7</sup> NATHANIEL HAWTHORNE, *THE HOUSE OF THE SEVEN GABLES* (Project Gutenberg ed., 2008), [http://www.gutenberg.org/catalog/world/readfile?fk\\_files=1441383](http://www.gutenberg.org/catalog/world/readfile?fk_files=1441383).

in the known works of each of the three candidate authors, respectively, we find that we can distinguish *Gables* from the works of each author with cross-validation accuracy of above 98%. If we were to conclude, therefore, that none of these authors wrote *Gables*, we would be wrong: Hawthorne, in fact, wrote it.

If we look closely at the models that successfully distinguish *Gables* from one of Hawthorne's other works (in this case, *The Scarlet Letter*), we find that only a small number of features distinguish between them. These features include "he," which appears more frequently in *The Scarlet Letter*, and "she," which appears more frequently in *Gables*. The situation in which an author will use a small number of features in a consistently different way between works is typical. These differences might result from thematic differences between the works, differences in genre or purpose, chronological stylistic drift, or deliberate attempts by the author to mask his or her identity.

Our main point is to show how this problem can be overcome by determining not only if *A* is distinguishable from *X*, but also how great the depth of difference between *A* and *X* is.<sup>8</sup> To do this, we use a technique that we call "unmasking."<sup>9</sup> The idea is to remove, by stages, those features that are most useful for distinguishing between *A* and *X* and to gauge the speed with which cross-validation accuracy degrades as more features are removed. Our main hypothesis is that if *A* and *X* are by the same author, then whatever differences are between them will be reflected in only a relatively small number of features, despite possible differences in theme, genre, and the like. Thus, for example, we expect that when comparing *Gables* to works by other authors, the degradation as we remove distinguishing features from consideration is slow and smooth but when comparing it to another work by Hawthorne, the degradation is sudden and dramatic.

Formally, our algorithm works as follows:

1. Determine the accuracy results of a ten-fold cross-validation experiment (using SVM as a learning algorithm and

---

<sup>8</sup> This material is adapted from an earlier work, Koppel et al., *supra* note 6.

<sup>9</sup> *Id.* at 1263–64.



the 250 most common words in the corpus as a feature set) for A against X.

2. For the model obtained in each fold, eliminate the  $k$  most strongly weighted positive features and the  $k$  most strongly weighted negative features.

3. Go to step 1.

In this way, we construct degradation curves for the pair  $\langle A, X \rangle$ .

In Figure 1, we show degradation curves obtained from comparing *Gables* to known works of Melville, Cooper, and Hawthorne, respectively. This graph bears out our hypothesis. Indeed, when comparing *Gables* to another work by Hawthorne, the degradation is far more severe than when comparing it to works by the other authors. Once a relatively small number of distinguishing markers are removed, the two works by Hawthorne become nearly indistinguishable.

This phenomenon is actually quite general. In fact, we have shown elsewhere<sup>10</sup> that we can distinguish same-author degradation curves from different-author degradation curves with accuracy above 90% in a variety of genres and languages. Unfortunately, unmasking does not work for short documents.<sup>11</sup> Below, we turn to the short-document problem.

### III. THE MANY-CANDIDATES PROBLEM FOR SHORT DOCUMENTS

Next, we consider cases in which there may be a very large number of candidate authors, possibly in the thousands. While most work has focused on problems with a small number of candidate authors, there has been some recent work on larger candidate sets.<sup>12</sup>

---

<sup>10</sup> *Id.* at 1264–67.

<sup>11</sup> Conrad Sanderson & Simon Guenter, *Short Text Authorship Attribution Via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation*, PROC. INT'L CONF. ON EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING, 2006, at 490, available at <http://itee.uq.edu.au/~conrad/papers.html>.

<sup>12</sup> See, e.g., Moshe Koppel et al., *Authorship Attribution with Thousands of Candidate Authors*, PROC. 29TH ANN. ACM & SIGIR CONF. ON RES. & DEV. ON INFO. RETRIEVAL, 2006, at 1–2, available at

We report here on a method we introduced in a previous paper.<sup>13</sup> The key insight is that a similarity-based approach can be used to identify the most likely authors, but the robustness of the similarity must be taken into account in order to filter false positive identifications.

We use a set of 10,000 blogs harvested in August 2004 from blogger.com.<sup>14</sup> The corpus is balanced for gender within each of a number of age intervals. In addition, each individual blog is predominantly in English and contains sufficient text, as will be explained. For each blog, we choose 2,000 words of known text and a *snippet*, consisting of the last 500 words of the blog, such that the posts from which the known text and the snippet are taken are disjoint. Our object is to determine which—if any—of the authors of the known texts is the author of a given snippet.

We begin by representing each text (both known texts and snippets) as a vector representing the respective frequencies of each *space-free character 4-gram*. For our purposes, a space-free character 4-gram is either (a) a string of characters of length four that includes no spaces or (b) a string of four or fewer characters surrounded by spaces. In our corpus, there are just over 250,000 unique (but overlapping) space-free character 4-grams. We select the 100,000 such features most frequent in

---

<http://www.csie.ntu.edu.tw/~r95038/paper/paper%20WebIR/p659-koppel.pdf> (demonstrating experiment with 10,000 authors); Kim Luyckx & Walter Daelemans, *Authorship Attribution and Verification with Many Authors and Limited Data*, PROC. 22ND INT'L CONF. ON COMPUTATIONAL LINGUISTICS, 2008, at 513, available at <http://www.clips.ua.ac.be/~kim/publications.php> (145 authors); David Madigan et al., *Author Identification on the Large Scale*, PROC. MEETING CLASSIFICATION SOC'Y N. AM., 2006, at 9, available at <http://dimacs.rutgers.edu/Research/MMS/PAPERS/authorid-csna05.pdf> (114 authors); Arvind Narayanan et al., *On the Feasibility of Internet-Scale Author Identification*, PROC. 33RD CONF. ON IEEE SYMP. ON SECURITY & PRIVACY, 2012, available at <http://www.cs.berkeley.edu/~dawnsong/papers/2012%20On%20the%20Feasibility%20of%20Internet-Scale%20Author%20Identification.pdf> (100,000 authors).

<sup>13</sup> Moshe Koppel et al., *Authorship Attribution in the Wild*, 45 LANGUAGE RESOURCES & EVALUATION 83, 86–87 (2011).

<sup>14</sup> This material is adapted from an earlier work, Moshe Koppel et al., *The “Fundamental Problem” of Authorship Attribution*, 93 ENG. STUD. 284, 286–88 (2012).

the corpus as our feature universe. Character  $n$ -grams have been shown to be effective for authorship attribution<sup>15</sup> and have the advantage of being measurable in any language without specialized background knowledge.

The methods we describe in Part I for authorship attribution were not designed for large numbers of classes, certainly not for 10,000 classes. Instead, we use a similarity-based method. Specifically, we use a common, straightforward information retrieval method to assign an author to a given snippet. Using cosine similarity as a proximity measure, we simply return the author whose known writing (considered as a single vector of space-free character 4-gram frequencies) is most similar to the snippet vector. Testing this rather naïve method on 1,000 snippets selected at random from among the 10,000 authors, we find that 46% of the snippets are correctly assigned. While this accuracy is perhaps surprisingly high, it is certainly inadequate for forensic applications. To remedy this problem, we adopt a previously devised approach,<sup>16</sup> which permits a response of “*Don't Know*” in cases where attribution is uncertain. The objective is to obtain high precision for those cases where an answer is given, while trying to offer an answer as often as possible.

The key to our new approach is the same as the underlying principle of unmasking. The known text of a snippet's actual author is likely to be the text most similar to the snippet, even as we vary the feature set that we use to represent the texts. Another author's text might happen to be the most similar for one or a few specific feature sets, but it is highly unlikely to be consistently so over many different feature sets.

This observation suggests using the following algorithm:

**Given:** snippet of length  $L_1$ ; known-texts of length  $L_2$  for each of  $C$  candidates

**Repeat**  $k_1$  times

Randomly choose some fraction  $k_2$  of the full feature set

Find top match using cosine similarity

---

<sup>15</sup> Efstathios Stamatatos et al., *Computer-Based Authorship Attribution Without Lexical Measures*, 35 COMPUTERS & HUMAN. 193, 207–08 (2001).

<sup>16</sup> Koppel et al., *supra* note 13; Koppel et al., *supra* note 14.

**For each** candidate author  $A$ ,  
 Score( $A$ ) = proportion of times  $A$  is top match  
**Output:**  $\arg \max_A \text{Score}(A)$  **if**  $\max \text{Score}(A) > \sigma^*$ ; **else**  
*Don't Know*

The idea is to check if a given author proves to be most similar to the test snippet for many different randomly selected feature sets of fixed size. The number of different feature sets used ( $k_1$ ) and the fraction of all possible features in each such set ( $k_2$ ) are parameters that must be selected. The threshold  $\sigma^*$ , which serves as the minimal score an author requires to be deemed the actual author, is a parameter that we vary for recall-precision tradeoff. We choose a high threshold if we wish to be cautious and avoid incorrect attributions, at the price of frequently returning *Don't Know*. We set the number of iterations ( $k_1$ ) to 100, the snippet length ( $L_1$ ) to 500, the known-text length for each candidate ( $L_2$ ) to 2000, and the fraction of available features used in the feature set ( $k_2$ ) to 40%. We consider how the number of candidate authors affects precision and recall. Figure 2 shows recall-precision curves for various numbers of candidate authors. Note that, as expected, accuracy increases as the number of candidate authors diminishes. The point  $\sigma^* = .90$  is marked on each curve. For example, for 1,000 candidates, at  $\sigma^* = .90$ , we achieve 93.2% precision at 39.3% recall.

#### IV. THE "FUNDAMENTAL PROBLEM" OF AUTHORSHIP ATTRIBUTION

The above method can serve as the basis for solving what we call the "fundamental problem" of authorship attribution: determining the authorship of two (possibly short) documents written by either the same or two different authors. Plainly, if we can solve this problem, we can solve the standard attribution problems considered above, as well as many other authorship attribution problems.

Our approach<sup>17</sup> to solving the fundamental problem is as follows: Given two texts,  $X$  and  $Y$ , we generate a set of

---

<sup>17</sup> Koppel et al., *supra* note 14.

impostors  $(Y_1, \dots, Y_n)$  and then use the above method to determine if  $X$  was written by the author of  $Y$  or any of the impostors or by none of them. If and only if we obtain a result that  $X$  was written by the author of  $Y$  with a sufficiently high score, we say that the two documents are by a single author. (Clearly, we can additionally, or alternatively, generate impostors  $X_1, \dots, X_n$  and compare them to  $Y$ .)

The crucial issues we must consider in order to adapt the above method to our problem are the following: How many impostors should be used? How should the impostors be chosen? What score should we require in order to conclude that two documents are by a single author?

We consider a test set consisting of 500 pairs of blog posts written by a single author and 500 pairs written by two different authors. Each post is truncated to exactly 500 words.

For each test pair  $\langle X, Y \rangle$ , we proceed as follows: Choosing from a very large universe of blog posts, we identify the 250 most similar blog posts to  $Y$  (to ensure that impostors at least roughly resemble  $Y$ ) and then randomly choose from among them 25 blog posts to serve as our impostors,  $Y_1, \dots, Y_n$ . We assign  $\langle X, Y \rangle$  to a single author if and only if  $Y$  is selected from among the set  $\{Y, Y_1, \dots, Y_n\}$  as most similar to  $X$  in at least 11 trials out of 100. (The threshold 11 was determined on a separate development set.)

Using this method, 87.3% of our 1,000 test pairs are correctly identified as *same-author* or *different-author*.

## V. DISCUSSION

To summarize, four distinct problems have been considered in this paper, roughly in order of difficulty. The ordinary attribution problem with a small, closed set of candidates is well understood and solvable with established machine-learning techniques. Authorship verification, in which we wish to determine if two documents are by the same author, can be solved using unmasking provided that the documents in question are sufficiently long. The case in which there are many candidate authors can be handled using feature randomization techniques with fairly high precision, but for many cases this

method will simply respond with “*Don’t Know.*” Finally, authorship verification for short documents can be handled by assembling an impostor set and then invoking the method used for the many-candidates problem. This method remains somewhat speculative.

In addition to the four problems discussed above, methods have been developed by the authors of this paper for profiling authors (in terms of gender, age, native language, and personality type).<sup>18</sup> Moreover, it has been shown by the authors that multi-author documents can be segmented into distinct authorial threads.<sup>19</sup>

Although in all these cases accuracy results on out-of-sample test sets have been provided, many methodological questions that are crucial in forensic contexts are left open. Are our test corpora comparable to the kinds of cases that arise in forensic contexts? Do we make hidden assumptions about the data that are not realistic? Do our methods allow us to tell a good enough story to persuade a judge or jury of the reliability of our conclusions?

These questions are probably best answered in cooperation with legal experts and are left open for discussion.

---

<sup>18</sup> Shlomo Argamon et al., *Automatically Profiling the Author of an Anonymous Text*, COMM. ACM, Feb. 2009, at 119.

<sup>19</sup> Moshe Koppel et al., *Unsupervised Decomposition of a Document into Authorial Components*, PROC. 49TH ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS, 2011, available at <http://www.aclweb.org/anthology-new/P/P11/P11-1136.pdf>.

FW	a list of 512 function words, including conjunctions, prepositions, pronouns, modal verbs, determiners, and numbers	Stylistic
POS	38 part-of-speech unigrams and 1,000 most common bigrams using the Brill (1992) part-of-speech tagger	Stylistic
SFL	all 372 nodes in SFL trees for conjunctions, prepositions, pronouns and modal verbs	Stylistic
CW	the 1,000 words with highest information gain (Quinlan 1986) in the training corpus among the 10,000 most common words in the corpus	Content
CNG	the 1,000 character trigrams with highest information gain in the training corpus among the 10,000 most common trigrams in the corpus (cf. Keselj 2003)	Mixed content and style

NB	WEKA's implementation (Witten and Frank 2000) of Naïve Bayes (Lewis 1998) with Laplace smoothing
J4.8	WEKA's implementation of the J4.8 decision tree method (Quinlan 1986) with no pruning
RMW	our implementation of a version of Littlestone's (1988) Winnow algorithm, generalized to handle real-valued features and more than two classes (Schler 2007)
BMR	Genkin et al.'s (2006) implementation of Bayesian multi-class regression
SMO	WEKA's implementation of Platt's (1998) SMO algorithm for SVM with a linear kernel and default settings

*Table 1: Feature types and machine-learning methods used in our experiments.*

features/learner	NB	J4.8	RMW	BMR	SMO
POS	61.0%	59.0%	66.1%	66.3%	67.1%
FW + POS	65.9%	61.6%	68.0%	67.8%	71.7%
SFL	57.2%	57.2%	65.6%	67.2%	62.7%
CW	67.1%	66.9%	74.9%	78.4%	74.7%
CNG	72.3%	65.1%	73.1%	80.1%	74.9%
CW + CNG	73.2%	68.9%	74.2%	83.6%	78.2%

**Table 2:** Accuracy on test set attribution for a variety of feature sets and learning algorithms applied to authorship classification for the email corpus.

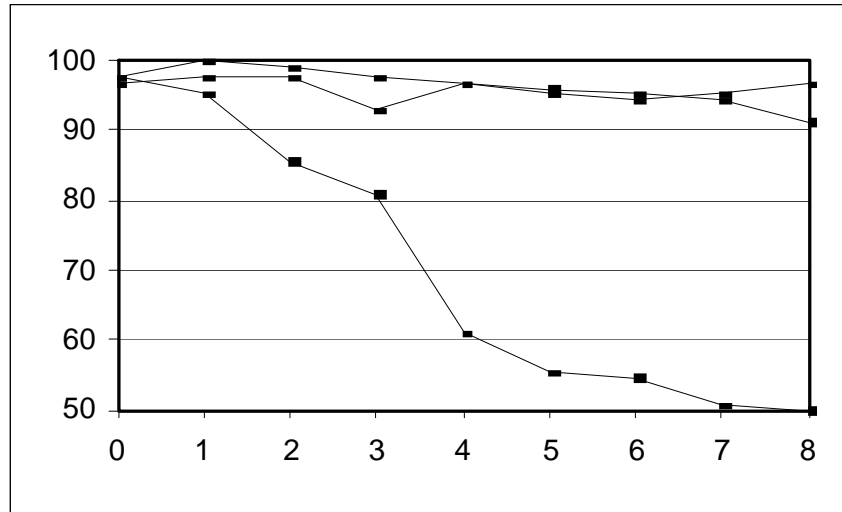
features/learner	NB	J4.8	RMW	BMR	SMO
FW	51.4%	44.0%	63.0%	73.8%	77.8%
POS	45.9%	50.3%	53.3%	69.6%	75.5%
FW + POS	56.5%	46.2%	61.7%	75.0%	79.5%
SFL	66.1%	45.7%	62.8%	76.6%	79.0%
CW	68.9%	50.3%	57.0%	80.0%	84.7%
CNG	69.1%	42.7%	49.4%	80.3%	84.2%
CW + CNG	73.9%	49.9%	57.1%	82.8%	86.3%

**Table 3:** Accuracy on test set attribution for a variety of feature sets and learning algorithms applied to authorship classification for the literature corpus.

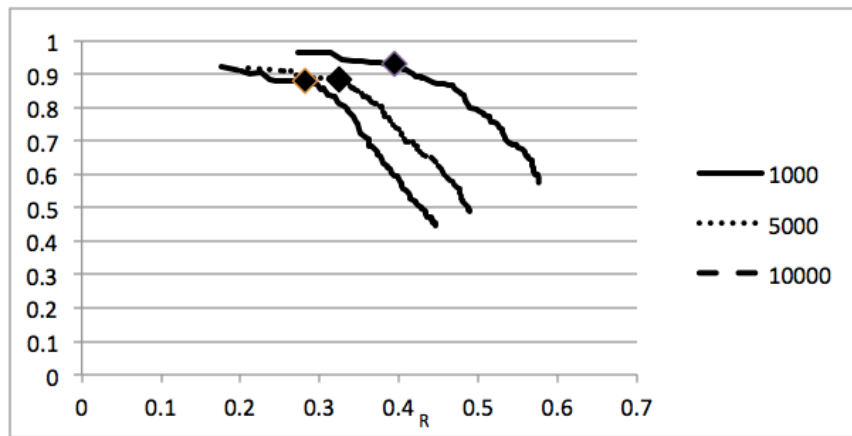
features/learner	NB	J4.8	RMW	BMR	SMO
FW	38.2%	30.3%	51.8%	63.2%	63.2%
POS	34.0%	30.3%	51.0%	63.2%	60.6%
FW + POS	47.0%	34.3%	62.3%	70.3%	72.0%
SFL	35.4%	36.3%	61.4%	69.2%	71.7%
CW	56.4%	51.0%	62.9%	72.5%	70.5%
CNG	65.0%	48.9%	67.1%	80.4%	80.9%
CW + CNG	69.9%	51.6%	75.4%	86.1%	85.7%

**Table 4:** Accuracy test set attribution for a variety of feature sets and learning algorithms applied to authorship classification for the blog corpus.





**Figure 1.** Ten-fold cross-validation accuracy of models distinguishing The House of the Seven Gables from each of Hawthorne, Melville, and Cooper. The x-axis represents the number of iterations of eliminating best features at previous iteration. The curve well below the others is that of Hawthorne, the actual author.



**Figure 2** Recall-precision for the many-candidates experiment (for various candidates set sizes).

# BEST PRACTICES AND ADMISSIBILITY OF FORENSIC AUTHOR IDENTIFICATION

*Carole E. Chaski\**

## I. INTRODUCTION

Forensic linguistics provides answers to four categories of inquiry in investigative and legal settings: (i) identification of author, language, or speaker; (ii) intertextuality, or the relationship between texts; (iii) text-typing or classification of text types such as threats, suicide notes, or predatory chat; and (iv) linguistic profiling to assess the author's dialect, native language, age, gender, and educational level. This article discusses author identification in relation to linguistics, research, and admissibility as evidence in U.S. courts.

Federal and states courts in the United States have undertaken three main approaches in determining whether to admit, partially admit, or exclude forensic authorship identification evidence. These three approaches are forensic computational linguistics, forensic stylistics, and stylometric computing. Each has a distinct origin. Forensic computational linguistics developed out of linguistic theory and computational linguistics.<sup>1</sup> Forensic stylistics developed out of traditional forensic handwriting identification.<sup>2</sup> The stylometric computing approach developed out of both literary authorship identification and machine-learning-based text classification.<sup>3</sup>

---

\* Institute for Linguistic Evidence; ALIAS Technology LLC, Georgetown, DE; Ph.D., Brown University.

<sup>1</sup> See Carole E. Chaski, *Who Wrote It? Steps Toward a Science of Authorship Identification*, NAT'L INST. JUST. J., Sept. 1997, at 15, 18 [hereinafter Chaski, *Who Wrote It?*].

<sup>2</sup> See GERALD R. MCMENAMIN, *FORENSIC STYLISTICS* 45–46 (1993).

<sup>3</sup> See Moshe Koppel & Jonathan Schler, *Exploiting Stylistic*

This article focuses upon the forensic computational linguistic approach and contrasts this approach to the forensic stylistics and stylometric computing approaches. In Section II, best practices for forensic linguistics are presented. The best practices provide an evaluative framework for the forensic computational linguistics approach, discussed in Section III; the forensic stylistics approach, discussed in Section IV; and the stylometric computing approach, discussed in Section V. In each section, a discussion of admissibility is included, since best practices should guide both judicial reasoning as well as scientific practice.

## II. BEST PRACTICES FOR FORENSIC LINGUISTICS

Best practices in forensic linguistics are essential to propel the field of authorship identification from an academic or law enforcement sideline consultancy to a real forensic science that is useful to the judicial system. Best practices include factors from both the legal standards for evidence, so as to be useful and address admissibility concerns, and scientific standards for research, so as to be reliable, replicable, and respectable.

Scientifically respectable and judicially acceptable methods for author identification should be:

- a. developed independent of any litigation;
- b. tested for accuracy outside of any litigation;
- c. tested for accuracy on “ground truth” data;
- d. able to work reliably on “forensically feasible” data;
- e. tested for known limits correlated to specific accuracy levels;
- f. tested for any errors of individual testing techniques that could cause accumulated error when combined with other techniques;
- g. replicable;
- h. related to a specific expertise and academic training;
- i. related to standard (“generally accepted”) techniques within the specific expertise and academic training; and

j. related to uses outside of any litigation in industries or fieldwork in the specific expertise.

By implementing these best practices, forensic computational linguistics is oriented primarily toward research- and empirically-driven protocols rather than expert-witnessing. In this way, forensic computational linguistics is a “normal science” subfield of computational linguistics and linguistic theory.<sup>4</sup> Accordingly, forensic computational linguistics belongs to a thriving community of academic and industry linguists with educational and industrial standards. These best practices go far toward “solving the ‘hired gun’ problem” that plagues American courts and universities—when academicians do not conduct research at all or research congruent with best practices but make themselves available as expert witnesses.<sup>5</sup>

---

<sup>4</sup> Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725 n.75 (2011). As an example of a research culture in forensic linguistics, the Institute for Linguistic Evidence, founded in 1998 through funding from the U.S. Department of Justice’s National Institute of Justice, is the first research organization devoted to validation testing for methods related to linguistic evidence. See INST. FOR LINGUISTIC EVIDENCE, <http://www.linguisticevidence.org> (last visited Apr. 18, 2013). ILE has embraced the forensic computational linguistic paradigm from its inception and over the years has averaged about five research associates working on an average of four research projects per year. See *id.* Academicians had been functioning as expert witnesses in forensic linguistics since the 1980’s. Professor Roger Shuy of Georgetown University was one of the earliest forensic linguistic experts and has described his cases prolifically, but has not sustained a research agenda in the field. Professor Gerald R. McMenamin, another early expert witness in forensic linguistics, has provided both case reports and descriptions of his method, but no testing of the method for error rate. Ironically, the “research culture” that Mnookin et al. fairly state as lacking in forensic science and crime labs is just as lacking for forensic linguistics in the halls of academe. See Mnookin et al., *supra*, at 765.

<sup>5</sup> The plague of “hired Guns” or “whores of the court” in the U.S. judicial system has been amply documented in PETER W. HUBER, *GALILEO’S REVENGE: JUNK SCIENCE IN THE COURTROOM* (1993); see also MARCIA ANGELL, *SCIENCE ON TRIAL* (1997); MARGARET A. HAGEN, *WHORE OF THE COURT: THE FRAUD OF PSYCHIATRIC TESTIMONY AND THE RAPE OF AMERICAN JUSTICE* (1997).

*A. Litigation Independence*

The implementation of these best practices involves litigation-independent development and testing of any method on a ground-truth dataset that contains forensically feasible data.<sup>6</sup> The researcher-forensic linguist runs experiments to test how well a method works outside of any litigation. The results are simply what they are, not favoring one side or the other of a legal dispute. Such a testing environment frees the researcher from confirmation bias because the results are simply what they are and enable the researcher to design the next set of experiments, as is usual in normal science.

*B. Ground-Truth Data*

For the testing to be meaningful, the experiments must be run on ground-truth data.<sup>7</sup> A ground-truth dataset contains known, verified examples with features relevant to the experiments being run.<sup>8</sup> For author identification, a ground-truth dataset typically contains text samples for which the authorship is known and verified.<sup>9</sup> For writer identification, a ground-truth dataset typically contains writing samples for which the hand writer is known and verified.<sup>10</sup> For linguistic profiling, a ground-truth dataset typically contains linguistic examples for which the demographics of each author/speaker are known and verified.

It is impossible to calculate a trustworthy accuracy rate if the researcher does not use ground-truth data. Determining a method's accuracy requires comparing the method's results to the correct answers. Correct answers can only arise from ground-truth data, where the dataset is known and verified. If

---

<sup>6</sup> Carole E. Chaski, *Author Identification in the Forensic Setting*, in THE OXFORD HANDBOOK OF LANGUAGE AND LAW 494, 494-99 (2011) [hereinafter Chaski, *Author Identification*].

<sup>7</sup> *Id.*

<sup>8</sup> *Id.*

<sup>9</sup> *Id.*

<sup>10</sup> Carole E. Chaski & Mark A. Walch, *Validation Testing for FLASH ID on the Chaski Writer Sample Database*, PROC. AM. ACAD. FORENSIC SCI. ANN. MEETING, 2009.

the researcher is using a dataset with 100 texts but an unknown number of authors, he will never know, with complete certainty, how many of those 100 texts his method correctly assigned to the actual author.<sup>11</sup> If the researcher is using a dataset containing 10,000 authors with demographic features, but the researcher has not verified those demographic features, he will never accurately know how many of those 10,000 authors his method assigned correctly to a gender, age group, or educational level.<sup>12</sup> Essentially, working without ground-truth data is a sophisticated form of guessing: it may look scientific, but it is not real science.

### *C. Forensically Feasible Data*

For the methods to work reliably in actual cases, ground-truth data must be forensically feasible, i.e., the same kind of data that is obtained in actual cases. In actual cases, writing exemplars are messy, ungrammatical, unedited, cross-genre, cross-register, and sparse because people write naturally, across a range of genres and registers. Accordingly, a forensically feasible dataset will contain business letters, love letters, angry rants, narratives, and essays so that the same author can be examined writing in different genres and registers. Each genre contributes something different to the dataset. For instance, business letters contain more formal word choice and more conventional spelling and punctuation patterns than personal e-mails, love letters, or angry blog posts. Even the writing medium—handwriting, typewriting, or computer keyboarding—can cause intra-author differences such that lexical, spelling, grammar, or punctuation patterns that occur in one medium typically do not occur in another.<sup>13</sup> In case data, the writing

---

<sup>11</sup> Chaski, *Author Identification*, *supra* note 6, at 494.

<sup>12</sup> *Id.*

<sup>13</sup> A nice example of how writing media can affect spelling comes from the *Van Wyk* case. See *infra* Part III.D. The contraction of [do not] occurred in two ways: in handwritten documents as [don't] and in typed documents as [don;t]. Typewriter and computer keyboards are different in the placement of the semicolon and apostrophe. The typewriter keyboard requires a shift to get the apostrophe, while a computer keyboard does not. The typist did not use

exemplars are typically not edited to any conventional, newspaper, academic, or industrial standards. If the researcher is not using a forensically feasible dataset to test his method, he might be misled into thinking that his method—built to assign clean, grammatical, edited business letters, newspaper articles, or novels—will work accurately on messy, ungrammatical, forensically significant texts. Essentially, building a method without testing it on forensically feasible data simply overgeneralizes a method's ability: it may look scientific because there are some validation tests to refer to, but the validation test results do not prove that the method can work on the data in the case or any forensically feasible data.

Research that focuses on literary classics or edited newspaper articles may develop accurate methods, but these methods must be tested on forensically feasible data before they are borrowed across-the-board for forensic authorship identification. In most cases, literary methods fail to work on forensic data simply because the literacy methods require far longer texts than the forensic case affords. Brevity is a fact of life inherent in forensic authorship identification that cannot be avoided or helped by research that focuses on texts that contains thousands, tens of thousands, or hundreds of thousands of words. Again, using methods that work well on literary texts or newspaper text banks, without independently testing the methods on forensically feasible data, may appear to be scientific because there is published literature in humanities computing to refer to about authorship identification in nonforensic settings,<sup>14</sup> but using such methods is akin to using a screwdriver on a nail—and an unvalidated screwdriver at that.

---

the shift key, producing a typical typing error for novices, while the handwriter never made that kind of mechanical error. The context of this difference was not noted in the forensic stylistics report by Agent Fitzgerald; instead he argued that [don;t] was a unique stylemarker.

<sup>14</sup> For instance, the Association for Computers and the Humanities publishes *Literary and Linguistic Computing*, a journal where authorship issues in literature, religion and other nonforensic settings are regularly discussed.

*D. Empirically Established Protocol*

In the research environment, the continual testing of a method of forensically feasible, ground-truth data empirically establishes the protocol for using the method in casework. First, a level of accuracy can be set: for instance, the method won't be used forensically until it reaches a certain accuracy level, such as eighty percent, ninety percent, or ninety-five percent. Second, the experiments are designed to control for variables such as the quantity of data, required number of authors, required number and types of linguistic features, and the required number and types of individual testing techniques that are combined in the method.

For the quantity of data, an important issue to resolve is the minimum number of words, sentences, or texts required for the method to obtain a certain level of accuracy.<sup>15</sup> For the number of authors, a method may require a minimum of two, five, or twenty-five suspects to obtain a certain level of accuracy. As in other pattern recognition techniques in forensic science, the number and type of features required for identification or elimination is established empirically by controlling the variable in a series of validation tests related to specific accuracy rates.<sup>16</sup> If fifteen linguistic features from syntactic analysis yield eighty-two percent accuracy, the next experiment will test sixteen, seventeen and so forth until the desired accuracy level is achieved. Those experiments empirically establish the number and type of features required for the method to obtain a specific accuracy level. Likewise, if measurement or feature selection techniques can be combined in a method (combining syntax with

---

<sup>15</sup> See Chaski, *Author Identification*, *supra* note 6, at 503.

<sup>16</sup> Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 INT'L J. SPEECH LANGUAGE & L. 1 (2001) [hereinafter Chaski, *Empirical Evaluations*]; Carole E. Chaski, *Who's at the Keyboard? Authorship Attribution in Digital Evidence Investigations*, INT'L J. DIGITAL EVIDENCE, Spring 2005, at 1 [hereinafter Chaski, *Who's at the Keyboard?*]; Carole E. Chaski, Presentation at the Eight Biennial Conference on Forensic Linguistics/Language and Law: Empirically Testing the Uniqueness of Aggregated Stylemarkers (July 14, 2007) [hereinafter Chaski, *Empirically Testing*].



other linguistic features or combining measurement based on word overlap with measurement based on *n*-grams), then experimental tests must be run to determine which techniques and how many techniques must be combined to reach a specific level of accuracy.

By working independently of any litigation and running experiments that control for different variables in how the method can be implemented, the researcher forensic linguist empirically establishes a protocol for each tested method. The protocol then becomes the guidelines for actually using the method in real casework. There will be cases where the tested methods cannot be used because data requirements cannot be met (i.e., a decedent cannot provide more writing samples), and there will be times when the tested methods can be used but only with the caveat that the data requirements for the most robust results are not met fully but are close to being satisfied (i.e., the decedent's writing samples are close to the required number). These types of situations should encourage additional research and not lead to abandonment of the research paradigm. In fact, the empirically established protocol prevents the researcher forensic linguist from becoming a "hired gun" who merely runs a method in whatever way to get the "desired result," rather than in accord with an empirically established protocol that provides a specific level of accuracy outside of litigation.

Note that "having worked a lot of cases" is not at all a substitute for empirically establishing a protocol. It simply means that a person has been hired a lot. The researcher forensic linguist has run a lot of experiments independent of litigation—a state that is far more valuable to developing forensic linguistics into a real and reliable science than a client list.

#### *E. Controlling Cumulative Error*

Most methods for forensic author identification require some tools for measurement or feature selection.<sup>17</sup> These tools can

---

<sup>17</sup> See Chaski, *Author Identification*, *supra* note 6, at 491–93.

produce errors in and of themselves; thus, an accuracy rate can be seriously affected by a series of accumulating errors in measurement or selection. For instance, off-the-shelf parsers developed in academia get very high accuracies for part-of-speech tagging on clean, edited data such as newspaper articles and novels. But these same off-the-shelf parsers often fail miserably on ungrammatical data. The problem of parsing ill-formed input or ungrammatical sentences was first discussed over thirty years ago,<sup>18</sup> and it has not been fully solved.<sup>19</sup> If the method uses an off-the-shelf parser and does not involve checking the parser results and correcting any errors of part-of-speech tagging or phrase chunking, then those errors pass through to the next step of the method. Another set of errors that can be created by software is the common practice of “preprocessing” texts to rid it of extra spaces, or to correct spellings, or insert punctuation. All of these preprocessing maneuvers actually change the original data and could remove some features that are actually useful for author identification. This kind of data handling is not scientifically acceptable even if it makes software run easily, and it undermines the accuracy of any methods that use the “preprocessed” data.

Another example is the interpretation of handwritten symbols: if a stroke is interpreted as an errant apostrophe but it is actually a low comma, this error of interpretation must be corrected, lest a later classification rely on the misinterpretation. As such errors accumulate, the linguistic analysis becomes less and less accurate, so that neither the method’s accuracy rate nor the final decision assigning texts to authors can be trusted.

---

<sup>18</sup> See K. Jensen et al., *Parse Fitting and Prose Fixing: Getting a Hold on Ill-Formedness*, 9 AM. J. COMPUTATIONAL LINGUISTICS 147 (1983); Ralph M. Weischedel & John E. Black, *Responding Intelligently to Unparsable Inputs*, 6 AM. J. COMPUTATIONAL LINGUISTICS 97 (1980); Ralph M. Weischedel & Norman K. Sondheimer, *Meta-Rules as a Basis for Processing Ill-Formed Input*, 9 AM. J. COMPUTATIONAL LINGUISTICS 161 (1983).

<sup>19</sup> See Jennifer Foster & Carl Vogel, *Parsing Ill-Formed Text Using an Error Grammar*, 21 ARTIFICIAL INTELLIGENCE REV. 269 (2004).

*F. Replicability*

The protocol developed through repeated validation testing must be repeatable by others who use it. Methods within the protocol also must be repeatable through the implementation in computer software or the strict operationalization of terms and procedures. Implementing a method in computer software is a sure way of providing objectivity and maintaining consistency. Systems can be designed so that each user can tweak parameters, thereby changing the algorithm. However, these tweaks might not be visible later. Accordingly, such systems do not maintain consistency in running a method, and the fact that a method is implemented in software does not necessarily guarantee that it is completely replicable.

*G. The Method's Relationship to Academic and Industrial Uses*

Finally, the research environment should be related to academia and/or industry by the sharing of knowledge, techniques, methods, or software. The researcher forensic linguist is part of a larger community of computational linguists, psycholinguists, corpus linguists, theoretical linguists, and computer scientists where forensic applications are just one application of common techniques, methods, and software put together in novel ways. For instance, text classification techniques were originally designed as part of summarization schemes but later became useful for finding plagiarism and duplicates within large electronic collections, just as DNA testing was originally used for paternity before it was applied forensically.

Forensic author identification methods should relate, in some recognizable way, to a theory of language, since the method is seeking to identify authorship based on language (rather than handwriting, ink, or IP address). Linguistics obviously offers the fullest theories of language, with the generative theory being the best developed. The generative theory of language includes

Chomsky's original transformational-generative grammar,<sup>20</sup> now known as Minimalism,<sup>21</sup> as well as its offshoots such as Lexical-Functional Grammar;<sup>22</sup> Generalized Phrase Structure Grammar;<sup>23</sup> Head Driven Phrase Structure Grammar;<sup>24</sup> and Construction Grammar.<sup>25</sup> What has been especially impressive about the generative theory of language is its ability to make predictions about linguistic structure, linguistic functions, and the psychological reality of linguistic structure. Other theories, such as Tagmemics<sup>26</sup> or Systemic Functional Grammar,<sup>27</sup> have remained primarily descriptive or taxonomic rather than predictive.

Prescriptive grammar—or school grammar—is taught in schools to indoctrinate students with the prestige or most socially desirable dialect and especially how to “use words correctly.” It teaches how a native speaker *should* speak rather than how a native speaker *actually* speaks. Prescriptive grammar is neither descriptive nor predictive, as it is not a scientific theory of language but is the standard approach to language for literary analysis and for anyone who has not studied linguistics. Prescriptive grammar is attractive to judges who typically write and speak a prestige dialect congruent with prescriptive grammar. However, research has demonstrated that prescriptive grammar is not an adequate theory of language for authorship identification.<sup>28</sup>

---

<sup>20</sup> See NOAM CHOMSKY, *ASPECTS OF THE THEORY OF SYNTAX* (1965).

<sup>21</sup> See NOAM CHOMSKY, *THE MINIMALIST PROGRAM* (1995).

<sup>22</sup> See JOAN BRESNAN, *LEXICAL FUNCTIONAL SYNTAX* (2001).

<sup>23</sup> See GERALD GAZDAR ET AL., *GENERALIZED PHRASE STRUCTURE GRAMMAR* (1985).

<sup>24</sup> See CARL POLLARD & IVAN A. SAG, *HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR* (1994).

<sup>25</sup> See THOMAS HOFFMAN & GRAEME TROUSDALE, *THE OXFORD HANDBOOK OF CONSTRUCTION GRAMMAR* (2013).

<sup>26</sup> See KENNETH L. PIKE, *LANGUAGE IN RELATION TO A UNIFIED THEORY OF THE STRUCTURE OF HUMAN BEHAVIOR* (1967); KENNETH L. PIKE, *LINGUISTIC CONCEPTS: AN INTRODUCTION TO TAGMEMICS* (1982).

<sup>27</sup> M.A.K. HALLIDAY & CHRISTIAN M.I.M. MATTHIENSEN, *AN INTRODUCTION TO FUNCTIONAL GRAMMAR* (3d ed. 2004).

<sup>28</sup> See Michael Brennan & Rachel Greenstadt, *Practical Attacks Against Authorship Recognition Techniques*, PROC. TWENTY-FIRST CONF. ON

Differences in academic training make the paradigm of experimental validation testing for forensic authorship identification more or less difficult to accept. Training in literary criticism does not focus on empirical methods, while pure computer science can bypass courses in experimental design. However, in most branches of linguistics, empirical work is mandatory. Psycholinguists design and run experiments testing the theoretical constructs posited by linguistic theory (usually from a generative theory), focusing on the cognition and memory required to produce and process human language. The validation testing described earlier is second nature to someone trained in psycholinguistics (including child language acquisition, psychology of literacy, and second language acquisition).

Even if the forensic linguist relates to the small community of sociolinguists, the methods that the forensic linguist develops should be recognizable as sociolinguistics. Historically, sociolinguistics introduced a quantitative approach to midcentury American linguistics and relied heavily on empirical data collection, phonetic measurements, and experimental research designs.<sup>29</sup> Therefore, when a forensic linguist asserts that his academic training is in sociolinguistics, but his method is neither quantitative, nor tested on ground-truth data, nor validated by experiments, the disconnect between the forensic activity and the academic world is startling to linguists, if invisible to attorneys or judges.

### III. THE FORENSIC COMPUTATIONAL LINGUISTICS APPROACH TO AUTHOR IDENTIFICATION

Work in forensic computational linguistics began in the mid-1990s, with funding from the National Institute of Justice.<sup>30</sup> By

---

INNOVATIVE APPLICATIONS ARTIFICIAL INTELLIGENCE (IAAI), 2009, at 60; Chaski, *Empirical Evaluations*, *supra* note 16; Koppel & Schler, *supra* note 3.

<sup>29</sup> Labov is considered the originator of sociolinguistics; his work is characterized by quantitative, statistical analysis of naturally collected or elicited linguistic behavior. *See generally* WILLIAM LABOV, *THE SOCIAL STRATIFICATION OF ENGLISH IN NEW YORK CITY* (2d ed. 2006).

<sup>30</sup> In 1995, I received a grant to validate linguistic methods for

the late 1990s, I had developed a method now known as SynAID (Syntactic Author Identification) within ALIAS (Automated Linguistic Identification and Assessment System).<sup>31</sup> This research has played a role in adjudicated cases in 1998, 2001, and 2008, discussed later.

Litigation-independent validation testing on forensically feasible ground-truth data is a core feature of the forensic computational linguistics approach. Implementation in software that is responsive to messy data is central to the forensic computational linguistics approach for both replicability and error control. Linguistic theory plays a central role in the forensic computational linguistics approach. These features distinguish the forensic computational linguistics approach in sometimes obvious, sometimes subtle ways from forensic stylistics and stylometric computing.

#### *A. Linguistic Theory Does Matter*

In linguistic theory, language is divided into levels for analytical purposes.<sup>32</sup> These levels are sound, word, and word combinations.<sup>33</sup> These levels, respectively, are analyzed in phonetics and phonology; morphology and the lexicon; syntax; semantics and pragmatics; and prosody.<sup>34</sup> These levels have

---

determining authorship, Grant ID 1995-IJ-CX-0012, Visiting Fellowship, Linguistics Methods for Determining Authorship.

<sup>31</sup> See Chaski, *Empirical Evaluations*, *supra* note 16; Chaski, *Empirically Testing*, *supra* note 16; Carole E. Chaski, Recent Validation Results for the Syntactic Analysis Method for Author Identification, International Conference on Language and Law (2004) [hereinafter Chaski, *Syntactic Analysis Method Identification*]; Chaski, *Who Wrote It?*, *supra* note 1; Chaski, *Who's at the Keyboard?*, *supra* note 16.

<sup>32</sup> This division of language into analytical levels is commonplace in standard textbooks in linguistics. See *e.g.*, RICHARD AKMAJIAN ET AL., *LINGUISTICS: AN INTRODUCTION TO LANGUAGE AND COMMUNICATION* (6th ed. 2001); EDWARD FINEGAN, *LANGUAGE: ITS STRUCTURE AND USE* (6th ed. 2012); VICTORIA FROMKIN ET AL., *AN INTRODUCTION TO LANGUAGE* (10th ed. 2013).

<sup>33</sup> See AKMAJIAN ET AL., *supra* note 32; FINEGAN, *supra* note 32; FROMKIN ET AL., *supra* note 32.

<sup>34</sup> See AKMAJIAN ET AL., *supra* note 32; FINEGAN, *supra* note 32;

different salience or prominence in processing and especially imitation of language. For instance, children acquire sounds and prosody before they acquire words.<sup>35</sup> Syntactic form—or the actual ordering and combination of words—is least salient and consequently least easy to imitate. There was a great deal of research in psycholinguistics starting in the 1960s, none of which has been refuted, about the way we remember the meaning of a statement while we forget how the statement was actually said.<sup>36</sup> In fact, in normal linguistic processing it appears that loss of syntactic structure occurs within milliseconds,<sup>37</sup> even in writing tasks.<sup>38</sup> Nonetheless, even though we do not remember the word order for long, syntactic structures are very real, albeit fragile and abstract. Again, a great deal of research in psycholinguistics and linguistic theory (starting with Fodor and Bever<sup>39</sup>) demonstrates the reality of syntactic structures, especially the edges of structures, like the beginnings and endings of noun phrases or clauses, because the edges are where most informative morphosyntactic elements appear, and also where the phrasal head—the dominant function—is placed. Therefore, the forensic computational linguistic approach focuses primarily on syntax because syntax would be more difficult to imitate than lexical choices or spelling and punctuation (the graphic correlate of phonetics and prosody).

---

FROMKIN ET AL., *supra* note 32.

<sup>35</sup> S. Katz-Gershon, *Word Extraction in Infant and Adult Directed Speech: Does Dialect Matter?* (2007) (unpublished Ph.D. dissertation, Wayne State Univ.) (on file with author).

<sup>36</sup> Philip N. Johnson-Laird & Rosemary Stevenson, *Memory for Syntax*, 227 *NATURE* 412 (1970) (citing Jacqueline S. Sachs, *Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse*, 2 *PERCEPTION & PSYCHOPHYSICS* 437, 437 (1967)).

<sup>37</sup> *Id.*

<sup>38</sup> See Holly P. Branigan et al., *Syntactic Priming in Written Production: Evidence for Rapid Decay*, 6 *PSYCHONOMIC BULL. & REV.* 635, 635–40 (1999).

<sup>39</sup> Jerry A. Fodor & Thomas G. Bever, *The Psychological Reality of Linguistic Segments*, 4 *J. VERBAL LEARNING & VERBAL BEHAV.* 414, 414–20 (1965).

Another aspect of linguistic theory essential to author identification is the theory of markedness.<sup>40</sup> In many human characteristics, there is an asymmetry in function of symmetrical design. Handedness and footedness are the obvious examples of this asymmetry, but the brain also has this kind of duality.<sup>41</sup> Language is permeated from phonetics through pragmatics with asymmetric oppositions, a fact that was first realized and articulated by the Prague School in the 1940s and then adopted within generative linguistics in phonology<sup>42</sup> and in syntax.<sup>43</sup> Markedness explains why some noun phrase structures are harder to process, produce, or find in high frequency while other nouns phrase structures are a dime a dozen, even in child language.<sup>44</sup> A noun phrase “*the tippy cup with your name on it that we found under the car seat yesterday*” is marked; the noun phrase “*your tippy cup*” is unmarked. Marked noun structures occur later in language acquisition and even in adult language are less frequent than unmarked noun structures.

In phonetics, normalization is the process of speaker recognition by which we come to recognize specific phonetic features in an individual’s voice—features that are consistent with the person but also different from someone else.<sup>45</sup> If recognition is possible at the phonetic level—and everyone has had the experience of recognizing a person by voice over the telephone—it is a testable hypothesis that a similar

---

<sup>40</sup> For an overview of markedness theory in linguistics, see generally EDWIN L. BATTISTELLA, *MARKEDNESS: THE EVALUATIVE STRUCTURE OF LANGUAGE* (1990).

<sup>41</sup> Kenneth Hugdahl, *Symmetry and Asymmetry in the Human Brain*, 13 *EUR. REV.* 119, 119–33 (2005).

<sup>42</sup> See NOAM CHOMSKY & MORISS HALLE, *THE SOUND PATTERN OF ENGLISH* (1968).

<sup>43</sup> Judith Aissen, *Markedness and Subject Choice in Optimality Theory*, 17 *NAT. LANGUAGE & LINGUISTIC THEORY* 673, 673–711 (1999); see also GERALD GAZDAR ET AL., *GENERALIZED PHRASE STRUCTURE GRAMMAR* (1985); CARL POLLARD & IAN A. SAG, *HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR* (1994).

<sup>44</sup> See BATTISTELLA, *supra* note 40.

<sup>45</sup> For an overview of speaker recognition, see Homayoon Beigi, *Speaker Recognition*, in *BIOMETRICS* 3, 3–29 (2011), available at <http://www.intechopen.com/books/biometrics/speaker-recognition>.



recognizability would be possible at the syntactic level. The issue is to find, again borrowing from phonetics, some invariant signal among the variation and noise (in an information theoretic sense).<sup>46</sup> Or borrowing from statistical terminology, what syntactic patterns would be distinctive enough among the potential note writers to differentiate intrawriter variation from interwriter variation?

Language is a conventional behavior where for the sake of mutual understanding we share the same code. In information theoretic terms, each of us is both sender and receiver. This is how we manage to finish each other's sentences: we are using the same code we share with another person in our linguistic circle. So the notion that individual language is unique, or that each of us has a unique linguistic behavior, is an idea that linguistics as a discipline denies by the very definition of language as a conventional behavior and shared code.

Even though linguistic behavior cannot be literally unique, it can and does show variation. By definition, dialect is the name for group-level linguistic behavior, where subgroups within the language can be determined. At the individual level, linguistics has posited the notion of idiolect, or a variation of language at the individual level.<sup>47</sup> Clearly, idiolect cannot be a unique language, or, again, the unique language would have a speaker of one, but variations at the individual level might still be discoverable. Idiolect was first posited at the phonetics level. The biological substrate of phonetic articulation certainly makes phonetic individual differences feasible.<sup>48</sup> Idiolect later became a useful theoretical term in recognizing syntactic variation between syntacticians. There is still no empirical method for demonstrating that each person has his or her own idiolectal variation that is uniquely identifiable, but author identification merely has to recognize intrawriter vs. interwriter variation

---

<sup>46</sup> Cf. CLAUDE E. SHANNON & WARREN WEAVER, *THE MATHEMATICAL THEORY OF COMMUNICATION* (1971).

<sup>47</sup> See, e.g., FROMKIN ET AL., *supra* note 32.

<sup>48</sup> See *id.*

strong enough to differentiate authors from each other and cluster documents by author.<sup>49</sup>

Finally, due to the brevity of the texts, a realistic forensic author identification method needs a way of measuring the texts to get as much information as possible out of them. Counting syntactic structure rather than words yields a higher count and makes statistical analysis possible. If a method only counts the words, the result is a long list of words with frequencies that are mostly one, and a few function words like [*the, a, of, with*] with slightly higher frequencies. But if the syntactic structures are counted, all the nouns in a sentence contribute to the noun category, all the determiners to the determiner category, and so forth. Likewise, by subcategorizing the noun phrases into marked and unmarked types, the frequency counts are divided into two separate measures for the marked and unmarked frequency of each syntactic category. The marked and unmarked subcategorization is a way to compare different authors' patterns of use for what is salient on the one hand (as marked patterns are salient by definition) but hard to imitate on the other (as syntactic structures are fragile in memory).

### *B. Ground-Truth Data*

The Chaski Writing Sample Database includes ten topics, listed in Table 1. The database makes cross-genre/register comparison possible for known authors who are not professional writers and produce unedited texts. With funding from the U.S. Department of Justice's National Institute of Justice, data were collected from students at a community college and a four-year college with a student body of both traditional students and returning adult students; the population provided a wide age range, males and females, and several races; Table 2 shows the demographics of an experiment that contrasted gender and controlled for race because race is highly correlated with some American English dialects.

---

<sup>49</sup> Chaski, *Who Wrote It?*, *supra* note 1, at 17.

Task ID	Topic
1.	Describe a traumatic or terrifying event in your life and how you overcame it.
2.	Describe someone or some people who have influenced you.
3.	What are your career goals and why?
4.	What makes you really angry?
5.	A letter of apology to your best friend
6.	A letter to your sweetheart expressing your feelings
7.	A letter to your insurance company
8.	A letter of complaint about a product or service
9.	A threatening letter to someone you know who has hurt you
10.	A threatening letter to a public official (president, governor, senator, councilman or celebrity)

Table 1: Topics in the Chaski Writing Sample Database

### C. Examples of Experimental Validation Testing

With forensically feasible ground-truth data on which to run experiments testing author identification methods, ten authors were selected from the Chaski Writing Sample Database, as shown in Table 2. Each author is represented in about 100 sentences and/or 2,000 words. This was a good starting point to consider how low we could go in terms of data requirements, far less than the literary methods use, and a number that can usually be obtained in real cases. Given ten authors, there were forty-five pairwise tests of each author paired with each other author ( $10 * 9 / 2 = 45$ ). At the time these experiments were run, most author identification tests were being run on two to four authors.<sup>50</sup> Some of the experiments reported here were first reported in my previous works.<sup>51</sup>

---

<sup>50</sup> Cf. O. de Vel et al., *Mining E-mail Content For Author Identification Forensics*, 30 ACM SIGMOD RECORD 55, 55–64 (2001); Efstathios Stamatatos et al., *Automatic Text Categorization in Terms of Genre and Author*, 26 COMPUTATIONAL LINGUISTICS 471, 471–95 (2000); Efstathios Stamatatos et al., *Computer-Based Authorship Attribution Without Lexical Measures*, 35 COMPUTERS & HUMAN. 193, 193–214 (2001) [hereinafter

Race, Sex	Topics by Task ID	Author ID Number	Number of Texts	Number of Sentences	Number of Words	Average Text Size (Min, Max)
WF	1 - 4, 7, 8	16	6	107	2,706	430 (344, 557)
WF	1 - 5	23	5	134	2,175	435 (367, 500)
WF	1 - 10	80	10	118	1,959	195 (90, 323)
WF	1 - 10	96	10	108	1,928	192 (99, 258)
WF	1 - 3, 10	98	4	103	2,176	543 (450, 608)
<b>WF Total</b>			<b>35</b>	<b>570</b>	<b>10,944</b>	
WM	1 - 8	90	8	106	1,690	211 (168, 331)
WM	1 - 6	91	6	108	1,798	299 (196, 331)
WM	1 - 7	97	6	114	1,487	248 (219, 341)
WM	1 - 7	99	7	105	2,079	297 (151, 433)
WM	1 - 7	168	7	108	1,958	278 (248, 320)
<b>WM Total</b>			<b>34</b>	<b>541</b>	<b>9,012</b>	
<b>Grand Total</b>			<b>69</b>	<b>1,111</b>	<b>19,956</b>	

Table 2: Authors and Texts

Stamatatos et al., *Computer-Based Authorship Attribution*].

<sup>51</sup> Chaski, *Syntactic Analysis Method Identification*, *supra* note 31, at 3–4; Chaski, *Who's at the Keyboard?*, *supra* note 16, at 3–11.

By the time these experiments were run, empirical work with a professional statistician had shown that linear discriminant function analysis (“LDFA”) was the best statistical procedure to use for classifying an unknown document based on quantitative comparisons of two sets of known documents. LDFA is used to generate a linear function which maximizes the difference between groups; the coefficients of this function can then be used to predict the group membership of new or holdout cases.<sup>52</sup> In these experiments, SPSS version 13 (“Statistical Package for the Social Sciences”) was used to run LDFA.

SPSS allows the user to select several variations on LDFA. The variables can be entered all together or stepwise. If the stepwise option is chosen, the user can select the number for entry or removal or use either of the defaults. The options include Wilks’ lambda, F ratio, and the Mahalanobis distance. The user can also request cross-validation using a leave-one-out process. Cross-validation shows how reliable the linear function determined by the original group members is when each member is left out of the group. SPSS also allows the user to select whether prior probabilities are computed from the group sizes or not. The specific options which were chosen for each variable set are described in the experiments, as these options provide, along with different linguistic features, a series of possible experiments to run.

*Experiment 1: Syntactically Classified  
Edge Punctuation Alone*

In this experiment, only the three variables relating to syntactically classified punctuation were used. The LDFA was run with all variables entering together, prior probabilities not computed from group size, and cross-validated using leave-one-out and Wilks’ lambda. Table 3 shows the cross-validation scores for each author-pair. The final row shows the average for each author. The grand average over all ten authors is 79.8% accuracy.

---

<sup>52</sup> SPSS, SPSS 13.0 BASE USER’S GUIDE (2004).

<b>Author</b>	<b>16</b>	<b>23</b>	<b>80</b>	<b>90</b>	<b>91</b>	<b>96</b>	<b>97</b>	<b>98</b>	<b>99</b>	<b>168</b>
16	X	100	94	93	92	94	92	80	93	93
23	100	X	93	93	91	93	91	67	83	92
80	94	93	X	72	75	65	81	86	71	65
90	93	93	72	X	64	66	86	75	80	47
91	92	91	75	64	X	50	58	90	54	62
96	94	93	65	66	50	X	75	86	70	77
97	92	91	81	86	58	75	X	80	85	85
98	80	67	86	75	90	86	80	X	82	91
99	93	83	71	80	54	70	85	82	X	86
168	93	92	65	47	62	77	85	91	86	X
<b>Author Average</b>	<b>92</b>	<b>89</b>	<b>78</b>	<b>75</b>	<b>71</b>	<b>75</b>	<b>81</b>	<b>81</b>	<b>78</b>	<b>78</b>

*Table 3: Cross-Validation Accuracy Scores for Three Edge-Punctuation Variables*

### *Experiment 2: Modifying the LDFA*

By running the LDFA in forward stepwise mode, using Mahalanobis distance and setting F to enter at 1.84 and F to remove at 0.71 (SPSS defaults), the accuracy scores improve, over all ten authors, to 85.9%, as shown in Table 4. In Pair 91/96, none of the variables met the F levels for entering and so no analysis was run (noted as “nqv” in the table, for “no qualifying variables”). In the average for this author-pair, the sums are divided by 8 for the eight comparisons that were possible (rather than 9).

Author	16	23	80	90	91	96	97	98	99	168
16	X	100	94	100	100	100	100	70	93	100
23	100	X	87	92	91	93	91	78	83	92
80	94	87	X	83	86	70	81	86	77	71
90	100	92	83	X	64	78	93	100	80	53
91	100	91	86	64	X	nvq	83	90	69	69
96	100	93	70	78	nvq	X	75	100	82	71
97	100	91	81	93	83	75	X	100	85	92
98	70	78	86	100	90	100	100	X	91	91
99	93	83	77	80	69	82	85	91	X	86
168	100	92	71	53	69	71	92	91	86	X
<b>Author Average</b>	<b>95</b>	<b>90</b>	<b>82</b>	<b>83</b>	<b>82</b>	<b>84</b>	<b>89</b>	<b>90</b>	<b>83</b>	<b>81</b>

Table 4: Cross-Validation Accuracy Scores for Three Edge-Punctuation Variables (Stepwise)

Even though these three edge-punctuation variables result in an accuracy score not far below the contemporaneous results from Stamatatos et al.,<sup>53</sup> Baayen et al.,<sup>54</sup> and Tambouratzis et al.,<sup>55</sup> Tables 3 and 4 also show that edge punctuation may be a very good discriminator for some authors, such as 16 and 23, but a rather poor discriminator for other authors, such as 91. Further, particular author pairs are very discriminable (such as 16/23, 91/98, 168/98) while other author pairs are hardly distinguishable (such as 90/168 and 91/96), and the function is classifying near or below chance level.

<sup>53</sup> See Stamatatos et al., *Computer-Based Authorship Attribution*, *supra* note 50, at 207.

<sup>54</sup> See Harald Baayen et al., *An Experiment in Authorship Attribution*, JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES, 2002, at 4.

<sup>55</sup> See George Tambouratzis et al., *Discriminating the Registers and Styles in the Modern Greek Language—Part 2: Extending the Feature Vector to Optimize Author Discrimination*, 19 LITERARY & LINGUISTIC COMPUTING 221 (2004).

*Experiment 3: Adding Markedness to  
Syntactically Classified Edge Punctuation*

In this experiment, the syntactically classified punctuation variables were combined with the marked and unmarked phrases. Given earlier results, the LDFA was run stepwise, using Mahalanobis distance, and the SPSS default settings for F to enter (at 3.84) and F to remove (at 2.71) were used. The cross-validation accuracy scores are shown in Table 5.

<b>Author</b>	<b>16</b>	<b>23</b>	<b>80</b>	<b>90</b>	<b>91</b>	<b>96</b>	<b>97</b>	<b>98</b>	<b>99</b>	<b>168</b>
16	X	100	100	100	100	100	100	70	100	100
23	100	X	100	100	100	100	100	89	83	92
80	100	100	X	83	nvq	70	81	100	77	82
90	100	100	83	X	71	78	100	100	87	87
91	100	100	nvq	71	X	81	92	100	nvq	nvq
96	100	100	70	78	81	X	75	100	85	100
97	100	100	81	100	92	75	X	100	85	100
98	70	89	100	100	100	100	100	X	91	100
99	100	83	77	87	nvq	82	85	91	X	93
168	100	92	82	87	nvq	94	100	100	93	X
<b>Author Average</b>	<b>97</b>	<b>96</b>	<b>85</b>	<b>88</b>	<b>89</b>	<b>85</b>	<b>92</b>	<b>98</b>	<b>85</b>	<b>93</b>

*Table 5: Cross-Validation Accuracy Scores for Markedness & Punctuation Variables*

Table 5 shows that the overall accuracy rate at 90.6% with the range from 85% to 98%. Note also that for three author pairs, these variables at these default settings for the stepwise procedure did not qualify for the analysis so that no analysis was done (noted as “nvq” in the table).

*Experiment 4: Syntactically Classified Edge  
Punctuation, Markedness, and Word Length*

In this experiment, the variable set included syntactically classified punctuation, phrase markedness and average word length. The LDFA was run stepwise, using Mahalanobis distance and the default settings for F to enter and F to remove.



Only one author pair had no variables qualify for the analysis under these settings.

Author	16	23	80	90	91	96	97	98	99	168
16	X	100	100	100	100	100	100	80	100	100
23	100	X	100	100	100	100	100	89	92	100
80	100	100	X	94	100	70	100	100	82	100
90	100	100	94	X	71	94	100	100	87	80
91	100	100	100	71	X	100	92	100	nvq	100
96	100	100	70	94	100	X	88	100	88	100
97	100	100	100	100	92	88	X	100	100	100
98	80	89	100	100	100	100	100	X	91	100
99	100	92	82	87	nvq	88	100	91	X	93
168	100	100	100	80	100	100	100	100	93	X
<b>Author Average</b>	<b>97</b>	<b>98</b>	<b>94</b>	<b>92</b>	<b>95</b>	<b>93</b>	<b>98</b>	<b>96</b>	<b>92</b>	<b>97</b>

*Table 6: Cross-Validation Accuracy Scores for Markedness, Edge Punctuation, and Average Word Length Variables*

Table 6 shows that the addition of word length in the variable set improves the overall accuracy rate to 95%, with individual authors' accuracy rates ranging from 92% to 98%. Note also that only one author pair was not analyzed due to "no qualifying variables" (or "nvq").

The kind of serial experimentation presented here empirically establishes a protocol, independent of any litigation, with data requirements, and known error rates that can be used in casework. One such protocol is presented below.

#### *D. Syntactic Method Protocol using SynAID*

0. Receive Q document and K documents of at least two suspects (the known authors), with approximately 100 sentences and/or approximately 2,000 words for each suspect.

1. Input Q and K documents in txt, rtf, Word format into ALIAS Documents Database.

2. Run the SynAID modules on all documents: Sentence Splitter, Tokenizer, Part-of-Speech Tagger.

3. Manually check each sentence and tag for accuracy.

4. Run the SynAID module: Markedness Subcategorizer.
5. Run the SynAID module: Punctuation Syntactic Edges Categorizer.
6. Manually check punctuation syntactic edges for accuracy.
7. Run SynAID's calculation of syntactic and punctuation quantification and average word length, for each text, normalizing so that texts of different sizes can be compared, and output the ALIAS Quantification vector for each text.
8. Input ALIAS Quantification output into SPSS (or DTReg or Weka or R)
9. If there are a large number (50+) of K documents or multiple Q documents, run K-means clustering for internal consistency testing. If K-means clustering of K documents shows maximal subsetting, split K if needed. If K-means clustering of Q documents shows minimum subsetting, group Q.
10. Run Linear Discriminant Function Analysis on pairwise K authors, with Q held out, using leave-one-out cross-validation and equal prior probability (not set to number of documents); use SPSS default options.
11. Check classification table.  
If the DFA returns high accuracy for differentiating K1 and K2,  
then report classification of Q and determinative features.  
If the DFA returns low accuracy for differentiating K1 and K2,  
then stop. Do not use low accuracy model for classifying Q.  
High accuracy is no lower than around 80% and is usually in the 90s.  
Average accuracy declines for multiple authors (3 or 4) than for author pairs.
12. Check documents for nonnative English or dialectal patterns and report.

### *E. Admissibility*

Methods in the forensic computational approach to author identification have been admitted as testimony in three trials,

discussed below, with unpublished rulings of the admissibility hearings. In each of these trials, testimony based on the method was admitted without any restrictions: the expert was allowed to state a conclusion about authorship. Since these three cases did not involve any opposing experts, a fourth case involving an opposition expert that settled before trial is also discussed.

In 1998, *Erdman v. Osborne* and *Zarolia v. Osborne/Buffalo Environmental Corp.* were heard in the Circuit Court for Anne Arundel County Maryland.<sup>56</sup> A *Frye* hearing (a.k.a. *Frye-Reed* in Maryland<sup>57</sup>) was conducted, and I was examined by the attorneys and judge outside the presence of the jury. Testimony included the investigative and experimental nature of the syntactic method (“SynAID”) in 1998, that the method was still being tested on a ground-truth database, and that there were current limitations still being experimentally tested. The method itself was described in detail and shown to follow standard analytical methods in linguistics and computational linguistics, as well as a common statistical procedure that was a standard technique in author identification at the time.

The court ruled that both my syntactic method for authorship identification and my analysis of second language interference were admissible without restrictions. In this case, the anonymous document could only have been written by a person in a small pool of suspects, five engineers. Writing samples from each one were analyzed using the syntactic method, and statistically, only one possible author was not differentiated from the questioned document. Also, the questioned document contained a typical first-language interference in English as a second language, i.e., the nonnative use of determiners such as [a, the]. Since many languages do not have the determiner grammatical category, using determiners such as [a, the] in the appropriate semantic places is difficult to do for nonnative speakers of English. It

---

<sup>56</sup> *Zarolia v. Buffalo Envtl.*, No. 1854 (Md. Ct. Spec. App. 1998); *Erdman v. Osborne*, No. 02C95025473 (Md. Cir. Ct. 1998), *appeal denied*, 729 A.2d 405 (Md. 1999).

<sup>57</sup> *See Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (establishing the general acceptance test used to determine the admissibility of scientific evidence); *Reed v. State*, 391 A.2d 364, 391 (Md. 1978) (adopting the test for admissibility established in *Frye*); *see also* MD. R. 5-702.

turned out that the lone engineer not statistically differentiated from the questioned document was a nonnative speaker of English, a native speaker of Gujarati, a language that does not have determiners.

In 2001, the United States District Court of the District of Columbia heard *Greene v. Dalton*.<sup>58</sup> Judge Henry Kennedy presided over a *Daubert* hearing. Testimony included reportage of validation testing results on ground-truth data, including the error rate, data requirements, and empirical standards for conducting a syntactic markedness analysis for authorship identification. Again, the SynAID method was described in detail and related to standard techniques in linguistics. I was permitted to testify about the authorship of a diary, without any restrictions on my ability to state a conclusion. The court admitted my syntactic method using SynAID without restrictions. Though the case was appealed, the diary evidence was not at issue.<sup>59</sup>

In 2008, the Fulton County Superior Court in Atlanta, Georgia heard *Arsenault-Gibson v. Dixon*.<sup>60</sup> Georgia follows the *Daubert* standard.<sup>61</sup> Opposing counsel filed a motion in limine regarding my syntactic method of authorship identification, so a *Daubert* hearing was conducted outside the presence of the jury. Testimony included a description of the method, error rate based on validation testing on ground-truth data outside of any litigation, and data requirements. The court rejected the motion in limine and ruled that testimony using SynAID about the authorship was admissible without restrictions.

Also in 2008, *Best Western International v. Doe*<sup>62</sup> was scheduled for hearing in the U.S. District Court for the District

---

<sup>58</sup> See *Greene v. Dalton*, No. CIV.A.96-2161 TPJ, 1997 WL 33475236 (D.D.C. Oct. 3, 1997), *aff'd in part, rev'd in part*, 164 F.3d 671 (D.C. Cir. 1999).

<sup>59</sup> See *Greene v. Dalton*, 164 F.3d 671, 674 (D.C. Cir. 1999).

<sup>60</sup> *Arsenault-Gibson v. Dixon*, No. 2004CV87715 (Ga. Super. Ct. 2008).

<sup>61</sup> A *Daubert* hearing is an evaluation by a trial judge on the admissibility of scientific evidence using the factors set forth in *Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579 (1993). See GA. CODE ANN. § 24-9-67.1 (2013).

<sup>62</sup> See *Best Western Int'l, Inc. v. Doe*, No. CV-06-1537-PHX-DGC, 2008 WL 4630313 (D. Ariz. Oct. 20, 2008).

of Arizona. DLA Piper, representing Best Western International, filed a motion in limine regarding my syntactic method of authorship identification. Before a *Daubert* hearing was conducted, the experts, myself for the defendants, and Robert Leonard for the plaintiff, were deposed. The main issue was the authorship of posts on a discussion board of Best Western International franchisees, including John Doe; there were over 100 questioned posts.

My deposition testimony included a detailed discussion of the method itself, how it relates to standard methods in linguistics and computational linguistics, and the error rate and data requirements from litigation-independent testing, including the use of computational linguistics outside of litigation in Internet search engines and text classification. Regarding the particular case analysis, deposition testimony included internal consistency testing results from the known authors and document classification based on known author statistical models, including one known author with two substyles from internal consistency testing. My conclusions included both litigation-independent error rate (five percent) and the particular error rates associated with each statistical model for a total case-document classification error rate, as well as evidence of native language interference from one known author whose native language, Polish, has a kind of prepositional ambiguity which causes a particular linguistic interference in English. Finally, the deposition testimony included a review of academic credentials, publications, conference presentations, and previous testimony and sworn reports.

In contrast to my deposition, Leonard's deposition testimony began with the fact that neither he nor his colleagues Roger Shuy and Benji Wald had conducted any analysis of the data; instead, he testified that my method had never been heard of and could not be understood by the three linguists Leonard, Shuy, and Wald, regardless of my publications.<sup>63</sup> In his deposition,

---

<sup>63</sup> See, e.g., Chaski, *Empirical Evaluations*, *supra* note 16; Chaski, *Empirically Testing*, *supra* note 16; Chaski, *Syntactic Analysis Method Identification*, *supra* note 31; Chaski, *Who Wrote It?*, *supra* note 1, at 15; Chaski, *Who's at the Keyboard?*, *supra* note 16, at 1.

Leonard described my method as only a “program” that “processes text” in a way that is not transparent because he was not able to find in the text features such as marked prepositional phrases or unmarked adjective phrases. (In fact, abstract syntactic structures are not found in the text itself but in the syntactic analysis of the text.) Leonard argued that in my method I do not analyze text as a linguist but just run a program. Curiously, when Leonard described his own method, which he called sociolinguistics, he testified that he also uses computer software written by someone else to create a concordance or word list. Further, to set his own method apart from other linguists, Leonard testified that his sociolinguistics method was not forensic stylistics, even though he concurrently mentioned that he used twelve of thirteen categories listed as potential stylemarkers in the primary texts on forensic stylistics.<sup>64</sup> When asked about the use of his sociolinguistic method outside of any litigation, Leonard testified that it could be used as the basis for scripts for movies and television shows.

After the depositions, and due to severe restrictions by the judge on what could be presented, DLA Piper withdrew its motion in limine to exclude my testimony and SynAID method. The court issued a summary judgment, which agreed with ninety-five percent of my report; the disagreement regarded documents I had not tested.<sup>65</sup> I was scheduled on a may-call list to testify, but the case settled with John Doe receiving \$2 million and no gag order, an important feature to John Doe and the reason why this settlement can be reported here.

#### IV. FORENSIC STYLISTICS APPROACH TO AUTHOR IDENTIFICATION

Forensic Stylistics is a method derived from handwriting identification, as mentioned by McMEnamin<sup>66</sup> who quotes the

---

<sup>64</sup> MCMENAMIN, *supra* note 2; GERALD R. MCMENAMIN, FORENSIC LINGUISTICS: ADVANCES IN FORENSIC STYLISTICS (2002) [hereinafter MCMENAMIN, ADVANCES].

<sup>65</sup> *Best Western*, 2006 WL 2091695.

<sup>66</sup> MCMENAMIN, *supra* note 2.

standard texts of traditional handwriting identification.<sup>67</sup> Among the methods tested and reported in prior work was forensic stylistics as described in McMEnamin.<sup>68</sup> McMEnamin's is the only text that describes the method and the categories of "stylemarkers," which are claimed to identify each person's unique writing style. As actually practiced in the reports by Professors McMEnamin, Shuy, Leonard, Coulthard, Grant, and a few other nonlinguists I have reviewed, the method consists of two steps:

1. Select stylemarkers by reading the questioned ("Q") and known ("K") documents;
2. Decide the authorship of the questioned document(s) based on the stylemarkers by listing similarities and/or differences and deciding which similarities and which differences are important or not.

The method offers:

- i. no protocol for the order of reading Q or K first, or back and forth between Q and K,
- ii. no protocol for internal consistency testing of K or Q documents, so that any number of Q documents can be put together, in violation of a standard forensic science principle of noncontamination;
- iii. no protocol for determining the importance or "significance" of stylemarkers,
- iv. no use of statistical analysis (in actual case reports); and
- v. no standard reference set of stylemarkers to be reviewed in each case.

---

<sup>67</sup> *Id.* at 113–20 (reviewing the use of linguistic features by handwriting examiners in ALBERT S. OSBORN (1910)); *see, e.g.*, JAMES V. P. CONWAY, EVIDENTIAL DOCUMENTS (1959); WILSON R. HARRISON, SUSPECT DOCUMENTS: THEIR SCIENTIFIC EXAMINATION (1958); ORDWAY HILTON, SCIENTIFIC EXAMINATION OF QUESTIONED DOCUMENTS (rev ed. 1982); ALBERT S. OSBORN, THE PROBLEM OF PROOF (1926); ALBERT S. OSBORN, QUESTIONED DOCUMENTS (2d ed. 1929); *see also* MCMENAMIN, ADVANCES, *supra* note 64, at 81–82 (attempting to distinguish the two fields of questioned document examination and forensic stylistics).

<sup>68</sup> *See* MCMENAMIN, *supra* note 2.

Number (v) is especially important because it means that the method allows the examiner to pick and choose stylemarkers without any predictability. This fundamental methodological flaw enables a host of problems, all rooted in subjectivity. On the one hand, it is essentially impossible to replicate a forensic stylistics analysis, while on the other hand, it is always possible to find an alternative analysis and opposing conclusion. This is the dilemma of any “pick and choose” method.

#### *A. Litigation Dependence*

Finegan documented a case in which five linguists were hired to conduct an authorship identification.<sup>69</sup> The five linguists each offered an opinion; each opinion used forensic stylistics to support the side which hired them. This is possible because each linguist picked stylemarkers, and each stylemarker could be deemed important or not by the linguist without any standard reference set. Finegan’s report of this case demonstrates that forensic stylistics suffers from a classic case of confirmation bias being built in to a method without litigation-independent validation testing.<sup>70</sup>

Without litigation-independent testing, the expert battles inside litigation are inevitable. Finegan predicted that this battle of the experts would occur and that it may be a good thing:

The expectation of expert rebuttal witnesses should contribute significantly to improvements in the quality of linguistic opinion available within the judicial system—and to justice.<sup>71</sup>

I would suggest that a better practice is litigation-independent validation testing, a controversial stance within the forensic stylistics community. In a recent recorded interview prior to deposition, Professor Leonard stated that he had “misgivings” about testing the method.<sup>72</sup>

---

<sup>69</sup> See Edward Finegan, *Variation in Linguists’ Analyses of Author Identification*, 65 AM. SPEECH 334 (1990).

<sup>70</sup> *Id.* at 339.

<sup>71</sup> *Id.* at 338.

<sup>72</sup> Interview with Dr. Robert Leonard (Feb. 22, 2011).



*B. Validation Testing*

Until my research was funded by NIJ, with subsequent publications,<sup>73</sup> there were no known error rates for the forensic stylistics method, because none of its proponents had ever tested the method on ground-truth data, independent of any litigation, and in a blind experimental method. My prior work reports testing several authorship identification techniques, including the most common stylemarkers of forensic stylistics.<sup>74</sup> My prior work followed a standard blind procedure.<sup>75</sup> A research intern selected four female authors, around the age of forty, from the Chaski Writing Sample Database; these writing samples were typed so that no handwriting could be used to sway the analysis of the linguistic features. The intern selected one of these writing samples as the questioned document and labeled the rest of the writing samples by the numerical identifier of the writers in the database. So, the research question was, which of the four authors authored the questioned document? Each author identification technique was applied to the known writing samples first, and then the questioned document and a statistical test ( $\chi^2$  or t-test) was applied to the analytical results. The actual author of the questioned document was not revealed until all the author identification techniques were tested, and the accuracy rate for each author identification technique was then calculated.

The testing procedure in my prior work added two pieces to standard forensic stylistics: first, the method was controlled by always testing the K before the Q document, and not going back and forth between K and Q; second, a simple statistical test was applied to results.<sup>76</sup> So even with this strengthening of the method (from the viewpoint of scientific procedure), most of the feature categories typically selected in forensic stylistic analyses were not reliable. The actual author of the questioned document

---

<sup>73</sup> See Chaski, *Empirical Evaluations*, *supra* note 16; Chaski, *Who Wrote It?*, *supra* note 1.

<sup>74</sup> Chaski, *Empirical Evaluations*, *supra* note 16, at 3.

<sup>75</sup> *Id.* at 44.

<sup>76</sup> See *id.* at 8.

was repeatedly not selected by a blind testing of stylemarker comparison.

One argument made against my prior work is that the stylemarkers were tested independently and not combined, but it is supposedly the combination of an unknown number of stylemarkers that supports the contention that each person has a unique authorial style.<sup>77</sup> However, anyone reading the test results could combine them, and when combined, the accuracy rate at identifying a questioned document to the real author in a pool of four authors for a combination of forensic stylistics stylemarkers is about fifty-two percent.

Forensic stylistics has very poor accuracy on ground-truth data where no one is preselected as author prior to K/Q feature selection. It is not a reliable method for authorship identification. The poor reliability of forensic stylistics, as reported in my prior article,<sup>78</sup> was later confirmed by validation testing using different ground-truth data by St. Vincent and Hamilton,<sup>79</sup> Koppel and Schler,<sup>80</sup> and Chaski.<sup>81</sup>

### *C. No Relationship to Standard Linguistic Methodology*

Crystal<sup>82</sup> provided a surprisingly caustic but accurate review of McMEnamin.<sup>83</sup>

M[cMenamin] talks in a semistatistical way (“It is extremely unlikely that this close lexical match in profanity could be due to chance coincidence . . .”) but he does not present the statistical analysis which would make such comparisons convincing. Indeed, at several points, one wonders whether it would in principle be

---

<sup>77</sup> *See id.*

<sup>78</sup> *See id.* at 3.

<sup>79</sup> *See* S. St. Vincent & T. Hamilton, *Author Identification with Simple Statistical Methods*, SWARTHMORE COLL., DEP’T OF COMPUTER SCI. (2001) (on file with author).

<sup>80</sup> Koppel & Schler, *supra* note 3.

<sup>81</sup> Chaski, *Empirically Testing*, *supra* note 16.

<sup>82</sup> David Crystal, Book Review, 71 *LANGUAGE* 381, 381–85 (1995) (reviewing MCMENAMIN, *supra* note 2).

<sup>83</sup> MCMENAMIN, *supra* note 2.

possible to do so, given the sample sizes, and the lack of lexical frequency norms. . . .

The conclusion, 'The above findings demonstrate an extraordinary level of stylistic similarity between the questioned diary and the known writings' might in the hands of a good lawyer convince a jury, but it would not be difficult for another good lawyer to question the supposedly 'scientific' basis of the argument. For instance, your honor, what norms are used as the baseline for the judgments? When M says, concerning the use of the percent sign and ampersand, that 'what . . . they have in common is their occasional use. Their use if not frequent or abnormal', or 'parenthesis . . . are used very frequently', or "The semicolon . . . occurs very frequently,' how are we to interpret these remarks? Is this linguistic SCIENCE? . . . .

The problem is, after reading this book, lawyers might be forgiven for thinking that this is an orthodox account of a domain of applied stylistics. It is not. It is an account which has been tailored to meet the traditions and expectations of the legal profession . . . . It may well do a service to jurisprudence; but I am not sure that it does a service to applied linguistics.<sup>84</sup>

I previously described problems with the forensic stylistics method and how misleading it might be to a jury who has no concept of linguistics.<sup>85</sup>

Goutsos also expressed disagreement with McMEnamin's subjective assessment method.<sup>86</sup> In his review of McMEnamin's work for *Forensic Linguistics: The International Journal of Speech, Language and Law*, the journal of the International Association of Forensic Linguists, he shows how McMEnamin's methodology does not follow normal linguistics methodology.

---

<sup>84</sup> Crystal, *supra* note 82, at 383-84.

<sup>85</sup> Chaski, *Who Wrote It?*, *supra* note 1; Carole E. Chaski, *Junk Science, Pre-Science and Developing Science*, NAT'L CONF. ON SCI. & L. PROC., 1999, at 97.

<sup>86</sup> Dionysis Goutsos, *Review Article: Forensic Stylistics*, 2 FORENSIC LINGUISTICS 99 (1995).

McMenamin evaluated as “odd” such spellings as [abit, a lot, anytime]. But when Goutsos used a typical linguistic methodology of checking for frequency in a corpus, in this case the ten-million word corpus of American English, the Bank of English Database, he found such spellings sufficient to comment: “this would imply that careful research must precede any prescriptive judgment.”<sup>87</sup> Indeed.

Certainly, in Professor McMenamin’s defense, his later book includes a chapter in which he does consider statistics that could be used in a forensic linguistics analysis.<sup>88</sup> Further, he does write about a corpus he is developing.<sup>89</sup> But there is still a real gap between the theory put forth in the book and the method and conclusions put forth in Professor McMenamin’s actual analyses and reports, as shown by Nunberg’s peer review.<sup>90</sup>

Nunberg prepared an affidavit in which he stated:

I believe I have a responsibility as a linguist to point out the deficiencies of Dr. McMenamin’s work, which misrepresents the methods of the discipline of linguistics. . . .

1. Professor McMenamin’s methods are not based on well-established theoretical principles nor are they consistent with rigorous practice in the statistical analysis of written texts. McMenamin has performed no statistical research that would give any scientific grounding to his conclusions. I would not classify McMenamin’s work as bad science; rather, it is not science at all.

2. Professor McMenamin’s choice of the features used in document comparison is arbitrary and subjective, and unmotivated by any empirical research; another set of features could well have been chosen that would have given very different results. His method could not pass the test of independent replicability.

---

<sup>87</sup> *Id.* at 105–06.

<sup>88</sup> MCMENAMIN, *ADVANCES*, *supra* note 64.

<sup>89</sup> *Id.*

<sup>90</sup> Statement of Geoffrey Nunberg, *In re* Marriage of Hargett, No. SDR-0017114 (Cal. Super. Ct. 2005) (on file with author).

3. Professor McMenamín's work is not accepted as sound science within the linguistic community. . . .

The process of authorship identification is predicated on the assumption that writers may betray their individuality by certain features, and that if two documents share certain features in common there may be grounds for assuming that they have the same author. Note however that a similarity in features is not by itself a ground for assuming that two documents have the same author. That depends, rather, on how widespread these features are in the population as a whole. . . .

It follows that if we have no information about the statistical frequency of various features of written texts, we can make no scientific assumptions as to whether they provide good evidence of authorship or not. . . .

McMenamin has not troubled to do the work of statistical analysis necessary to teach scientific conclusions about the authorship of documents—neither in his report or in his published writings on the subject. . . .

In the absence of a prior statistical analysis, McMenamín has no scientific basis for distinguishing those features of a document that are likely to be likely cues of authorship, nor does he have any grounds for assuming that the appearance of the same feature . . . in two texts offers significant evidence of common authorship. In effect, he has no way of distinguishing left-handed redheads from right-handed brunettes. Scientifically speaking, McMenamín's analyses are worthless.<sup>91</sup>

These reviews of forensic stylistics from other academically degreed linguists suggest two important points for judges to consider. First, forensic stylistics is not considered standard linguistics by well-established, highly regarded linguists. Second, there is certainly no general acceptance of the method,

---

<sup>91</sup> *Id.*

as represented by McMenemy's work, the best exposition of the method, or Leonard's testimony in the BWI deposition.

*D. Admissibility*

In *United States v. Van Wyk*,<sup>92</sup> Judge Bassler reasoned that intuition-based forensic linguistics had never been tested for its reliability, so no one knows how well or how poorly it actually works, and no one knows how much writing is required for it to work, or whether it works well or poorly at identifying authors. This lack of scientific rigor falls short of Federal Rule of Evidence 702.<sup>93</sup> As the court put it:

Although Fitzgerald employed a particular methodology that may be subject to testing, neither Fitzgerald nor the Government has been able to identify a known rate of error, establish what amount of samples is necessary for an expert to be able to reach a conclusion as to probability of authorship, or pinpoint any meaningful peer review. Additionally, as Defense argues, there is no universally recognized standard for certifying an individual as an expert in forensic stylistics. Various judicial decisions regarding handwriting analysis, while not identical to text analysis, are instructive because handwriting analysis seems to suffer similar weakness in scientific reliability, namely the following: no known error rate, no professional or academic degrees in the field, no meaningful peer review, and no agreement as to how many exemplars are required to establish the probability of authorship.<sup>94</sup>

---

<sup>92</sup> *United States v. Van Wyk*, 83 F. Supp. 2d 515 (D.N.J. 2000).

<sup>93</sup> FED. R. EVID. 702.

<sup>94</sup> *Van Wyk*, 83 F. Supp. 2d at 522 (citing *United States v. Hines*, 55 F. Supp. 2d 62, 69 (D. Mass. 1999); *see also* *United States v. Santillan*, No. CR-96-40169 DLJ, 1999 WL 1201765, at \*2 (N.D. Cal. Dec. 3, 1999); Pre-Trial Transcript, *United States v. McVeigh*, No. 96-CR-68, 1997 WL 47724 (D. Colo. Feb. 5, 1997).

However, Judge Bassler believed that Fitzgerald's expertise in text analysis enabled him to know more about the frequency of items than the juror or judge might know.

Unlike his opinion on authorship, Fitzgerald's expertise in text analysis can be helpful to the jury by facilitating the comparison of the documents, making distinctions, and sharing his *experience as to how common or unique a particular "marker" or pattern is*. Therefore, the Court is satisfied that Fitzgerald's testimony as to the specific similarities and idiosyncracies between the known writings and questioned writings, as well as testimony regarding, for example, *how frequently or infrequently in his experience, he has seen a particular idiosyncrasy, will aid the jury* in determining the authorship of the unknown writings.<sup>95</sup>

Unfortunately, Judge Bassler assumed that a person's experience as to the frequency of a previously undefined "marker" is trustworthy.<sup>96</sup> He assumed that a person's experience is sufficient so that he can evaluate a "marker" as idiosyncratic or unique. Nothing more than the expert's personal experience is offered or expected.

Judge Bassler had access to Fitzgerald's report and the book Fitzgerald relied upon, McMenamin.<sup>97</sup> Defense did not produce other documentation or an opposing expert, so Judge Bassler was not provided any reviews of forensic stylistics by linguists. He might have reconsidered some of his ruling if he had seen peer reviews that speak directly to the particular issue of frequency estimation in intuition-driven forensic linguistics, especially Crystal.<sup>98</sup>

Closely following the *Van Wyk* ruling, testimony based on forensic stylistics has been partially admitted, with the expert not allowed to state an opinion about authorship, in New Jersey<sup>99</sup>

---

<sup>95</sup> *Van Wyk*, 83 F. Supp. 2d at 524 (citations omitted).

<sup>96</sup> *See id.*

<sup>97</sup> MCMENAMIN, *supra* note 2.

<sup>98</sup> *See* Crystal, *supra* note 82.

<sup>99</sup> *State v. McGuire*, 16 A.3d 411, 430 (N.J. Super. Ct. App. Div. 2011).

and Utah.<sup>100</sup> More in line with the scientific community's estimate of forensic stylistics, testimony based on forensic stylistics has been excluded by trial judges in California<sup>101</sup> and New York.<sup>102</sup> Testimony based on forensic stylistics has been withdrawn after a rebuttal report, depositions, affidavit, or evidence hearings in Virginia,<sup>103</sup> Washington,<sup>104</sup> and California.<sup>105</sup>

In a case currently under appeal, testimony based on forensic stylistics was admitted without a *Frye* hearing because the plaintiff argued that the method was not scientific and therefore not subject to *Frye*, but still presented an expert for opinion testimony.<sup>106</sup>

#### IV. STYLOMETRIC COMPUTING APPROACH TO AUTHOR IDENTIFICATION

Stylometric disputes in literature trace their roots to the Shakespeare, Pauline, and Federalist Papers controversies. Stylometry is the measurement of style, which has a long history since the 1880s of quantifying features of written language that are easy to measure, such as sentence length, word frequency, or common words among texts. Traditional stylometric features are grounded in literary criticism, not linguistics. This kind of analysis is based on school grammar, rhetoric, and textual criticism, not linguistic theory.

---

<sup>100</sup> *United States v. Zajac*, 748 F. Supp. 2d 1340, 1353 (D. Utah 2010).

<sup>101</sup> *People v. Flinner*, No. SCE211301, 2003 WL 24306950 (Cal. Super. Ct.); *Beckman Coulter v. Dovatron Flextronics*, No. 01CC08395 (Cal. Super. Ct. 2003).

<sup>102</sup> *Padiyar v. Yeshiva Univ.*, No. 110578/05 (N.Y. Sup. Ct. filed Jan. 3, 2006).

<sup>103</sup> *Lesnick v. Mathews*, No. CL01009530-00 (Va. Cir. Ct. Nov. 21, 2003), *aff'd sub nom. Lindeman v. Lesnick*, 604 S.E.2d 55 (2004).

<sup>104</sup> *State v. Preston*, No. 02-1-03082-4 (Wash. Super. Ct. Nov. 17, 2004).

<sup>105</sup> *In re Marriage of Isaacs*, No. BD403783 (Cal. Super. Ct. L.A. Cnty. 2011); *Hanus v. Hale*, No. GIC867514 (Cal. Super. Ct. June 14, 2006).

<sup>106</sup> Respondent's Exceptions to Referee's Report and Brief on the Merits at 49, *In re Masters*, No. S130495 (Cal. Jan. 12, 2012).



With large literary datasets and the advent of computer science, stylometric computing offers more sophisticated, statistical procedures for use in comparing documents than traditional stylometry. Computer science offers, for instance, machine-learning methods for text classification. But like traditional stylometry, stylometric computing uses language features that are not grounded in linguistic theory but are easy for a computer to work with, such as character strings, words, word frequency, and common words among texts.

Recently, several researchers such as Koppel, Argamon, Juola, Chen, and their students have begun to use stylometric computing for forensic author identification.<sup>107</sup> In light of the best practices for forensic author identification and a recent admissibility ruling, stylometric computing currently needs to incorporate at least three of these best practices.

#### A. Ground-Truth Data

Ground-truth data are all too often overlooked or undervalued in stylometric computing. One intriguing study of the “writeprint” claimed a high degree of accuracy at identifying the authorship of emails, with over ninety-seven percent accuracy for English and over ninety-two percent accuracy for Chinese.<sup>108</sup> This impressive result, however, is undermined by the fact that the dataset was not ground-truth data, as revealed by the researchers’ comment about a substudy of three authors in their English dataset: “Clearly, Mike’s distinct writeprint from the other two indicates his unique identity. The high degree of similarity between the writeprints of Joe and Roy suggests these two IDs might be the same person.”<sup>109</sup> Joe and Roy’s “writeprints” are almost identical. Yet it is also possible

---

<sup>107</sup> See, e.g., Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL’Y 299 (2013); Patrick Juola, *Stylometry and Immigration: A Case Study*, 21 J.L. & POL’Y 287 (2013); Moshe Koppel et al., *Authorship Attribution: What’s Easy and What’s Hard?*, 21 J.L. & POL’Y 317 (2013).

<sup>108</sup> Jiexun Li et al., *From Fingerprint to Writeprint*, 49 COMM. ACM 9, 9–10 (2006).

<sup>109</sup> *Id.* at 82.

that Joe and Roy are distinct people, and the method cannot clearly recognize the difference between Joe's and Roy's documents. We will never know which explanation is correct because a dataset of ground-truth data was not used. If a ground-truth dataset had been used, if known authors were attached to one or more screennames before validation testing was begun, the accuracy of the method could have been legitimately tested.

Ground-truth data must be verified. Scraping data from the web is a fast way of collecting a lot of data, but the data are not at all easily verifiable. Koppel and his colleagues harvested a dataset of blog posts from approximately 19,000 bloggers, which is available for research.<sup>110</sup> The bloggers are identified by a numerical identifier, gender, age, industry, and zodiacal sign. As with any data collected from the web, there is an assumption that the screenname belongs to one person at the keyboard, but this assumption is not trustworthy, since most web-based author identification disputes focus on the facts that screennames are not reliable indicators of textual ownership. Further, ages and gender can be falsely reported and are typically not verified in any way on blog postings, or even in blog ownership.

### *B. Forensically Feasible Data*

Traditional literary and recent computer-science-based stylometry have focused on literary texts, religious texts, and scholarly publications in science for electronic librarianship. All of the text types contain edited, rhetorically sophisticated, and highly stylized or formulaic language. These texts are also typically long, with tens of thousands of words.

In fact, using techniques that work well on tens of thousands of words is not at all a guarantee that it works on a few thousand (or hundred) words in an actual case of forensic author identification. Even computer tools for part-of-speech tagging that have been built on traditional "novels and newspaper"

---

<sup>110</sup> Jonathan Schler et al., *Effects of Age and Gender on Blogging*, AAAI SPRING SYMPOSIUM: COMPUTATIONAL APPROACHES TO ANALYZING WEBLOGS (2006).

corpora will not perform well on messy, unedited texts found in forensic author identification.

*C. Empirically Established Protocol*

Stylometric computing methods that work on literary texts or large collections of electronic text (as in electronic librarianship) are still untested on forensically feasible data. Bringing these methods wholesale into the forensic author identification problem is not the same as empirically establishing a protocol using these methods on forensically feasible data. The stylometric computing methods must be tested on forensically feasible ground-truth data for us to know how well they really work.

Further, it is essential to make sure that the stylistic features that are being used in different components of the techniques and then subjected to the statistical multiplication rule are truly independent features. The independence of linguistic features can really only be determined by a linguistic theory, not by school grammar or literary criticism. The counting of words alone and the counting of the same words in  $n$ -grams are not independent counts. However, since stylometric features are so unsophisticated linguistically, these kinds of dependencies are both common and not taken into consideration in the statistical manipulations.

Finally, the number of texts required for a technique, the number of component statistical tests (with truly independent features in them, if the multiplication rule is applied), and the ability to reach a high level of accuracy on forensically feasible ground-truth data all must be established empirically before a forensic author identification method based in stylometric computing is both legally and scientifically acceptable. Fancy statistics and vague references to “research has shown” when the statistics are ill-applied and the references refer to nonforensic research could very well overwhelm a judge or jury with the aura of expertise, but it may also be seen as smoke and mirrors and not a reliable method when the smoke clears.

#### *D. Admissibility*

In *United States v. Fresenius* in the District Court for the Western District of Texas,<sup>111</sup> the court ruled in favor of Fresenius's motion in limine to exclude stylometric computing testimony regarding the authorship of medical records. The proffered method focused on words, a standard stylometric analytical level. The statistical techniques included the Bernoulli mixture method. Yet even with a standard word-based stylometry and sophisticated statistical analysis, Judge Martinez ruled the testimony inadmissible because the expert, a professor of computational linguistics at the University of Texas, whose credentials were duly noted as impressive, could not offer any error rate or any verification of his method, while also maintaining that his method was 100% accurate. Judge Martinez's ruling warns us that sophisticated statistical analysis does not replace the need for empirically established protocols with known error rates through validation testing of each method on forensically feasible ground-truth data.

#### VI. CONCLUSION

Some scholars cast these three approaches, in a binary distinction, as intuition versus algorithm or nonquantitative versus quantitative.<sup>112</sup> From this perspective, forensic stylistics (the nonquantitative, intuitive approach) stands in contrast to forensic computational linguistics and stylometric computing (both of which are algorithmic and quantitative). I would suggest that there are two other binary distinctions to be considered in evaluating current approaches to forensic author identification.

First is the role of linguistics: is the approach linguistics or not? Forensic computational linguistics is grounded in linguistic theory, implements linguistic analysis in software, and uses standard linguistic methodology not only for analytical

---

<sup>111</sup> *United States ex rel. Gonzalez v. Fresenius Med. Care N. Am.*, 748 F. Supp. 2d 95 (W.D. Tex. 2010), *aff'd sub nom. Gonzalez v. Fresenius Med. Care N. Am.*, 689 F.3d 470 (5th Cir. 2012).

<sup>112</sup> See, e.g., Lawrence M. Solan, *Intuition Versus Algorithm: The Case of Forensic Authorship Attribution*, 21 J.L. & POL'Y 551 (2013).

techniques but also for data collection and research methodology. Neither forensic stylistics nor stylometric computing is grounded in linguistic theory. Instead, both forensic stylistics and stylometric computing are grounded in conceptions of language that are common in prescriptive grammar and literary criticism or focused on naïve conceptions of language as a list of words or a list of function words. So considering the “linguistics” in forensic linguistics, of which author identification is a primary task, forensic computational linguistics employs standard linguistics, while forensic stylistics and computer science neither use linguistics in analytical techniques nor theoretical underpinnings.

Second is the role of research in the approaches. In order for the *Daubert* factors to be met, litigation-independent validation testing on forensically feasible “ground-truth” data must be conducted. Forensic computational linguistics has met this challenge directly through the use of forensically feasible “ground-truth” datasets such as the Chaski Writer Sample Database. Independent of any litigation, validation tests have been conducted, as reported earlier in this paper. These tests have been run on forensically feasible data—that is, documents which are short, in several types of genre and register, and without any correction to grammar, spelling, or prescriptive conventions about writing. Further, the data are ground-truth data, where the authorship of each document is known; there is no possibility that someone else was using a screenname or posting blogs under a pseudonym. Finally, the validation test research has resulted in a known protocol for what is needed to apply the forensic computational linguistic methods; the test results empirically limit the amount of data required. It is hoped that both forensic stylistics and stylometric computing will conduct the kind of research that forensic computational linguistics performs, so that reliable methods of forensic authorship identification can be offered to our courts.

# ATTRIBUTION OF MUTUAL UNDERSTANDING

*Carl Vogel\**

## INTRODUCTION

This article explores advances in a method of analysis of conversational interaction, as recorded through text transcripts, for evidence of grounding in order to quantify certainty of mutual understanding. It is necessary to take into account aspects of communication in which certainty of having arrived at a common understanding of dialogue content must be pessimistically assessed. It may be that in many or even most contexts, the urgency of the linguistic elements of communication is negligible. It is a relatively rare event for linguistic acts, independently of other forms of communication, to have a distinctive, measurable, impact on human survival, and therefore, where such events exist, they tend to be spectacular.<sup>1</sup> That a failure to achieve mutual understanding does not typically result in catastrophic events does not constitute

---

\* Computational Linguistics Group, O'Reilly Institute, Trinity College Dublin, Dublin 2, Ireland (vogel@tcd.ie). Thanks are due to Lawrence Solan for organizing the Brooklyn Law School Authorship Attribution Workshop and his acknowledgement of the breadth of problems that fall into the category of linguistic attribution in legal contexts. Additionally, I am very grateful to Carole Chaski for discussions of this material and for making me aware of the work of Linda Smith.

<sup>1</sup> See, for example, the discussion of Flying Tiger Flight 66, on February 19, 1989, in which the air traffic control directive, “descend two four zero zero” was evidently understood as having a preposition, “to” rather than a numeral “two,” with the consequence that the aircraft flying at an altitude of 400 feet crashed into a hill 437 feet above sea level, killing all four people on the aircraft. STEPHEN CUSHING, FATAL WORDS: COMMUNICATIONS CLASHES AND AIRCRAFT CRASHES 14 (1994).

sufficient evidence to suggest that conversation is generally successful. It appears to be largely a matter of doctrine whether communicators are assumed to generally understand each other and signify the same ideas with the same language—this is known as the doctrine of intersubjective conformity.<sup>2</sup> The skeptical position that neither interlocutors (the participants in dialogues) nor external observers can ever completely verify whether the interlocutors have really understood each other is so robust that this skeptical position cannot be refuted. However, a weaker notion of intersubjective conformity is available. The weaker notion is that dialogue participants or observers may pragmatically behave as if there has been mutual understanding unless contrary evidence emerges. Nonetheless, in some contexts, forensic ones in particular, it seems safer to adopt the null hypothesis about communicative success which holds that communication attempts have been unsuccessful unless positive evidence exists that mutual understanding has emerged sufficiently to make one reject the null hypothesis and accept the alternative hypothesis that communication has been successful, rather than to assume success by default.

This article focuses on the development and testing of objective measures for assessing the likelihood of shared understanding of linguistic communication in contexts where shared understanding has a critical role, such as in forensic interrogations or other courtroom interactions. After motivating a principle of skepticism in assessing the likelihood of mutual understanding emerging for all participants in any given dialogue, a method is described which is deployed here to quantify levels of engaged interaction as a proxy measure for mutual understanding. The basic idea is that where positive evidence is needed to assert that dialogue participants have understood each other, levels of interaction that are statistically significant in divergence from random interactions provide a tangible basis for asserting that corresponding levels of mutual understanding may have been achieved. The method is illustrated with respect to an excerpt of a dialogue transcript

---

<sup>2</sup> TALBOT J. TAYLOR, MUTUAL MISUNDERSTANDING: SCEPTICISM AND THE THEORIZING OF LANGUAGE AND INTERPRETATION 29 (1992).

about which it is relatively easy for third party observers to form opinions about the communicative success achieved, then the method is applied to two transcripts available in public records of cases in which levels of mutual understanding has been contested. It is argued that the method of analysis is able to contribute useful facts to debates about the level of mutual understanding achieved in dialogues in which that form of communicative success matters.

Many contexts of legal interpretation are primed by principles associated with criminal trials, rigorously applied (i.e., “presumed innocent, unless proven guilty”). Presently, it is argued that the null hypothesis regarding the success of linguistic communication is that language use is ineffective unless proven effective.<sup>3</sup> This statement is jarring on first encounter because language use in communication is largely taken for granted as being as effective as the use of language in thought. However, one need only reflect on the many sorts of ambiguity that exist in language (i.e., sonic, syntactic, semantic) as well as their potential for combinatoric increase in the number of potential meanings to realize how great the chances are for miscommunication to arise through linguistic channels.<sup>4</sup> Indeed, much literature about theories, models, and simulations

---

<sup>3</sup> See *id.* at 30. Taylor presents such arguments at a meta level, in relation to possible rebuttals and resolutions; here, the proposal is to make do with this skeptical position rather than to argue against it. See generally *id.*

<sup>4</sup> Consider an example:

Suppose a sentence has three ambiguous lexical items and two (disjoint) places with attachment ambiguities; even if each ambiguity allows only two possibilities, the sentence will have, in principle,  $2^5 = 32$  interpretations. A simple example satisfying this description is given in (7); others would be easy to construct.

7. Old friends and acquaintances remembered Pat’s last time in California.

Here old can mean aged or long-term (or former) and can modify either friends and acquaintances or just friends; last can mean final or previous; time can mean occurrence or duration (e.g. if Pat was a racer), and in California can modify remembered or time.

Thomas Wasow et al., *The Puzzle of Ambiguity*, in MORPHOLOGY AND THE WEB OF GRAMMAR: ESSAYS IN MEMORY OF STEVEN G. LAPOINTE (C. Orhan Orgun & Peter Sells eds., 1998).



of language evolution explores the emergence of linguistic systems from the assumption that the first communicators began with shared thoughts and built language on prior, shared, thought.<sup>5</sup> When the context of communication is legal, given its conventional location of burden of proof, it can be more important to attempt to quantify the level of understanding that could have been achieved by linguistic means, for example, during testimony, particularly when nonnative speakers of a language are involved.<sup>6</sup>

Consider the following transcript of a courtroom dialogue<sup>7</sup>:

---

<sup>5</sup> Of course, this bootstrapped semantic infallibility, or “telepathy” approach is not universal in the language evolution literature. *See, e.g.*, Andrew D.M. Smith, *The Inferential Transmission of Language*, 13 ADAPTIVE BEHAV. 311, 311–23 (2005); Martin Bachwerk & Carl Vogel, *Establishing Linguistic Conventions in Task-Oriented Primeval Dialogue*, PROC. COST 2102 INT’L CONF., 2010, at 48–55; Martin Backwerk & Carl Vogel, *Language and Friendships: A Co-Evolution Model of Social and Linguistic Conventions*, 9TH INT’L CONF. ON EVOLUTION LANGUAGE, 2012, at 34–41; James R. Hurford, *Biological Evolution of the Saussurean Sign as a Component of the Language Acquisition Device*, 77 LINGUA 187, 187–222 (1989) (analyzing the biological evolution of the Saussurean sign as a component of the language acquisition device); James R. Hurford & Simon Kirby, *Co-Evolution of Language-Size and the Critical Period*, in SECOND LANGUAGE ACQUISITION AND THE CRITICAL PERIOD HYPOTHESIS 39, 39–63 (David Birdsong ed., 1999); James Hurford, *The Evolution of the Critical Period for Language Acquisition*, 40 COGNITION 159, 159–201 (1991); Simon Kirby, *Syntax Without Natural Selection: How Compositionality Emerges from Vocabulary in a Population of Learners*, in THE EVOLUTIONARY EMERGENCE OF LANGUAGE: SOCIAL FUNCTIONS AND THE ORIGINS OF LINGUISTIC FORM 303, 303–23 (James Hurford et al. eds., 2000).

<sup>6</sup> The burden of proof is not identically located in all legal contexts. In contract law, it is typical to presume that an individual understands the contractual issues, including the fine print; for example in relation to arguing the invalidity of a contract on the basis that it presupposes mistaken assumptions, “[t]he basic rule is that a mistaken assumption must be shared and be of fundamental importance if it is to be the basis for setting aside an otherwise valid contract. This is rarely satisfied.” STEPHEN A. SMITH, CONTRACT THEORY 283 (1993). Extra measures are needed to establish the invalidity of a contract due to deliberate drafting obfuscation, but this can be achieved. Melvin A. Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 STAN. L. REV. 211, 240–49 (1995).

<sup>7</sup> *People v. Herrero*, 756 N.E.2d 234, 242 (Ill. App. Ct. 2001).

- BASTOUNES Mr. Herrero is a Spanish speaking individual. We ordered an interpreter and he isn't here. He understands well enough that if you want you can admonish him on the record. He understands well enough what is going on in terms of picking this jury. He doesn't have a problem and wants to proceed this way. Perhaps we should put that on the record and tomorrow morning when we do opening statements and evidence
- COURT Mr. Herrero would you step up here, please.
- BASTOUNES I'm for sure that we would have an interpreter.
- BRODE We would like to put on the record that neither one of these individuals ever needed an interpreter.
- BASTOUNES I think the first time I was here with Mr. Herrero at the bond hearing didn't we? I just wanted to be sure that the record is clear and that there is no alleged error later on we will want an interpreter for the trial and it should be no problem getting one tomorrow.
- BASTOUNES For the record I have discussed with my client Mr. Herrero his desire to proceed this afternoon with picking the jury and he has indicated to me that he understands and wish well enough for that portion of the trial and wants to proceed.
- BASTOUNES Mr. Herrero, is it your desire now to proceed with picking the jury?
- HERRERO Yes.
- BASTOUNES Without an interpreter?
- HERRERO Yes.

- BASTOUNES Do you understand what I'm saying to you now, is that correct?
- HERRERO Yes.
- BASTOUNES Okay. Judge, if you want to inquire further.  
\* \* \*
- COURT Mr. Herrero, have you understood the conversation that has taken place in the last ten minutes or so?
- HERRERO I understand a little bit.
- COURT Mr. Herrero, do you have any objection to picking the jury now without the interpreter?
- HERRERO No.
- COURT Okay. All right, then we can proceed. We're going to try to get an interpreter.
- BASTOUNES I did try earlier.
- COURT We're trying now, I put in the request.

This transcript was reviewed during an appeal by Hector Herrero of a drug possession conviction and twenty-five year sentence.<sup>8</sup> The crux of the appeal was since Herrero had a poor understanding of the English language and was not aided by a translator, he was effectively absent from the trial.<sup>9</sup> While the appeal was not successful, it was not due to Herrero's claims about linguistic ability.<sup>10</sup> Even though the question of whether the transcript reveals that the defendant did not understand English sufficiently to be deemed present at the trial did not

---

<sup>8</sup> *Id.* at 242-43.

<sup>9</sup> *Id.* at 242-45.

<sup>10</sup> *See id.* at 243-44 ("The decision to appoint an interpreter is within the trial court's discretion, and a conviction will be reversed only when an abuse of the court's discretion deprived the defendant of some basic right. . . . Defendant did not ask for an interpreter at any time during jury selection. Nor did defendant raise this issue in his motion for new trial. Consequently, the issue is waived . . . . Even if we were to find that this issue was not waived, defendant would still not prevail.").

determine the appeal in this case, it would be beneficial in related circumstances to be able to quantify certainty of interlocutor involvement in the dialogue. Interlocutor involvement is the degree to which dialogue participants are engaged in the conversation; this is more than the frequency and balance of turn-taking, since a virtual monologue with interleaved vocalizations such as “uh-huh” can demonstrate ample turn-taking with little engagement. Notably, in this example, the defendant provided answers to yes-no questions that were coherent as a set. This appears to evidence a greater level of engagement than if those turns were uniformly “yes.” On the other hand, the defendant says little in each question-and-answer to suggest that there was genuine understanding of what the binary responses entailed, except the hedge, “I understand a little bit.” There is no restatement in the defendant’s own words of what was understood. It is precisely a method of quantifying likelihood of understanding that is described and argued appropriate for such forensic contexts in this article. Part III returns to this transcript.

In assessing mutual understanding in conversation, it is important to be clear about what constitutes a null hypothesis and where the burden of proof lies in establishing an alternative hypothesis. Its relevance is illustrated by the fact that the question of whether the evolutionary niche of language is as a cognitive tool for thought or as a cognitive tool for communication remains a topic of debate. This is part of a debate about whether language is a socially evolved construct or a biologically evolved one. It is natural to compare human language and human vision in this context. The eye is a delicate and highly functional product of biological evolution. In contrast, flaws of the linguistic system, including ambiguity at every level of linguistic description, leave language as a poor medium for communication. If one were to design a visual system from scratch, the eye as it is now would likely be a part. If one were trying to evolve a system as well-suited to communication as the eye is for vision, one would strive for telepathy rather than human language.<sup>11</sup>

---

<sup>11</sup> Cf. Steven Pinker & Paul Bloom, *Natural Language and Natural*

On the other hand, language is a very good system in which to conduct thought, and it serves very well for mental representation and reasoning about a plethora of complex matters. While a speaker may be sometimes vague, typically one who utters an ambiguous sentence has an intended reading in mind. Some topics of potential thought remain notably ineffable, (i.e., thoughts of extreme pain or pleasure or profoundly spatial topics such as geographic directions—precious few people are adept at expressing in words only, without recourse to gesture or maps, how to navigate from one point to another in a city that does not have a grid-based street system), but for the most part, it is difficult to imagine human thought without language.<sup>12</sup>

Given the fundamental flaws of human languages as media for communicating toward mutual understanding, there are strong reasons to view the null hypothesis about human communication in a pessimistic light. In the absence of strong evidence to the contrary, human interaction through dialogue does not reach mutual understanding of the language each other has used in dialogue to describe the world, much less mutual agreement that the world is (or should be) the way that interlocutors understand each other to describe it. A shared understanding of the world may come from common embodiment, the fact that humans share much of their genetic constitution and occupy the same niche in the ecosystem with each other, independently of agreement arising from communication, or from communication nurtured without language used in the process.<sup>13</sup> I claim that the appropriate null hypotheses about the outcome of language use is not that utterances were interpreted as uttered for all parties to a conversation and agreed in their truth relations to the described world; rather, the null hypothesis pertaining to ordinary dialogue is that communication did not make obvious the existence of disagreement about meanings and the relations between those claims and the world.

---

*Selection*, 13 BEHAV. & BRAIN SCI. 707, 784 (1990).

<sup>12</sup> See MICHAEL NEWTON, SAVAGE GIRLS AND WILD BOYS: A HISTORY OF FERAL CHILDREN 20–21 (2002).

<sup>13</sup> See STEVEN STROGATZ, SYNC: HOW ORDER EMERGES FROM CHAOS IN THE UNIVERSE, NATURE, AND DAILY LIFE 264 (2003).

Literature written about pragmatic theory<sup>14</sup> notes that interlocutors may arrive at working hypotheses by coming to the realization that they have been talking about the same things, understanding each others' comments about those things, and agreeing about the propositional contents put forward on all sides.<sup>15</sup> Much of this literature appeals to processes of *grounding*, which provides a foundation for the speculative conclusions that interlocutors have understood each other.<sup>16</sup> Grounding ultimately is anchored in repetition of words, phrases, and syntactic structures among interacting agents.<sup>17</sup> Detailed conversation analysis has been deployed in legal contexts to emphasize the significant effort necessary to achieve the effect that clients feel understood by their lawyers, for example.<sup>18</sup> Repetitions provide linguistic mechanisms that may be used to develop confidence that conversation has not resulted in misunderstanding.<sup>19</sup> Differential use of repetition according to authority and expertise among interlocutors has been pointed out,<sup>20</sup> but in general for all parties in conversation, repeated information is taken to be more securely placed in common ground.<sup>21</sup>

---

<sup>14</sup> Herbert H. Clark & S.A. Brennan, *Grounding in Communication*, in PERSPECTIVES ON SOCIALLY SHARED COGNITION 127–29 (Lauren B. Resnick et al. eds., 1991). See generally DEBORAH TANNEN, *TALKING VOICES: REPETITION, DIALOGUE, AND IMAGERY IN CONVERSATIONAL DISCOURSE* (2007).

<sup>15</sup> It is one thing to mutually *understand* that the content of some utterance is the proposition  $p \rightarrow q$ , and it is another to *agree* that the proposition  $p \rightarrow q$  is true.

<sup>16</sup> See generally Paul Vogt & Federico Divina, *Social Symbol Grounding and Language Evolution*, 8 INTERACTION STUD. 31–32 (2007).

<sup>17</sup> Clark & Brennan, *supra* note 14; TANNEN, *supra* note 14.

<sup>18</sup> See, e.g., Linda F. Smith, *Always Judged—Case Study of an Interview Using Conversation Analysis*, 16 CLINICAL L. REV. 423 (2010).

<sup>19</sup> Cf. PATRICK HEALEY, *COMMUNICATION AS A SPECIAL CASE OF MISUNDERSTANDING: SEMANTIC COORDINATION IN DIALOGUE* (1996).

<sup>20</sup> Eve V. Clark & Josie Bernicot, *Repetition as Ratification: How Parents and Children Place Information in Common Ground*, 35 J. CHILD LANGUAGE 349, 364 (2008).

<sup>21</sup> Clark & Brennan, *supra* note 14; TANNEN, *supra* note 14.

In certain contexts of urgency, grounding mechanisms are part of the ritual of communication designed to avoid miscommunication.<sup>22</sup> The rituals of air traffic communication emphasize repetition in order to reduce potential misunderstanding arising from conversation.<sup>23</sup> Repetition of words and phrases has been analyzed as providing a means for interlocutors to increase their involvement in dialogues.<sup>24</sup> Accordingly, it follows that enhancing the involvement of aircraft cockpit personnel via repetition increases the chances of a shared understanding of the matter being spoken of by increasing joint immersion in the context at hand. Conscious repetition incrementally eliminates chances that the interlocutors are focused on distinct perspectives on the immediate context.

This article describes and evaluates a method of analysis that can be used to measure engagement in interaction. Where interactions are assessed with respect to these measures, it is possible to quantify certainty that interlocutors have successfully communicated. A growing body of research develops automated and semiautomated methods of measuring synchronization among dialogue participants in terms of such analysis.<sup>25</sup> This measure is argued here to correlate with mutual understanding. Some scholars have examined laboratory constructed task-based dialogues in order to correlate effects associated with repetition

---

<sup>22</sup> CUSHING, *supra* note 1, at 40.

<sup>23</sup> *Id.*

<sup>24</sup> See TANNEN, *supra* note 14, at 84.

<sup>25</sup> See, e.g., Junko Itou & Jun Munemori, *Repetition of Dialogue Atmosphere Using Characters Based on Face-to-Face Dialogue*, 6278 KNOWLEDGE-BASED & INTELLIGENT INFORMATION & ENGINEERING SYS. 527 (2010); Fabian Ramseyer & Wolfgang Tschacher, *Nonverbal Synchrony or Random Coincidence? How to Tell the Difference*, 5967 DEVEL. MULTIMODAL INTERFACES: ACTIVE LISTENING & SYNC. 182 (2010); David Reitter et al., *Computational Modelling of Structural Priming in Dialogue*, PROC. HUM. LANGUAGE TECH. CONF., N. AM. CHAPTER ASS'N FOR COMPUTATIONAL LINGUISTICS ANN. MEETING, 2006, at 121; David Reitter & Johanna D. Moore, *Predicting Success in Dialogue*, PROC. 45TH ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS, 2007, at 808; Carl Vogel & Lydia Behan, *Measuring Synchrony in Dialog Transcripts*, 7403 COGNITIVE BEHAV. SYS. 73 (2012).

with task-oriented success.<sup>26</sup> In the present work, transcripts of dialogues from outside laboratory settings are analyzed. Rather than considering repetition counts up to a point in time from the beginning to an evaluation point in the dialogue, repetitions of tokens as a proportion of total tokens that could have been repeated between an utterance and immediately preceding utterances are considered.<sup>27</sup> The level of mutual understanding experienced by the interlocutors is in all cases studied here subjectively assessed, independently through the sources from which the data is drawn. In cases where the method does not support the conclusion that mutual understanding has been achieved, the independent assessments historically provided appear to agree with the conclusions drawn through analysis using the method. The critical cases are those where it is a main issue whether one of the participants understood what was going on, and outside the laboratory environment, it is seldom possible to obtain independent measures of mutual understanding among dialogue participants. Thus, the role of this article is to provide evidence from relatively clear cases that the measurements suggested are valid as a proxy for assessing mutual understanding and to show their efficacy by pointing out the contributions they make in cases that are open to greater debate about the levels of mutual understanding that were likely to have been experienced.

It is not possible to directly measure the actual degree of mutual understanding—neither as a dialogue participant nor as an outside observer. However, the extent to which synchrony and grounding behaviors indicate mutual understanding is the extent to which it may be quantified. If there is no evidence of synchronized engagement, the basis for certainty that there is mutual understanding is undermined. Where even low levels of synchrony are evident, the level of certainty that mutual understanding is in place is correspondingly increased. The method of quantification is to quantify levels of repetition in dialogue. Where repetition differs from chance expectations,

---

<sup>26</sup> See Reitter et al., *supra* note 25, at 122; Reitter & Moore, *supra* note 25, at 809.

<sup>27</sup> See Vogel & Behan, *supra* note 25, at 75.



pragmatic inferences may be drawn. The occurrence of such instances has led to the development of a typological theory of functions of repetition.<sup>28</sup> The distinction between allo-repetition and self-repetition corresponds to distinctions in the pragmatics of attending to others and attending to oneself. However, this correspondence is not total in that, for example, allo-repetition serves (at least) the purpose of reassuring an initial speaker that a message has been heard but also increasing confidence for the speaker who repeats that the message was at least correctly heard. Self-repetition has a strong relationship with persistence in a dialogue plan. Inversely, allo-repetition avoidance (in its strongest form, a refusal to adopt the language of one's dialogue partners) manifests a focus on the self, and self-repetition avoidance can indicate the absence of an independent dialogue plan. The absence of a significant difference in repetition between actual and randomized dialogues entails a lack of engagement according to that measure and retention of the null hypothesis that mutual understanding was not achieved. This conclusion is based on studies that quantify repetition in mutually engaged conversation<sup>29</sup> and communication that leads to success in collaborative tasks.<sup>30</sup>

This discussion began with an argument that the null hypothesis about linguistic success in ordinary conversation should be that language did not yield mutual understanding; however, in certain legal contexts, the normal burden of proof necessitates even greater caution in assuming that linguistic communication has been successful. The methods discussed here can be used to mitigate risks associated with inappropriately rejecting that null hypothesis. In what follows, a range of dialogues are used to illustrate a theory of repetition in dialogue. The ramifications of the research are as relevant to forensic

---

<sup>28</sup> Note that this is on a different scale of analysis than that of Deleuze of the general epistemological effects of repetition, but is rather restricted to linguistic pragmatics, and as such is a contribution to the framework associated with Tannen. *See generally* GILLES DELEUZE, *DIFFERENCE & REPETITION* (Paul Patton, trans., Athlone Press 2004) (1968); TANNEN, *supra* note 14.

<sup>29</sup> *See generally* Vogel & Behan, *supra* note 25, at 73–88.

<sup>30</sup> Reitter & Moore, *supra* note 25.

contexts as to safety in air traffic communications: measures of conversational synchrony can be used to mitigate doubt about the extent to which, for example, interrogated individuals have understood the nature of conversations in which they participate, including, for example, the extent of understanding about *Miranda* rights.<sup>31</sup>

The paper is structured as follows. First, a method is described for analyzing interaction in dialogue with respect to interlocutor alignment. The algorithmic core of this method has been employed in the analysis of a range of natural dialogues as recorded in available textual transcripts.<sup>32</sup> The essence of the method is the evaluation of the degree to which the various forms of repetition are visible between actual dialogues and a number of randomized reorderings of the dialogue turns. Where precise temporal alignment information is available, overlap and other temporal features of synchronization may also be measured. Lydia Behan and I illustrated how the measures are manifest in natural conversations representative of types: one in which discussion is casual and mutually supportive among participants of equal social standing and another in a crisis situation with a clearly defined leader (an aircraft crash transcript).<sup>33</sup> The work here extends the methods of statistical analysis further and in directions that support forensic deployment of the method in the attribution of interlocutor engagement and understanding of critical legal discourse, for purposes such as police interview, courtroom testimony, and cross-examination. The theory developed here is that where actual repetitions do not exceed random counterparts at all, there is reason to think that the dialogue exemplifies lack of engagement and misunderstanding (or rather, there is no reason to reject the null hypothesis in such a case). Similar assessments are considered with respect to individual participants within the dialogues.

---

<sup>31</sup> See generally *Miranda v. Arizona*, 384 U.S. 436, 467 (1966) (holding that an accused person must be made aware of his or her constitutional rights upon arrest, including the right to remain silent).

<sup>32</sup> See generally Vogel & Behan, *supra* note 25, at 73–88.

<sup>33</sup> *Id.* at 77–87.

## I. METHODS

In all cases, the dialogues analyzed here have already been independently transcribed and are available on the web or in publications cited. The importance of this, given the intention to study repetition across successive dialogue turns, is that the turn-taking structure of dialogue has been determined independently, without evidence of prior reflection on the possibility of this sort of analysis being undertaken. Temporally, the transcripts are partially ordered given that contributions of interlocutors are interleaved; however, temporal overlap analysis is not conducted. Ideally, one would have available not just textual transcripts that indicate the sequence of turns but also the timing of those turns so that temporal overlap of turns can be taken into account. However, as with the dialogues analyzed here, one cannot be guaranteed the availability of timing information.

A decision has to be made with respect to the level of linguistic description at which to consider repetition (tokenization): morphemes, words, part-of-speech (“POS”) labels, concepts, etc. or combinations thereof. The units of representation decided on are types, and their instances are tokens. At this stage, punctuation marks are disregarded. Representation of semantic information that is not directly lexically encoded is not made, since it is not safe to conclude that speakers accept as true all logically valid consequences of their assertions. The text is individuated as words and restricted part-of-speech labeling. POS labeling is only used for personal pronouns to capture the fact that, ordinarily, they are not repeated verbatim but with complementarity, in dialogue that proceeds successfully. Thus, the sole other treatment of the data analyzed here (apart from ignoring punctuation) is to transform dialogues in the form of examples like those numbered below (1) and (3) into those like (2) and (4), respectively; that is, complementary first-person and second-person personal pronouns are replaced with a single item (“IY,” regardless of grammatical number). No deeper parsing is deployed and no other POS labels are used; even third-person pronouns are left intact. Avoiding parsing is desirable to ensure that the methods

are replicable and not dependent on any particular theory of natural language syntax.

- (1) A: Do you understand?  
B: I understand.
- (2) A: Is this your address?  
B: Yes, this is our address
- (3) A: Do IY understand?  
B: IY understand.
- (4) A: Is this IY address?  
B: Yes, this is IY address

The tokens counted in the analysis are sequences of words and POS labels of this form; these are known as *n-grams*. The value of *n* varies between one and three. Thus, in the treated dialogue fragment (2) above, B is regarded as repeating two unigrams and one bigram from A's utterance. In (4), B is counted as repeating four unigrams and one bigram. Since they are *sequences*, word order matters, and "this is" does not count as a repetition of "is this," even though there is a natural *syntactic* complementarity between English subject-auxiliary verb inversion in polar questions and the canonical ordering of the subject and verb in answers, just as there is lexical complementarity in pronouns. The equation of the two bigram forms is not made here because of the decision to avoid the need to parse texts. If one were to take syntactic structures into account, then one could consider structural complements as well.

For each transcript processed, the algorithm for data extraction designates a location of memory called a "register" for each speaker. The register, which is initially empty, eventually contains the contents of the most recent contribution of the corresponding speaker.<sup>34</sup> A generalization of the method would afford each speaker a vector of registers in order to evaluate repetitions arbitrarily far back in a conversation. In the work described here, the single register for each speaker is initially empty, but it subsequently records the last contribution

---

<sup>34</sup> Register machines or abacus machines with registers form a standard theoretical model of computation. See GEORGE S. BOLOS ET AL., COMPUTABILITY AND LOGIC 45-62 (4th ed. 2002).

made by that speaker. This remains a structural definition, since it is not constrained temporally.

For each utterance, the tokens<sup>35</sup> are compared with those recorded in the register for each actor, counting those tokens that are shared between the utterance and its speaker's own register (selfshared; AKA self-repetition) and counting those tokens which are shared with utterances recorded in the registers of each of the other speakers (othershared; AKA allo-repetition). Then the speaker's register is updated to contain the most recently processed utterance. The token counts are conducted for each level of *n-gram*. Repetitions of *n-grams* (again,  $1 \leq n \leq 3$ ) are recorded as counts with respect to the values in the registers as either "SelfShared" or "OtherShared" tokens. In measuring the degree of sharing for a turn  $u_i$ , these figures are regarded as proportions of the total number of *n-grams* for each level of  $n$  that could have been shared, given the length of  $u_i$ , between the turn and the immediately preceding turns as recorded in the registers for each actor. In analyzing the three levels of  $n$  it is useful to think of there actually being two levels, lexical (unigrams) and phrasal (bigrams and trigrams), and therefore this factor is accordingly presented in terms of a derived factor *Nbar* with levels "1" and "2+."

The actual repetition values are then compared with those derived from some number (ten, in each of the experiments here) of randomized reorderings of the turns (AKA contributions). The constituent words within any individual contribution are left intact in their original order; the reordering is of entire turns with respect to other turns into a random partial ordering. This method supports the level of repetition analysis by speaker or aggregated across speakers comparing self-repetition and self-repetition of sequences as manifest in actual dialogues and their turn-randomized counterparts. Forms of repetition are deemed significant in their visibility when the difference between the actual and randomized counterparts is statistically significant.

---

<sup>35</sup> The tokens are in sequences—sequences of words or sequences of POS tags, etc., depending on the choice of tokenization.

Dialogues are reordered by generating new start-times and durations for each utterance as turn indices. The times and durations are selected using random generators based on parameters that depend on the values in the original conversation. Thus, for each utterance  $u_i$  a re-indexing  $\hat{u}_i$  is constructed. The  $\hat{u}$  are sorted on their temporal indices. Analysis of duration of overlaps is enabled by this framework but not conducted here given that the transcripts addressed are not annotated temporally beyond the relative order of turns. Where temporal annotations are available, the analysis of synchronization may have greater depth with the inclusion of consideration of temporal overlap. In the reordered dialogue, counts of allo-shared tokens and self-shared tokens are recorded, just as with actual dialogue. The variables measured and analyzed here are as specified in Table 2. The results allow for the depiction of many contrasting proportions; however, the specific contrasts of interest are whether actual repetition of unigrams and  $n$ -grams for larger values of  $n$  exceeds the random counterparts for any speaker. Thus, the null hypotheses tested in each dialogue are as in (5) and (6).

$$(5) \text{ Randomized.Speaker.1} - \text{Actual.Speaker.1} \geq 0$$

$$(6) \text{ Randomized.Speaker.2+} - \text{Actual.Speaker.2+} \geq 0$$

The data is analyzed in each case using a generalized linear model with a binomial error family.<sup>36</sup> Adjustments are made for multiple comparisons using directed tests for significance, wherein the null hypothesis essentially is that where `DialogType = Randomized` repetition will equal or exceed repetition for the corresponding Actual case.<sup>37</sup>

---

<sup>36</sup> Within R, this is using the following:

```
glm(OSprop ~ DialogType*Speaker*Nbar,family=binomial) and
glm(SSprop ~ DialogType*Speaker*Nbar,family=binomial).
```

<sup>37</sup> With the R multcomp package, the following representative constructs are used:

```
fos <- interaction(DialogType,Speaker,Nbar),
mfos <- glm(OSprop ~ fos,family=binomial),
mfos.mc <- glht(mfos,linfct = mcp(fos = "Tukey"),alternative="1").
```

See FRANK BRETZ ET AL., MULTIPLE COMPARISONS USING R (2011). Subsequently, all tests are discarded which do not hold constant Speaker and

Table 2: Variables Analyzed

<i>Variable</i>	<i>Interpretation</i>
DialogType	actual vs. randomized
OtherSpeakers	total number of participants, minus one
$n$	length of $n$ -grams (1, 2 or 3)
$n$ -bar	$n$ -gram length as a two-level factor (“1” or “2+”)
Nbar	$n$ -bar
NGrams	total number of $n$ -grams in a turn
SelfShared	count of tokens from turn shared with own prior contribution
OtherShared	count of tokens from turn shared with prior contribution of other
ss	SelfShared
os	OtherShared
SSrel	SelfShared / NGrams
OSrel	OtherShared / (NGrams * OtherSpeakers)
NonSelfShared	NGrams-SelfShared
NonOtherShared	(NGrams * OtherSpeakers)-OtherShared
nss	NonSelfShared
nos	NonOtherShared
SSprop	SelfShared, NonSelfShared
OSprop	OtherShared, NonOtherShared

## II. A CLEAR FAILURE TO COMMUNICATE

The method of analysis advocated here provides an index of synchronized engagement in conversation. This section addresses an example of conversation that has been independently transcribed and analyzed in order to provide a reference point for assessing levels of engagement as a proxy measure of mutual understanding in cases that are selected as more contentious and discussed in Part III. The example is one in which opinion evidently converges on the notion that the conversation does not exemplify mutual understanding.

---

Nbar and vary solely DialogType.

In a BBC television interview in 1997, Jeremy Paxman interviewed Michael Howard, the former Home Secretary in the UK government.<sup>38</sup> The text, prior to treatment as described in Part I, is included in Appendix A. The transcript of the Howard interview is a relatively famous example of an interview in which questions were asked directly, but did not yield responsive answers. If the interlocutors were not both native speakers of English, one might reach the generous conclusion that Howard did not understand the question asked by Paxman. Abstracting over the context, with knowledge of the roles involved—news presenter and politician—one is more likely to infer evasiveness on the part of the politician, since the language used is not manifestly complex. In any case one can conclude from the text of the discussion that the conversation did not instantiate a collaborative flow of information. This conversation is a useful one to show how the proposed measures fare in assessing the level of mutual engagement exemplified. This is demonstrated below.

The mean counts of shared tokens by levels of  $n$  are provided in Table 3.

Table 3: Shared token counts in the Paxman-Howard interview

$n$ -bar	<u>OtherShared</u>		<u>SelfShared</u>	
	1	2+	1	2+
Actual	4.71	1.94	3.04	.06
Randomized	4.24	1.58	2.88	.01

The statistical significance of repetitions is assessed by comparing the proportions in token counts that were shared between each utterance and its immediately preceding utterance of the totals that could have been shared in each case. For neither speaker, for neither level of  $n$  in the comparisons here is

---

<sup>38</sup> See Interview by Jeremy Paxman with Michael Howard, former U.K. Home Sec’y, *Newsnight* (BBC television broadcast May 13, 1997), available at <http://www.youtube.com/watch?v=Uwlsd8RAoqI>.



the actual repetition (OS,SS) a significantly greater proportion of total tokens that could have been shared (OS+NOS,SS+NSS) for actual dialogue than in the randomized case. The proportions in the comparisons here are depicted as follows: allo-repetition in Figure 1 and self-repetition in Figure 3. The graphs depict the relevant proportions. Of the two sorts of dialogue type, the area occupied by the counts for “Randomized” dialogue is necessarily larger than the area for “Actual” dialogue because there are ten random reorderings of the actual dialogue. Within the sorts of “Other-Sharing” (for either of the two dialogue types) the instances of sharing of items that are shared (“OS”) tends to be much smaller than the number of items that could have been, but were not, “other-shared” (hence, the label, “NOS”). It is apparent that Howard spoke more than Paxman, but the contrast of interest is in the proportions shared and not shared between the actual and randomized conditions for the two individuals. Thus, the mosaic plot<sup>39</sup> in Figure 1 does not show any significant difference in allo-repetition for either speaker between the actual and randomized dialogues. The same information, with an additional contrast, is shown in Figure 2: here, the proportion of shared and nonshared unigrams and  $n$ -grams, for values of  $n > 1$  aggregated, are shown to illustrate the proportions as they depend on the length of expressions, for allo-repetition in this dialog. Recall that the precise statistical tests are used probe ((5) and (6)) for each level of  $n$ -bar throughout; however, the graphs which do not separate the levels of  $n$ -bar demonstrate the main relationships discussed more clearly. Figure 3, which shows the same proportions for self-repetition, looks different to Figure 1, because Paxman repeated more of his own utterances (“SS”) in relation to his own unrepeated items (“NSS”) than Howard repeated of his own utterances. However, for neither Paxman nor Howard is the difference significantly greater for the actual dialogue than the randomized counterparts.

---

<sup>39</sup> David Meyer et al., *The Strucplot Framework: Visualizing Multi-Way Contingency Tables with VCD*, J. STAT. SOFTWARE, Oct. 2006, at 1 (“A mosaic plot is basically an area-proportional visualization of (typically, observed) frequencies, composed of tiles (corresponding to the cells) created by recursive vertical and horizontal splits of a rectangle.”) (citations omitted).

Table 4: 1997 Paxman-Howard interview: counts of shared and nonshared tokens by dialogue-type, speaker, and n-bar

Speaker	DialogType	OtherSharing				SelfSharing			
		Actual		Randomized		Actual		Randomized	
		Nbar	OSNOS OS	NOS	SS NSS	SS	NSS		
Howard	1	43	154	435	1435	46	151	423	1547
	2+	1	357	4	3576	12	346	125	3455
Paxman	1	30	91	256	954	67	54	594	616
	2+	2	204	3	2057	81	125	631	1429

Figure 1: Allo-repetition by Paxman & Howard: shared vs. nonshared tokens by speaker in actual and randomized dialogue

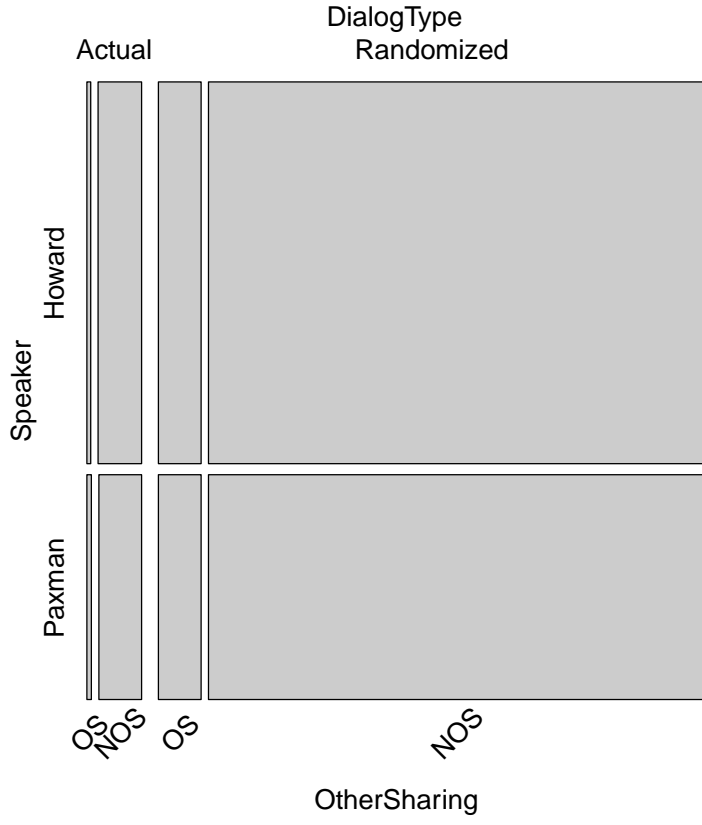


Figure 2: Allo-repetition by Paxman & Howard: shared vs. nonshared tokens by speaker and n-bar in actual and randomized dialogue

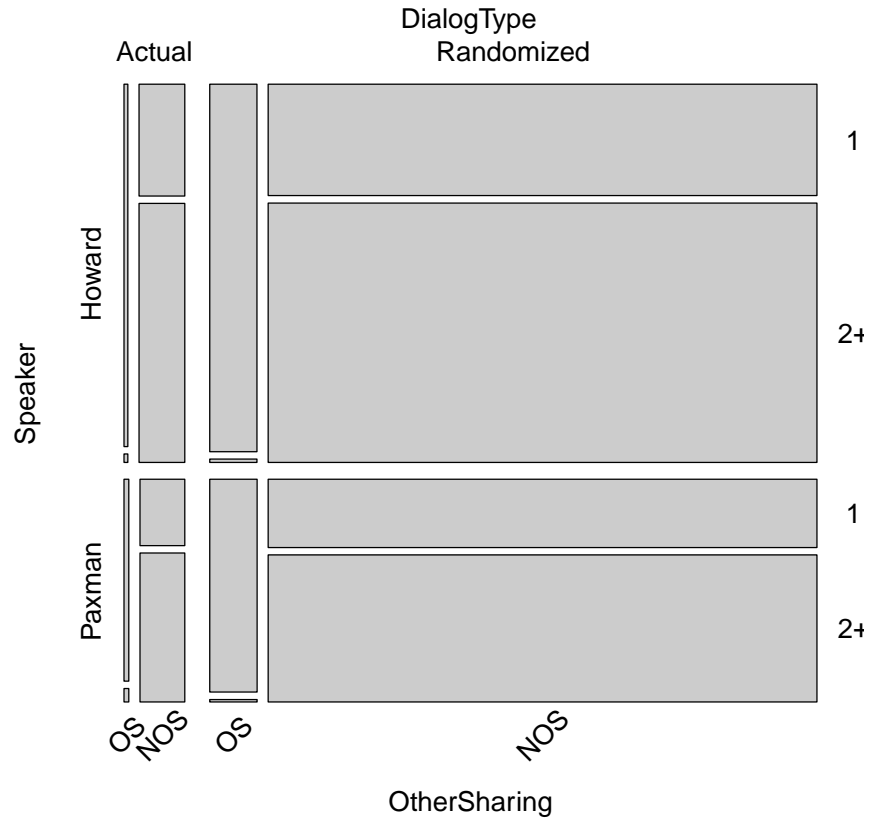
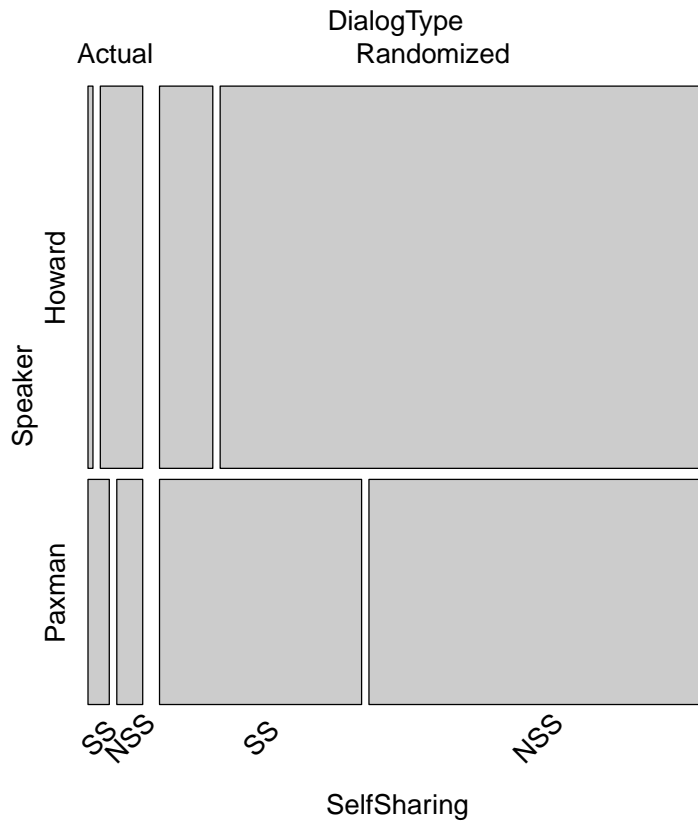


Figure 3: Self-repetition by Paxman & Howard: shared vs. nonshared tokens by speaker in actual and randomized dialogue



The lack of significance in the difference between the actual and randomized dialogues in the direction taken as of interest here suggests that the dialogues might as well have had the turns randomly reordered to obtain the same overall effect of engagement. Certainly, it is clear that the repetitions of the Paxman questions makes self-repetition high for Paxman but not substantially different in the ten randomized reorderings. Similarly, Howard's repetition of terms used by Paxman is relatively impervious to reordering. The statistical effects reveal that at the level of textual content, there is little engagement exhibited. Accordingly, this leaves open that an analysis including gesture, timing of utterances, or overall energy measurable in the scene during the flow of the dialogue could still detect involvement and engagement at that level,<sup>40</sup> just as speakers of mutually unintelligible languages may interact with engagement but without full understanding. The lack of significance in the contrasts of interest here implies that, although it seems that willful avoidance might be at issue, one cannot say for certain that Paxman and Howard reached an understanding of each other.

It is interesting to note that it is relatively difficult to find transcripts of naturally, publicly occurring dialogues in which third party observers conclude that the interlocutors do not understand each other. Fabrications of such are the stuff of comedy, such as the "Who's on first" routine of Abbott and Costello. The Paxman-Howard example was selected as a relatively famous example of failure to communicate. With more successful dialogues, it is generally not true that in all cases all parties will support measurements according to the methods used here in which the actual dialogue differs in the hypothesized direction from the ten turn-randomized alternatives. Where there is significant self-repetition but no significant allo-repetition for any of the participants in a dialogue, it would seem that there is evidence of persistence but not of linguistic engagement. Conversely, where allo-repetition effects are significant, but not self-repetition, there is evidence of understanding and engagement. Where neither of the effects is visible for any of

---

<sup>40</sup> Ramseyer & Tschacher, *supra* note 25.

the participants, the conversation is difficult to classify as successful. It is not safe, in such a case, to make judgments about the level of mutual understanding achieved among the participating parties.

### III. LEVELS OF MUTUAL UNDERSTANDING IN SAMPLE CASES

The next two sections present analyses of transcripts which represent the sorts of situations in which one might expect the methods described here to have benefit. In the present context, they constitute evaluations of the method to the extent that the conclusions supported by the method are in agreement with independent historical assessments of the conversations.

#### A. *Case Study 1: People v. Herrero*<sup>41</sup>

The data the first case study relies upon has been described in the introduction.<sup>42</sup> The data is treated and processed according to the methods specified, with the actual dialogue giving rise to ten counterparts in which the turns have been randomly reordered with respect to each other.

##### 1. *Results*

As before, the proportions of allo-repetition<sup>43</sup> and self-repetition<sup>44</sup> are analyzed. Notably, Herrero has higher levels of both OS and SS in the randomized versions of the dialogue than in the actual dialogue. Furthermore, recall that the testing conducted is directional and the null hypothesis is that random repetition will exceed or equal actual repetition in the proportion measures ((5) and (6)).

---

<sup>41</sup> *People v. Herrero*, 756 N.E.2d 234, 242–43 (Ill. App. Ct. 2001).

<sup>42</sup> *See supra* Introduction.

<sup>43</sup> *See infra* Figure 4.

<sup>44</sup> *See infra* Figure 5.

Figure 4: Allo-repetition in the Herrero case: shared vs. nonshared tokens by speaker in actual and randomized dialogue



Figure 5: Self-repetition in the Herrero case: shared vs. nonshared tokens by speaker in actual and randomized dialogue

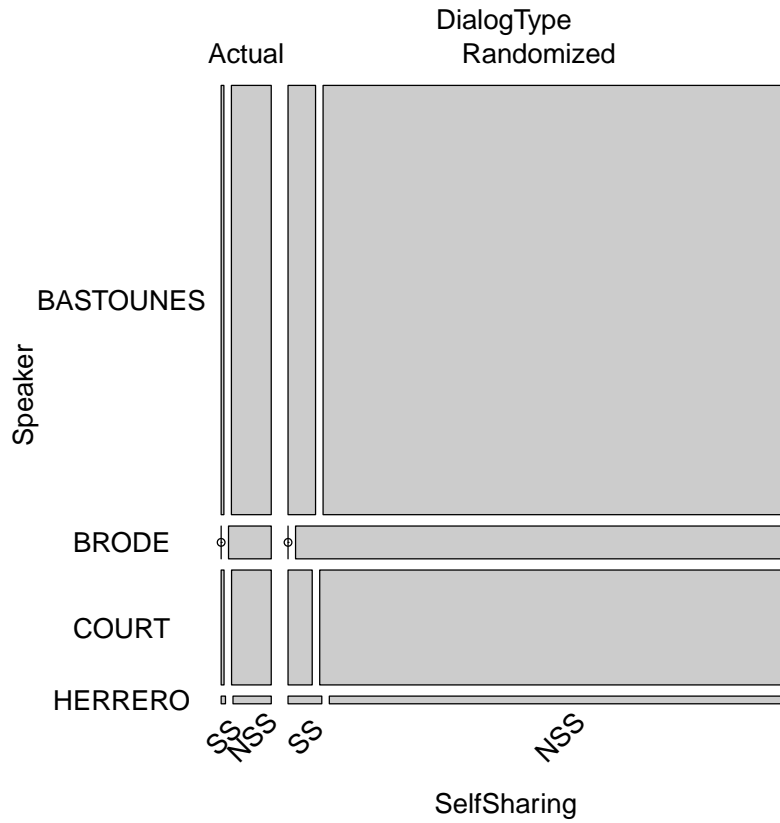




Table 5: Herrero case: counts of shared and nonshared tokens by dialogue-type, speaker, and n-bar

Speaker	DialogType Nbar	<u>Other Sharing</u>				<u>Self Sharing</u>			
		Actual		Randomized		Actual		Randomized	
		OS	NOS	OS	NOS	SS	NSS	SS	NSS
BASTOUNES1		63	613	737	6073	52	175	490	1780
	2+	7	2450	105	24465	16	803	92	8098
BRODE	1	7	47	92	448	0	18	0	180
	2+	2	184	14	1846	0	62	0	620
COURT	1	34	164	288	1692	17	49	124	536
	2+	2	640	223	6398	3	211	13	2127
HERRERO	1	3	24	27	243	2	7	13	77
	2+	0	30	0	300	0	10	0	100

Recall that the null hypothesis asserts that there is no more repetition in actual dialogue than in randomized counterparts. The contrasts between actual and randomized dialogues are not significant enough to allow rejection of the null hypothesis for any participant nor for any level of  $n$ -bar, whether for self-repetition or allo-repetition.

## 2. Discussion

The results of applying the method suggest that the null hypothesis must be retained: the dialogue does not present sufficient repetition of words or phrases to suggest that the interlocutors have engaged sufficiently to achieve mutual understanding. Despite the fact that the answers to the questions posed provided by Herrero are rational and mutually consistent, particularly given that the individual is a nonnative speaker of English and given that the answers are all one-word responses to polar interrogatives, there simply is not sufficient evidence here and on these measures to support the claim that Herrero understood the proceedings. In the actual legal case, as discussed in the introduction, the final decision did not hinge on the answer to the question of whether the defendant understood

the proceedings. The dialogue was produced to the Illinois Appellate Court, First District, with an implicit argument that the responses of the defendant were contextually appropriate, but with a null hypothesis about whether this is evidence of mutual understanding that is at odds with the arguments given in the introduction.<sup>45</sup>

*B. Case Study 2: State v. Cunningham*<sup>46</sup>

The Supreme Court of North Carolina heard an appeal in which the defendant sought a new trial because of faulty jury selection. The transcript of jury selection was considered to assess whether a particular potential jury member, Carnes, whose exclusion from the jury was not permitted, had been predisposed towards a particular verdict or understood the notion of presumed innocence. Defense counsel Murphy tried to explain the notion of presumed innocence and attempted to ascertain whether Carnes understood. Occasionally, Wolfe, for the prosecution, and the Court intervened. As an independent reader of the transcript of the relevant jury selection process, it is easy to form the opinion that Carnes did not understand what it meant to accept the concept of presumed innocence.<sup>47</sup>

*1. Results*

Table 6 shows the distribution of token counts across the categories studied. Figure 6 depicts the relative proportions of shared vs. nonshared tokens in the case of allo-repetition. None of the contrasts of interest are statistically significant for allo-repetition (i.e., actual vs. randomized tokens produced by each speaker for each level of *n*-bar, repeating tokens from the last turns of all of the other speakers). Figure 7 shows the same

---

<sup>45</sup> *Herrero*, 756 N.E.2d 234.

<sup>46</sup> *State v. Cunningham*, 474 S.E.2d 772 (N.C. 1996). The transcript is included in App. B. The data used here is taken from the public records of the State of North Carolina. *State v. Cunningham No. 232A91*, IBIBLIO, <http://www.ibiblio.org/pub/docs/nc-supreme-court/jun0493/cunningham.asc> (last visited Apr. 17, 2013).

<sup>47</sup> See *State v. Cunningham No. 232A91*, *supra* note 46.

proportions for self-repetition. Self-repetition effects are significant for Murphy (the defense attorney questioning the potential juror) and Carnes (the juror being questioned) in that actual repetition exceeds repetition in randomized dialogues, for both unigrams and 2+ grams (adjusted  $p \leq .01$ ). The effect also exists for the court with respect to unigrams (adjusted  $p \leq .01$ ). Actual repetition of 2+ grams is not significantly in excess of randomized repetition for the court, and neither level of  $n$ -gram yields significant actual self-repetition for Wolfe (the prosecutor).

Figure 6: Allo-repetition in the Cunningham case: shared vs. nonshared tokens by speaker in actual and randomized dialogue



Figure 7: Self-repetition in the Cunningham case: shared vs. nonshared tokens by speaker in actual and randomized dialogue

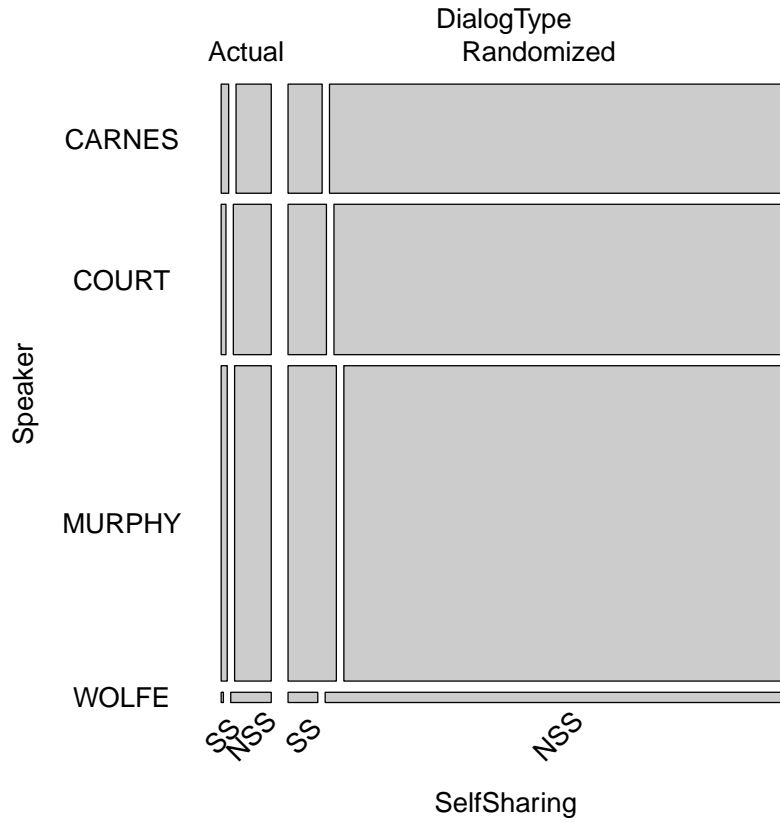


Table 6: State v. Cunningham: counts of shared and nonshared tokens by dialog-type, speaker, and n-bar

Speaker	DialogType Nbar	Other Sharing				Self Sharing			
		Actual OS	NOS	OS	NOS	Actual SS	NSS	SS	NSS
CARNES	1	159	933	1627	9293	110	254	558	3082
	2+	30	1827	285	18285	65	554	119	6071
COURT	1	175	1238	2029	12101	110	361	761	3949
	2+	32	2623	315	26235	42	843	295	8555
MURPHY	1	261	2679	4023	25377	324	656	2462	7338
	2+	23	5575	451	55529	80	1786	316	18344
WOLFE	1	11	100	163	947	4	33	46	324
	2+	1	167	14	1666	1	55	10	550

## 2. Discussion

Applying the method proposed to the data examined here supports the conclusion that the prosecutor and potential juror persisted in their respective communication strategies. The fact that allo-repetition effects that distinguish the actual dialogue from its ten randomized counterparts do not appear strengthens the argument that mutual understanding did not emerge during the interaction. This analysis coincides with the determination on appeal that the jury member did not unambiguously demonstrate understanding of the concept of presumed innocence.

## IV. CONCLUSIONS

While the current work is in the spirit of the traditions in statistical methods for authorship attribution,<sup>48</sup> it is focused on

<sup>48</sup> See generally GEORGE U. YULE, *THE STATISTICAL STUDY OF LITERARY VOCABULARY* (1944) (examining word-distribution from different portions of author's works throughout history); Harald Baayen et al., *Outside the Cave of Shadows: Using Syntactic Annotation to Enhance the Authorship Attribution*, 11 *LITERARY & LINGUISTIC COMPUTING* 121, 121-32 (1996) (reporting an experiment in which statistical measures and methods previously applied to words and their frequencies of use are applied to rewrite rules);

the attribution of mutual understanding and engagement in dialogue. The distinguishing features sought are not linguistic (words and phrases) in themselves but patterns of their use. Hence, the method can be considered to provide a tool for quantifying pragmatics. Other algorithmic means have been used to assess levels of understanding and awareness in dialogue. Apart from the explicit use of lexicalized feedback tags (e.g., “eh?”, “I see”), information contained in prosody has also been studied.<sup>49</sup> In analysis of other dialogues, the methods underlying

---

Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 INT’L J. SPEECH LANGUAGE & L. 1, 1–65 (2001) (testing language-based author identification techniques based on syntactic analysis, syntactically classified punctuation, sentential complexity, vocabulary richness, readability, content analysis, and errors); Carole E. Chaski, *Who’s at the Keyboard? Authorship Attribution in Digital Evidence Investigations*, 4 INT’L J. DIGITAL EVIDENCE, Spring 2005, at 1, 1–13 (applying computational, stylometric authorship attribution methods to crimes involving digital evidence); Jack Grieve, *Quantitative Authorship Attribution: An Evaluation of Techniques*, 22 LITERARY & LINGUISTIC COMPUTING 251, 251–70 (2007) (comparing thirty-nine different types of textual measurements commonly used in authorship attribution studies to determine which measurements are the best indicators); David I. Holmes, *Authorship Attribution*, 28 COMPUTERS & HUMAN. 87, 87–106 (1994) (quantifying literary style and looks at several variables to find the stylistic “fingerprints” of a writer); Kim Luyckx & Walter Daelemans, *Shallow Text Analysis and Machine Learning for Authorship Attribution*, PROC. FIFTEENTH MEETING COMPUTATION LINGUISTICS IN NETH., 2005, at 149–60, *available at* <http://lotos.library.uu.nl/publish/articles/000139/bookpart.pdf> (reporting on the use of syntax-based features as possible predictors for an author’s style and token-based features that are predictive to author style); Harold Somers & Fiona J. Tweedie, *Authorship Attribution and Pastiche*, 37 COMPUTERS & HUMAN. 407, 407–29 (2003) (testing whether authorship attribution techniques can distinguish between a deliberate imitation and its model); Carl Vogel & Gerald Lynch, *Computational Stylometry: Who’s in a Play?*, in VERBAL AND NONVERBAL FEATURES OF HUMAN-HUMAN AND HUMAN-MACHINE INTERACTION 1, 169–86 (2008) (applying automatic text classification techniques to quantifying strength of characterization within plays); George U. Yule, *On Some Properties of Normal Distributions, Univariate and Bivariate, Based on Sums of Squares of Frequencies*, 30 BIOMETRIKA 363, 363–90 (1938) (evaluating univariate and bivariate distributions and squaring every ordinate).

<sup>49</sup> See generally Jens Edlund et al., *The Effects of Prosodic Features on the Interpretation of Clarification Ellipses*, 2005 PROC. INTERSPEECH 2389

the current work yield significant differences between actual and randomized repetitions, which support the inference that significant engagement has happened, giving more certainty to attributions of mutual understanding,<sup>50</sup> thereby supporting the face validity of the methods. Further, recall that face validity is found elsewhere, in the correlation between repetition levels and task success in task-based dialogue.<sup>51</sup>

The capacity to assess and attain mutual understanding from a position external to a dialogue is important outside forensic contexts. The capacity to reliably and objectively make such attributions is also relevant in clinical diagnostics. Moreover, even if repetition behaviors do not figure explicitly into Diagnostic and Statistical Manual diagnoses of schizophrenia, analysis of repetition figures into current concepts of schizophrenia,<sup>52</sup> including attention to whether such analyses can lead to inappropriate diagnoses.<sup>53</sup> In preliminary study in this

---

(describing an experiment in which subjects listened to short dialogue fragments and judged what was actually intended by the elliptical clarification, based on prosodic features); Jens Edlund et al., *User Responses to Prosodic Variation in Fragmentary Grounding Utterances in Dialog*, PROC. INTERSPEECH 2006, at 2002 (testing whether subjects change their behavior to different fragmentary grounding utterances in a human-computer dialogue setting).

<sup>50</sup> Vogel & Behan, *supra* note 25.

<sup>51</sup> See Reitter & Moore, *supra* note 25, at 808–15.

<sup>52</sup> See generally Michael Covington et al., *Schizophrenia and the Structure of Language: The Linguist's View*, 77 SCHIZOPHRENIA RES. 85, 85–98 (2005) (evaluating schizophrenic language impairments to see how schizophrenia affects phonology, syntax and semantics); Gina R. Kuperberg, *Language in Schizophrenia Part 1: An Introduction*, 4 LANGUAGE & LINGUISTICS COMPASS 576, 576–89 (2010) (discussing language output in schizophrenia and the theory that language dysfunction in schizophrenia arises from abnormalities in semantic memory and/or working memory and executive function); Gina R. Kuperberg, *Language in Schizophrenia Part 2: What Can Psycholinguistics Bring to the Study of Schizophrenia . . . and Vice Versa?*, 4 LANGUAGE & LINGUISTICS COMPASS 576, 590–604 (2010) (applying online psycholinguistic methods to schizophrenic language).

<sup>53</sup> See Susan Trumbetta et al., *Language-Related Symptoms in Persons with Schizophrenia and How Deaf Persons May Manifest These Symptoms*, 1 SIGN LANGUAGE STUDIES 214, 228–53 (2001).

area,<sup>54</sup> transcripts previously analyzed by Steuber<sup>55</sup> in the identification of linguistic features that discriminate transcripts of interviews with individuals diagnosed with either schizophrenia or depression have been considered, finding no significant level of repetitions of others, but individual persistence via self-repetition. The potential use of the methods proposed in the suite of tools for diagnosis of syndromes with distinctive accompanying effects on conversational linguistic abilities is an area ripe for deeper exploration using the methods demonstrated here.

This article has introduced a method of interaction analysis based on repetition analysis that is distinct in analytical details from other analytical methods in the literature. The use of the methods has been demonstrated by analyzing transcripts that are freely available and with respect to which it is possible to draw upon independent assessments of the degree to which the transcribed conversations demonstrate engaged interaction and mutual understanding. Allo-repetition effects are taken to be those where the repetitions of tokens by an individual of dialogue partners immediately preceding contributions, summed over the conversation, significantly exceed in actual conversation the same measurement averaged over turn-randomized treatments of the conversation. Self-repetition effects are those where in the cumulative counts of repetitions of a speaker's immediately prior contribution significantly exceed for actual conversation the averaged accumulated counts for randomized counterparts. Self-repetition effects are taken to be indicative of speaker persistence with dialogue plans. Allo-repetition effects are taken to be signals of mutual engagement in dialog, and the conversations where these effects appear accord with independent intuition about the level of engagement and mutual understanding (distinct from mutual agreement) achieved within

---

<sup>54</sup> Carl Vogel, *Quantifying Interaction Synchrony as Evidence of Mutual Understanding*, 49 CORTEX (forthcoming 2013).

<sup>55</sup> Lucas C. Steuber, *Disordered Thought, Disordered Language: A Corpus-Based Description of the Speech of Individuals Undergoing Treatment for Schizophrenia* (2011) (unpublished M.A. thesis, Portland State Univ.), [http://dr.archives.pdx.edu/xmlui/bitstream/handle/psu/7087/Steuber\\_psu\\_0180\\_E\\_10321.pdf?sequence=1](http://dr.archives.pdx.edu/xmlui/bitstream/handle/psu/7087/Steuber_psu_0180_E_10321.pdf?sequence=1).



the conversation. Thus, the method is put forward for further exploration in contexts that require quantified analysis of attributions of mutual understanding in linguistic interaction. The argument made here is that where actual allo-repetition in a dialogue is not significantly in excess of its counterpart measure in randomized versions of dialogue, the level of engagement is insufficient to make confident attributions of mutual understanding.

*Appendix*

*A. Transcript of Interview with Michael Howard<sup>56</sup>*

Paxman Did you threaten to overrule him?  
 Howard I was not entitled to instruct Derek Lewis and I did not instruct him.  
 Paxman Did you threaten to overrule him?  
 Howard The truth of the matter is that Mr. Marriot was not suspended  
 Paxman Did you threaten to overrule him?  
 Howard I did not overrule Derek Lewis  
 Paxman Did you threaten to overrule him?  
 Howard I took advice on what I could or could not do  
 Paxman Did you threaten to overrule him?  
 Howard and acted scrupulously in accordance with that advice. I did not overrule Derek Lewis  
 Paxman Did you threaten to overrule him?  
 Howard Mr. Marriot would not suspend him  
 Paxman Did you threaten to overrule him?  
 Howard I have accounted for my decision to dismiss Derek Lewis  
 Paxman Did you threaten to overrule him?  
 Howard in great detail before the House of Commons  
 Paxman I note that you're not answering the question whether you threatened to overrule him.  
 Howard Well, the important aspect of this which it's very clear to bear in mind  
 Paxman I'm sorry, I'm going to be frightfully rude but—I'm sorry—it's a straight yes-or-no question and a straight yes-or-no answer did you threaten to overrule him?  
 Howard I discussed the matter with Derek Lewis. I gave him the benefit of my opinion. I gave him the benefit of my opinion in strong language, but I did not instruct him because I was not, er, entitled to instruct him. I was entitled to express my opinion and that is what I did.

---

<sup>56</sup> Yuri Prime, *Newsnight*, EVERYTHING2, <http://everything2.com/title/Newsnight> (last visited Apr. 17, 2013).

Paxman With respect, that is not answering the question of whether you threatened to overrule him.

Howard It's dealing with the relevant point which was what I was entitled to do and what I was not entitled to do, and I have dealt with this in detail before the House of Commons and before the select committee.

Paxman But with respect you haven't answered the question of whether you threatened to overrule him.

Howard Well, you see, the question is. . . .

*B. Transcript of Contested Jury Member Selection in State v. Cunningham.*<sup>57</sup>

MURPHY Do you understand, Ms. Carnes, that we have at law what is called the presumption of innocence, that is, a person who is charged with a criminal offense is presumed to be innocent until and unless the State can prove that person's guilt beyond a reasonable doubt?

CARNES Yes, sir.

MURPHY You understand that, don't you?

CARNES Yes, sir, I do.

MURPHY And, of course, you understand that the charge in this particular case is first-degree murder. It involves the shooting of a police officer. Do you understand that?

CARNES Yes, I do.

MURPHY And one of the things that you will be called upon to do is to apply the principles that we were talking about to this particular case if you sit as a juror.

CARNES Yes, I do.

MURPHY Now, it is one thing, of course, to say that you can do something and it may be entirely different.

CARNES Yes.

MURPHY That is, that you actually be able to do that, and that is really what I want you to search yourself about. I

---

<sup>57</sup> State v. Cunningham No. 232A91, *supra* note 46.

want you to think about that. You seem to be one who holds your opinion strong, and that's fine. Given that you have such a strong feeling about the death penalty in your statement that if a person takes another life, they should be put to death, given that Mr. Cunningham is charged with first-degree murder, as you sit there today, can you honestly say to yourself, not to me necessarily but to yourself, that you are able to presume Mr. Cunningham innocent?

CARNES Until he is proven guilty.

MURPHY Do you expect that to happen?

WOLFE Object.

COURT Sustained.

CARNES I don't know.

COURT Don't answer the question when I sustain it.

MURPHY I understand that if he is proven guilty of first-degree murder, then that would remove the presumption of innocence, but that is really not what I am asking you. Okay? What I am really asking you at this point is can you honestly, as he sits there right now, and as you sit in that seat right now, and nobody knows this any better than you, I'm just asking, can you honestly presume him to be innocent?

CARNES Yes, because I don't know what happened.

MURPHY Now, part and parcel of the principle of the presumption of innocence is the defendant's right not to testify, not to present any evidence if he doesn't want to, because he doesn't have that burden. The State has the entire burden of proof in a criminal case to satisfy you beyond a reasonable doubt of a person's guilt, if they can do that. Okay? Now, would it present a problem for you in returning a verdict of not guilty if the State fails to prove to you beyond a reasonable doubt the defendant's guilt and Mr. Cunningham didn't testify?

CARNES I'm not sure I follow that.

MURPHY Okay. If Mr. Cunningham doesn't testify in this

case, in your mind does that make the State's job any more difficult or easier?

CARNES I would think it would be more difficult.

MURPHY If he does not testify?

CARNES Yes, because they have to prove him innocent or guilty. I would think that he would have to testify, or need to.

MURPHY Okay. You understand that the State only has to prove guilt. They don't have to prove innocence.

CARNES Yes.

MURPHY And is it your expectation or is it your thinking now that we would have to prove that Mr. Cunningham is innocent?

CARNES Do I think you would have to prove it?

MURPHY Yes.

CARNES Yes, I thought that is what you would be trying to do.

MURPHY Trying to prove that he's innocent?

CARNES Yes.

MURPHY Do you understand that the burden of proof is on the State?

CARNES Yes.

MURPHY Not us?

CARNES Yes.

MURPHY You would still expect him, or us, to prove Mr. Cunningham is innocent. Correct?

CARNES Yes.

WOLFE I would ask for a clarification on the law on that, your Honor.

COURT Ms. Carnes—

CARNES He's getting me very confused.

COURT Okay. Let me explain to you. I think I told you that Mr. Cunningham has entered a plea of not guilty.

CARNES Yes.

COURT And under the law of North Carolina, the fact that he has been charged with a crime is not evidence of his guilt. He is not required to prove his innocence; he is presumed to be innocent.

CARNES Okay.

COURT The State of North Carolina has the burden of proof, and that burden is to prove each element of the offense of which he is charged beyond a reasonable doubt. Now, the law also says that Mr. Cunningham does not have to testify in his own behalf. He doesn't have to call any witnesses or present any other form of evidence, and that you cannot hold that against him. Do you understand that?

CARNES Yes, ma'am.

COURT Can you follow that law?

CARNES Yes, ma'am.

COURT Mr. Murphy?

MURPHY All right. Now, that's what I'm asking you, Ms. Carnes. The judge told you what the law is, and I think the district attorney also said the same thing to you. I thought I had explained that. I thought I said that. Now, the question is your ability to follow that law.

CARNES Yes.

MURPHY And that's what I'm asking you. That given your understanding at this point—and I trust that that is clear – is it your feeling that Mr.—we at this table would have to prove to you that Mr. Cunningham is innocent of this offense?

CARNES Yes.

MURPHY We offer her for cause.

WOLFE Object, Your Honor.

COURT Did you understand the explanation?

CARNES Yes, ma'am.

COURT And in light of my explanation that he is presumed to be innocent and is not required to prove his innocence, you would still require him to testify or to prove his innocence?

CARNES Right now he is innocent, or he is innocent until proven guilty. I understand that. But you are saying I need to—I'm sorry, I'm not sure.

COURT You need to slow down just a little bit.

CARNES He is innocent until proven guilty. I understand that, until he is proven guilty, before we can say he is

guilty. That, I understand.

COURT Which part is it that you don't understand?

CARNES Well, I thought I understood everything.

COURT Well, I told you that he is not required to prove his innocence.

CARNES Then I guess that means his attorney will have to prove that he is not guilty? He doesn't have to prove his innocence then—is that what you're saying—since he's innocent until proven guilty?

COURT Let me start over. I told you that the fact that he has been charged with an offense is not evidence of his guilt. You can't consider it as evidence of his guilt. I told you also that he is presumed to be innocent and is not required to prove his innocence. The State of North Carolina, represented by Mr. Wolfe and Ms. Brown, has the burden of proof. That burden is to prove each element of the offense with which Mr. Cunningham is charged beyond a reasonable doubt. The State has to carry that burden of proof and convince all twelve jurors beyond a reasonable doubt of each element of the offense before the jury may return a verdict of guilty. Mr. Cunningham is presumed to be innocent, and that presumption stays with him throughout the course of the trial unless the jury finds after they go into the jury room to deliberate that the State has carried its burden of proof. The law also says that Mr. Cunningham does not have to testify. He does not have to call any witnesses on his behalf or present any evidence. He is not required to prove his innocence. And that you, as a juror, cannot hold that against him. Do you understand that?

CARNES Yes.

COURT Were you confused?

CARNES Yes.

COURT I'm going to deny the challenge for cause at this point.

MURPHY Okay. Ms. Carnes, it is not my purpose to try to confuse you. That's why I want you to stop me

when we go along. If you don't understand anything that I have said, or if you need further clarification, stop me and we will ask the judge to do that because we don't want a confused juror. We want a juror who is clear with what they have to do. Okay?

CARNES Okay.

MURPHY Now, I do, however, want to pursue that with you just a little bit because I want to know how you feel about the matter and not just telling me things because you think that's what I want to hear. Okay? Because it's not what I want to hear; it's how you honestly feel about things. And what I want you to tell me is that if you would require the defendant to prove his innocence to you.

CARNES No.

MURPHY You would be satisfied then just to hear from the State and rely, if necessary, just on what the State presents to you on the guilt or innocence phase before you would return a verdict. Is that correct?

WOLFE Object.

MURPHY Well, the State has the burden.

WOLFE That is an improper statement of the law.

COURT Sustained as to form.

MURPHY I will rephrase the question. Can you require the State to prove to you, if they can, Mr. Cunningham's guilt beyond a reasonable doubt?

CARNES Well, if I understand what they are saying, they have to prove he is guilty and not require his innocence to be proven. He doesn't have to prove his innocence, I guess, is what I'm trying to say.

MURPHY And would you accept that? I mean—

CARNES Yes, if he doesn't want to prove his innocence, I would have to accept that.

MURPHY Okay.

WOLFE May we approach the bench just a minute, Your Honor?

COURT Yes.

MURPHY I guess I'm a little bit confused myself at this point, Ms. Carnes. Let's see if we can understand each



other. Okay? You had indicated something to the effect that if we didn't want to prove his innocence, that you would accept what the State offered?

CARNES I understand that he is innocent right now until proven guilty. So if they prove him guilty, I would accept the fact that he is guilty, if they prove him to be guilty.

MURPHY Okay. I guess I didn't understand what you meant when you said if we didn't want to prove his innocence.

CARNES And then I said until they prove him guilty. When they prove him guilty, then he is guilty, when they prove he is guilty.

MURPHY Okay. When you say when they prove him guilty, what do you mean?

CARNES When they prove that he did it, when they come up with all of the evidence that he did it.

MURPHY I suppose I'm having some problems with that. It sounds like you expect them to do that.

WOLFE Object.

CARNES No, I don't. I said—

COURT Sustained.

CARNES Well, I should have said if they do.

COURT Ms. Carnes, let me say when there is an objection, you need to stop talking.

# ON THE ROBUSTNESS OF AUTHORSHIP ATTRIBUTION BASED ON CHARACTER N-GRAM FEATURES

*Efstathios Stamatatos\**

## ABSTRACT

*A number of independent authorship attribution studies have demonstrated the effectiveness of character n-gram features for representing the stylistic properties of text. However, the vast majority of these studies examined the simple case where the training and test corpora are similar in terms of genre, topic, and distribution of the texts. Hence, there are doubts whether such a simple and low-level representation is equally effective in realistic conditions where some of the above factors are not possible to remain stable. In this study, the robustness of authorship attribution based on character n-gram features is tested under cross-genre and cross-topic conditions. In addition, the distribution of texts over the candidate authors varies in training and test corpora to imitate real cases. Comparative results with another competitive text representation approach based on very frequent words show that character n-grams are better able to capture stylistic properties of text when there are significant differences among the training and test corpora. Moreover, a set of guidelines to tune an authorship attribution model according to the properties of training and test corpora is provided.*

---

\* Assistant Professor, University of the Aegean.

## I. INTRODUCTION

Authorship attribution is the line of research dealing with the identification of the author of a text under investigation given a set of candidate authors (e.g., suspects) and samples of known authorship for each one of them. Indeed, in many forensic examinations, part of the evidence refers to texts (e.g., notes, e-mail messages, SMS messages, written reports, etc.). The ability to verify that a text was written by one of the suspects could be crucial to support a case. During the last decades, significant progress has been achieved in the automation of this procedure by incorporating statistical and/or machine learning techniques (i.e., algorithms that can learn from data).<sup>1</sup> There is strong potential for this technology to be used as evidence in a judicial process, given that it provides effective results in well-designed experimental tests. So far, a primitive and controversial technique has been used in British courts.<sup>2</sup> In addition, Chaski discusses examples of the use of a semiautomated author identification method in U.S. courts.<sup>3</sup>

From the machine-learning point of view, authorship attribution can be viewed as a multiclass, single-label classification problem (i.e., there may be multiple suspect authors, one of whom must be selected) and can be studied

---

<sup>1</sup> See Patrick Juola, *Authorship Attribution*, 1 FOUND. & TRENDS IN INFO. RETRIEVAL 234, 235, 284–86 (2006); Moshe Koppel et al., *Computational Methods in Authorship Attribution*, 60 J. AM. SOC'Y FOR INFO. SCI. & TECH. 9, 10–13 (2009); Efstathios Stamatatos, *A Survey of Modern Authorship Attribution Methods*, 60 J. AM. SOC'Y FOR INFO. SCI. & TECH. 538, 538 (2009).

<sup>2</sup> R.A. Hardcastle, *CUSUM: A Credible Method for the Determination of Authorship?*, 37 J. FORENSIC SCI. SOC'Y 129, 137–38 (1997).

<sup>3</sup> See Carol E. Chaski, *Who's at the Keyboard? Authorship Attribution in Digital Evidence Investigations?*, INT'L J. DIGITAL EVIDENCE, Spring 2005, at 9, 10–11 (providing examples of cases in which the syntactic analysis method of authorship identification has been used in U.S. courts); Carol E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 FORENSIC LINGUISTICS 1, 1–2 (2001) (discussing the admissibility of FBI forensic stylistics methods in a federal district court case).

along the lines of other text categorization tasks.<sup>4</sup> However, there are some properties of authorship attribution that differentiate it from other text categorization tasks.<sup>5</sup> First, and perhaps most important, the stylistic choices of an author are far more difficult to capture and quantify in comparison to topic-related information. Stylistic information is usually based on very frequent patterns that are encountered in texts by the same author. On the other hand, it is preferable to focus on stylistic choices that are unconsciously made by the author and remain stable over the text length. To this end, a very large number of such features have been proposed, including measures about the length of words or sentences, vocabulary richness measures, function word frequencies, character *n*-gram<sup>6</sup> frequencies, and syntactic-related or even semantic-related measures.<sup>7</sup> In several independent studies, it has been demonstrated that function words (defined as the set of the most frequent words of the training set) and character *n*-grams are among the most effective stylometric features, though the combination of several feature types usually improves the performance of an attribution model.<sup>8</sup>

Practical applications of authorship attribution usually provide a limited number of samples of known authorship unevenly distributed over the candidate authors. Therefore, it is essential for the attribution model to be able to handle limited and imbalanced training sets.<sup>9</sup> Moreover, the availability of many samples for one candidate author does not necessarily increase the probability that the author is the true author of

---

<sup>4</sup> See Fabrizio Sebastiani, *Machine Learning in Automated Text Categorization*, ACM COMPUTING SURVEYS, Mar. 2002, at 5 (listing “author identification for literary texts of unknown or disputed authorship” as an application of text categorization).

<sup>5</sup> See Stamatatos, *supra* note 1, at 553.

<sup>6</sup> For example, the character 3-grams of the beginning of this footnote would be “For”, “or ”, “r e”, “ ex”, etc.

<sup>7</sup> See Stamatatos, *supra* note 1, at 539–44.

<sup>8</sup> KIM LUYCKX, SCALABILITY ISSUES IN AUTHORSHIP ATTRIBUTION 124–26 (2010); Jack Grieve, *Quantitative Authorship Attribution: An Evaluation of Techniques*, 22 LITERARY & LINGUISTIC COMPUTING 251, 266–67 (2007).

<sup>9</sup> See Efstathios Stamatatos, *Author Identification Using Imbalanced and Limited Training Tests*, PROC. EIGHTEENTH INT’L WORKSHOP ON DATABASE & EXPERT SYS. APPLICATIONS: DEXA 2007, at 237, 237–41.

another text. This is in contrast to other text categorization tasks (e.g., thematic classification of texts) where well-represented classes have high prior probability.<sup>10</sup> In addition, in authorship attribution applications it is probable to have samples of known authorship on a certain thematic area (e.g., politics) while the unknown texts are on another thematic area (e.g., sports). The same can be said about the genre (e.g., known samples are scientific papers while the unknown texts are e-mail messages). In other words, in authorship attribution it is very likely to have heterogeneous training and test sets in terms of distribution of samples over the training authors, topic of texts, and genre of texts. Note that in text categorization research, it is usually assumed that the test set follows the properties of the training set.<sup>11</sup>

Most of the authorship attribution studies examine the simple case where the topic and genre are controlled in both the training and the test corpus.<sup>12</sup> While this differs from most practical applications, it aims at ensuring that the authorial style will be the crucial factor responsible for the differences among texts. In some cases, a variety of topics are covered but the

---

<sup>10</sup> See Stamatatos, *supra* note 1, at 540, 553.

<sup>11</sup> See Sebastiani, *supra* note 4, at 19.

<sup>12</sup> See Stamatatos, *supra* note 9 (addressing the problem of author identification); Moshe Koppel et al., *Authorship Attribution in the Wild*, 45 LANGUAGE RESOURCES & EVALUATION 83, 83–94 (2011) (explaining how similarity-based methods can be used with “high precision” to attribute authorship to a “set of known candidates [that is] extremely large (possibly many thousands) and might not even include the actual author”); Moshe Koppel et al., *Measuring Differentiability: Unmasking Pseudonymous Authors*, 8 J. MACHINE LEARNING RES. 1261, 1261–76 (2007) (presenting “a new learning-based method for adducing the ‘depth of difference’ between two example sets and offer[ing] evidence that this method solves the authorship verification problem with very high accuracy”); Efstathios Stamatatos et al., *Automatic Text Categorization in Terms of Genre and Author*, 26 COMPUTATIONAL LINGUISTICS 471, 471–95 (2000) (presenting “an approach to text categorization in terms of genre and author for Modern Greek”); Hans van Halteren et al., *New Machine Learning Methods Demonstrate the Existence of a Human Stylome*, 12 J. QUANTITATIVE LINGUISTICS 65, 65–77 (2005) (explaining how the ability to distinguish between writings of less experienced authors “implies that a stylome exists even in the general population”).

same topics may be found in both the training and test set.<sup>13</sup> Although this setting makes sense in laboratory experiments, it is rarely the case in practical applications where usually the available texts of known authorship and the texts under investigation are completely different with respect to thematic area and genre. The control for topic and genre in training and test sets provide results that may overestimate the effectiveness of the examined models in more difficult (but realistic) cases. In a recent study,<sup>14</sup> the authors present a cross-genre authorship verification experiment where the well-known *unmasking* method<sup>15</sup> is applied on pairs of documents that belong to two different genres (e.g., prose works and theatrical plays) and the performance is considerably decreased in comparison to intragenre document pairs. In order for authorship attribution technology to be used as evidence in courts, more complicated tests should be performed to verify the robustness of this technology under realistic scenarios.

In this paper, an experimental authorship attribution study is presented where authorship attribution models based on character *n*-gram and word features are stress-tested under cross-topic and cross-genre conditions. In contrast to the vast majority of the published studies, the performed experiments better match the requirements of a realistic scenario of forensic applications where the available texts by the candidate authors (e.g., suspects) may belong to certain genres and discuss specific topics while the texts under investigation belong to other genres and are about completely different topics. We examine the case where the training set contains texts on a certain thematic area

---

<sup>13</sup> LUYCKX, *supra* note 8, at 96–99.

<sup>14</sup> Mike Kestemont et al., *Cross-Genre Authorship Verification Using Unmasking*, 93 ENG. STUD. 340, 340 (2012).

<sup>15</sup> See generally Koppel et al., *Measuring Differentiability*, *supra* note 12, at 1264 (“The intuitive idea of unmasking is to iteratively remove those features that are most useful for distinguishing between A and X and to gauge the speed with which cross-validation accuracy degrades as more features are removed. . . . [I]f A and X are by the same author, then whatever differences there are between them will be reflected in only a relatively small number of features, despite possible differences in theme, genre and the like.”).

or genre while the test set includes texts on another thematic area or genre. Moreover, we make sure that the distribution of texts over the candidate authors differs in training and test sets, again to imitate realistic conditions. Two of the most successful stylometric features are tested: frequent words and character *n*-grams. Moreover, it is demonstrated that, when training and test corpora have significant differences, the most crucial decision concerns the appropriate selection of the representation dimensionality (i.e., number of features). Based on the experimental results, a set of general guidelines is provided to tune an attribution model according to specific properties of training and test corpora.

The next section compares the stylometric features we examine. Section III describes the corpus used in this study while Section IV includes the performed experiments. Finally, Section V summarizes the main conclusions and proposes future work directions.

## II. FREQUENT WORDS VERSUS CHARACTER *N*-GRAMS

An intuitive way to quantify a text is based on frequencies of occurrence of words. For authorship attribution, as well as any style-based text categorization task, the most frequent words have proved to be the most useful features.<sup>16</sup> Interestingly, in topic-related text categorization, very frequent words (e.g., articles, prepositions, conjunctions, etc.) are usually excluded since they carry no semantic information. Hence, they are frequently called “stopwords” or function words. There are two main methods to define a set of such words to be used in an authorship attribution model: 1) using a predefined list of words belonging to specific closed-class parts of speech, such as articles, prepositions, etc.,<sup>17</sup> or 2) using the most frequent words

---

<sup>16</sup> Stamatatos, *supra* note 1, at 540.

<sup>17</sup> Shlomo Argamon et al., *Stylistic Text Classification Using Functional Lexical Features*, 58 J. AM. SOC'Y INFO. SCI. & TECH. 802, 803 (2007); see also Ahmed Abbasi & Hsinchun Chen, *Applying Authorship Analysis to Extremist Group Web Forum Messages*, IEEE INTELLIGENT SYS., Sept. 2005, at 67, 68 (focusing on the use of lexical, syntactic, structural, and content-specific features).

of the training corpus.<sup>18</sup> In the latter case, the top words with respect to their frequency correspond to function words. As we descend the ranked list, we encounter more and more nouns, verbs, and adjectives (possibly related with thematic choices). One disadvantage of lexical features is that they fail to capture any similarity in cases of noisy word forms (probably the result of errors in language use). For example, “stylometric” and “stilometric” are considered two different words. Another shortcoming is that in some languages, mostly East Asian ones, it is not easy to define what a word is.

Nowadays, character *n*-grams provide a standard approach to represent texts. Each text is considered as a mere sequence of characters. Then, all the overlapping sequences of *n* consecutive characters are extracted. For example, the character 3-grams of the beginning of this sentence would be “For,” “or,” “r e,” “ex,” etc. Character *n*-gram features have several important advantages: simplicity of measurement; language independence; tolerance to noise (“stylometric” and “stilometric” have many

Figure 1: An example of an online article and the extracted main text.

The screenshot shows a web browser displaying an article on 'theguardian.com'. The page title is 'With honourable friends like these ...'. The author is Roy Hattersley, dated Saturday 17 May 2008. The article title is 'A Question of Honour' by Lord Michael Levy. The main text is highlighted with a red box. The text reads: 'A long life in politics is, inevitably, punctuated with regrets. I now must add to my failure to lead the Labour party or hold one of the great offices of state the comforting words which I spoke to Lord Levy during an afternoon at the height of the honours for sale controversy. No doubt the Metropolitan Police were right to conclude, as I thought at the time, that Tony Blair's fund-raiser-in-chief had no case to answer. But if disloyalty was an ineluctable offence, he would spend the rest of his life in Westminster. So this is a novel of a writer's...'. The page also features navigation links, social media sharing options, and a search bar.

<sup>18</sup> J.F. Burrows, *Not Unless You Ask Nicely: The Interpretative Nexus Between Analysis and Information*, 7 LITERARY & LINGUISTIC COMPUTING 91, 91–109 (1992).



common character 3-grams); effectiveness in authorship attribution tasks, as has been proven in several studies and competitions;<sup>19</sup> and they require a high-dimensional representation based on information difficult to understand by humans, so deception attempts are less likely to be successful. On the other hand, the high dimensional representation requirement means that they can only be used in combination with certain classification algorithms able to support thousands of features. Furthermore, they capture small pieces of stylistic information, making the interpretation of the stylistic property of text very difficult if not impossible. Such an interpretation is crucial in case the authorship attribution technology is used as evidence in a judicial process.

Another common intuition is that character *n*-grams unavoidably capture thematic information in addition to the stylistic information. Under the assumption that all the available texts are on the same thematic area, this property of character *n*-grams can be viewed as an advantage since they provide a richer representation including preference of the authors on specific thematic-related choices of words or expressions (e.g., vehicle vs. automobile). However, when the available texts are not on the same thematic area, a topic-independent approach to represent texts, like the use of a few dozen function words, sounds more promising. In this paper we examine this assumption and show that, contrary to intuition, character *n*-grams are more robust features than frequent words when the thematic area or the genre of the texts is not controlled.

### III. THE GUARDIAN CORPUS

The corpus used in this study is composed of texts published in *The Guardian* daily newspaper. The texts were downloaded using the publicly available API<sup>20</sup> and preprocessed to keep the unformatted main text.<sup>21</sup> An example is depicted in Table 1.

---

<sup>19</sup> See Grieve, *supra* note 8, at 259; Vlado Keselj et al., *N-Gram-Based Author Profiles for Authorship Attribution*, PROC. PAC. ASS'N FOR COMPUTATIONAL LINGUISTICS, 2003, at 255, 255–64; Stamatatos, *supra* note 1, at 538–56; Stamatatos, *supra* note 9, at 237–41.

<sup>20</sup> *Open Platform*, GUARDIAN, <http://explorer.content.guardianapis.com/>

Table 1: The Guardian corpus.

Author	Opinion articles				Book reviews
	Politics	Society	World	UK	
CB	12	4	11	14	16
GM	6	3	41	3	0
HY	8	6	35	5	3
JF	9	1	100	16	2
MK	7	0	36	3	2
MR	8	12	23	24	4
NC	30	2	9	7	5
PP	14	1	66	10	72
PT	17	36	12	5	4
RH	22	4	3	15	39
SH	100	5	5	6	2
WH	17	6	22	5	7
ZW	4	14	14	6	4
<b>Total:</b>	254	94	377	119	160

The majority of the corpus comprises opinion articles (comments). The newspaper describes the opinion articles using a set of tags indicating its subject. There are eight top-level tags (World, U.S., U.K., Belief, Culture, Life&Style, Politics, Society), each one of them having multiple subtags. It is possible (and very common) for an article to be described by multiple tags belonging to different main categories (e.g., a specific article may simultaneously belong to U.K., Politics, and Society). In order to have a clearer picture of the thematic area of the collected texts, we only used articles that belong to a single main category. Therefore, each article can be described by multiple tags, all of them belonging to a single main category. Moreover, articles coauthored by multiple authors were discarded.

In addition to opinion articles on several thematic areas, the presented corpus comprises a second text genre—book reviews. The book reviews are also described by a set of tags similar to the opinion articles. However, no thematic tag restriction was taken into account when collecting book reviews, since our main concern was to find texts of a specific genre that cover multiple

---

(last visited Mar. 2, 2013).

<sup>21</sup> Titles, names of authors, dates, tags, images, etc. were removed.

thematic areas. Note that since all texts come from the same newspaper, they are expected to have been edited according to the same rules, so any significant difference among the texts is not likely to be attributed to the editing process.

Table 1 shows details about *The Guardian Corpus* (“TGC”). It comprises texts from thirteen authors selected on the basis of having published texts in multiple thematic areas (Politics, Society, World, U.K.) and different genres (opinion articles and book reviews). At most 100 texts per author and category have been collected—all of them published within a decade (from 1999 to 2009). Note that the opinion article thematic areas can be divided into two pairs of low similarity, namely Politics-Society and World-U.K. In other words, the Politics texts are more likely to have some thematic similarities with World or U.K. texts than with the Society texts.

TGC provides texts on two different genres from the same set of authors. Moreover, one genre is divided into four thematic areas. Therefore, it can be used to examine authorship attribution models under cross-genre and cross-topic conditions.

#### IV. EXPERIMENTS

Two types of text representation features are examined—namely, words and character 3-grams. In both cases, the features are selected according to their total frequency of occurrence in the training corpus, a method proven to be suitable for authorship attribution tasks.<sup>22</sup> Let  $V$  be the vocabulary of the training corpus (the set of different words or character 3-grams) and  $F = \{f_1, f_2, \dots, f_i, \dots, f_v\}$  be the set of features ordered in decreased frequency of occurrence in the training corpus. Given a predefined threshold  $t$ , the feature set  $F_t$  includes all the features with  $f_i \geq t$ . The higher the  $t$ , the lower the dimensionality of the representation and vice versa. Therefore, it is possible to examine different sizes of the feature

---

<sup>22</sup> John Houvardas & Efstathios Stamatatos, *N-Gram Feature Selection for Authorship Identification*, in *ARTIFICIAL INTELLIGENCE: METHODOLOGY, SYSTEMS, AND APPLICATIONS* 77, 82–84 (Jérôme Euzenat & John Domingue eds., 2006).

set by modifying  $t$ . In this study, the following frequency threshold values were used: 500, 300, 200, 100, 50, 30, 20, 10, 5, 3, 2, 1.

The well-known Support Vector Machine (“SVM”) classifier<sup>23</sup> is used. It is a powerful classification model that can handle high dimensional and sparse data, and it is considered one of the best algorithms for text categorization tasks. The linear kernel (which is used to produce a linear boundary between the classes) is used since the dimensionality of the representation is usually high, including several hundreds or thousands of features.<sup>24</sup> There is no attempt to optimize the classification model by using different classification algorithms, since our aim is to highlight the capability of text representation features to remain robust in cross-topic and cross-genre conditions.

In each experiment, we follow the procedure described below:

- An attribution model is learned based on SVM and texts from a single topic category of TGC (e.g., Politics). At most, ten texts per author are used in the training phase. This provides an imbalanced training corpus.
- The learned classifier is applied to the texts of a category of TGC. Again, at most ten texts per author are used. If the selected category is Politics, that is the same as the topic category used in the training phase (intratopic attribution). The first ten texts are skipped, so there is no overlapping with the texts used in the training corpus. If the selected category is U.K., World, Society (cross-topic attribution) or Books (cross-genre attribution), then an imbalanced test corpus is compiled. Note that the distribution of the training corpus over the candidate authors is not necessarily the same with the corresponding distribution of the test corpus. This ensures that in case the attribution model favors the authors with the most training texts, it will produce many errors.

---

<sup>23</sup> See Corinna Cortes & Vladimir Vapnik, *Support-Vector Networks*, 20 MACHINE LEARNING 273, 274–75 (1995).

<sup>24</sup> See Thorsten Joachims, *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*, MACHINE LEARNING: ECML-98: 10TH EUR. CONF. ON MACHINE LEARNING, 1998, at 137.

### A. Intratopic Attribution

In the first experiment, we examine the simplest (but unrealistic) scenario that all texts included in both training and test corpora belong to the same genre and the same thematic area. That way, the personal style of the author is more likely to be the most significant factor for discriminating between texts. Using TGC, the texts of the Politics thematic category were used for both training and test (recall, there is no overlap between training and test texts). The distribution of test texts over the candidate authors is unavoidably similar to the corresponding distribution of the training texts.

The classification accuracy results are shown in Figure 2 for models based on frequent words and character 3-grams with a varying number of features (acquired by the different values of the frequency threshold). As can be seen, the models based on character 3-grams are far more effective than models based on words and achieve perfect classification accuracy. Their performance seems to increase with the dimensionality of the representation. This indicates that even the most rare character  $n$ -grams carry information that help the classifier to discriminate between author choices. Since all the texts are on the same thematic area, these choices also include preferences of the authors on specific thematic-related words or phrases.

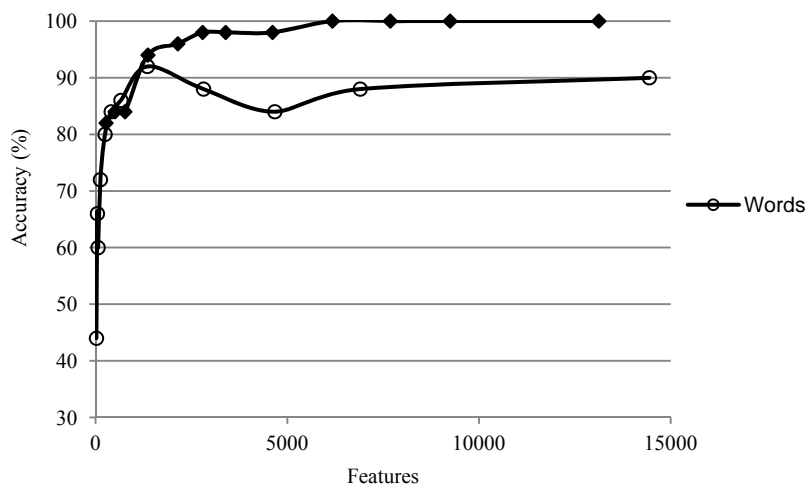


Figure 2: Performance of the intratopic attribution models (training on Politics, test on Politics).

As concerns models using word features, their performance constantly increases until about 1,500 features, then drops a little bit and then increases again. Hence, low-frequency words, probably associated with thematic-related choices, provide useful information to the classifier. In conclusion, when all the texts are controlled in terms of genre and topic, it seems that a very high dimensionality of the representation is a reliable option for both character  $n$ -gram and word features.

### *B. Cross-Topic Attribution*

Next, and more interestingly, we examine the cross-topic scenario where the classifier is trained using the Politics texts and then applied to the other thematic categories (that is, Society, World, and U.K.) of the same genre. Recall that the test texts distribution over the candidate authors does not follow the corresponding distribution of the training texts. The results are shown in Figures 3, 4, and 5, respectively.

In all three cases, character 3-gram features are significantly more effective than words. When the topic of the test texts is distant with respect to training texts (i.e., Society), the performance steadily increases until about 3,500 features and then significantly drops. In the cases of thematic areas unrelated with the training texts (i.e., World and U.K.), there is a similar pattern but the performance does not drop so much when the dimensionality increases. This indicates that low frequency features found in the training corpus (usually associated with thematic information) should be avoided when the thematic area of the test corpus is distant with respect to the thematic area of the training corpus. On the other hand, these rare features are not so crucial when the thematic area of the test corpus is not specifically related to that of the training corpus. The best performance is acquired by different frequency thresholds. In the World texts the performance peak is at about 6,000 features while in the U.K. texts the peak is at about 2,500 features. Therefore, it seems that one very crucial decision in cross-topic attribution to achieve high performance is the appropriate selection of the number of features.

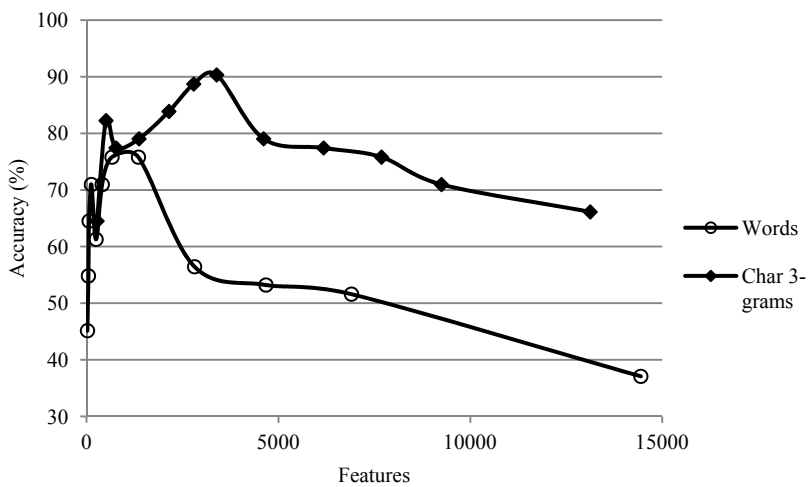


Figure 3: Performance of the cross-topic attribution models (training on Politics, test on Society).

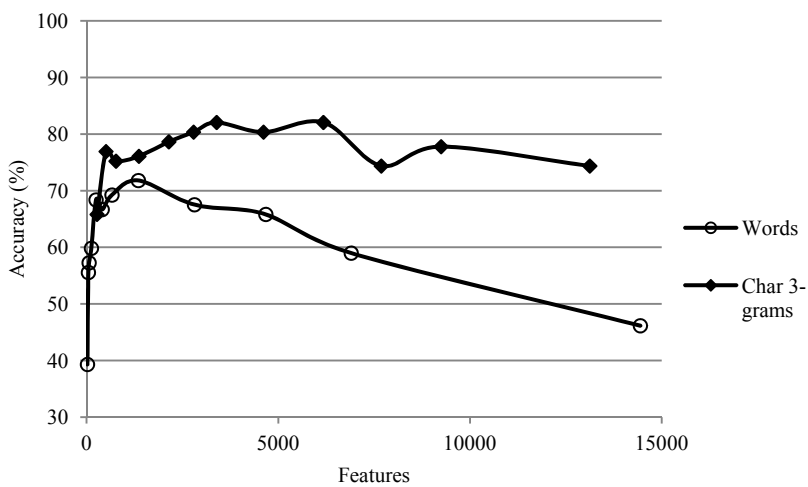


Figure 4: Performance of the cross-topic attribution models (training on Politics, test on World).

The performance of the models based on word features has similar characteristics. It steadily grows or remains practically stable until about 1,500 features and then drops significantly. The drop is much more abrupt in the case of Society texts

indicating that thematic-related words have a very negative effect when the test texts are about a topic distant from that of the training texts. In comparison to character  $n$ -grams, the word features are far more vulnerable by low frequency features in cross-topic conditions. Moreover, the models based on word features achieve their best performance with about 1,000 features (Society), 1,500 features (World), and 250 features (U.K.). Again, the appropriate selection of the dimensionality of the representation seems to be crucial. In comparison to character  $n$ -grams, word features need lower dimensionality to achieve good results in cross-topic attribution.

### *C. Cross-Genre Attribution*

Finally, we applied the classifier learned on opinion articles about Politics to texts of another genre, book reviews. As with the cross-topic experiments, the test set is imbalanced but its distribution over the candidate authors does not follow that of the training texts. The classification accuracy results for attribution models based on word and character 3-gram features are shown in Figure 6.

Again, character  $n$ -gram representation seems to be far better than the word representation. The best achieved performance is lower than all the best performances for the three cross-topic experiments, indicating that cross-genre attribution is a more difficult case. However, the average performance of the cross-genre models is very close to the average performance of the cross-topic models. Another interesting point is that the best performance is achieved with considerably higher dimensionality (about 9,000 features) with respect to the best performance of the cross-topic attribution models. It seems that low frequency features, probably related to thematic information, are helpful in cross-genre conditions. Some of the book reviews included in the test corpus may refer to books about Politics. Hence, when text genre varies between training and test corpora, topic-related choices may assist the attribution model.



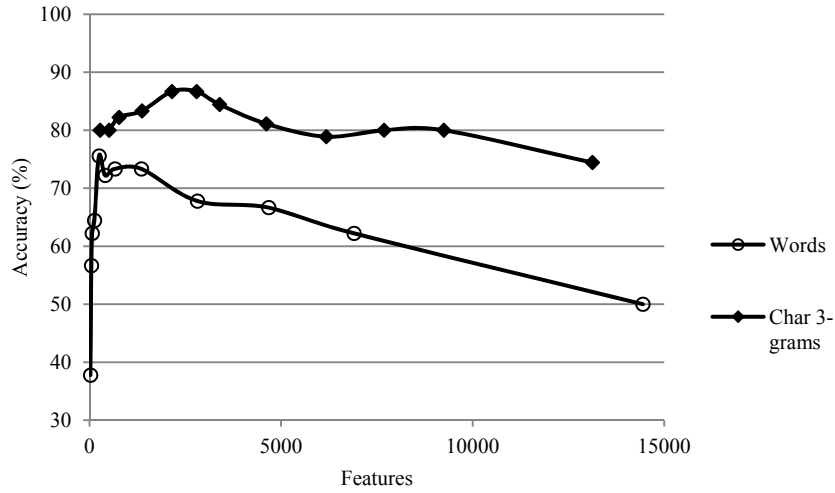


Figure 5: Performance of the cross-topic attribution models (training on Politics, test on UK).

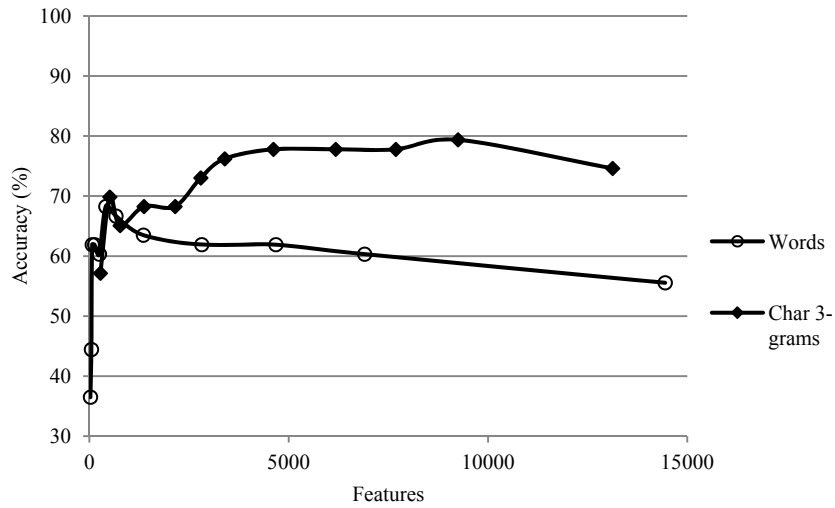


Figure 6: Performance of the cross-genre attribution models (training on Politics, test on Book reviews).

The models based on word features achieve their best performance at about 400 features, far lower than the character  $n$ -gram representation. However, the performance of the models based on more features does not drop dramatically as happens in cross-topic experiments. Again, this confirms the above conclusion about the usefulness of thematic-related information

in cross-genre attribution. On the other hand, the appropriate selection of the number of features is very important to achieve the best possible results.

## V. DISCUSSION

One main conclusion of this study is that, in addition to the simple intratopic attribution, character  $n$ -grams produce models more effective and robust than those based on word features in both cross-topic and cross-genre conditions. In general, models based on words require fewer features to achieve their best results, but they are significantly inferior to the best models based on character  $n$ -grams. An authorship attribution model based on character 3-grams in combination with a SVM classifier with linear kernel, although simple, proves to be very effective and can be used as a baseline approach, with which every new or advanced model should be compared.

The simple scenario of intratopic (in combination with intragenre) attribution seems to be a relatively tractable problem for current technology. The performance based on both character  $n$ -grams and words is very high, and unlikely to be matched by human experts, even when there are multiple candidate authors and relatively short texts. However, taking into account only such cases, the accuracy of the attribution models may be overestimated.<sup>25</sup> The presented cross-topic and cross-genre experiments show that the performance is affected sometimes considerably when topic and genre of training and test texts are not controlled. On the other hand, in such difficult cases, if the models are fine-tuned to the appropriate dimensionality of the representation, then the classification results remain surprisingly high. Hence, in the general case of applying authorship attribution technology to real world applications, a one-model-fits-all approach is not adequate. According to the properties of the texts of known authorship and the texts under investigation, one should fine-tune the attribution models appropriately to maintain a high level of effectiveness.

---

<sup>25</sup> See LUYCKX, *supra* note 8, at 4; Kestemont et al., *supra* note 14, at 343.

Several observations from the performed experiments may be used as guidelines for tuning an attribution model:

- In intratopic attribution, a very high dimensionality of the representation is advisable. Surely, high frequency features are the most important. However, it seems that low frequency features also contribute to the discrimination ability of the model.
- In cross-topic attribution, if the topic is distant from the topic of the training texts (e.g., Politics vs. Society, World vs. U.K.), low frequency features should be avoided. Since they are closely related with nuances of thematic choices, they harm the effectiveness of the attribution models. The crucial decision is the appropriate selection of the representation dimensionality.
- In cross-topic attribution, if the topic is not specifically associated to the topic of the training texts (e.g., Politics vs. World), low frequency features are not so harmful. However, it is better to exclude them, and again there is a crucial decision about the appropriate selection of the representation dimensionality.
- In cross-genre attribution, a high representation dimensionality seems to be advisable, especially when topic similarities are likely to be found in training and test texts.

An interesting conclusion that can be drawn from this study is that cross-topic attribution where the topic of the training and test texts can be regarded as highly dissimilar (e.g., Politics vs. Society) may be more challenging than cross-genre attribution. Additionally, in cross-genre attribution, perhaps counterintuitively, models based on thousands of features (both character  $n$ -grams and words) are either better than or competitive with ones that use only a few hundreds of features.

Surely, more experiments are needed to verify all these conclusions. An interesting direction for future work is to explore the role of the candidate set size and how it affects the appropriate representation dimensionality. The combination of different feature types should also be examined since this approach usually improves the performance of the attribution models, as is exemplified by some of the most successful participant methods in the recently organized competitions on

authorship attribution.<sup>26</sup> Finally, a missing block in the authorship attribution research that is necessary to use this technology as evidence in court is the ability to explain the automatically derived decisions. In the case of attribution models based on low-level information, like character  $n$ -grams, that seem to be the most robust and effective approach, what is needed is a way to associate this highly dimensional information to some human interpretable high-level features.

---

<sup>26</sup> See Shlomo Argamon & Patrick Juola, *Overview of the International Authorship Identification Competition at PAN-2011* (Sept. 19–22, 2011), <http://www.uni-weimar.de/medien/webis/research/events/pan-11/pan11-papers-final/pan11-authorship-identification/juola11-overview-of-the-authorship-identification-competition-at-pan.pdf>; Patrick Juola, *An Overview of the Traditional Authorship Attribution Subtask Notebook for PAN at CLEF 2012* (Sept. 17–20, 2012), <http://www.uni-weimar.de/medien/webis/research/events/pan-12/pan12-papers-final/pan12-author-identification/juola12-overview-of-the-traditional-authorship-attribution-subtask.pdf>.

## ON ADMISSIBLE LINGUISTIC EVIDENCE

*Malcolm Coulthard\**

### PREAMBLE

This is a very unconventional journal article, the likes of which I have never before written. It is based on a paper that was conceived of and written for the Authorship Attribution Workshop (“Workshop”) hosted at Brooklyn Law School in October 2012 with the intention of exploring the boundaries of admissibility of linguistic evidence in U.S. courts. This paper focuses on admissible linguistic evidence in an English court and explores whether some or all of it would be accepted in a U.S. court, where the *Daubert* acceptability criteria,<sup>1</sup> particularly information about known rates of error, are more rigorous than the criteria currently in force in the U.K. Interestingly, it is likely that *Daubert*-like criteria will be introduced into the U.K. in the not too distant future, so it was not just academic curiosity that that led me to inquire whether my evidence would be admissible. Specifically, I wondered if in the U.S. I would be permitted to express an opinion on the evidence or only to act as a “tour guide,”<sup>2</sup> simply presenting the linguistic evidence to the court without evaluation. The general consensus of the Workshop’s evidence experts was that most of my evidence would indeed be allowable in a U.S. court.

Comments made during the Workshop about my presentation and analytic advances outlined by Dr. Tim Grant during his presentation have led me to revise and add to my analysis. As a consequence, what you will read below is, I hope, a more

---

\* Federal University of Santa Catarina, Brazil.

<sup>1</sup> See *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579 (1993).

<sup>2</sup> See generally Lawrence Solan, *Linguistic Experts as Semantic Tour Guides*, 5 FORENSIC LINGUISTICS 87 (1998).

convincing and more soundly based analysis of the evidence, and one that would comply better with the *Daubert* criteria. I leave it to you, the reader, to reach your own decision on admissibility. Interestingly, and again uniquely in my own experience, I will be able to present my evidence in court for a second time later this year because the first trial ended with a hung jury.<sup>3</sup>

#### INTRODUCTION

Professors Peter Tiersma and Larry Solan note that although “[U.S.] courts have allowed linguists to testify on such issues as the probable origin of a speaker, the comprehensibility of a text, whether a particular defendant understood the *Miranda* warning, and the phonetic similarity of two competing trademarks,” in other linguistic areas the situation is more problematic, as the system now requires evidence to conform to the *Daubert* principles.<sup>4</sup> Solan notes,

it must be conceded that, in cases where conclusions depend on observations about the frequency or rarity of particular linguistic features in the texts under examination, many linguists would have considerable difficulty in stating a “known rate of error” for their results, even if this phrase is interpreted as a likelihood ratio. It is for this reason that some linguists will be forced to change their way of reaching and presenting their opinions, while others may choose to see their role more as that of “tour guides” than opinion givers.<sup>5</sup>

Solan goes on to address the problem that is unique to experts in linguistics—the fact that the judges of fact, whether they be actual judges or jury members, are seen for most

---

<sup>3</sup> As the case is still ongoing, I have changed the names of all of the participants.

<sup>4</sup> Peter Tiersma & Lawrence M. Solan, *The Linguist on the Witness Stand: Forensic Linguistics in American Courts*, 78 *LANGUAGE* 221, 221 (2002).

<sup>5</sup> MALCOLM COULTHARD & ALISON JOHNSON, *AN INTRODUCTION TO FORENSIC LINGUISTICS: LANGUAGE IN EVIDENCE* 210 (2007) (citing Solan, *supra* note 2).

purposes to be their own experts in the area of language use and interpretation. The law is, much of the time, concerned with the meaning(s) that ordinary speakers attach to words and expressions.<sup>6</sup> Even so, Solan argues that there is still a role for the linguist, which is to explain and elucidate facts *about* language and usage as a result of which judge and jury will then be in the same position as the linguist and so can make linguistically informed decisions.<sup>7</sup> He explains that his linguistic training has made him “more sensitive to possible interpretations that others might not notice” and as a consequence he can point these out to the jury. However, he adds, “[O]nce I point these out and illustrate them clearly, we should start on an equal footing.”<sup>8</sup>

One of Solan’s points that is crucially relevant to what follows is that, although juries and judges may well be able to process words, phrases, and even sentences as well as any professional linguist, they may have problems with long documents or with a series of related documents because they may not be able to make the necessary links: “Of course a jury can read the document[s]. . . . But not all jurors, without help, can focus on a phrase in paragraph 24 of a contract that may have an impact on how another word should be interpreted in paragraph 55.”<sup>9</sup>

To facilitate a discussion of Solan’s points, I present below an edited version of an expert report I wrote where there was one questioned email and tens of thousands of emails available for searching written by many authors whose authorship was unchallenged. As a Coda, I add a new analysis produced as a consequence of the stimulating discussion at the Workshop.

## I. EXPRESSING OPINIONS

The lawyers in the case I discuss below wanted me to express my opinions using degrees of likelihood: “it is

---

<sup>6</sup> Solan, *supra* note 2, at 91.

<sup>7</sup> *Id.* at 92.

<sup>8</sup> *Id.*

<sup>9</sup> *Id.* at 94.

(quite/very) (un)likely that *X* is the author of the email.” However, as Philip Rose argues convincingly, expressing an opinion in this form is tantamount to expressing an opinion on the likelihood of the accused being guilty, which is the exclusive role of the judges of fact.<sup>10</sup> All that a linguist can comment on is the degree of similarity or difference between linguistic choices in the questioned and the known texts. Rose supports his argument by pointing out that no expert can make an estimate of the likelihood of guilt or innocence on the basis of the linguistic evidence alone; only those with access to all the available evidence can assess the value of each piece of it.<sup>11</sup> For this reason, I prefer to approach questions of authorship attribution as a two-stage process, asking first if the choices in the questioned document are *compatible* with choices made by the potential authors in their known documents. If the choices are not compatible, no further analysis is undertaken. Then, as a second stage for those candidate author(s) for whom the choices are indeed compatible, one comments on how *distinctive* the particular linguistic choices are, on a five-point scale from *not distinctive* to *exceptionally distinctive*.

## II. THE BRIEF

I was asked to express an opinion on the likely authorship of a questioned email sent from the email account of a Mr. Stephen Goggin to a Mr. Denis Juola at 16.30 on July 23, 2004. I was briefed that, given the timing and content of the email, in particular the knowledge of and explicit reference to an earlier phone call to Mr. Juola timed at 14.50, only a small number of people—Mr. Goggin; Mr. Tim Widdowson, the CEO; Mr. John Shuy, the Finance Director of MaxiSoft; and possibly their PA, Ms. Janet Gavalda—could have been in a position to author and type the email. I was asked to proceed on the assumption that, although the email was sent from Mr. Goggin’s e-account, it may not have been physically typed on his computer, because

---

<sup>10</sup> PHILIP ROSE, FORENSIC SPEAKER IDENTIFICATION 76 (James Robinson ed., 2002).

<sup>11</sup> *Id.* at 68.



Ms. Gavalda had authorized access, which included the facility to send emails in his name from her own machine.

### III. TEXTS

#### *A. Emails*

I was given electronic access to a large, though selective, database of some 190,000 emails and other texts, including all those authored by Goggin, Widdowson, and Shuy. During my analysis, it became evident that it would have been useful to be able to search in addition a corpus of emails written by Ms. Janet Gavalda in her own voice. However, there was no separate collection of her output available, and so it was only possible to examine those occasional emails authored by her which happened to have been reproduced in other emails sent or received by Goggin, Widdowson, and Shuy, or by other authors included in the database.

My initial analysis focused on three emails: the questioned email sent at 16.30 on July 23rd, and two undisputed emails, one sent by Goggin to Juola at 17.02 and another sent by Widdowson on August 18th to Shuy and Gavalda titled "Chief Exec's Update."

#### *B. Minutes*

In addition, I examined eight sets of contemporaneous committee meeting minutes that had been produced by Ms. Gavalda over a fourteen-month period from April 2003 until June 2004.

#### *C. Handwritten Notes*

I was also provided with both scanned and transcribed versions of two handwritten entries for July 23rd in a notebook belonging to Mr. Goggin:

an untimed entry headed "Audit committee report" and consisting of brief notes of a telephone conversation with Widdowson and possibly also Shuy, concerning both an

“Audit committee report” that had been leaked to the *Guardian* newspaper and an article that was anticipated to appear shortly in another newspaper the *Sunday Times*. This conversation preceded the 14.50 phone call; a later entry in the notebook headed “D Juola 14.50, 23/07/04” consisting of notes of the topics covered during the 14.50 telephone call.

At a later date, I was provided with notes made by a financial analyst, Caldas, of a telephone conversation with Widdowson two days earlier, on July 21st, also discussing the leak to the *Guardian*.

#### IV. LINGUISTIC UNDERPINNING

My analysis will focus on linguistic choices and is based on the premise that all language production is rule governed. The underlying linguistic theory is that all speaker/writers of a given language have their own personal form of that language, technically labeled an *idiolect*. A speaker/writer’s *idiolect* will manifest itself in distinctive and cumulatively unique rule-governed choices for encoding meaning linguistically in the written and spoken communications they produce. For example, in the case of vocabulary, every speaker/writer has a very large learned and stored set of words built up over many years. Such sets may differ slightly or considerably from the word sets that all other speaker/writers have similarly built up, in terms both of stored individual items in their passive vocabulary and, more importantly, in terms of their preferences for selecting and then combining these individual items in the production of texts.<sup>12</sup>

Thus, whereas any speaker/writer can use any word at any time, what in fact happens is that they make typical and repeated selections and coselections of preferred words, which

---

<sup>12</sup> See, e.g., COULTHARD & JOHNSON, *supra* note 5, at 161; Malcolm Coulthard, *Author Identification, Idiolect and Linguistic Uniqueness*, 25 APPLIED LINGUISTICS 431 (2004); Timothy D. Grant, *Text Messaging Forensics: TXT 4N6: Idiolect Free Authorship Analysis?*, in THE ROUTLEDGE HANDBOOK OF FORENSIC LINGUISTICS 508, 508–09 (Malcolm Coulthard & Alison Johnson eds., 2010).

collectively constitute a kind of linguistic fingerprint. Admittedly, this analogy is not precise since a single fingerprint sample has all the necessary information, whereas a single piece of language data has only a minute fraction of the total.

Linguists divide all words into two groups, which they call content, or *lexical*, and formal, or *grammatical*. *Lexical* words are nouns, verbs, adverbs, and adjectives, and it is these words that carry almost all of the message or content of a text, as well as the features of the idiolectal distinctiveness of the author. The *grammatical* words are rather like cement or glue and bind the lexical words together. There are very large numbers of lexical words but only a few hundred grammatical words—thus, a speaker has a very wide choice of content words but a very limited choice of grammatical words. For this reason, linguistic authorship attribution, particularly when the texts involved are short, tends to focus on variation in the selection of the lexical words and on how much overlap there is between authorial choices in known and questioned texts.<sup>13</sup>

Complicating and partly determining the selection of individual lexical words is topic. Given the same basic topic, different speakers/writers will still choose to mention and/or omit different aspects and choose differing lexis to encode any given topic item. Thus, while the occurrence of individual lexical items shared between topically related texts is significant

---

<sup>13</sup> There is, of course, another tradition of authorship attribution represented in this volume by the papers written by Argamon, Juola, Koppel, and Stamatatos. Those works analyze almost exclusively high frequency items, which tend to be word fragments and short grammatical words. See Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL'Y 299 (2013); Patrick Juola, *Stylometry and Immigration: A Case Study*, 21 J.L. & POL'Y 287 (2013); Moshe Koppel et al., *Authorship Attribution: What's Easy and What's Hard?*, 21 J.L. & POL'Y 317 (2013); Efstathios Stamatatos, *On the Robustness of Authorship Attribution Based on Character N-Gram Features*, 21 J.L. & POL'Y 421 (2013). This type of analysis works well with long texts and large collections of texts, as a reading of the articles will confirm, but is unable to cope with very short texts like the questioned email in this case. See, e.g., Argamon & Koppel, *supra*. Both methods have strengths and weaknesses, but I have no doubt that in the future a much more successful method that combines the two will emerge.

in authorship attribution, much more significant is the shared occurrence of coselected items or what linguists call *collocates*, as for instance when *employee* is coselected or collocated with *disgruntled* and/or with *former*.

For example, the questioned email, which is presented in full below (and with the original typos), sets out a situation in which MaxiSoft is ***under attack*** by means of ***rumours*** that are being ***peddled*** by either ***disgruntled employees*** or ***competitors***, these ***rumours*** being concerned with ***revenue*** which, it is claimed, should not have been ***recognised*** and costs which have not been ***fully expensed***.

As we discussed on the telephone, it would appear that MaxiSoft is currently ***under attack*** from some quarter. There are various rhumours flying around that we anticipate will receive some press coverage over the comming days. We do not know the source of these rhumours, which may be from ***disgruntled*** (current/former) ***employees*** or unsuccessful ***competitors***.

One of the ***rhumours*** being ***peddled*** is that because of the delay in the finalisation of the HIS contract, we may have ***recognised*** some ***revenue*** assocaited with that work. However, I reassure you that such allegations are completely false and that we will refute and defend any such allegations. In addition, all the cost of supporting the HIS bid to date have been ***fully expensed***. This issue may not be raised in the press, but I thought I would let you know just in case.

*Text 1: Questioned email sent on July 23, 2004 at 16.30*

As I noted above, any speaker/writer can use any word at any time and thus for the vast majority of words we can find many instances of their use by large numbers of authors. For simplicity's sake, I will use the Google search engine to illustrate this observation. If we take the eleven word forms I have bolded in the questioned email above and use the Google search engine, we find that all of them are common, some extremely so—there are many millions of hits for most of the items, and even the least used of the group, *peddled*, occurs some 1.5 million times. In other words, none of these word

forms is in any sense rare. See Table 2 below for rounded occurrence figures:

<b>Word</b>	<b>Google Occurrences</b>
Under	5 billion
Attack	823 million
Disgruntled	13 million
Employees	727 million
Competitors	185 million
Rumours	50 million
Peddled	1.5 million
Recognised	85 million
Revenue	454 million
Fully	1.2 billion
Expensed	1.8 million

*Table 2: Google Word Frequency Searches on Feb. 29, 2012*

However, as noted above, what distinguishes speakers/writers and the texts they produce is their coselections. Thus, when we look at some of the coselections in the production of word sequences, we note how quickly the frequency of occurrence decreases as a given phrase lengthens. Here are two examples chosen from the end of the first paragraph of the questioned email:

<b>Words and Phrases</b>	<b>Google Occurrences</b>
competitors	185,000,000
unsuccessful competitors	16,100
or unsuccessful competitors	639
employees or unsuccessful competitors	0
* * * * *	
Disgruntled	12,800,000
disgruntled current	16,800
disgruntled current former	2,570
disgruntled current former employees	55
disgruntled current former employees or	1
disgruntled current former employees or unsuccessful	0

*Table 3: Google Word and Phrase Searches, on Feb. 29, 2012*

We find this same phenomenon of rapidly reducing numbers of occurrences when we examine the co-occurrence of individual words and short phrases which, although they have not been coselected in a strict linear sequence like those above, still co-occur in the same text. Again, as one would expect, the number of texts sharing a given set of co-occurring items decreases, often dramatically, each time one more item is added. Below, as exemplification, are the cumulative occurrence figures for the first three pairs of collocates pairs that I highlighted in the questioned email. I have presented the search figures in the sequence in which the collocate pairs occur in the email—note an “\*” has been used to indicate that I am also including instances where other words occur between the chosen pair of collocates.

Words and Phrases	Google Cumulative Occurrences
Under attack	18,000,000
+ Disgruntled * employees	5,500,000
+ Rumours * peddled	0

Table 4: Google Cumulative Searches on Feb. 29, 2012

It is very clear, without needing to include in the search any of the further narrowing coselections of *competitors*, *recognise* + *revenue* and *fully* + *expensed*, that the questioned email has a unique set of lexical coselections—they did not occur together in any of the billions of texts that Google searched.

Thus, we can see clearly that, although in theory anyone can use any word at any time, the topics they choose, the aspects of the topic they decide to focus on, and their preferred linguistic realizations ensure that texts quickly become linguistically unique. This raises the question of who in the software company conceptualized and then linguistically encoded the press problems in ways similar to those used by the author of the questioned email.

A search in the database yielded examples of Widdowson using most of the distinctive vocabulary items in a series of emails written over the period July 16 to August 19, 2004. All of these emails are concerned with the problems raised by the *Guardian* journalist.

In the case of the questioned email, we must also deal with features of typing and copyediting. Some typists are more accurate than others and, because typing is a semiautomated, learned activity, it is possible to characterize less competent typists by the kinds of fingering mistakes they make; I myself frequently missequence, or *metathesize*, letters, and *teh* in particular is a very common mistake in my typing. In addition to typing *mistakes*, i.e. misfingerings, which the typist will recognize as incorrect if s/he rereads what s/he has typed, texts also include what linguists distinguish as *errors*. Errors are nonstandard spellings and grammatical and punctuation choices which the typist does not recognize as such, of which *rhumours*, misspelled identically three times in the questioned email, is an example.

Potentially masking all this idiolectal evidence about a typist is the word-processor's spell-checker, which can save even a poor typist who doesn't proofread and makes not simply typing mistakes but also errors from betraying her/his incompetence. For instance, my spell-checking program automatically corrected the *teh* example above, not once but twice and also warned me that *rhumours* is a nonstandard spelling. Of course, another personal variable is if, when, and to what extent an individual typist actually bothers to use the spell-checker.

## V. ANALYSES

### A. Stephen Goggin as a Candidate Author

#### 1. Orthography

For its length, the questioned email has a comparatively large number of typing mistakes—four—and one repeated spelling error. There are several categories of mistake and some words have been categorized twice in the listing below because there are alternative possible explanations for the form which has been typed. The first four categories are typing *mistakes*, and the fifth is a spelling *error*:

1. metathesis of letters: *assocaited*, *currentlty*
2. omission of letter: *becase*
3. double keying: *comming*
4. additional letter: *currentlty*
5. spelling error: *rhumours*, *comming*

I have highlighted these items in bold in the email reproduced below:

As we discussed on the telephone, it would appear that MaxiSoft is **currentlty** under attack from some quarter. There are various **rhumours** flying around that we anticipate will receive some press coverage over the **comming** days. We do not know the source of these **rhumours**, which may be from disgruntled (current/former) employees or unsuccessful competitors.



One of the **rumours** being peddled is that **because** of the delay in the finalisation of the HIS contract, we may have recognised some revenue **associated** with that work. However, I reassure you that such allegations are completely false and that we will refute and defend any such allegations. In addition, all the cost of supporting the HIS bid to date have been fully expensed. This issue may not be raised in the press, but I thought I would let you know just in case.

An examination of emails which Goggin affirmed that he had sent from his computer around the period of the questioned email shows that they are completely error free. In particular, the 17.02 email, sent a mere thirty minutes after the questioned email, has no spelling or keying mistakes. In other words, Goggin did not send mistake- or error-filled emails from his computer.

## *2. Opening and Closing*

The questioned email has an in-text opening heading of “Strictly Private and Confidential” in bold. There are no examples of this heading in any Goggin emails. The message closes with “Best Regards,” yet the message sent to the same recipient, Juola, only half an hour later at 17.02 and accepted as authentic by Goggin ends simply with “Regards.” Indeed, an analysis of all the emails sent by Goggin to Juola in the preceding six months shows that some eighty percent of them end simply with “Steve,” and in the twenty percent of emails where there is a closing, it is, as in the 17.02 authentic email, invariably an unmodified “Regards.” There are no examples of “Best Regards.” In other words, neither the opening nor the closing of the questioned email were choices that Goggin made in his emails at the time.

## *3. Lexical Choices*

Three distinctive lexical choices in the questioned email are *disgruntled*, *peddled*, and *under attack*; none of them occur in any emails Goggin accepts as authentic. Neither does Goggin,

who was a salesman, not an accountant, send any emails with the phrases *recognising revenue* or *fully expensed*.

*4. Finding Regarding Goggin  
as a Candidate Author*

The linguistic choices made by the author of the email are not consistent with those instanced in Goggin's other emails.

*B. Others as Candidate Authors*

*1. Content and Expression*

The language of the questioned email has significant lexical links with that of the person(s) who briefed Goggin in the earlier telephone call already mentioned above, which was recorded in his notebook as "audit committee report." This person must have been Widdowson or Shuy because Goggin says they were the only other participants. Relevant words and phrases in Goggin's notes on this briefing are highlighted in bold in the extract below and can be compared with the same items occurring in the immediately following extracts taken from the questioned email:

Someone trying to suggest that **we have recognised revenue**

Take so long – **delay**

**Under attack**                      **competitor/disgruntled employee**

As we discussed on the telephone, it would appear that MaxiSoft is currently **under attack** from some quarter. There are various rumours flying around that we anticipate will receive some press coverage over the coming days. We do not know the source of these rumours, which may be from **disgruntled** (current/former) **employees** or unsuccessful **competitors**.

One of the rumours being peddled is that because of the **delay** in the finalisation of the HIS contract, **we**

may **have recognised** some **revenue** associated with that work.

We can see, highlighted in the text of the questioned email, all the important lexical items from the briefing notes not simply recurring but recurring in the same collocational groupings. In other words, the author(s) of these two messages which are closely related in time, the one spoken and the other written or dictated, is/are choosing to present the company's problem with the press within the same conceptual framework: that is, not as a legitimate, although admittedly annoying and distracting, investigation by a journalist but as a motivated "attack" either by aggrieved insiders or by those competing for contracts. Not only is the conceptualization of the problem in the email the same as in the telephone briefing but so also is its lexical encoding: "under attack," "disgruntled employees/competitors," "delay," and "we have recognised revenue."

These linguistic facts strongly suggest the possibility of single authorship; in other words, whoever briefed Goggin earlier in the day also authored the questioned email. A search of Shuy's emails did not produce examples of him using any of the central lexis used in the questioned email. Widdowson, however, does use much of this vocabulary.

Two days before the telephone briefing of Goggin, Widdowson briefed company analyst Caldas. In this briefing, the company is also presented as *under attack*, an attack which is characterized as *malicious* and which involves someone who is *feeding* to the *press* claims about *revenue* having been *recognised* before a contract has been signed. Caldas's notes include the following items

**disgruntled employee** dismissed False letter to **GRD**  
[*Guardian*]

*feeding* to *jornos*

why **rev recognised** before signed?

subject direct *malicious attack*

also signed & **RR'd** [revenue recognised]

co **under attack**

In an email sent to a market analyst on August 13th, Widdowson again refers to the problems with *The Guardian* and

again characterizes the encounter as *malicious* and as an **attack**: “[t]he last few weeks have really been quite extreme and we appreciate the quality of the advice provided and your dogged determination to see off this *malicious attack*.”

Five days later on August 18th, Widdowson circulated a text entitled “CEO Statement” in which he referred again to the problems with the *Guardian* journalist and used six of the lexical items that occurred in the questioned email, including the same collocations in the same close proximity:

Having had the initial *malicious rumour* planted . . . .

Our response to this direct **attack** was however measured. . . .

[T]here is little evidence that the *malicious rumours peddled* by the *Guardian* journalist have had any material effect on the perception of MaxiSoft in the healthcare IT supply market with either existing or prospective customers. It is an interesting contrast to note that most in the supply market see straight through the recent newspaper ‘noise’, speculating that it emanates from a **disgruntled former employee** seeking to further a particular selfish personal agenda.

We can compare this lexical encoding with the questioned email:

We do not know the source of these **rumours**, which may be from **disgruntled** (current/former) **employees** or unsuccessful competitors.

One of the **rumours** being **peddled** is that because of the delay in the finalisation of the HIS contract, we may have recognised some revenue associated with that work.

These particular lexical items do not co-occur in any other company emails, let alone in such close proximity to each other.

Widdowson also uses *peddle* on other occasions to disparage communications: in an October 1st email he refers to information having “been *peddled* around already” and on October 12th he characterises a Mr. Steer as “*peddling*.”

In addition, Widdowson, an accountant, unlike Goggin, does write frequently about *recognising revenue* and uses the expression “*fully expensed*.” In an email sent to Goggin on August 6th and titled “Message re *Guardian* Update,” Widdowson writes, “The balance of the SPfiN-related **revenue recognised** in 04 was in respect of earlier deliverables of existing product and services,” and on July 16th, a week before the questioned email was sent, in an email entitled “draft script for our friend at the *Guardian*,” Widdowson included the observation that “the value of R+D spend is confirmed as **fully expensed**.” Finally, while the heading of the questioned email **Strictly Private and Confidential** is very rare in company emails, it does occur in another email about this same *Guardian* investigation sent by Widdowson to Gavalda and then forwarded by Gavalda to the Executive Board on August 13, 2004:

MaxiSoft - THE HEALTH iNNOVATOR

**Strictly private and confidential**

In other words, all of the core vocabulary that is highlighted in the questioned email below is vocabulary that Widdowson also uses in other emails concerned with the problem of press coverage:

Strictly private and confidential

As we discussed on the telephone, it would appear that MaxiSoft is currently under attack from some quarter. There are various rumours flying around that we anticipate will receive some press coverage over the coming days. We do not know the source of these rumours, which may be from disgruntled (current/former) employees or unsuccessful competitors.

One of the **rumours** being **peddled** is that because of the delay in the finalisation of the HIS contract, we may have **recognised** some **revenue** associated with that work. However, I reassure you that such allegations are completely false and that we will refute and defend any such allegations. In addition, all the cost of supporting the HIS bid to date have been **fully expensed**. This

issue may not be raised in the press, but I thought I would let you know just in case.

To summarize: six central vocabulary choices made by the author of the questioned email occur in other emails on the same topic written by Shuy and three of them also occur in both the Goggin notes of the telephone conversation and in Caldas's notes. By contrast, there are no examples of Goggin making any of these vocabulary choices in his emails at this time.

Words and Phrases	Goggin emails	Goggin Notes	Questioned email	Widdowson <i>Guardian</i> emails	Caldas notes
attack	NO	YES	YES	YES	YES
Recognise(d) + revenue	NO	YES	YES	YES	YES
Disgruntled + employee(s)	NO	YES	YES	YES	YES
Peddle + rumour(s)	NO	NO	YES	YES	NO
fully expensed	NO	NO	YES	YES	NO
Strictly private and confidential	NO	NO	YES	YES	NO

Table 5: Comparison of Occurrences of Six Crucial Linguistic Encodings

## 2. Finding Regarding Others as Candidate Authors

Significant lexical choices in the questioned email are **consistent** with choices Widdowson makes elsewhere, particularly in emails about the problem with the *Guardian* journalist. In addition, these coselections do not occur in emails sent by anyone else and so are **distinctive**.

### 3. Orthography

While the content and expression of the questioned email share important features with other texts authored by Widdowson, the frequency of mistakes is certainly atypical of his normal production, which displays only the occasional mistake like “furture” in the August 13th document. Thus, Widdowson is not an obvious candidate for typist of the email.

I was asked to consider the possibility that the questioned email had been dictated to Ms. Gavalda and, as noted above, I was provided with a set of her minutes. The task of typing a dictated email is in some ways very similar to taking minutes—in both cases, it is the conversion of the spoken content of others into typewritten form.

A comparison of the type and frequency of the mistakes in the questioned email with those in a randomly selected set of Ms. Gavalda’s minutes produced in September 2003 identifies her as a candidate typist. Below are some mistakes and errors from these minutes. It will be seen that she makes mistakes in all of the five categories identified above:

1. metathesis of letters: **palce**; **strentghs**; **addiotnal**; **terroritires**; **surpiring**; **abiltiy**; **juen**; **fari**;
2. omission of letters: **announcment**; **arrangemnt**; **launcing**; **takig**; **dicussion**; **acountable**; **terminte**; **rsourece**; **postion**; **stategy**; **surpiring**; **expections**; **rining**; **contractr**;
3. double keying: **haave**; **theem**;
4. additional letters: **decfision**; **etec**; **meetinig**; **damanges**; **incentivised**; **analystst**; **finajncial**; **happending**; **rsourece**; **announcmenet**; **renvenue**; **prodocuct**;
5. spelling error: **hussle**, (**hustle**); **disbute**, (**dispute**); **pharse**, (**farce**);

To convey an impression of the sheer frequency of Ms. Gavalda’s mistakes, I have pasted below an extract from another set of her minutes dated April 7, 2004:

PM – updated on the TAW note. Have asked for the **fucnational** heads to prepare a little script and have had two in, awaiting the rest.

**Discssuion** whether one or individual – one but will include individual ones as well.

Can get **stared** on the employee representatives, ought to get going – RR taking that forward around payrolls. Can be used for redundancy as well.

Making good progress with carrying on the **templte** meetings (**identifiyinig** headcount reductions). Driven by accounts -0 drop date 26 April 2004. Needs to be done within the next week. Meetings agreed. Still waiting for date from RK. Can it be done virtually – Tuesday via telephone with. Sibsons are over in India – can do it over in Chennai.

TAW – make sure everyone is clear on the process. **Logalical** process of – database – mapped everyone to the new structure, all arrived on Monday. TAW, SPG and PM – biggest concern is in respect of NPfIT engagement and RK spoken to DR – **thinking** moving forward – major conflicts and outstanding issues – who is involved where and what does this mean in respect of the mappings.

NP **struutre** needs to encompass the central solution team (software delivery team) that sits between the **rpodocut** business and NP team (**deploymnete** or **engagement** team)

Confirm structures and names against the structures

TAW – np **strucurre** – most difficult area – where are we up to and when will it be finished. DR spoke to PM – RK, DR and RB – main area with regard to product delivery components. Central solution team is now effectively in 3 **component**, solution definition (identifying futures and obligations), manufacture and design and two **componesnt** solution delivery and support.



*4. Finding Regarding Ms. Gavalda  
as a Candidate Typist*

The range and nature of the mistakes in the questioned email are **compatible** with the mistakes that Ms. Gavalda makes in her contemporaneous minutes. In addition, the frequency is **distinctive**.

VI. OPINIONS

**Opinion 1:** The distinctive linguistic features of the questioned email are **not compatible** with Mr. Shuy's usage in other attested emails.

**Opinion 2:** The distinctive linguistic features of the questioned email are **not compatible** with Mr. Goggin's usage in other attested emails.

**Opinion 3:** The linguistic features of the questioned email are **compatible** with Mr. Widdowson's usage in other attested emails and with items in the notes made by recipients of two telephone conversations. These linguistic features are **distinctive**.

**Opinion 4:** The orthographic features of the questioned email are **compatible** with Ms. Gavalda's usage in contemporaneous minutes. These features are **distinctive**.

VII. CODA

Essentially, my expert report ended at this point, and the evidence I gave in court was based closely on it. However, I was unhappy that my evidence lacked any discussion of the frequency or rarity of the linguistic items I had claimed were crucial to the attribution of authorship. The analysis therefore was vulnerable to a cross-examiner suggesting that my analysis was not replicable and thus its credibility depended too much on my own credibility as an expert.

By a fortunate coincidence after I wrote the draft of my Workshop paper, I became aware of the work of doctoral student David Wright, who is using the Enron email database to develop computerized authorship attribution tools. Like me,

Wright is interested in the classificatory and attributory value of lexical as opposed to grammatical items. Thus his analyses, like mine, exclude function words such as articles, determiners, pronouns, and prepositions, which figure prominently in the analytic tools of many of the other authors in this volume.

Wright set out to investigate the degree of lexical similarity between different datasets and authors by examining the number of lexical types shared in the emails of selected Enron employees and then using the simple similarity metric Jaccard's coefficient<sup>14</sup> to evaluate the significance of his findings.

In an early exploratory study, he focused on the emails produced by a closed set of four Enron traders.<sup>15</sup> He found:

[Even though] the writers were all men of working age, all shared occupational and institutional goals, were writing on largely the same topics and within the same register, when [their sets of emails] were compared with each other the Jaccard similarity scores were low. [This clearly indicated] that, despite being socially and professionally very similar, the four authors had their own distinctive and identifiable lexicons.<sup>16</sup>

Blind testing demonstrated that the four authors could indeed be distinguished from each other by means of their individual lexical choices. This clearly has important implications for forensic authorship identification and attribution. Wright tested his method by setting out to match sets of 100 emails to the original author and was able to do so with a very high success rate.<sup>17</sup> In my case, there were by this point only two potential authors, Widdowson and Goggin (Shuy having already been

---

<sup>14</sup> This method is discussed in some detail in Grant's paper. Tim Grant, *TEXT 4N6: Method, Consistency, and Distinctiveness in the Analysis of SMS Text Messages*, 21 J.L. & POL'Y 467, 482 n.44 (2013).

<sup>15</sup> David Wright, *Existing and Innovative Techniques in Authorship Analysis: Evaluating and Experimenting with Computational Approaches to "Big Data" in the Enron Email Corpus*, 3D EUR. CONF. INT'L ASS'N FORENSIC LINGUISTS, Oct. 2012.

<sup>16</sup> David Wright, *Measuring Lexical Similarity for Authorship Identification: An Enron Email Case Study*, 28 LITERACY & LINGUISTIC COMPUTING (forthcoming 2013).

<sup>17</sup> *Id.*

discounted) and there is only one email, so the statistical route is not open to me. However, the question remains of whether the single email contains sufficient distinctive lexical information to make an attribution.

In undertaking this later analysis, I drew on a methodology proposed in Grant's article in this volume—a methodology which he developed for categorizing text messages.<sup>18</sup> Like me, Grant was working on a case with only two possible authors, but his data consisted of text messages.<sup>19</sup> Working from the known to the unknown, he took the two sets of known text messages and examined them in order to discover "whether there were features that discriminated consistently to some degree between the two writers in their known texts."<sup>20</sup> Grant only focused on features which were used predominantly by one author or the other and used "a rate of more than sixty-six percent of its total occurrence" as his criterion.<sup>21</sup>

Because in my case there was only one questioned email but vast numbers of comparison emails, I decided to restrict analysis to all and only the emails sent during a seven-month period, three months before and three months after the month in which the questioned email was sent. What I set out to do was, like Wright, to discover whether the lexical selections made by the author of the email were compatible with the usage of Goggin or of Widdowson. I decided to use Grant's criterion of majority usage to classify those items that occurred in both sets of emails as being characteristic of the usage of one of the authors, but I raised the required classificatory level of usage to a minimum of seventy-five percent.

My task was further complicated because while Grant had roughly equivalent sets of texts to compare, Goggin had produced over 2.5 times as many emails as Widdowson in the seven-month period—3,150 as compared with 1,234. For this reason, the raw scores for Goggin were reduced by sixty percent to normalize the frequencies before the comparison was made.

---

<sup>18</sup> Grant, *supra* note 14.

<sup>19</sup> *Id.*

<sup>20</sup> *Id.* at 480.

<sup>21</sup> *Id.*

Then the usage scores for all of the lexical items in the questioned email were compared. The scores for some items showed little difference in usage, but the relative frequencies of others were markedly different. Table 6 below shows first the items that were used only or more frequently by Widdowson (indicated in bold), then the Goggin items. It will be evident that there are many more distinctively Widdowson items in the list, and it becomes clear that the questioned email was composed using many more Widdowson than Goggin items.

Features (Normalized)	Total in 1243 TW emails	40% of Total in SG emails	Total	Percent in TW emails	Percent in SG emails
<b>Recognise + revenue</b>	<b>7</b>	<b>0</b>	<b>7</b>	<b>100</b>	<b>0</b>
<b>Peddle</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>100</b>	<b>0</b>
<b>Attack</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Coming Days</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Competitor</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Disgruntled</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Former employee</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Fully expensed</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Rumour</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>Strictly Private and Confidential</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>100</b>	<b>0</b>
<b>It would appear</b>	<b>7</b>	<b>0.4</b>	<b>7.4</b>	<b>95</b>	<b>5</b>
<b>To date</b>	<b>14</b>	<b>1.2</b>	<b>15.2</b>	<b>92</b>	<b>8</b>
<b>Delay</b>	<b>15</b>	<b>2</b>	<b>17</b>	<b>88</b>	<b>12</b>
<b>Best Regards</b>	<b>3</b>	<b>0.4</b>	<b>3.4</b>	<b>88</b>	<b>12</b>
<b>Press coverage</b>	<b>2</b>	<b>0.4</b>	<b>2.4</b>	<b>88</b>	<b>12</b>
<b>In addition</b>	<b>10</b>	<b>2.4</b>	<b>12.4</b>	<b>80</b>	<b>20</b>
<b>Currently</b>	<b>16</b>	<b>4</b>	<b>20</b>	<b>80</b>	<b>20</b>
<b>Employee</b>	<b>3</b>	<b>0.8</b>	<b>3.8</b>	<b>79</b>	<b>21</b>
<b>Issue + raise</b>	<b>3</b>	<b>0.8</b>	<b>3.8</b>	<b>79</b>	<b>21</b>
Just in case	1	3.6	4.6	22	78
Reassure	0	0.4	0.4	0	100
Completely	0	2	2	0	100

Table 6: Preferred Vocabulary Items for Widdowson and Goggin

What is evident in the highlighted version of the questioned email below is that a significant amount of the lexis is lexis that occurs predominantly in emails written by Widdowson (indicated in **bold**), whereas only three items are typical Goggin items, (indicated in *italic*).

Strictly private and confidential

As we discussed on the telephone, it would appear that MaxiSoft is currently under attack from some quarter. There are various rumours flying around that we anticipate will receive some press coverage over the coming days. We do not know the source of these rumours, which may be from disgruntled (*current/former*) employees or unsuccessful competitors.

One of the rumours being peddled is that because of the delay in the finalisation of the HIS contract, we may have recognised some revenue associated with that work. However, I *reassure* you that such allegations are *completely* false and that we will refute and defend any such allegations. In addition, all the cost of supporting the HIS bid to date have been fully expensed. This issue may not be raised in the press, but I thought I would let you know *just in case*.

Best Regards

#### CONCLUSION

Unlike Forensic Phoneticians, forensic linguists are never going to have reliable population statistics to enable them to talk about “the frequency or rarity of particular linguistic features.” I would argue, however, that the work of Wright and Grant opens a way to derive reliable and usable data about individual linguistic usage that can be applied in cases of authorship attribution. With tools like these, linguists can begin to make statements about frequency and likelihood of occurrence and, in cases where the data permits a Jaccard analysis, provide rigorous probability statistics.

# **TXT 4N6: METHOD, CONSISTENCY, AND DISTINCTIVENESS IN THE ANALYSIS OF SMS TEXT MESSAGES**

*Tim Grant\**

## INTRODUCTION

This paper presents a case study in forensic authorship analysis for SMS text messages. The case involves a domestic murder where the husband attempted to disguise the timing and mode of his wife's death through sending a series of SMS text messages from her phone.<sup>1</sup> Late in the evening on Sunday, January 17, 2009, the fire brigade was called to the home of Christopher and Amanda Birks in Stoke-on-Trent, UK. After phoning the emergency services, Christopher Birks entered the house and rescued his sleeping children from the lower floors. On the arrival of the firefighters, he informed them that his wife had gone to bed in the attic bedroom and that she must be trapped there. Placing themselves at considerable risk, firefighters entered the building to attempt a rescue from the top floor bedroom, but were only able to recover Amanda Birks' severely burned body. Apparently Amanda had been in bed

---

\* Centre for Forensic Linguistics, Aston University.

<sup>1</sup> Details of the case as described in this introduction were supplied to me in my role as External Expert Advisor to Staffordshire Constabulary and were provided in the course of a series of police briefings and prosecution case conferences. The case received limited local and national news coverage in the United Kingdom. See, e.g., *Businessman Admits Murdering Wife and Setting Fire to Her Home*, TELEGRAPH (Nov. 3, 2009, 7:00 AM), <http://www.telegraph.co.uk/news/uknews/crime/6488982/Businessman-admits-murdering-wife-and-setting-fire-to-her-home.html> [hereinafter *Businessman Admits Murdering Wife*].

when the fire began, and there was no indication that she had woken during it or made any attempt at escape.

Subsequent forensic examination showed that fibers recovered from Amanda's body were from her daytime clothes, and toxicology reports indicated that Amanda's lungs contained little or no carbon monoxide. Those findings indicate that when Amanda went to bed, she had not changed into her nightclothes and that when she was burnt, she had not been breathing and thus had not inhaled fumes from the rising fire. As can be typical where a body is badly damaged by fire, no precise cause of death could be determined.

Christopher Birks asserted that Amanda had been at home during most of the day on January 17. Midmorning, an employee stopping by to collect wages had seen Christopher and Amanda at the house, and when Christopher was out during the afternoon, a series of SMS text messages were sent from Amanda's phone. These messages, sent to Christopher, employees of their joint business, and other family members, indicated that Amanda had been going about her normal business at home. They also suggested that Amanda had had important discussions with Christopher about the state of their relationship and that she was going to bed early "relaxing with candles" in the attic bedroom. Since Amanda's phone was destroyed in the fire, these messages were collected from the various recipients' phones.

Christopher Birks' account, however, seemed contradictory to the evidence from the domestic burglar alarm. Expert technical examination of the alarm showed that it had been fully set during the afternoon of January 17 but also showed that it had not recorded Amanda's movements about her home.

A circumstantial case was built against Christopher Birks, which included the evidence of the textile fibers on Amanda's body, the lack of carbon monoxide in her lungs, the burglar alarm log, and finally, forensic linguistic evidence concerning the authorship of the text messages. Christopher Birks was charged with the murder of his wife and with the endangerment of the lives of his children and of the firefighters.

This article will describe the linguistic analysis carried out and consider the methodological and theoretical basis for that analysis. The method employed in this specific case was purely



descriptive. This article, however, uses that method as a starting point from which to discuss and develop a general statistical method for forensic analysis of text messages. This article posits that such a method will be a helpful tool in future cases for analyzing text messages and other short form messages.<sup>2</sup>

## I. FORENSIC AUTHORSHIP ANALYSIS

### A. *Stylometric Versus Stylistic Approaches to Analysis*

Forensic authorship analysis of written texts is achieving increasing acceptance in the United Kingdom's courts. Academically, there is a significant literature developing around the discussion of the theoretical presuppositions and implications of this work, the necessity and limits of quantification in the field,<sup>3</sup> and the law and application of the law concerning admissibility of such evidence.<sup>4</sup>

While much of this discussion is beyond the scope of this article, it is relevant to note that in the UK, admissibility is subject to review by the UK Law Commission.<sup>5</sup> On the basis of the published Law Commission report, it seems that the UK

---

<sup>2</sup> Such short form messages include Twitter feeds, Blackberry Messenger communication, and Facebook status updates.

<sup>3</sup> See generally Tim Grant, *Text Messaging Forensics: TXT 4N6: Idiolect Free Authorship Analysis?*, in THE ROUTLEDGE HANDBOOK OF FORENSIC LINGUISTICS 508, 508–22 (Malcolm Coulthard & Alison Johnson eds., 2010) [hereinafter Grant, *TXT 4N6*] (discussing the theoretical assumptions about the nature of an individual's linguistic personae and how such assumptions are expressed through methods adopted in authorship analysis work). See also Tim Grant, *Quantifying Evidence for Forensic Authorship Analysis*, 14 INT'L J. SPEECH LANGUAGE & L. 1, 1–25 (2007) [hereinafter Grant, *Quantifying Evidence*]; Moshe Koppel et al., *Authorship Attribution in the Wild*, 45 LANGUAGE RESOURCES & EVALUATION 83, 83–94 (2011).

<sup>4</sup> See Lawrence M. Solan & Peter M. Tiersma, *Author Identification in American Courts*, 25 APPLIED LINGUISTICS 448, 448–65 (2004); Blake Stephen Howald, *Authorship Attribution Under the Rules of Evidence: Empirical Approaches in the Layperson Legal System*, 15 INT'L J. SPEECH LANGUAGE & L. 219, 222–24 (2009).

<sup>5</sup> The UK Law Commission is a statutory body, independent of Parliament, whose function is to monitor and review laws and, where appropriate, make proposals for reform.

courts are likely to require a showing of the scientific validity of techniques not dissimilar to the *Daubert* criteria currently applicable in United States federal courts.<sup>6</sup> Although scientific validity and reliability do not require quantification, and quantification is by no means sufficient to demonstrate appropriate scientific status, there appears to be a presupposition in some literature that an appropriate quantified method can make it easier to demonstrate that a method is both reliable and valid.<sup>7</sup>

Following previous work,<sup>8</sup> I here draw a distinction between *stylometric* and *stylistic* approaches to authorship analysis. Stylometric approaches exemplified by scholars across the field seek to find or describe quantifiable markers of authorship, which in the general sense vary more between authors than within authors.<sup>9</sup> Typical stylometric markers include relative frequencies of different word classes or even nonword letter clusters referred to as *n*-grams. The demonstration of the usefulness of a stylometric marker of authorship requires that, for almost any sampled set of authors, there be significant differences in the occurrence of the marker between authors, regardless of other textual variables such as topic, register, or

---

<sup>6</sup> THE LAW COMMISSION, EXPERT EVIDENCE IN CRIMINAL PROCEEDINGS IN ENGLAND AND WALES 65–67 (2011), available at [http://lawcommission.justice.gov.uk/docs/lc325\\_Expert\\_Evidence\\_Report.pdf](http://lawcommission.justice.gov.uk/docs/lc325_Expert_Evidence_Report.pdf); cf. *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 579–95 (1993).

<sup>7</sup> See Solan & Tiersma, *supra* note 4, at 454; see also Howald, *supra* note 4, at 236.

<sup>8</sup> Grant, *TXT 4N6*, *supra* note 3, at 510–13.

<sup>9</sup> See Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL'Y 299 (2013); John Burrows, *Questions of Authorship: Attribution and Beyond*, 37 COMPUTERS & HUMAN. 5, 5–13 (2003); Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 INT'L J. SPEECH LANGUAGE & L. 1, 2–8; Carole E. Chaski, *Best Practices and Admissibility of Forensic Author Identification*, 21 J.L. & POL'Y 333 (2013); Grant, *Quantifying Evidence*, *supra* note 3, at 1–5; David I. Holmes et al., *Stephen Crane and the New York Tribune: A Case Study in Traditional and Non-Traditional Authorship Attribution*, 35 COMPUTERS & HUMAN. 315, 315–31 (2001); Patrick Juola, *Stylometry and Immigration: A Case Study*, 21 J.L. & POL'Y 287 (2013); Moshe Koppel et al., *Authorship Attribution: What's Easy and What's Hard?*, 21 J.L. & POL'Y 317 (2013).

genre. This I call *population-level distinctiveness*.<sup>10</sup> With longer texts involving some thousands of words, such approaches can be used to make successful attributions.<sup>11</sup> Accordingly, a good stylometric marker should demonstrate that, on that marker, an individual is distinctive against the background population from which it is drawn. The presuppositions of this research are revealed through the language of some of its proponents who refer to linguistic fingerprinting<sup>12</sup> or the discovery of a stylome.<sup>13</sup> This is the language of a forensic science discipline which can provide individuation—the discrimination of one individual from any other in a population.

In contrast to these stylometric approaches, forensic practitioners working on shorter and sometimes fragmentary texts have tended to use more stylistic approaches.<sup>14</sup> Such approaches do not assume that the discovery of population-level discriminants is necessary to authorship analysis but focus on variation between specific individuals.<sup>15</sup> Furthermore, that variation is understood as being created by habitual choice across a wide and unpredictable range of features.<sup>16</sup> Thus, one author might fall into a habit of using unusual punctuation whereas another author might exhibit a preference for elaborate adjective use. Because, before examining a text, one does not know precisely what sort of feature one is looking for, quantitative methods tend to be less well defined

---

<sup>10</sup> See Grant, *TXT 4N6*, *supra* note 3, at 515.

<sup>11</sup> See, e.g., Holmes et al., *supra* note 9, at 322–28.

<sup>12</sup> See, e.g., Sebastian Bernhardsson et al., *The Meta Book and Size-Dependent Properties of Written Language*, *NEW J. PHYSICS* 6 (Dec. 10, 2009), [http://iopscience.iop.org/1367-2630/11/12/123015/pdf/1367-2630\\_11\\_12\\_123015.pdf](http://iopscience.iop.org/1367-2630/11/12/123015/pdf/1367-2630_11_12_123015.pdf).

<sup>13</sup> See Hans van Halteren et al., *New Machine Learning Methods Demonstrate the Existence of a Human Stylome*, *12 J. QUANTITATIVE LINGUISTICS* 65 (2005).

<sup>14</sup> See generally GERALD R. MCMENAMIN, *FORENSIC STYLISTICS* 161 (1993) [hereinafter MCMENAMIN, *FORENSIC STYLISTICS*]; GERALD R. MCMENAMIN, *FORENSIC LINGUISTICS: ADVANCES IN FORENSIC STYLISTICS* (2002) [hereinafter MCMENAMIN, *ADVANCES*].

<sup>15</sup> MCMENAMIN, *FORENSIC STYLISTICS*, *supra* note 14, at 161; MCMENAMIN, *ADVANCES*, *supra* note 14, at 171–72, 174.

<sup>16</sup> MCMENAMIN, *FORENSIC STYLISTICS*, *supra* note 14, at 162–70; MCMENAMIN, *ADVANCES*, *supra* note 14, at 45–65.

within this approach. At a population level, it has been suggested by McMenamin,<sup>17</sup> among others, that a naïve Bayesian approach could be used to quantify stylistic features, and such an approach would seem to be appropriate. However, distributional assumptions and assumptions of independence are hard to demonstrate in the linguistic context, and it is difficult to model how their indeterminacy or violation affects the posterior probability of assigning a text to a particular author.<sup>18</sup> In the UK legal context, a recent judgment criticized a Bayesian approach to footprint analysis, in part because of the lack of certain population data.<sup>19</sup> Linguistic stylistic analysis might also stumble at this hurdle.

One aim of this paper is to demonstrate how it is possible to derive a methodologically rigorous approach to stylistic authorship analysis that can result in statistically described results. The approach is based on the analysis of vocabulary choices and morphological features and is particularly tailored to text messages and allied genres (such as Twitter feeds), though it may be generalized to other text types and other features. Rather than focusing on population-level distinctiveness, this approach gives primacy to *pairwise distinctiveness* between

---

<sup>17</sup> MCMENAMIN, *ADVANCES*, *supra* note 14, at 171–72.

<sup>18</sup> A full discussion of Bayesian approaches to forensic science problems is beyond the scope of this article, but a good introductory discussion is provided by DAVID LUCY, *INTRODUCTION TO STATISTICS FOR FORENSIC SCIENTISTS* 6, 108–12 (2006). The essence of the Bayesian approach is to ask, “What is the probability of the evidence given two opposing hypotheses?” The ratio of the two probabilities (in this case, the probability that the text was authored by AB over the probability that the text was authored by CB) gives a measure of the weight of evidence in favor of one or the other of the probabilities. In authorship problems this likelihood ratio can be applied to the occurrence of each individual feature and summed to provide an overall weight of evidence for authorship given a specified basket of features.

In discussion at this symposium, and through a useful commentary provided by discussant Professor Ed Cheng, it was advocated that I might use a Bayesian approach to examine the case described in this paper. Pressure of time and space meant that this was not pursued here, but a paper examining this should follow.

<sup>19</sup> *R v. T*, [2010] EWCA (Crim) 2439, [86], [2011] 1 Crim. App. 9 (Eng.).

potential authors of messages, thus avoiding some of the population sampling issues which otherwise arise.

*B. Consistency and Distinctiveness*

All comparative authorship analysis depends upon two theoretical assumptions.<sup>20</sup> The first assumption is that there is a sufficient degree of consistency of style within relevant texts by an individual author. The second assumption is that this consistency of style inherent in an author's writings is sufficiently distinctive to discriminate the one author from other relevant authors. Ultimately, the idea that comparative authorship analysis rests upon a strong theoretical assertion of an idiolect is false. The empirical discovery of consistency and distinctiveness can, however, be a sufficient foundation for such work.<sup>21</sup>

The first assumption, that there is "a sufficient degree of consistency of style within relevant texts," requires further discussion. It is not necessary to identify features of an author's language that are wholly consistent. As shall be seen in the Birks case, a weight of evidence for authorship may be built upon a *degree* of consistency. It must be recognized, however, that the greater the degree of consistency in any comparison corpus, the greater the weight of evidence there will be for an attribution. Identifying consistency within relevant texts also requires the creation of a linguistically *relevant* comparison corpus, which accounts for genre<sup>22</sup> as well as other sources of linguistic variation. For example, it must take account of accommodation effects between different recipients of messages and between the possible modes of production—whether the text message was created using a twelve-key alphanumeric system (as is found on more old-fashioned phones), a touch sensitive qwerty keyboard such as is found on an iPhone, or even through a speech-to-text system.

The second assumption raises different considerations. There may be *degrees* of distinctiveness between pairs of individuals or

---

<sup>20</sup> Grant, *TXT 4N6*, *supra* note 3, at 509.

<sup>21</sup> *Id.* at 521–22.

<sup>22</sup> To avoid comparison, for example, of text messages with emails.

within smaller or larger groups. It seems obvious that some pairs of authors will produce writings wherein the authors are easily distinguished, whereas other authors may generally produce texts that are stylistically similar. In the former case, a comparative authorship analysis will be easier, whereas in the latter case, it may be impossible to distinguish between the authors' writings. Therefore, it may not be necessary to show a writer's distinctiveness against all possible authors; it may only be only necessary to compare one author with other relevant authors in the case. A linguistic fingerprint or stylome may be a holy grail for some stylometric researchers; but, should this grail prove as elusive as the Arthurian Holy Grail, comparative authorship analysis can still proceed and provide useful forensic evidence. Generally, investigators or the circumstances of a case will provide the definition of the relevant set of authors, and, as will be demonstrated in the Birks case, in some circumstances it can be sufficient to provide evidence of distinctive style between authors without hazarding to provide evidence of author identification.

*C. Linguistic Analysis of SMS Text Messaging in Previous Cases*

Increasingly, linguists—interested in describing the nature of text messaging as textual, functional, and social phenomena—are studying the language variety used to communicate with mobile telephones and similar devices.<sup>23</sup> Text messaging is shown to cross age, gender, and cultural boundaries. Stylistically, text messages generally are not full of “texting language”—abbreviations and initialisms. In fact, these tend to comprise less than twenty percent of vocabulary choices in text messages.<sup>24</sup> What is characteristic is that there is little or no censure for nontraditional spelling variants or for syntactic ellipses (such as omission of articles, auxiliaries, and other parts of speech), and

---

<sup>23</sup> See DAVID CRYSTAL, *TXTNG: THE GR8 DB8 37–62* (2008) (identifying six principal distinctive features of text messages); see also David Bamman et al., *Gender in Twitter: Styles, Stances, and Social Networks*, 2–6, 30–31 (2012), <http://www.cc.gatech.edu/~jeisenst/papers/GenderInTwitter923.pdf>.

<sup>24</sup> CRYSTAL, *supra* note 23, at 22, 156.

thus in text messaging, creative language play is positively reinforced.<sup>25</sup> This creativity results in an area of linguistic production where idiosyncratic use can flourish, and this can be advantageous for the forensic analyst.

There have been several UK criminal cases involving text messaging which have not yet been directly reported in the research literature.<sup>26</sup> One example is Professor Malcolm Coulthard's involvement in the Hodgson case.<sup>27</sup> Coulthard assisted the successful prosecution of David Hodgson for the murder of Jenny Nicholl, even though her body has never been found.<sup>28</sup> Coulthard observed a style shift in the sequence of text messages sent from Nicholl's phone.<sup>29</sup> Specifically, the latter messages were not compatible with Nicholl's previous texting style but were compatible with the previous style of Hodgson.<sup>30</sup> Coulthard's method involved discriminating between the two potential writers by identifying consistent and distinctive alternate vocabulary choices.<sup>31</sup> He shows that Jenny Nicholl, in her undisputed messages, tended to write "my" and "myself" whereas the defendant, David Hodgson, tended to follow the North English pronunciation and use "me" and "meself."<sup>32</sup> With

---

<sup>25</sup> *Id.* at 74.

<sup>26</sup> Examples of cases include one brought against Stuart Campbell for the murder of Danielle Jones, *Text Messages Examined in Danielle Case*, BBC NEWS (Oct. 9, 2002, 7:27 PM), [http://news.bbc.co.uk/2/hi/uk\\_news/england/2314389.stm](http://news.bbc.co.uk/2/hi/uk_news/england/2314389.stm) (describing the case as the first UK case to turn on text messaging evidence, with analysis provided by Professor Malcolm Coulthard), and a case of murder against Margaret James, *Margaret James Fails to Overturn Plot Conviction*, BBC NEWS (May 15, 2012, 1:49 PM), <http://www.bbc.co.uk/news/uk-england-cornwall-18078998> (discussing the background of the case). The prosecution in the James case initially introduced text messaging analysis but later withdrew it on the sight of expert response reports provided by Professor Coulthard and myself.

<sup>27</sup> See Owen Amos, *The Text Trap*, N. ECHO (Feb. 27, 2008, 11:38 AM), <http://www.thenorthernecho.co.uk/news/2076811.print/>; see also Grant, *TXT 4N6*, *supra* note 3, at 508–09.

<sup>28</sup> Amos, *supra* note 27.

<sup>29</sup> *Id.*

<sup>30</sup> *Id.*

<sup>31</sup> Grant, *TXT 4N6*, *supra* note 3, at 515–17.

<sup>32</sup> *Id.* at 516.

this and eight additionally selected opposing features, Coulthard demonstrates degrees of distinctive and consistent use in the undisputed messages and then goes on to show that the disputed messages are inconsistent with Nicholl's previous style.<sup>33</sup> In a similar way, he also concludes that the disputed messages are compatible with David Hodgson's known style.<sup>34</sup>

While Coulthard's work in this case was admitted at trial and upheld at an appeal hearing,<sup>35</sup> his methodology can be developed and broadened in two respects. First, notwithstanding Coulthard's considerable linguistic skills and reputation, his method depends upon the expert identification and selection of potential distinctive vocabulary choices. This leaves open the possibility that confirmation bias may lead to the selection of features favorable to an analyst's implicit or explicit expectations about a case.<sup>36</sup> Describing a replicable process for feature selection may mitigate against this possibility to some degree. Second, while Coulthard's method is wholly descriptive rather than statistical, development of a quantified method may lead to a better demonstration of the reliability of the conclusions and the validity of the methods. The development of methods for the Birks investigation was intended to safely build on Coulthard's successful methods while simultaneously addressing these issues. The method described here draws on well-established methodological and statistical approaches used in behavioral case linkage as undertaken by forensic psychologists.<sup>37</sup>

---

<sup>33</sup> *Id.*

<sup>34</sup> *Id.* at 515.

<sup>35</sup> See *The Failed Appeal*, JENNY NICHOLL (June 29, 2009), <http://jennynicholl.blogspot.co.uk/search?updated-min=2009-01-01T00:00:00Z&updated-max=2010-01-01T00:00:00Z&max-results=2>.

<sup>36</sup> See, e.g., Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74, 76-77 (2006) (finding that fingerprint experts made different judgments when misled with extraneous contextual information).

<sup>37</sup> See, e.g., Jessica Woodhams et al., *From Marine Ecology to Crime Analysis: Improving the Detection of Serial Sexual Offences Using a Taxonomic Similarity Measure*, 4 J. INVESTIGATIVE PSYCHOL. & OFFENDER PROFILING 17, 17-27 (2007); Jessica Woodhams et al., *The Psychology of Linking Crimes: A Review of the Evidence*, 12 LEGAL & CRIMINOLOGICAL PSYCHOL. 223, 223-49 (2007).



In conclusion, the aim in the Birks analysis was to provide a sound evidentiary analysis based on a pairwise discrimination of Amanda and Christopher Birks' respective vocabulary choices in their text messages. My method seeks to avoid, as much as possible, my own biases affecting the selection of features, the analysis, or the drawing of conclusions. In addition, in the analysis presented in this article, I shall extend the actual approach taken in the case to demonstrate a statistical approach that can further support future analyses.

## II. METHODS

This section follows with a description of the available data and the analytic approach taken in the Birks analysis. Justifications are provided for the decisions taken.

### A. Data

The text message data were provided by Staffordshire Police as part of their investigation into the death of Amanda Birks, and I was a paid expert witness for the prosecution.<sup>38</sup> Case details and the forensic analysis were released to the UK media.<sup>39</sup>

Two small corpora of text messages were provided in the form of spreadsheets identifying phone numbers, names of recipients, sending times, and text for each SMS message. The first spreadsheet contained collated text messages sent by Amanda Birks ("AB") to eleven separate recipients over a period of several days prior to January 17, 2009, the date of her death. As AB's phone was destroyed in the house fire, police collected this information from reports produced by forensic telecommunication engineers who downloaded the messages from the recipients' telephones.<sup>40</sup> After a number of automated

---

<sup>38</sup> Subsequent to the conclusion of the case, permission was obtained from the police to use these materials in teaching and research.

<sup>39</sup> See, e.g., *Businessman Admits Murdering Wife*, *supra* note 1.

<sup>40</sup> I was also provided with the raw telecommunications engineers' forensic reports, but I used these only to check the accuracy of the spreadsheets.

and repeat messages were removed from the spreadsheet, 204 text messages apparently sent by AB remained.

On a preliminary version of this spreadsheet, investigators indicated that a shift in texting style occurred on January 17, 2009 at 12:07 p.m. After that point, investigators believed the messages sent from AB's phone were not crafted by AB. The basis of this hypothesis was essentially a shift in the manner of signing off the messages using "kiss" marks. Prior to this time, the assertion was that AB tended to sign off using "Xxx" or "xxx" with no spacing. From about midday, the style shifts to include spaces in the sign off—"X x." When investigators brief a forensic expert, it is common for them to have already formulated a hypothesis based on such observations. Since confirmation bias may affect expert findings, this is not ideal. In the interest of full disclosure, the investigators' hypothesis and the concerns it raises were reflected in the final forensic report.

In addition to the AB data, a parallel corpus was provided containing undisputed text messages sent by Christopher Birks ("CB") to ten separate recipients. CB had deleted all sent messages from his phone such that they could not be recovered, and so these messages were also collected from recipients. The spreadsheet also contained a small number of unsent messages from CB's draft message folder, which had not been deleted. After a number of automated messages and repeat messages were removed from this spreadsheet, there remained 203 text messages apparently sent by CB.

Where a message had been automatically split into two messages for sending because the original was greater in characters than the permitted SMS length, these were left as two messages.

### *B. Feature Analysis, Coding, and Preliminary Selection*

The method depended upon the identification of vocabulary choices with their associated spelling variants. As the intention was to avoid selectivity, all word forms were listed using corpus linguistics software WordSmith tools<sup>41</sup> and coded as being

---

<sup>41</sup> See generally Mike Scott, *WordSmith Tools Manual, Version 3.0*,

present or not present in every message. In addition to coding these lexical items, a number of further features of texting abbreviations were also coded. These features included the use or absence of spacing between words in certain contexts; letter and number substitutions (using “c” for “see” or “4” for “for”) for syllables and whole words; accent stylizations (“ad” for “had” or “cuz” for “because”); initialisms (“imho” for “in my honest opinion”); and so on. In total a basket of 154 features were developed, and every text message was scored as containing or not containing each of these features.

As the demonstration of consistency requires a certain amount of repetition of a feature, the next decision made was to reject any feature that had fewer than ten occurrences across the 407 text messages. As might be expected, this resulted in the removal of the majority of features from the analysis, leaving a reduced set of just twenty-eight features that formed the basis of the ongoing analysis.

### *C. Identifying Undisputed Text Messages*

The text messages were then examined more closely to determine which messages could be considered to be, without dispute, of known authorship of either Amanda Birks or Christopher Birks. As described above, the police had indicated that they were suspicious of messages sent after midday on January 17, 2009. Taking a precautionary approach, all texts written after midnight on January 16, 2009 were considered disputable and set aside.<sup>42</sup> In addition to this, two further texts apparently sent by CB but from AB’s phone were discarded—one was explicitly signed “Chris” and the other from its content appeared to have been sent by CB. Removing these messages left a total of 165 messages, and it was taken to be a reasonable but not infallible assumption that these messages had indeed been sent by AB.

---

LEXICALLY.NET (1998), <http://www.lexically.net/wordsmith/version3/manual.pdf> (explaining how software analyzes word behavior in texts).

<sup>42</sup> It was only later learned that AB had been seen alive by independent witnesses at approximately 11:00 on the morning of the January 17, 2009.

Examining CB's messages in a similar way, there was no indication for removal of any messages, and this left the original set of 203 messages intact.

*D. Identifying Features with Discriminative Power*

The selected undisputed texts and features were further examined to determine whether there were features that discriminated consistently to some degree between the two writers in their known texts. Features were only retained in the analysis if one author used them in at least twice as many messages as the other—that is to say where one author used a feature at a rate of more than sixty-six percent of its total occurrence. This left a set of just eighteen features that are tabulated in Table 1.

This frequency table based on the known messages of AB and CB clearly indicates, in a general sense, that CB and AB write text messages using distinctive styles, and for some features the degree of distinctiveness is absolute. AB never, in this data, writes "with" as "wiv" and CB never writes "had" as "ad." Other features are features of predominately one author; CB tends to use commas, and AB does so very rarely. This contrasts, for example, with AB's tendency to use "t" for "the," which is only rarely used by CB. For each author, lists can thus be created of features characteristic of their text messaging style; ten features for AB and eight features for CB. In contrast to Coulthard's method in the Jenny Nicholl inquiry, these feature lists are not necessarily reciprocal alternates. For two words, "don't" and "with," each writer has as a feature a preferred spelling variant of that word, but for the other features this is not the case.

For neither author can these lists be considered identifying in an absolute sense. The features contained in these lists are not linguistic "fingerprints" identifying individuals against a population. Rather, they demonstrate a relative consistency of habit and a pairwise distinctiveness which thus can be used to stylistically discriminate between messages of the suspect and the victim in this case. As we do not have good knowledge of the distribution of texting features across the population of all

texters, we cannot say with certainty how many other people will share each of these sets of features. I did test these feature sets against a corpus containing ten messages from each of 500 texters, and no other individuals demonstrated the use of either complete set of features. Such information is useful but cannot be employed in statistical calculations, as theoretical linguistic difficulties remain over how any such reference corpus can be considered representative of the population of texters.<sup>43</sup>

Table 1: Frequency distribution of elicited features

	Feature	# in CB texts	# in AB texts	Total	% in AB texts	% in CB texts
Features characteristic of AB's texting	"ad" for "had"	0	13	13	100%	0%
	"dont" for "don't"	0	9	9	100%	0%
	"t" for "the"	1	15	16	93.8%	6.3%
	"bak" for "back"	1	10	11	90.9%	9.1%
	"av" for "have"	1	9	10	90.0%	10.0%
	"wud" for "would"	2	9	11	81.8%	18.2%
	"w" for "with"	3	10	13	76.9%	23.1%
	"y" for "yes"	2	6	8	75.0%	25.0%
	"wil" for "will"	4	9	13	69.2%	30.8%
	"wen" for "when"	4	9	13	69.2%	30.8%
Features characteristic of CB's texting	"dnt" for "don't"	8	0	8	0%	100%
	"jst" for "just"	12	0	12	0%	100%
	"wiv for "with"	15	0	15	0%	100%
	4 for "for" with no trailing space	35	0	35	0%	100%
	2 for "to" with no trailing space	58	0	58	0%	100%
	Use of comma	87	5	92	5.4%	94.6%
	"4get" for "forget"	15	1	16	6.3%	93.8%
	"thanx" for "thanks"	16	2	18	11.1%	88.9%

<sup>43</sup> See, e.g., Grant, *Quantifying Evidence*, *supra* note 3, at 6–9, 7 fig.1 (discussing issues of population sampling for authorship analysis work).

*E. Quantification of Distinctiveness*

In the analysis that contributed to the investigation, Table 1 was used to demonstrate the distinctiveness between the two authors. In developing the model further, it is possible to follow methods rooted in forensic psychology to demonstrate the distinctiveness statistically. Each feature is marked as present or absent in each text message by using a “one” for a present feature and a “zero” for an absent feature. The resulting array of zeros and ones can then be used to compare messages using a binary correlation known as Jaccard’s coefficient.<sup>44</sup> Pairs of messages were placed into three categories: the first two categories were within-author pairings, each comprising two texts by AB and two texts by CB. Using the random case selection feature within SPSS statistical analysis software, a sample of 100 within-author pairings was taken for each author.<sup>45</sup>

*Table 2: Jaccard values for linked and unlinked pairs of messages*

	AB-AB pairs	CB-CB pairs	AB-CB pairs
Mean Jaccard	0.195	0.199	0.09
Standard Deviation	0.24	0.25	0.12

<sup>44</sup> Jaccard’s coefficient is a correlation for binary values and can be typically read as a distance measure. Results vary between zero and one where one indicates that two binary arrays are identical and zero indicates that they are completely different. Decimals between zero and one indicate variation between these two extremes. One advantage of Jaccard is that it does not inflate similarity on the basis of two absences. Absence of evidence of a stylistic feature in a particular text message is not evidence of its absence from that individual’s stylistic range when texting generally, and thus using Jaccard does not risk overstating the explanatory power of a single text.

<sup>45</sup> A within-author pairing comprised either two AB texts or two CB texts; a between-author pairing comprised one AB text and one CB text. Given 165 AB texts and 203 CB texts were used, this sample was taken from more than 13,000 potential AB pairings and more than 20,000 potential CB pairings.

In addition, a further random sample of 100 between author pairings was taken. For each of these three sets of pairings, Jaccard's coefficients were calculated.

As Table 2 suggests, the mean Jaccard values for linked pairs of messages show roughly similar levels of within-author consistency; Jaccard values for AB-AB pairs of messages and CB-CB pairs of messages are relatively close. The Jaccard values for unlinked pairs, each containing an AB and a CB message, tend to have considerably lower Jaccard scores. That is to say that linked pairs appear to have a greater degree of stylistic similarity than unlinked pairs. On further examination, however, it can be seen that the Jaccard values fall into nonnormal distributions (Kolmogorov-Smirnov  $Z$  values, respectively: AB-AB pair = 2.30, CB-CB pair = 2.19, AB-CB pair = 3.21;  $N = 100$  in each group;  $p < 0.0005$ ) indicating a nonparametric approach is required.<sup>46</sup>

Using Mann-Whitney  $U$  comparisons, pairs of messages containing only AB texts can be shown to be significantly more similar than pairs of messages each containing a CB and AB text ( $U = 3832$ ;  $N = 200$ ;  $p = 0.002$ ). Additionally, pairs of messages containing only CB texts can be shown to be significantly more similar than pairs of messages each containing a CB and AB text ( $U = 3730$ ;  $N = 200$ ;  $p = 0.001$ ). These findings both demonstrate a significant degree of consistency of style within each author and at the same time distinctiveness between the authors' text messaging styles.

#### *F. Disputed Text Messages*

Having determined distinctiveness in style between the two authors by examining the undisputed material, it is possible to analyze the disputed text messages as a group. The disputed messages are hereafter referred to as queried or Q messages.

---

<sup>46</sup> Many statistical tests (such as  $t$ -tests) require data to approximate a normal distribution and the Kolmogorov-Smirnov  $Z$  test can be used to test whether a distribution is normal. Where data is not normally distributed, one must choose an alternative test. The Mann-Whitney  $U$  is in essence equivalent to a  $t$ -test but can be used on nonnormal, nonparametric data.

We can follow the quantified method described above. Jaccard values were obtained for a random sample of 100 AB-Q pairs of messages and 100 CB-Q pairs. The AB-Q messages were compared with AB only linked pairs. The Mann-Whitney comparison showed a significant lower Jaccard value for the AB-Q pairs ( $U = 4119$ ;  $N = 200$ ;  $p = 0.016$ ). In contrast, no significant difference can be shown between the CB-only linked pairs and the Jaccard values for pairs of messages containing a CB and a Q message ( $U = 3572$ ;  $N = 200$ ;  $p = 0.203$ ). Taken as a set, the stylistic choices made in the disputed messages show significant dissimilarity from the stylistic choices in Amanda Birks's undisputed messages but no equivalent difference can be shown for the undisputed messages of Christopher Birks.

### *G. Examining the Sequence of Individual Messages*

The quantified method can demonstrate the general point of consistent style within authors, which can discriminate between them, but this takes one only so far. Appropriate caution, however, requires that the queried texts not be treated as a homogenous group but rather that the authorship of each queried message be considered on its own merits. Quantified classification of individual data points in a nonnormal distribution would create a considerable statistical challenge, and because of this, a more qualitative descriptive analysis is preferred.

The queried messages fall into a rough time-ordered sequence. Caution is required—since the messages were collected from recipients' phones, the time tagged on each messages may not be an accurate indication of when the message was sent.

To demonstrate the descriptive method, a test case is provided by the data. A message sent on January 13 was removed from the spreadsheet of known texts attributed to AB because, although it was sent from AB's phone, it was signed "Chris." The message was longer than the permitted SMS length, and so sent as two messages, but in full it reads:



I think u were abit hard on me earlier, 1. I wasnt l8 purposely, 2, i nd a car 4assessments. Iv had a luk at myself in mirror n undastand 1, y u dnt fancy me n 2, y u dnt like me. Sory 4all the pain iv caused u, i love u amanda..... Chris. Xxx

Messages sent 01/13/09

Qualitatively, this message can be analyzed for features consistent with AB's texting style and features consistent with CB's texting style. Of the eight features considered characteristic of AB's text messaging style indicated in Table 1, none of them appear in this message. Further, given that "had," "the," and "don't" are used in the message, these might be considered opportunities to display these features. On this descriptive basis, the analysis suggests this message can be considered inconsistent with AB's style.

In contrast, the message contains features considered characteristic of CB's style: six commas, the use of "dnt" for "don't," and two uses of the digit "4" for the word "for" where there is no trailing space, i.e., "4assessments" and "4all." On this basis one can reasonably conclude that the message is consistent with CB's texting style.

In the forensic context the expression of results is important. In this case the language used is that of consistency and distinctiveness; thus, initial opinions in the forensic report for this case are as follows:

- i. Some of the messages sent from Amanda Birks phone . . . on the 17 January 2009 are stylistically distinctive from messages known to have been sent by Amanda Birks before that date.
- ii. Analysis of text messages known to have been written by Christopher Birks has given rise to a description of his habitual style which is described in this report. This described style will be shared by a limited number of people and is distinctive from the habitual style of Amanda Birks.
- iii. Some of the messages sent from Amanda Birks phone . . . on the 17 January 2009 show stylistic consistency

with text messages known to have been sent by Christopher Birks on and before that date.

The absence of identifying language is deliberate and significant. I am able to say that there are “a limited number of people” who may share Christopher Birks’s texting style because I have tested the feature set against my independent corpus of 500 texters, but it is a weakness of my position that I have no substantial basis to say how many individuals in a population will share this feature set. Another way of thinking of this is that the question before the court is, “How likely is it that Christopher Birks wrote the text messages on the 17 January 2009?” I do not answer this question; rather, I would respond that “as the texts are distinctive from AB’s historic style and consistent with CB’s historic style, it is more likely to have been Christopher Birks than Amanda Birks who wrote those texts.” While this information is useful to the court considered in conjunction with other evidence, it is by no means identification evidence standing alone.

#### *H. Messages of January 17, 2009 Sent from AB’s Phone*

Turning to the list of messages sent from AB’s phone on January 17, 2009, each message was evaluated in turn to arrive at a qualitative conclusion for stylistic consistency with each of CB’s and AB’s previously described style.

*Table 3: Examination of disputed messages*

<b>Time</b>	<b>Sent To</b>	<b>Message</b>	<b>Comment</b>
00:40	Friend 1	I love u my gorgeous sexy babe! Xxx	Contains none of the specific features listed for either AB or CB.
10:04	Friend 1	Got go fetch milly. Val cant cope w her x	Contains “w” for “with,” identified as a feature of AB’s style. CB tends to use “wiv” but does use “w” on occasions.

*AB was seen alive at 11 a.m.*

11:39	Friend 1	Wen r u up 4a repeat performance? X x	Digit “4” used for “for” with no trailing space is used only by CB. “Wen” is used by AB with twice the frequency of CB.
11:49	Friend 1	Wot do think? Cant believe theresa! X x	Contains none of the identified features. The use of “wot” is not discriminating between AB & CB.
12:07	Friend 1	Txt me, talkin with chris. X x	Contains a comma, which is rare in AB’s texts. Slight inconsistency with AB, consistent with CB.
12:10	Friend 1	About your route, spk lata, talkin with chris. X x	Contains commas, which are rare in ABs texts. “Spk” and “lata” are only used in the disputed messages.
12:39	Friend 1	U wen u filled ur application in. X x	Contains “wen” used twice as often by AB than CB.
12:39	Friend 1	Am talkin wiv chris, am confused. Ur 2 young 4me. X x	Contains the use of commas, the use of “wiv” rather than “w” or “with” and the use of “4” without a trailing space. <b><i>First text to be judged inconsistent with AB and consistent with CB.</i></b>
12:41	Friend 1	Txt u lata. X x	Contains none of the identified features.
12:54	Friend 1	Chris is sayin the same, giv me space, u know wot i think of u. X	Contains commas, which are rare in AB’s texts.
13:02	Friend 2	Not sure yet, am jst talkin wiv chris so will txt u lata, dont worry.	Contains commas, the use of “wiv” rather than “w” or “with,” and the use of “jst.” Also contains “dont” (with no apostrophe) which otherwise is used only by AB.

14:03	Friend 1	Dnt kno, need think, am goin relax in bath then go sleep, really tired. X	Contains commas and "dnt." <b>Judged inconsistent with AB; consistent with CB.</b>
14:05	Friend 1	Will do when i wake up, so tired. Need talk 2 chris. Still hav feelins 4him. X	Contains commas and the use of "4" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
14:12	Friend 1	Just give me time and space please, im tired, confused n wud like runaway, x	Uses "wud" which is and unusual for CB and "just" rather than "jst."
14:51	Friend 3	Iv bin out all week, me n chris not been gettin on so probably stay in 2talk.x	Contains commas and the use of "2" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
16:58	Friend 1	Just need get my head straight, maybe 2mro	Contains comma but also "just."
17:00	Friend 1	I can try?	Contains none of the identified features.
17:06	Friend 4	Talkin with chris, tryin c if theres a way 4ward 4us. Love u.	Contains commas and the use of "4" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
17:27	CB	Take 1day at at time but u still mean alot 2me n i do love u. X	Contains "2" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
17:32	Friend 2	Sorry just want time 2think. Been talkin 2chris so tryin get my head 2getha! R u out with wayne 2nite? X	Contains "2" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
17:37	Friend 2	Thanx tim, i just want chris talk 2me right, he needs learn not take his stresses out on me, then we can b happy. X	Contains commas and the use of "2" without a following space. <b>Judged inconsistent with AB; consistent with CB.</b>
18:11	CB	Goin stay in and look	Contains comma and the use

		through my of “2” without a following aromatherapy books, space. <b><i>Judged inconsistent with AB; consistent with CB.</i></b> 2tired u hav fun. Xx
19:34	Friend 3	Dont really want talk, am shattered, nan says i shud sort it, we had a talk n probably will, he just needs talk 2me better! Xx Contains comma and the use of “2” without a following space. But also contains “dont,” otherwise used entirely by AB.
19:48	Friend 3	Weve come so far, had a great xmas, things just ontop of us, think we need a holiday. Xx Contains commas.

In Table 3, each message sent on January 17, 2009 is treated individually, but it is possible to consider the set of messages as a roughly sequentially ordered group. The early text messages in this series demonstrate more consistency with AB’s known texting style than with CB’s known texting style. The later messages demonstrate a reciprocal pattern of more consistency with CB’s style and more inconsistency with AB’s style. This pattern, however, is not absolute.

The first message that is clearly and substantially inconsistent with AB and consistent with CB was received at 12:39 p.m. and reads:

*Am talkin wiv chris, am confused. Ur 2 young 4me. X x*

Of the seven messages that are timed before this message, two are entirely consistent with AB’s style; two are consistent with CB’s style and are inconsistent with AB’s style; and two demonstrate none of the identified features at all. One message in this earlier set demonstrates consistencies and inconsistencies with both authors.

Of the sixteen messages timed as occurring after this point, eleven messages demonstrate consistent features with CB’s texting style and or features inconsistent with AB’s texting style. No messages are wholly consistent with AB’s texting style. Three messages demonstrate consistencies and inconsistencies with both authors. Finally, two messages demonstrate none of the identified features.

Explanations for this slightly unclear picture are both linguistic and technical. The technical issues concern the fact that the timings of these messages are from the recipient phones. It is possible that delays in the telephone network created a time gap between sending and receipt. In addition to technical issues, the simple issue of linguistic variability has to be taken into account. Although I have attempted to describe levels of consistency within the known texts of AB and CB respectively, this does not preclude the possibility of shifts in texting style by either author for unknown reasons. Even where a consistency of style has been demonstrated over a stretch of two hundred messages, it must be considered that such a pattern could change.

### III. DISCUSSION

I have presented here a method for the forensic authorship analysis of SMS text messages. In some ways, the case is straightforward: the police evidence indicates a pair of candidate authors. Although a further author cannot be precluded as a possibility, the presence of a pair of candidate authors makes the analytic task easier.<sup>47</sup>

Framing the task in terms of consistency and distinctiveness allows for a combination of statistical and descriptive methods. Describing the points of consistency in the two corpora of undisputed messages allows one to quantify what is essentially a stylistic description and thereafter conclude statistically that a pairwise discrimination can be obtained between them. Avoiding claims about any population distribution of the identified features limits the conclusions that can be drawn. The *R v. T* case<sup>48</sup> suggests that quantification of identification requires some approximate knowledge of distributional data, and this is not available or perhaps even not obtainable for language data.<sup>49</sup> Given these concerns, it is not possible to identify Christopher Birks as the sender of the last messages from his wife's phone,

---

<sup>47</sup> In my experience, many comparative, forensic authorship analyses are similar comparisons between small sets of potential authors.

<sup>48</sup> *R v. T*, [2010] EWCA (Crim) 2439, [86], [2011] 1 Crim. App. 9 (Eng.).

<sup>49</sup> Grant, *Quantifying Evidence*, *supra* note 3, at 14.

but it is possible to assert that his style is distinctive from his wife's and that the last messages are inconsistent with her previously described style and compatible with his style. As a contribution to a wider criminal case involving other forms of evidence, this is strong information and can be useful enough to contribute to the evidence in the case.

With regard to methodology, one of the perceived weaknesses of stylistic analysis can be an overreliance on subjective expertise and an apparent lack of method in the identification of features. The method employed here attempts to limit that subjectivity. A further remedy would be the explicit statement of a protocol for feature identification and analysis, which could be designed and stated in advance of approaching an individual case. Casework invariably involves working with awkward situations and imperfect data. Consequently, one aspect of practitioner expertise, generally underreported, is the negotiation of this real world difficulty.

*A. Proposed Protocol for Stylistic Analysis in  
Classification Problems<sup>50</sup>*

*1. Try to Know as Little as Possible  
About the Wider Details of the Case.*

The aim here is to mitigate the well-documented cognitive biases that occur across forensic disciplines.<sup>51</sup>

*2. Describe the Features of the Known Texts First.*

Once it has been established that the known texts are

---

<sup>50</sup> I divide cases into classification, inclusion, and exclusion problems. Classification problems take the form, "Which of these set of authors is the most likely to have written the query text?" The definition of the set of potential authors will be defined by nonlinguistic evidence, and it must be explicitly stated that linguistic conclusions presuppose the soundness of this evidence. This protocol is only for such classification problems.

<sup>51</sup> See Itiel E. Dror et al., *Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison*, 208 FORENSIC SCI. INT'L 10 (2011); Dror et al., *supra* note 36, at 74.

linguistically relevant to the queried texts, the queried texts should be put aside. Ideally this preliminary examination might be carried out by a first analyst who then provides just the known texts and very limited information to a second analyst. In classification problems, this contrast provides a detailed but unbiased description of the known texts. The analysis must allow for possible outcomes where no distinction can be drawn between authors' styles in the known texts and for the possible outcome that the query texts are distinctive from all known authors' previously described styles. Ordering the analysis in this way allows for these important possible outcomes.

*3. The Contrastive Analysis Should Elicit Consistent and Distinctive Features Within the Known Texts.*

*a. Within-Author Consistency*

This requires several, and sometimes many, texts for each possible author. The number of texts of known authorship that make good comparison documents in terms of genre, recipient effects, and other linguistic variables becomes key in determining whether a case should be taken.

*b. Between-Author Distinctiveness*

It seems likely that stylistic distinctiveness can only be demonstrated pairwise or for small groups. That we do not have population distributions of stylistic features is not just a question of inadequacy or a lack of effort in carrying out linguistic surveys; linguistic complexity in the sources of language variation may mean that it is not possible to collect representative population samples of stylistic features.

The output of the contrastive analysis becomes a "locked" feature list that cannot be altered hereafter.

*4. Carry out an Examination of the Query Texts for the Identified Features*

If at this stage further features are found which seem useful



that are not contained in the locked feature list, it is too late as any such features cannot be included in the final analysis.

On the other hand there will be some features in the locked feature list that do not occur in the query texts. The significance of any such features should be considered as possible evidence that none of the included authors wrote the query texts.

*5. Draw Conclusions Based on the Consistency  
and Distinctiveness of the Query Texts with  
Each Set of Texts of Known Authorship.*

The Birks case was brought to me by Staffordshire police in the absence of any such protocol; as such, there were numerous times when my analysis did not follow this outline. For example, early on the police explained to me their hypothesis, which they without a doubt hoped my analysis would support. This is not a criticism of the police, who will have little knowledge of the needs of forensic linguistic analysis, but it may point to the need for an intermediary between the investigators and the forensic analyst such that the intermediary might control the information and data that reach the analyst.

No claim is made that the features that distinguish Amanda and Christopher's text messages will be useful in distinguishing between other pairs of authors or for the same authors in other genres. They are not population-level stylometric markers of authorship. Furthermore, and in contrast to Coulthard's analysis of the Nicholl-Hodgson case,<sup>52</sup> the selection of markers did not depend on my individual skill in linguistic observation; rather, the features were elicited from the data according to a set of linguistically and statistically justifiable criteria. The method can be (and has been) developed and tested in other similar cases—and in a more recent case involving email analysis, the use of two analysts has proved invaluable. No claim, however, is made for the reliability of the specific set of markers used, and there are no grounds to generalize their use to other cases.

---

<sup>52</sup> See *supra* Part I.C.

## IV. CONCLUSION

Ultimately, the theoretical importance of the distinction between population level distinctiveness and pairwise or small group distinctiveness is crucial to the success of this approach and to the statistical method employed. While this limits the opinions of the expert to considerations of consistency and distinctiveness, given the richness of linguistic variation, such a limitation is appropriate to forensic authorship analysis and provides an expert with more certain and more credible evidence to offer the courts.

On November 2, 2009, Christopher Birks was due to be tried. On the morning before trial, he changed his pleas to “guilty” of the murder of his wife, “guilty” of arson, and “guilty” of the endangerment of his children and of the firefighters. He was subsequently handed down a life sentence with a minimum term of incarceration of nineteen years.<sup>53</sup>

---

<sup>53</sup> *Man Jailed over Wife Fire Murder*, BBC NEWS (Dec. 11, 2009, 7:28 PM), [http://news.bbc.co.uk/2/hi/uk\\_news/england/staffordshire/8408020.stm](http://news.bbc.co.uk/2/hi/uk_news/england/staffordshire/8408020.stm).

# TOWARDS AN INDEX OF IDIOLECTAL SIMILITUDE (OR DISTANCE) IN FORENSIC AUTHORSHIP ANALYSIS

*M. Teresa Turell\* and Núria Gavalda\**

## I. INTRODUCTION

Forensic linguistics is a discipline concerned with the study of language in any judicial context. The framework for the present article is the area of forensic linguistics known as Language as Evidence, where a sample or several samples of oral or written linguistic productions of one or more individuals may constitute evidence in a judicial process. In these cases, linguists acting as expert witnesses in court must compare two (sets of) samples, i.e., the nondisputed sample, the authorship of which cannot be questioned, and the disputed sample, the authorship of which is questioned, to determine the linguistic differences and similarities that the samples show and to try to reach a conclusion regarding the possibility that they have been produced by the same individual.

Linguistic evidence is not like other kinds of evidence such as DNA or fingerprints, in the sense that language is intrinsically variable. Sociolinguists have shown for decades that languages are in a state of constant change and that any language is intrinsically variable in all its levels, even at the idiolectal level.<sup>1</sup> In other words, the linguistic production of a single

---

\* ForensicLab, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra (Barcelona, Spain).

<sup>1</sup> See, e.g., WILLIAM LABOV, SOCIOLINGUISTIC PATTERNS 122, 127, 271–72, 319–25 (1972); see also J.K. CHAMBERS, SOCIOLINGUISTIC THEORY: LINGUISTIC VARIATION AND ITS SOCIAL SIGNIFICANCE 33–37 (2009); M. Teresa Turell Julià, *La base teòrica i metodològica de la variació lingüística*, in LA SOCIOLINGÜÍSTICA DE LA VARIACIÓ 17, 20–22 (M. Teresa Turell ed.,

speaker or writer will generally show some variation. Consequently, when comparing two samples, the expert witness must ponder whether the degree of variation present is likely to be due to interspeaker/writer differences or to intraspeaker/writer differences. To do this, the linguist must analyze as many linguistic parameters as possible in order to reliably reach such conclusions.

Research in the last forty years has successfully identified parameters that can contribute to this endeavor. In the field of forensic speech comparison, where oral samples (recordings) are analyzed, both acoustic and linguistic parameters are normally considered. On the one hand, phoneticians analyze the acoustic nature of individual sounds (vowels and consonants) together with parameters related to the fundamental frequency (related to the pitch of the voice), voice quality, and suprasegmental patterns such as intonation or linguistic rhythm.<sup>2</sup> On the other hand, phonological variables are related to individual choices that each individual makes depending on their place of origin and other social factors such as gender, education, and class.<sup>3</sup> Moreover, variables related to the particular syntactic, morphological, or lexical patterns that an individual shows can also shed light on the differences or similarities between oral samples. In the field of forensic text comparison, or authorship analysis, where written texts are analyzed, variables related to lexical density, lexical richness, and syntactic and morphological patterns have been proven to be reliable markers of authorship.<sup>4</sup>

---

1995).

<sup>2</sup> See, e.g., Peter French, *An Overview of Forensic Phonetics with Particular Reference to Speaker Identification*, 1 FORENSIC LINGUISTICS 169, 174–76, 178 (1994); see also Erika Gold & Peter French, *International Practices in Forensic Speaker Comparison*, 18 INT’L J. SPEECH LANGUAGE & L. 293, 295–96 (2011).

<sup>3</sup> See, e.g., Paul Foulkes & Peter French, *Forensic Phonetics and Sociolinguistics*, in CONCISE ENCYCLOPEDIA OF SOCIOLINGUISTICS 329, 330 (Rajend Mesthrie ed., 2001).

<sup>4</sup> See, e.g., David Woolls & Malcolm Coulthard, *Tools for the Trade*, 5 INT’L J. SPEECH LANGUAGE & L. 33, 37 (1998); see also Harald Baayen et al., *Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution*, 11 LITERARY & LINGUISTIC COMPUTING 121, 128 (1996); M. Teresa Turell, *Textual Kidnapping Revisited: The Case of*

Also, other features related to the deep structure of language, such as the analysis of parts of speech via *n*-grams,<sup>5</sup> have also been shown to account for idiosyncratic characteristics.

This article proposes an Index of Idiolectal Similitude (or Distance) (hereinafter IIS) as a new tool to carry out forensic speech and text comparison.<sup>6</sup> Part II provides some of the premises and hypotheses underlying the study of forensic linguistics. Part III contains an overview of the study, including descriptions of its objectives, theoretical framework, hypotheses, and methodology. Finally, Part IV presents the result of the study and is followed by an assessment of the results and discussion on the future of the study.

## II. PREMISES AND HYPOTHESES

The study of idiolectal similitude or distance is based on two fundamental premises: 1) language provides oral and written

---

*Plagiarism in Literary Translation*, 11 INT'L J. SPEECH LANGUAGE & L. 1, 19–20, 24 (2004).

<sup>5</sup> *N*-grams are sequences of grammatical categories. For example, “the man” is a bigram (sequence of two grammatical categories (article + noun)) and “the man is” is a trigram (sequence of three parts of speech (article + noun + verb)). See, e.g., Maria S. Spassova & M. Teresa Turell, *The Use of Morpho-syntactically Annotated Tag Sequences as Forensic Markers of Authorship Attribution*, PROCEEDINGS OF THE SECOND EUROPEAN IAFL CONFERENCE ON FORENSIC LINGUISTICS / LANGUAGE AND THE LAW 229, 229–37 (2007); see also Maria Stefanova Spassova, *El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español* 59–63 (2009) (unpublished Ph.D. dissertation, Universitat Pompeu Fabra), available at <http://tesisenred.net/bitstream/handle/10803/7512/tmss.pdf.pdf?sequence=1>.

<sup>6</sup> The research presented in this article is based on the findings of two research projects, *Idiolectometría aplicada a la lingüística forense*, funded by the Spanish Ministry of Science and Education (EXPLORA-HUM2007-29140-E; PI: M. Teresa Turell, 2007–08), and the FFI project, *Idiolectometría forense e Índice de similitud idiolectal*, funded by the Spanish Ministry of Science and Innovation (FII2008-03583/FILO; PI: M. Teresa Turell, 2008–11). See generally FORENSIC LAB—UNITAT DE VARIACIÓ LINGÜÍSTICA, FORENSIC IDIOLECTOMETRY AND INDEX OF IDIOLECTAL SIMILITUDE (2013), [http://www.iula.upf.edu/rec/forensic\\_isi/docums/forensic\\_isi\\_en.pdf](http://www.iula.upf.edu/rec/forensic_isi/docums/forensic_isi_en.pdf).

information of several kinds and can reveal an individual's socio-individual and socio-collective traits; and 2) each individual seems to have a unique idiosyncratic use of language that distinguishes him or her from the rest of language users in his or her community. This individual use of language has traditionally been referred to by forensic linguists as "idiolect."<sup>7</sup> This article follows the more recent concept of "idiolectal style" proposed by Turell, which is defined as follows:

[The] concept "idiolectal style," following the use of the term "style" in pragmatics, is proposed as a notion which could be more relevant to forensic authorship contexts. "Idiolectal style" would have to do primarily, not with what system of language/dialect an individual has, but with a) how this system, shared by lots of people, is used in a distinctive way by a particular individual; b) the speaker/writer's production, which appears to be "individual" and "unique" (Coulthard 2004)<sup>8</sup> and also c) Halliday's (1989) proposal of "options" and "selections"<sup>9</sup> from these options.<sup>10</sup>

Regarding forensic authorship analysis, there have been some recent objections to current work, in particular with approaches involving qualitative analyses of the data. These objections deal with the fact that qualitative approaches may be considered nonscientific and subjective, that they are rarely testable, and that their rate of error has never been established.<sup>11</sup>

---

<sup>7</sup> See J.R. Baldwin, *Phonetics and Speaker Identification*, 19 MED. SCI. & L. 231, 231 (1979); see also GERALD R. McMENAMIN, FORENSIC LINGUISTICS: ADVANCES IN FORENSIC STYLISTICS 53-54, 112 (2002); Malcom Coulthard, *Author Identification, Idiolect, and Linguistic Uniqueness*, 25 APPLIED LINGUISTICS 431, 431 (2004).

<sup>8</sup> Coulthard, *supra* note 7, at 445.

<sup>9</sup> M.A.K. HALLIDAY & RUQAIYA HASAN, LANGUAGE, CONTEXT AND TEXT: ASPECTS OF LANGUAGE IN A SOCIAL-SEMIOTIC PERSPECTIVE 55-56, 113-15 (1989).

<sup>10</sup> M. Teresa Turell, *The Use of Textual, Grammatical and Sociolinguistic Evidence in Forensic Text Comparison*, 17 INT'L J. SPEECH LANGUAGE & L. 211, 217 (2010).

<sup>11</sup> See, e.g., Carole E. Chaski, *Empirical Evaluations of Language-Based Author Identification Techniques*, 8 FORENSIC LINGUISTICS 1, 2 (2001); see

In this sense, if we compare this area with other forensic linguistic sciences, such as forensic phonetics and acoustics, forensic authorship analysis does not count on a common framework regarding the definition of the nature, number, and size of the samples to be used before one can attribute authorship safely. Moreover, it is also necessary to agree on what comparison baseline is needed before one can achieve degrees of reliability. Thus, there is a general need in all languages, as well as in all operational areas of Language as Evidence, to be able to count on corpora consisting of all possible existing spoken or written idiolectal styles of each speaker or writer, even if this is a daunting, almost impossible, endeavor.

Meanwhile, forensic authorship analysis can benefit from a complementary combination of both qualitative and quantitative methods.<sup>12</sup> In other words, until the Likelihood Ratio framework<sup>13</sup> for written texts can be adopted in forensic authorship analysis, among other quantitative methods, different approaches that complement each other—i.e., cumulative evidence—will have to be used in the comparison of disputed and nondisputed texts. Studies have shown that there are several techniques that can be used in forensic authorship analysis,

---

also Tim Grant & Kevin Baker, *Identifying Reliable, Valid Markers of Authorship: A Response to Chaski*, 8 FORENSIC LINGUISTICS 66, 68–76 (2001).

<sup>12</sup> See Turell, *supra* note 10, at 218, 220.

<sup>13</sup> The Bayesian likelihood ratio represents the framework within which other forensic sciences such as analysis of DNA are being developed. This statistical method calculates the probability of the evidence considering the hypotheses given by both the defense and the prosecution. However, one of the most important limitations by which this method cannot be used in present-day authorship analysis is that it needs a Base Rate Knowledge of population distribution in order to make decisions regarding how significant certain differences and similarities between linguistic samples are, which is only available for very limited linguistic features. This Base Rate Knowledge implies the collection of data regarding the general usage of the linguistic parameters being considered by a relevant population, or group of language users from the same linguistic community, with which the specific behavior of the speakers or writers under comparison can be compared.

including textual qualitative analytical procedures,<sup>14</sup> the analysis of lexical density and lexical richness,<sup>15</sup> and the use of reference corpora to account for the rarity of linguistic variables.<sup>16</sup> Furthermore, the use of semiautomatic analyses of “deep-structure” linguistic variables (such as Discriminant Function Analysis of sequences of annotated linguistic categories) has also proved to be a reliable technique.<sup>17</sup> Finally, the measurements of idiolectal similitude/distance such as those involved in the use of the IIS proposed here may also be a good approach to carry out forensic authorship analysis.

---

<sup>14</sup> See, e.g., Ol’ga Feiguina & Graeme Hirst, *Authorship Attribution for Small Texts: Literary and Forensic Experiments*, PROC. SIGIR’07 INT’L WORKSHOP ON PLAGIARISM ANALYSIS, AUTHORSHIP IDENTIFICATION, & NEAR-DUPLICATE DETECTION, 2007, at 236, 236–39; David I. Holmes, *Authorship Attribution*, 28 COMPUTERS & HUMAN. 87, 87–106 (1994); Spassova & Turell, *supra* note 5, at 229–37; Hans van Halteren et al., *Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution*, 11 LITERARY & LINGUISTIC COMPUTING 18, 18–24 (1996).

<sup>15</sup> See, e.g., Woolls & Coulthard, *supra* note 4, at 37–38 (describing a method of authorship identification that focuses on lexical richness, average sentence length, and grammar); see also Coulthard, *supra* note 7, at 435 (discussing the value of measuring the percentage of lexical types in detecting plagiarism); Turell, *supra* note 4, at 24 (summarizing findings measuring uniqueness of used terms and phrases by measuring density); M. Teresa Turell, *The Disputed Authorship of Electronic Mail: Linguistic, Stylistic and Pragmatic Markers in Short Texts* (2004) (unpublished conference paper).

<sup>16</sup> See, e.g., Malcom Coulthard, *On the Use of Corpora in the Analysis of Forensic Texts*, 1 FORENSIC LINGUISTICS 25, 28–29 (1994) (explaining how corpora may be used to, for example, determine how likely it is for a word to occur, both individually and with other words); see also Turell, *supra* note 10, at 216, 218 (describing linguistic variables and their influence on forensic text comparison).

<sup>17</sup> See, e.g., Spassova, *supra* note 5; see also Núria Bel et al., *The Use of Sequences of Linguistic Categories in Forensic Written Text Comparison Revisited*, PROC. INT’L ASS’N FORENSIC LINGUISTS’ TENTH BIENNIAL CONF., 2012, at 192, 192–93, 197–98, 200, available at <http://www.forensiclinguistics.net/iafl-10-proceedings.pdf> (reporting positive findings through the use of qualitative and semi-automatic and quantitative approaches, based on various analyses, including Discriminant Function Analysis); Feiguina & Hirst, *supra* note 14.



## III. THE STUDY

## A. Main Objectives

This article presents a study that explores and develops the possibility of measuring the linguistic differences existing between idiolectal styles and each individual's idiolectal similitude or distance, with the aim of establishing an IIS which will compare several linguistic samples and calculate the linguistic distance between them. The main objective of the establishment of the IIS is to a) create a technique that allows researchers to compare several linguistic samples in terms of the variables that the protocol contemplates, b) calculate the linguistic similitude or distance between them, and c) determine what kind of idiolectal similitude is needed in order to say as definitively as possible that two linguistic samples have been

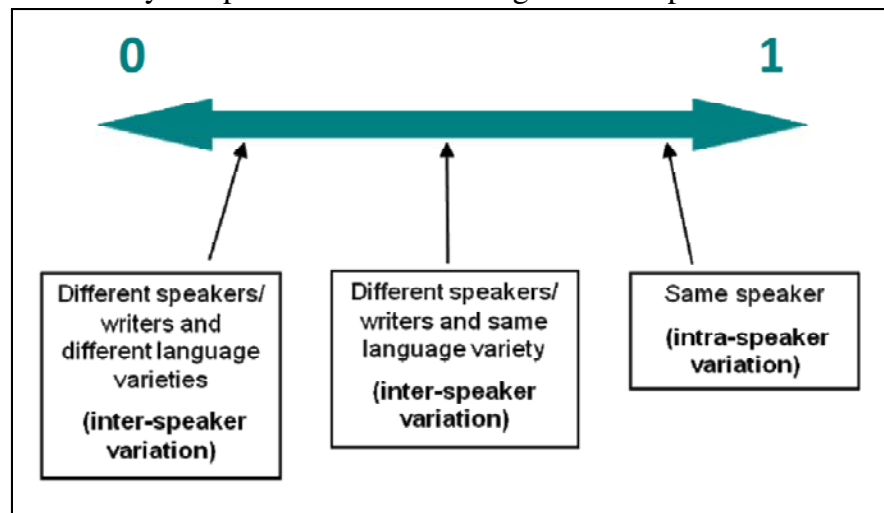


Figure 1: Representation of the IIS as a continuum

produced, or not, by the same individual. The final aim of this project is to be able to apply the IIS methodology to real forensic cases, where instead of comparing two samples from real world data, i.e., where we know who the authors or speakers are, one would compare one disputed and one nondisputed sample or several disputed and nondisputed sample sets.

The IIS is conceived as a continuum (see Figure 1) between 0 and 1, where 0 indicates maximum difference and 1 indicates minimum difference. According to this concept, when two (sets of) linguistic samples, either oral or written, are compared, and the IIS is applied, a result closer to 0 indicates that the two samples under comparison were produced by different individuals and that these samples exhibit interspeaker/writer variation. A value at an intermediate position along the continuum indicates that there is also interspeaker/writer variation, but the slight increase in similarity may indicate that the two individuals share the same linguistic variety. Finally, a value close to 1 would mean that there exists an expected intraspeaker/writer variation but would lead the expert to conclude that the two samples are so similar that they could have been produced by the same individual.

### *B. Theoretical Framework*

The theoretical framework behind the IIS proposal draws from the Theory of Language Variation and Change (“TLVC”) developed by William Labov during the 1960s. The TLVC maintains that language is in a state of constant change and that changes in language can be perceived synchronically by means of variation present at all levels of language. In this sense, linguistic variation was demonstrated not to be random, as previous theories of language had maintained, but proved to be systematic and patterned. This correlates to internal linguistic characteristics such as the particular phonetic context in which a specific sound appears and also external social factors such as gender, age, social class, and level of income.<sup>18</sup> Labov’s theory

---

<sup>18</sup> See, e.g., WILLIAM LABOV, *SOCIOLINGUISTIC PATTERNS* 111, 120–21,

is, according to Turell, “theory building” in terms of three main dimensions: first, in terms of the basic aim stated, which is to describe linguistic variation and change; second, regarding the data it analyzes, which is an individual’s most spontaneous variety, (that is, his or her vernacular); and third, as regards the methods it applies in order to measure this variation, namely observation, description, and explanation.<sup>19</sup> The TLVC studies both individual and group (speech community) variation.<sup>20</sup> This individual–speech community binomial has proved to be very useful, not only in studies of linguistic variation but also in other areas of applied linguistics such as the linguistic profiling aspects of forensic linguist expert witness work. For the purposes of further applications of the IIS to real forensic data, one relevant issue drawn from this theory is the exploration of single dimensions of variation through the binary division of linguistic internal factors, and when relevant, of social factors as well.<sup>21</sup> Also of relevance are the use of multivariate analyses to show the simultaneous effect of all relevant independent variables and the use of cross-tabulation to give a more refined view of the distribution of the data and the degree of independence of intersecting variables.<sup>22</sup>

---

161 (1972) (providing an overview of factors impacting linguistic variation); *see also* 1 WILLIAM LABOV, PRINCIPLES OF LINGUISTIC CHANGE: INTERNAL FACTORS 5 (1994) (“To explain a finding about linguistic change will mean to find its causes in a domain outside of linguistics . . . .”); 2 WILLIAM LABOV, PRINCIPLES OF LINGUISTIC CHANGE: SOCIAL FACTORS 74–75 (2001) (distinguishing between former and current approaches to assessing variation).

<sup>19</sup> *See* M. Teresa Turell, William Labov Laudatio, Universitat Pompeu Fabra (June 15, 2012), *available at* [http://www.upf.edu/enoticies/1112/\\_pdf/laudation\\_turell\\_angles\\_.pdf](http://www.upf.edu/enoticies/1112/_pdf/laudation_turell_angles_.pdf).

<sup>20</sup> *See, e.g.*, WILLIAM LABOV ET AL., ATLAS OF NORTH AMERICAN ENGLISH 69, 157, 285, 303 (2006).

<sup>21</sup> *See, e.g.*, LABOV, *supra* note 1, at 110–121, 160–182 (examining the relationship of sociology and linguistic variations). *See generally* 1 LABOV, *supra* note 18 (discussing the internal factors affecting linguistic variation); 2 LABOV, *supra* note 18 (noting the role of socioeconomics on changes in linguistics).

<sup>22</sup> *See, e.g.*, LABOV, *supra* note 1, at 7–8, 11, 41, 72, 108, 226 n.30 (presenting studies of linguistic variables and the sociolinguistic characteristics these variables reveal); *see also* WILLIAM LABOV, WHAT IS A

In addition to this, drawing from what is now known as forensic sociolinguistics, it can be stated that the linguistic production of an individual can provide clues regarding social factors such as their age, gender, occupation, education, religion, political background, their geographical origin, their ethnicity or race,<sup>23</sup> their nonnativeness when using a second or foreign language, and a variety of language reflecting markers of language contact.<sup>24</sup>

### C. Hypotheses

The working hypotheses to be tested through the analysis of the observed linguistic parameters and variables are the following:

1. Interspeaker/writer variation will be higher than intraspeaker/writer variation. In this sense, IIS results obtained when comparing samples from the same speaker or writer should be closer to 1 than those obtained when comparing samples from different individuals.

2. Despite the existing intraspeaker/writer variation, an individual's idiolectal style will be quite stable throughout time. Consequently, IIS results should be close to 1 when comparing two samples from the same individual from different measurement times.

3. An individual's idiolectal style will also remain relatively stable despite the use of different genres or textual registers but possibly not as stable as it might be throughout time. Therefore, when comparing samples from the same individual involving

---

LINGUISTIC FACT? 12 (1975) (noting the need for improvement in linguistic data methodology as well as the scope of linguistic variation).

<sup>23</sup> See Sharon S. Smith & Roger W. Shuy, *Forensic Psycholinguistics: Using Language Analysis for Identifying and Assessing Offenders*, FBI L. ENFORCEMENT BULL., Apr. 2002, at 16–21, available at <http://diogenesllc.com/statementlinguistics.pdf> (noting the ability of language to reveal characteristics of the speaker).

<sup>24</sup> Turell, *supra* note 10, at 220–25 (noting the ability to use linguistic production to identify users from different geographical regions and users whose first language is not Spanish).

different genres, IIS should also be close to 1 (but perhaps not as close as results in hypothesis 2).

#### *D. Methodology*

The analysis of idiolectal distance that is presented here is based on research carried out in two projects.<sup>25</sup> Each project involved several stages where different numbers of subjects and methods were analyzed. This article presents final results obtained in the last stage, in which six individuals were studied per each module, and a final list of variables, ranging between 10 and 18 depending on the module, were selected after some preliminary studies where some other variables were discarded. Moreover, a total of four different methods were explored, but only three were involved in the final stage. The remaining method, which was based on the Euclidean distance, was finally discarded, and it is not included in this account.

##### *1. Linguistic Modules and Variables*

The protocol devised to calculate the IIS has explored, so far, three different linguistic levels, or modules: the phonological module, the morphosyntactic module, and the discourse-pragmatic module. The phonological module involves the analysis of phonological processes related to insertion, elision, or change of sounds, such as yod-coalescence in English (a process by which a word like *duke* can be pronounced [dju:k] or [dʒu:k]). The morphosyntactic module considers variables related to morphological and syntactic patterns, such as the presence or absence of the conjunction *that* in a sentence like *I thought (that) it was nice*. Finally, the discourse-pragmatic module considers discursive and pragmatic phenomena, such as the choice of the intensifier *really* in contrast with other intensifiers such as *absolutely* or *completely*, as in *I was really/absolutely/completely terrified*.

---

<sup>25</sup> See *supra* note 6.

Regarding the variables, the IIS is concerned with discrete variables<sup>26</sup> that occur in the idiolectal style of the two speakers or writers under analysis, and they all show variation, which is structured in two main variants, either variant A or B, or the presence or lack of the process, following the most standard formulations of linguistic variation analysis.<sup>27</sup> For example, the variable that deals with the process of yod-coalescence explained above contemplates two variants: 1) the presence of the process, by which all instances where yod-coalescence occurs are calculated and 2) the lack of process, by which all the instances where yod-coalescence could occur but does not, are calculated.

## 2. Corpora

Different corpora have been used to test the formulated hypotheses, and all, in one way or another, have involved the elicitation of semispontaneous speech,<sup>28</sup> except for the morphosyntactic module of Spanish, which was analyzed by using a written corpus. Moreover, all the corpora (except that of the discourse-pragmatic module of Spanish) contain data from the same adult men and women collected in two measurement times (“MT1” and “MT2,” respectively) with a lapse of ten to twenty years depending on the module, in order to investigate the subjects’ idiolectal style throughout time.

The corpus of study for the Catalan modules contains data on Eastern Catalan and consists of sociolinguistic interviews recorded in La Canonja, a Catalan speech community in the

---

<sup>26</sup> In statistics, variables may be a) discrete, meaning that they take a limited number of values, such as gender (either male or female) or social class; and b) continuous, which implies any value within a range of values on a scale, such as age, for example.

<sup>27</sup> See, e.g., LABOV, *supra* note 1, at 192–93; WILLIAM LABOV, *THE SOCIAL STRATIFICATION OF ENGLISH IN NEW YORK CITY* 31 (2d ed. 2006).

<sup>28</sup> Semispontaneous speech implies the speech resulting from an interview, where the electronic equipment such as microphones or cameras may make the speaker aware of the situation and inhibit them from using completely spontaneous speech, or their *vernacular*, as it is referred to in sociolinguistics.

Tarragona area in a real-time, Labovian study.<sup>29</sup> In the Spanish modules, several corpora were used: the Mexican Spanish HETA corpus<sup>30</sup> was used to analyze the phonological module, the written Peninsular Spanish corpus<sup>31</sup> was used to analyze the morphosyntactic module and, finally, the Peninsular Spanish corpus,<sup>32</sup> only available for MT1, was used to analyze the discourse-pragmatic module. Regarding the English modules, a corpus containing data on Southern British English in MT1 and MT2 was compiled by means of radio and TV interviews, and the subjects are world-known artists, whose recordings are available online.

### 3. Methods

The three phonological modules in Catalan, Spanish, and English were analyzed following the auditory-acoustic approach,<sup>33</sup> and the three morphosyntactic and discourse-

---

<sup>29</sup> Oral corpus of La Canonja (1987–92), compiled by Juan José Pujadas, Mercè Pujol, and M. Teresa Turell, through 2 CICYT research projects (PBS90-0580 and SEC93-0725).

<sup>30</sup> Fernanda López, *El análisis de las características dinámicas de la señal de habla como posible marca para la comparación e identificación forense de voz: Un estudio para el español de México* (2010) (unpublished Ph.D. dissertation, Universitat Pompeu Fabra), *available at* <http://www.tdx.cat/bitstream/handle/10803/42940/tfle.pdf;jsessionid=AA3F9AC40961A1652DA3E5E543E32BD9.tdx2?sequence=1>.

<sup>31</sup> Maria S. Spassova, *Las marcas sintácticas de atribución forense de autoría de textos escritos en español* (May 2006) (unpublished PhD dissertation, Universitat Pompeu Fabra).

<sup>32</sup> PRESEEA, <http://presea.linguas.net/> (last visited Feb. 6, 2013).

<sup>33</sup> The auditory-acoustic approach to forensic phonetics is the combination of two main approaches. On the one hand, in order to carry out an auditory analysis, phoneticians make use of their knowledge about general phonetics and phonology and the phonetics and phonology of the linguistic system at hand for the interpretation of the acoustic samples being analyzed. On the other hand, an acoustic analysis involves the use of specially developed techniques—normally involving specialized computer software aimed at the acoustic analysis of speech—together with the phonetician's knowledge of physics and the acoustic properties of the speech signal, especially those characteristics most relevant to the language under analysis. For further information, see Francis Nolan, *Speaker Recognition and*

pragmatic modules in these same languages were coded for the different linguistic variables that had been located by their discreteness. Method 1 involves the calculation of an average of the difference in the percentage of occurrences of each variant. On the other hand, method 2 is based on the Adjusted Residual Value ("ARV") obtained after running cross-tabulations, which is a number indicating the difference in the distribution of the variables in the samples compared. Finally, method 3 is based on the Phi Coefficient, which is a coefficient that ranges from 0 to 1 and provides an indication of the strength of the relationship between the variables considered.

#### IV. RESULTS

The results obtained by using the three methods were very similar. However, method 3, which is based on the Phi Coefficient, proved better at accounting for intra- and interspeaker/writer results.

Regarding the phonological modules, hypothesis 1, which stated that intraspeaker results would be higher in the IIS continuum than interspeaker results, is confirmed by all three methods in all three languages. In this article, only results from method 3 will be shown and discussed for all the modules. Figure 2 shows interspeaker IIS results with method 3, where each point in the graph corresponds to an IIS value after comparing samples from two different speakers. Results show that all interspeaker IIS values are relatively low in general (between 0.2 and 0.8), which is an expected result considering that, except for the Catalan corpus, all speakers belong to the same dialectal area. Method 3 has proved useful in the case of the phonological module of Catalan in order to observe that when the IIS is calculated between speakers of different varieties, the interspeaker IIS values are lower than when the speakers compared belong to the same dialectal area, a result

---

*Forensic Phonetics*, in THE HANDBOOK OF PHONETIC SCIENCES 744, 744–67 (William Hardcastle & John Laver eds., 1994); French, *supra* note 2, at 295–96.



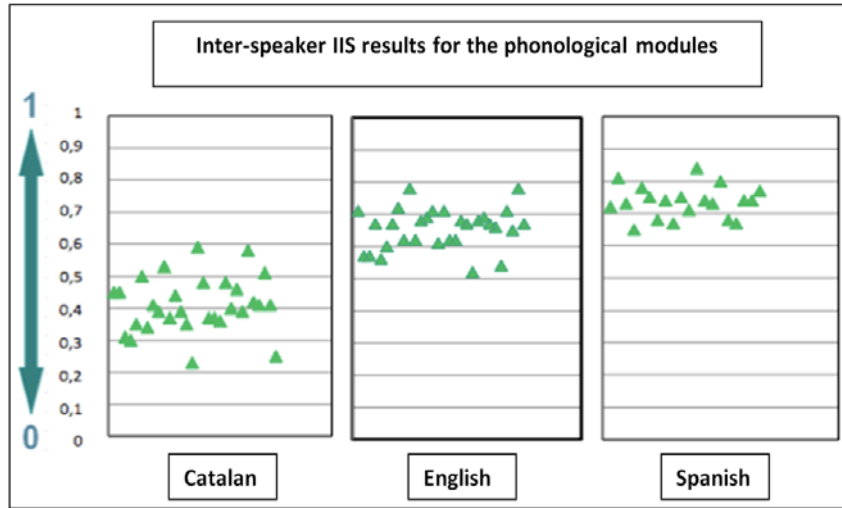


Figure 2: Interspeaker IIS results for the phonological modules

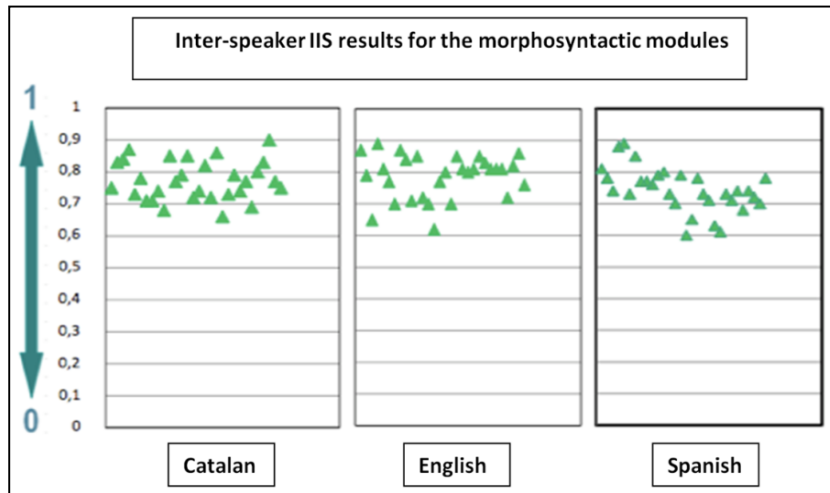


Figure 3: Interspeaker IIS results for the morphosyntactic modules

which is very relevant in real forensic cases concerned with linguistic profiling.

Hypothesis 1 is also confirmed by all 3 methods for the morphosyntactic modules (Figure 3) and the discourse-pragmatic modules (Figure 4). In both modules in the three languages, all interspeaker/writer IIS values are relatively low in general (they range between 0.6 and 0.8), which is an expected result considering that all the subjects belong to the same dialectal area.

Hypothesis 2 stated that an individual's idiolectal style would stay relatively stable despite the course of time. In order to

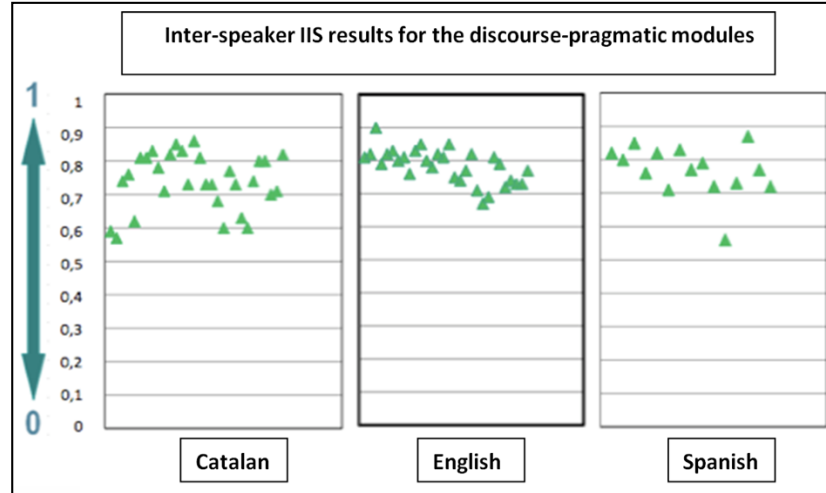


Figure 4: Interspeaker IIS results for the discourse-pragmatic modules.

confirm this hypothesis, samples from the same individual in MT1 and MT2 were compared with each other. In Figures 5–7, points in the graph indicate an intraspeaker/writer IIS result, i.e., an IIS value after comparing samples from the same subjects in two separate points in their lives.

Results show that this second hypothesis is confirmed for both the phonological and the morphosyntactic modules. Figures 5 and 6 illustrate results in these two modules for the three languages. As can be seen, IIS results for all the modules range between 0.8 and 0.9, which is high, as expected, since 1 on the IIS continuum means maximum similarity.

With regard to the discourse-pragmatic modules, hypothesis 2 could only be tested for the Catalan and English modules, since the Spanish corpus for this module did not contain data in two measurement times. Hypothesis 2 is also confirmed with all three methods of Catalan and English. With method 3 (Figure 7), all IIS values are quite high, as expected, with the majority ranging between 0.9 and 0.7.

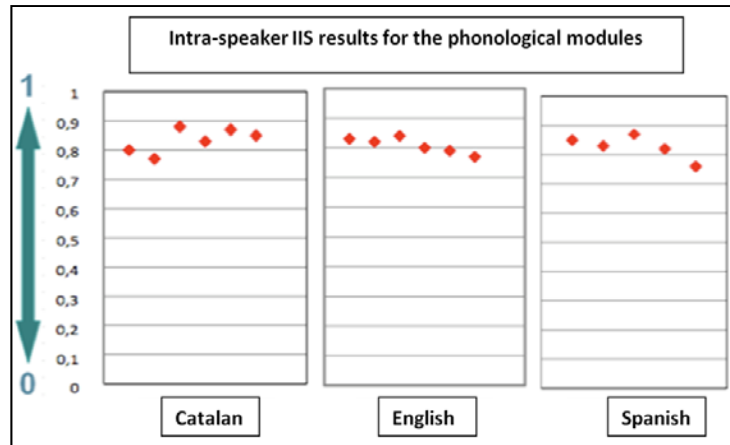


Figure 5: Intraspeaker IIS results for the phonological modules.

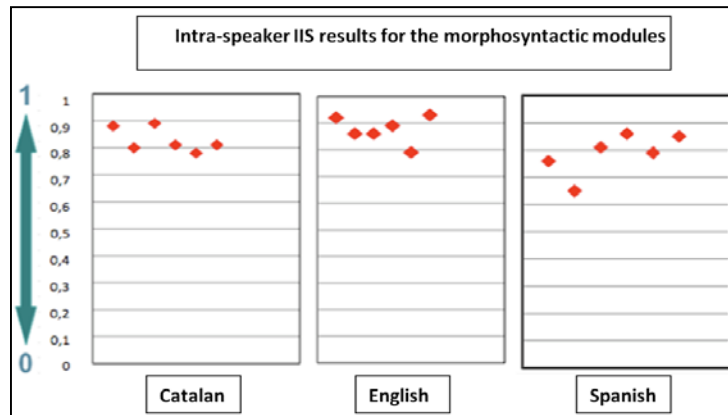


Figure 6: Intraspeaker IIS results for the morphosyntactic modules.

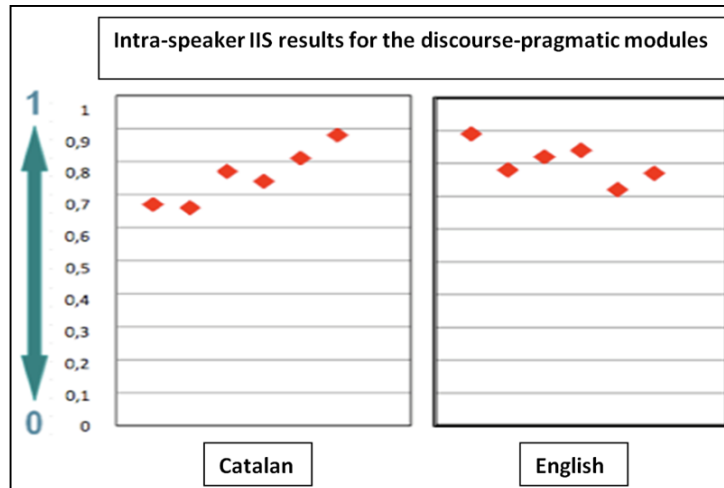


Figure 7: Intraspeaker IIS results for the discourse-pragmatic modules.

## CONCLUSIONS

The first conclusion that we can draw from our results, which has already been noted above,<sup>34</sup> is that method 3, based on the Phi Coefficient, turned out to be the most reliable method, in the sense that it triggered the most robust results for both intra- and interspeaker/writer variation, and in particular in the phonological modules, although with some exceptions. Moreover, hypothesis 1 is confirmed for all modules and languages in that there seems to be more variation, and thus more idiolectal distance, between different individuals than between two samples of the same individual. Also, hypothesis 2 is also confirmed in that samples from the same individual at two measurement times seem to show pretty stable patterns, which would seem to confirm that an individual's idiolectal style (spoken or written) does not appear to vary much throughout time.

If we look more closely into interspeaker/writer IIS results, some IIS values seem to be too high, or at least higher than expected, especially for the morphosyntactic and the discourse-pragmatic modules. In this sense, it should be borne in mind that, except for the phonological module of Catalan, all the subjects considered belong to the same language variety; therefore, high results placed at a middle point along the IIS continuum were expected. However, it is true that in some cases, the IIS methodology does show unexpected results in that some of these interspeaker/writer values are certainly as high as intraspeaker/writer results. We believe that these unexpected results have to do with certain methodological difficulties that we encountered in the process of our research. First, the sample stratification regarding genre, time, language variety, and gender might have had some influence. Not all corpora were stratified for different genres (and at the same time, for different measurement times), and so, for the time being, it has not been possible to test hypothesis 3, which stated that an individual's idiolectal style should be quite stable in spite of the use of different genres. This hypothesis will be explored in the future.

---

<sup>34</sup> See *supra* Part IV.

Regarding time, the phonological module of Spanish only had five speakers in MT2, whereas the discourse-pragmatic module of Spanish contained data in MT1 for all the speakers. As for language variety (or dialect), even if it was not formulated as a hypothesis, the analysis of the phonological module of Catalan, stratified with speakers from two dialects, has proven very robust in its ability to account for interspeaker variation, so it would be desirable to be able to count on all the other modules stratified by language variety. Finally, as regards gender, for the IIS itself and also in order to contribute to the Base Rate Knowledge of population distribution, it would be interesting to test whether there is more interspeaker/writer variation when all speakers are considered together or when a distinction is made in the comparison between female and male speakers or writers.

Another difficulty for comparative purposes—naturally not exclusively related to the IIS measure but which could affect the internal validity of results—has to do with the nature of the variables, namely the different nature that morphosyntactic and discourse-pragmatic variables have in comparison with phonological variables. On the one hand, morphosyntactic and discourse-pragmatic variables have a lower frequency than phonological variables, which could affect final results. On the other hand, the discreteness of morphosyntactic and discourse-pragmatic variables (i.e., their capacity for being formulated as discrete variables with two variants) is much more difficult to establish than that of phonological variables.

Furthermore, it is also possible that the nonparallel nature of the corpora under analysis may have had an effect on the final results. Only in the case of the English (internet TV/radio samples) and the Catalan (La Canonja) IIS calculation, the same corpus was used to analyze the three modules under investigation, while the three linguistic modules of Spanish each contemplated different corpora.

Robust results seem to be associated with the choice of the variables, the establishment of their discreteness, and the number of variables. The more variables, the better IIS results seem to be. The robustness of the IIS will be better grasped when other relevant results are tabulated (for example, when pattern similarity in all modules for each pair of speakers or

writers compared is applied). In other words, for two samples to be attributed to the same speaker or writer, the IIS values must all be near 1 in all modules; for two samples to be attributed to different speakers or writers, IIS values must all be between 0.7 and 0.5 (same speech variety) or between 0.5 and 0.3 (different speech varieties).

The disparity of results obtained in some of the IIS values has had a direct effect on the design of further experiments and on future data collection. Future research will focus on increasing the number of languages as objects of analysis (e.g., Arabic), the sample size (i.e., more subjects for each language), and also on the stratification of the corpora by genre in order to confirm hypothesis 3. Additionally, other indicators such as gender, age, or educational level will be examined to contribute to the Base Rate Knowledge of population distribution.

In conclusion, the IIS measure can provide reliability to the concept of idiolectal similitude or distance, and once the protocol for its calculation is consolidated, the IIS measure may be successfully complemented with other approaches to forensic speech and text comparison to be used in real forensic cases. In addition to this, research towards the establishment of the IIS measure can also provide forensic linguistics with a Base Rate Knowledge of population distribution as regards several linguistic variables for the three modules and the three languages under study, which is a fundamental issue in current forensic linguistic work.

# LINGUISTIC CONFUSION IN COURT: EVIDENCE FROM THE FORENSIC SCIENCES

*Jonathan J. Koehler\**

## INTRODUCTION

When fingerprint evidence was approved for admission in U.S. courts in 1911,<sup>1</sup> the approving court noted that “[e]xpert evidence is admissible when the witnesses offered as experts have peculiar knowledge or experience not common to the world, which renders their opinions, founded on such knowledge or experience, an aid to the court or jury in determining the questions at issue.”<sup>2</sup> In other words, expert testimony is appropriate when a qualified witness has something to say that helps a fact finder in the instant case. One hundred years later, the sentiment expressed in *Jennings* appears in the Federal Rule of Evidence pertaining to the admissibility of expert testimony.<sup>3</sup> But how can we know when expert testimony is helpful rather than unhelpful or even harmful?

I submit that the *specific language* used in court by experts can be the difference between testimony that is truly helpful and testimony that is confusing or unhelpful. This idea is particularly germane to scientific testimony in cases where the triers of fact have a limited understanding of the principles and methods from which the testimony derives its strength.

---

\* Beatrice Kuhn Professor of Law, Northwestern University School of Law.

<sup>1</sup> *People v. Jennings*, 96 N.E. 1077, 1081–82 (Ill. 1911).

<sup>2</sup> *Id.* at 1083.

<sup>3</sup> FED. R. EVID. 702(a) (stating that expert testimony is admissible when it “will help the trier of fact to understand the evidence or to determine a fact in issue”).

This simple point, which has obvious implications for the law, should also be of interest to the forensic linguistics community. Forensic linguists are sometimes called to testify about document content analysis, speaker identification, and authorship, among other things. How, for example, should forensic linguists testify about their analyses of who wrote a particular text message? At present, forensic scientists in general, and forensic linguists in particular, take one of two very different approaches. One approach is testimony that culminates in the expert's *subjective source opinion*. For example, those who practice "forensic stylistics" commonly offer their opinions about who wrote (or who did not write) a document after taking account of such stylistic characteristics as document format, spelling, capitalization, abbreviations, punctuation, word choice, and syntax.<sup>4</sup> Among the more established forensic sciences, fingerprint analysis offers a similarly subjective conclusion.<sup>5</sup> The second approach culminates in a *quantitative statement about the degree of observed correspondence* between an unknown target and a known reference. In forensic linguistics, this approach is favored by computational linguists because it fits well with the field's tendency to identify statistical models for language use. However, there are few, if any, databases from which to generate quantitative statements. DNA analysis, which does rely on large databases to generate probability assessments, offers similarly data-driven probabilistic conclusions. For example, a DNA analyst will commonly report the frequency with which a matching DNA profile exists in a reference population (i.e., the "random match probability").<sup>6</sup>

Regardless of which approach is used at trial, there is a significant risk that expert testimony on scientific and technical matters will confuse or mislead triers of fact.<sup>7</sup> This risk is particularly important whenever material is unfamiliar or

---

<sup>4</sup> ANDRE E. MOENSSENS ET AL., *SCIENTIFIC EVIDENCE IN CIVIL AND CRIMINAL CASES* 252 (6th ed. 2013).

<sup>5</sup> NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD* 139 (2009) ("[F]riction ridge analysis relies on subjective judgments by the examiner.").

<sup>6</sup> MOENSSENS ET AL., *supra* note 4, at 862.

<sup>7</sup> FED. R. EVID. 403.



complex. In these cases, the specific language used by legal actors may be the difference between testimony that is truly helpful to the trier of fact and testimony that is misleading and unhelpful.<sup>8</sup> In this paper, I address issues related to how scientific and technical information should and should not be communicated in court. Because confusion in the DNA and fingerprint areas has been documented and is relatively common, my comments focus on linguistic problems in DNA and fingerprint expert testimony in hopes that forensic linguists can avoid the testimonial traps and errors that plague these forensic scientists.

Section I of this paper examines DNA match statistics and describes the confusion that legal actors experience when dealing with conditional probabilities. Section II examines statistical inverse errors in the 2010 U.S. Supreme Court case *McDaniel v. Brown*.<sup>9</sup> Section III examines a seductive, but faulty, statistical assumption that commonly arises in paternity cases. Section IV examines the role of error rates in forensic sciences and concludes that identifying those error rates is particularly important in fields that rely on highly discriminating statistical techniques. Section V offers an illustration of the crucial role ill-defined language can play in a legal proceeding. Standard and precise terms are recommended. The paper concludes with a section identifying implications for the forensic linguistics and authorship attribution communities.

## I. DNA MATCH STATISTICS

When an expert testifies about DNA evidence found at a crime scene, the punch line is usually statistical. Specifically, after the expert testifies to a “match” (or inclusion) between the DNA profile of an evidentiary sample and the DNA profile of a particular individual, the strength of that match is often described by the random match probability (“RMP”). The RMP

---

<sup>8</sup> The Federal Rules of Evidence require that expert testimony must “help the trier of fact to understand the evidence or to determine a fact in issue.” FED. R. EVID. 702(a).

<sup>9</sup> *McDaniel v. Brown*, 130 S. Ct. 665 (2010).

is a statistic that describes the frequency of a DNA profile in a population.<sup>10</sup> Other things being equal, smaller RMPs (such as one in one billion) indicate a stronger DNA match than larger RMPs (such as one in one hundred) because the chance that the match is purely coincidental in the former instance is much less likely.<sup>11</sup>

---

<sup>10</sup> MOENSSENS ET AL., *supra* note 4, at 863 (“[T]he ‘random match probability’ (RMP) is the probability that a randomly selected, unrelated individual in the relevant population would have a particular DNA profile.”).

<sup>11</sup> Importantly, there are at least two circumstances in which the RMP provides a misleading indicator of the strength of a DNA match. The first circumstance is when the potential source population includes close relatives of the putative source. The chance that a putative source will share a DNA profile with a close relative is usually much larger than the RMP, and therefore the chance of a coincidental match with the crime scene sample is larger as well. See NAT’L RESEARCH COUNCIL OF THE NAT’L ACADS., *THE EVALUATION OF FORENSIC DNA EVIDENCE* 123 (1996). The second circumstance in which the RMP provides a misleading indicator of the strength of a DNA match is when the risk of laboratory error is substantially larger than the RMP. COLIN AITKEN & FRANCO TARONI, *STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENCES* 425 (2004) (“If the probability of an error . . . is much greater than the probability of matching profiles . . . then the latter probability is effectively irrelevant to the weight of the evidence.”); DAVID J. BALDING, *WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES* 35 (2005) (“If the false-match probability (ii) is judged to be much larger than the chance-match probability (i), then the latter probability is effectively irrelevant to evidential weight . . . . [I]t is not the absolute but the *relative* magnitude of the false-match to the chance-match probabilities that determines whether the former can be safely neglected.”); Jonathan J. Koehler et al., *The Random Match Probability (RMP) in DNA Evidence: Irrelevant and Prejudicial?*, 35 *JURIMETRICS J.* 201 (1995) (“RMPs contribute little to an assessment of the diagnostic significance of a reported DNA match beyond that given by the false positive laboratory error rate when RMPs are several orders of magnitude smaller than this error rate.”); Richard Lempert, *After the DNA Wars: Skirmishing with NRC II*, 37 *JURIMETRICS J.* 439, 447 (1997) (“the probative value of a DNA match is always limited by the chance of false positive error”); William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 *J. FORENSIC SCI.* 1, 1 (2003) (“[H]aving accurate estimates [of] the false positive probabilities can be crucial for assessing the value of DNA evidence.”). Laboratory error includes all types of error that might result in a reported match on a person who is not, in fact, the source of the evidentiary item.

The strength of a DNA match may also be given by a likelihood ratio (“LR”).<sup>12</sup> A LR is a ratio of conditional probabilities that examines the probability of observing evidence under two competing hypotheses.<sup>13</sup> The LR technique allows experts to determine how much more (or less) the evidence favors one hypothesis over the other. Ignoring for the time being the twin issues of close relatives and laboratory error identified in footnote 10, the LR is approximately the inverse of the RMP (i.e., 1/RMP).<sup>14</sup> The numerator is approximately 1 (or 100%) because if the putative source is, in fact, the actual source of the evidentiary item, then he or she will share a common DNA profile with the evidentiary item.<sup>15</sup> Similarly, if the putative source is not, in fact, the actual source, then he or she will not share a DNA profile with the true source unless he or she, by sheer coincidence, has the same DNA profile.<sup>16</sup> The RMP captures the chance of a coincidental match and is commonly

---

<sup>12</sup> AITKEN & TARONI, *supra* note 11, at 153–55.

<sup>13</sup> NAT’L RESEARCH COUNCIL OF THE NAT’L ACADS., *supra* note 11, at 31. (“The LR is the ratio of the probability of a match if the DNA in the evidence sample and that from the suspect came from the same person to the probability of a match if they came from different persons.”).

<sup>14</sup> *Id.* (“Since the probability of a match when the samples came from the same person is one (unless there has been a mistake), the likelihood ratio is simply the reciprocal of the match probability.”).

<sup>15</sup> This assumes, of course, that a person’s DNA profile remains constant across time. In rare cases, an individual’s DNA may change. *See, e.g.*, Cai Wenjun, *Rare Mutation Solves Murder*, SHANGHAI DAILY (Nov. 12, 2012), <http://www.shanghaidaily.com/nsp/National/2012/11/12/Rare%2Bmutation%2Bsolves%2Bmurder/> (discussing a rare mutation that aided police in identifying a suspect from a pair of identical twins).

<sup>16</sup> The chance of a coincidental match depends, in large part, on how many loci are examined. Today, thirteen loci are most commonly used, and the resultant random match probabilities are on the order of one in billions, trillions, and quadrillions. *See* NAT’L RESEARCH COUNCIL OF THE NAT’L ACADS., *supra* note 5, at 3–12; NAT’L RESEARCH COUNCIL OF THE NAT’L ACADS., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 142 (3d ed. 2011), *available at* [http://www.au.af.mil/au/awc/awcgate/fjc/manual\\_sci\\_evidence.pdf](http://www.au.af.mil/au/awc/awcgate/fjc/manual_sci_evidence.pdf); WILLIAM C. THOMPSON, THE POTENTIAL FOR ERROR IN FORENSIC DNA TESTING (AND HOW THAT COMPLICATES THE USE OF DNA DATABASES FOR CRIMINAL IDENTIFICATION) 6–7 (2008), *available at* <http://www.councilforresponsiblegenetics.org/pageDocuments/H4T5EOYUZI.pdf>.

inserted into the denominator of the LR.<sup>17</sup> Thus, when the RMP is 1 in 3,000,000, the corresponding LR is often reported as 3,000,000:1. This means that the matching DNA profile is 3,000,000 times more likely under the hypothesis that the defendant is the source of the evidentiary item than under the hypothesis that the defendant is not the source.

What this does not mean, however, is that the defendant is 3,000,000 times more likely to be the source of the evidentiary item than not to be the source. Most people, experts included, would be hard-pressed to explain why this is so. But a careful review of the relevant conditional probabilities provides insight. The LR describes  $P(\text{Evidence} \mid \text{Source}) / P(\text{Evidence} \mid \text{Not Source})$ . However, the statement “the defendant is 3,000,000 times more likely to be the source of the evidentiary item than not to be the source,” describes the posterior odds ratio  $P(\text{Source} \mid \text{Evidence}) / P(\text{Not Source} \mid \text{Evidence})$ . The posterior odds ratio is the inverse of the LR. Those who confuse the LR with the posterior are committing a transposition error or “inverse fallacy.”<sup>18</sup> This error is no mere technicality. Just as we may not assume that the probability that Jack will eat a hot dog given that he is at the ball game (very high probability) is the same as the probability that Jack is at a ball game given that he is eating a hot dog (much lower probability), we may not assume that  $P(\text{Source} \mid \text{Evidence}) = P(\text{Evidence} \mid \text{Source})$  or that  $P(\text{Not Source} \mid \text{Evidence}) = P(\text{Evidence} \mid \text{Not Source})$ .

Nonetheless, people often commit inverse errors when dealing with conditional probabilities.<sup>19</sup> People also confuse

---

<sup>17</sup> NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra* note 11, at 31.

<sup>18</sup> D.H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J. ROYAL STAT. SOC'Y SERIES A 75, 77–78 (1991).

<sup>19</sup> Ward Cascells et al., *Interpretation by Physicians of Clinical Laboratory Results*, 299 NEW ENG. J. MED. 999, 1001 (1978) (showing 45% inverse errors among Harvard physicians); Leda Cosmides & John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty*, 58 COGNITION 1, 25 (1996) (showing 56% inverse errors among Stanford students); Kaye & Koehler, *supra* note 18, at 77 (reviewing inversion fallacy data in pre-DNA mock juror studies conducted in the 1980s); Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies*,

conditional probabilities with joint probabilities,<sup>20</sup> and are less likely to engage in sound probabilistic reasoning when using conditional probabilities than when those probabilities are converted into frequency form.<sup>21</sup> These problems may have significant consequences for legal cases that involve scientific and statistical testimony. Jurors who make these mistakes may believe that the RMP identifies the probability that the defendant is innocent. This belief is known as the “prosecutor’s fallacy.”<sup>22</sup> There is evidence that experts, attorneys, and other legal actors fall prey to this fallacy in actual cases.<sup>23</sup> Similarly, legal actors fall prey to the source probability error,<sup>24</sup> which involves equating the RMP with the probability that the putative source is not the source of the evidentiary item in question. This latter error is so tempting that the RMP is *routinely* described in court

---

*Likelihood Ratios and Error Rates*, 67 COLO. L. REV. 859, 877–78 (1996) (noting that people treat LRs much as they treat posterior odds ratios); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence*, 52 LAW & CONTEMP. PROBS. 9 (1989); Christopher R. Wolfe, *Information Seeking on Bayesian Conditional Probability Problems: A Fuzzy-Trace Theory Account*, 8 J. BEHAV. DECISION MAKING 85, 97 (1995) (noting that 77% of college students verbally confused LRs with posterior odds ratios).

<sup>20</sup> Stephen E. Edgell et al., *Base Rates, Experience and the Big Picture*, 19 BEHAV. & BRAIN SCI. 21, 21 (1996); Gerd Gigerenzer & Ulrich Hoffrage, *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, 102 PSYCHOL. REV. 684, 694 (1995).

<sup>21</sup> Cosmides & Tooby, *supra* note 19, at 25 (comparing errors among Stanford students and finding a 56% rate for inverse errors but only 5% rate when frequencies used); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defense Attorney’s Fallacy*, 11 LAW & HUM. BEHAV. 167, 172–76 (1987) (noting that 22% committed inverse fallacy on blood matching evidence in the context of a hypothetical robbery case when the evidence was presented in P(E | -G) form, whereas a frequency presentation of the blood evidence produced inverse fallacies only 4% of time).

<sup>22</sup> Thompson & Schumann, *supra* note 21, at 171.

<sup>23</sup> *McDaniel v. Brown*, 130 S. Ct. 665, 672–73 (2010); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 JURIMETRICS J. 21, 32 (1993).

<sup>24</sup> *McDaniel*, 130 S. Ct. at 673; AITKEN & TARONI, *supra* note 11, at 81–82; Koehler et al., *supra* note 11, at 212.

opinions as the chance that someone other than the defendant is the source of the genetic evidence.<sup>25</sup>

## II. STATISTICAL INVERSE ERRORS: *MCDANIEL V. BROWN*

Probability errors took center stage, at least in defense filings and an amicus brief,<sup>26</sup> in the U.S. Supreme Court case *McDaniel v. Brown*.<sup>27</sup> In *McDaniel*, Troy Brown was tried and convicted of a rape in Nevada largely based on DNA evidence. Renee Romero, a criminalist for the county, discovered semen on the victim's underwear that matched Brown's DNA profile. On direct exam, Romero estimated the frequency of the DNA profile to be "one in 3 million."<sup>28</sup> When the prosecutor asked "[s]o that means that only one in 3 million people will share the same genetic code?" Romero correctly answered in the affirmative.<sup>29</sup>

The Supreme Court described Romero's testimony on this matter as follows: "The State's expert, Renee Romero, tested the [blood stain] and determined that the DNA matched Troy's and that the probability another person from the general population

---

<sup>25</sup> *State v. Reaves*, No. COA10-1246, 716 S.E.2d 441, at \*3 (N.C. Ct. App. Oct. 4, 2011) (unpublished table decision) ("The lowest probability that someone other than Defendant in the North Carolina African American population contributed the DNA discovered on Ms. Curtis' steering wheel was one in 147,000."); *State v. Timm*, No. 13-11-23, 2012 WL 367589, at \*2 (Ohio Ct. App. Feb. 6, 2012) ("Mr. Weiss testified that the statistical probability that someone other than Timm could be the source of the DNA in the sperm fraction extracted from the shorts was less than one in more than 6.5 billion."); *Murga v. State*, No. 05-10-01237, 2012 WL 807081, at \*2 (Tex. Ct. App. Mar. 13, 2012) ("The third analysis showed a one in 11.1 billion possibility that someone other than appellant had a DNA profile that matched appellant's.").

<sup>26</sup> Brief for 20 Scholars of Forensic Evidence as Amici Curiae Supporting Respondents, *McDaniel*, 130 S. Ct. 665 (No. 08-559), reprinted in Erin Murphy & William C. Thompson, *Common Errors and Fallacies in Forensic DNA Statistics: An Amicus Brief in McDaniel v. Brown*, 46 CRIM. L. BULL. 5 (2010).

<sup>27</sup> *McDaniel*, 130 S. Ct. at 671.

<sup>28</sup> Jury Trial Transcript Day 3, September 29, 1994, *State v. Brown*, No. 5833 (Nev. Dist. Ct. 1994), reprinted in 2 Joint Appendix at 330, 437, *McDaniel*, 130 S. Ct. 665 (No. 08-559) [hereinafter *McDaniel Transcript*].

<sup>29</sup> *Id.* at 438.

would share the same DNA (the ‘random match probability’) was only 1 in 3,000,000.”<sup>30</sup> The Court’s characterization of Ms. Romero’s testimony is ambiguous. When the Court says, “the probability another person from the general population would share the same DNA . . . was only 1 in 3,000,000,”<sup>31</sup> it might mean (a) the chance that *any* person in the general population (The U.S.? The world?) would share the DNA profile in question is 1 in 3,000,000, or (b) the *frequency* with which people in the general population share the DNA profile in question is 1 in 3,000,000. The latter interpretation is the correct interpretation of what Ms. Romero actually said,<sup>32</sup> but one cannot know this from the Court’s imprecise language. The ambiguity in the Court’s restatement here is ironic given that a central issue raised by the defendant in the appeal was the use of imprecise language concerning the DNA evidence at trial and its impact on those who heard it.

#### A. Source Probability Error

In all likelihood, the prosecutor in *McDaniel* wanted the expert to state the RMP in a more dramatic fashion. That is, he probably wanted Romero to describe it as a source probability. Of course, the RMP cannot be converted to a source probability. But the prosecutor nonetheless attempted to extract one from Romero. This following exchange between the prosecutor and Romero ensued:

Q: Now, as far as a—for my benefit, we’re looking at a one in 3 million statistic. Is there another way to show that statistic? In other words, what—let’s say 100 percent—what is the likelihood that the DNA found in the panties is the same as the DNA found in the defendant’s blood?

---

<sup>30</sup> *McDaniel*, 130 S. Ct. at 668.

<sup>31</sup> *Id.*

<sup>32</sup> We know that the latter interpretation is the right interpretation of what Romero said because she agreed with the prosecutor that the one in three million statistic meant that “only one in 3 million people will share the same genetic code.” *McDaniel* Transcript, *supra* note 28, at 438.

A: Paternity testing uses percentages.

Q: Okay.

A: Not the way forensics likes to look at it. We prefer the one in 3 million.

Q: I understand that, but for just another way to look at it, what would that percentage be?

A: It would be 99.99967 [*sic*]<sup>33</sup> percent.<sup>34</sup>

When the prosecutor asks “what is the likelihood that the DNA found in the panties is the same as the DNA found in the defendant’s blood?” he appears to be asking for a source probability. Though hard to interpret, he seems to want Ms. Romero to identify the probability that the DNA in the panties and the DNA in the defendant’s blood share a common source. In other words, he seems to be asking Ms. Romero to identify P(Defendant is the Source of the Recovered DNA | The Recovered DNA Matches the Defendant). Ms. Romero’s initial answer—“Paternity testing uses percentages”—is not responsive. Ms. Romero’s second answer—“we prefer the one in 3 million,” is odd for several reasons. First, “one in 3 million” is a frequency, not a percentage. This contradicts her immediately preceding statement about using “percentages” in paternity testing. Second, it is a simple matter to translate a frequency into a probability. Here, for example, a frequency of one in 3,000,000 is mathematically identical to .000033%.

When the prosecutor presses Ms. Romero further by asking for “another way to look at it,” the “it” transforms from an RMP of 1 in 3,000,000 (or .000033%) to “1 – RMP” or 99.999967%.<sup>35</sup> Obviously .000033% is not the same as 99.999967%. Therefore, it is apparent that Ms. Romero was not referring to the RMP at all when she offered the 99.999967% figure.

Having succeeded in extracting the 99.999967% source probability estimate from his expert, the prosecutor next attempts

---

<sup>33</sup> According to the *McDaniel* transcript, Ms. Romero said “99.99967 percent.” *Id.* at 458. However, she presumably meant (or said) “99.999967 percent” which is the outcome of 100 percent minus .000033 percent.

<sup>34</sup> *Id.*

<sup>35</sup> *Id.*



to get Romero to restate the one in three million RMP as the probability that the defendant is not the source of the DNA, i.e.,  $P(\text{Not Source} \mid \text{Match})$ .

Q: So, the—would it be fair to say, based on that that the chances that the DNA found in the panties—the semen in the panties—and the blood sample, the likelihood that it is not Troy Brown would be .000033?

A: I'd prefer to refer to it as the one in 3 million.

Q: All right. But from a mathematical standpoint, would that be inaccurate?

A: Repeat the question, please.

Q: Would it be fair, then, to say that with that mathematical calculation there, that the likelihood that the DNA extracted from the semen in the panties and the DNA extracted from the blood that the likelihood that it's not Troy Brown, that it's not a match is .000033?

Mr. Lockie [Defense Counsel]: Your Honor, I'm going to object on relevance. The witness is testifying that it's not scientifically valid in her opinion. So it's not relevant.

The Court: Well, I don't know that—

[Mr. Smith (Prosecutor)]<sup>36</sup>: That's just a subtraction problem.

The Court: Let's go back. I don't think that's what she said. I don't think that's what she said. Let's go back a step and find out. I don't think that's what she said.

By Mr. Smith [Prosecutor] (continuing)

Q: Now, I understand that—and what I'm trying to do is make this into a percentage where I can understand it. And so I recognize that as far as your testing, you would prefer to have it as a one in 3 million, but just as another

---

<sup>36</sup> The trial transcript indicates that Mr. Lockie (defense counsel) makes this comment, but it seems unlikely that he would contradict his own objection by stating that this is “just a subtraction problem.” *Id.* at 460–61. The statement was probably made by the prosecutor in response to defense counsel's objection as indicated in the text above.

way of looking at it, would it be inaccurate to state it that way?

A: It's not inaccurate, no.

Q: All right. Then in response to my question, would the likelihood that the semen from the DNA found in the panties and the blood from Troy Brown, that it's not the same, would it be—the chances that they are not a match would be .000033?

A: Yes. That's the way the math comes out.

Q: All right.

THE COURT: Let's make sure. It's the same thing—it's the same math just expressed differently. Is that correct?

THE WITNESS: Yes. Exactly, your Honor.

THE COURT: Thank you.<sup>37</sup>

As before, Romero initially resists the prosecution's efforts to turn an RMP into a source probability by stating a preference for expressing the DNA statistic as a frequency rather than as a probability. But Romero's resistance misses the mark. As noted above, it makes no mathematical difference whether a frequency statistic is expressed as a frequency or as its equivalent probability (decimal) value.<sup>38</sup> One in three million may be described as .00000033 or as its percentage equivalent, .000033%.

What Romero *should have* resisted was the prosecutor's attempt to convert the .000033% RMP statistic<sup>39</sup> into a posterior

---

<sup>37</sup> *Id.* at 460–62.

<sup>38</sup> Of course, although frequencies and their corresponding probabilities are mathematically equivalent, people may respond differently to the form of the presentation. Indeed, there is evidence that people respond differently to frequencies and their mathematically equivalent probabilities in the context of DNA statistics. *See generally* Jonathan J. Koehler & Laura Macchi, *Thinking About Low-Probability Events: An Exemplar-Cuing Theory*, 15 *PSYCHOL. SCI.* 540 (2004) (finding that people were less persuaded by low probability DNA evidence when it was presented in an exemplar-conducive way than when it was not).

<sup>39</sup> The prosecutor omits the “percent” on the .000033% RMP statistic. Although surely unintentional, this omission inflates the RMP from one in 3,000,000 to one in 30,000. McDaniel Transcript, *supra* note 28, at 460–62.

probability that the matchee is not the source of the evidence. The conversion of an RMP into a posterior probability is not simply “another way of looking at it,”<sup>40</sup> as the prosecutor suggested. It is a fallacious maneuver, albeit one that the prosecutor may not have realized was fallacious. Indeed, if one were to assign blame for the statistical confusion on this point, it must fall squarely on the shoulders of the expert witness, Romero. When the prosecutor committed the inverse fallacy and then directly asked Romero, “[W]ould it be inaccurate to state it that way?”<sup>41</sup> Ms. Romero erroneously replied, “It’s not inaccurate, no.”<sup>42</sup> She affirmed this error repeatedly in this exchange both with the prosecutor and then with the trial judge.<sup>43</sup>

We should expect more from forensic science experts who offer statistical testimony. They must know what the inverse fallacy is, they must correct the error when it is made by judges or attorneys, and they certainly must not promote it in their own testimony. When experts commit the error that Ms. Romero committed, they elevate the risk that jurors will believe that the evidence is stronger than it really is.<sup>44</sup>

---

<sup>40</sup> *Id.* at 461.

<sup>41</sup> *Id.*

<sup>42</sup> *Id.* at 462.

<sup>43</sup> Another noteworthy aspect of the exchange above is defense counsel’s objection to the prosecutor’s attempt to lure Ms. Romero into approving and committing a source probability error. Defense counsel objects on grounds of “relevance,” not misstatement of fact. *Id.* at 461.

He protests that the inversion is not relevant because Romero testified that it’s not “scientifically valid.” *Id.* However, as the judge correctly notes, Romero did not expressly reject the inversion as scientifically invalid (as she should have). Instead, Romero only expressed an unjustified preference for expressing the RMP in a particular way. *Id.*

<sup>44</sup> Having identified this risk, I should also note that empirical studies with mock jurors frequently show that jurors undervalue DNA evidence relative to Bayesian norms. *See, e.g.,* David H. Kaye et al., *Statistics in the Jury Box: How Jurors Respond to Mitochondrial DNA Match Probabilities*, 4 J. EMPIRICAL LEGAL STUD. 797, 802 tbl.1 (2007). However, the Bayesian norms generally ignore the role of close relatives and, more importantly, laboratory error. But if one assumes that jurors tend to undervalue DNA evidence, it is possible that source probability errors such as those made by Romero in *McDaniel* may actually *increase* the chance that jurors will give

*B. Prosecutor's Fallacy*

As egregious as Romero's statistical testimony was, the prosecutor committed an even more serious error in his closing argument when he converted the RMP into a probability that the defendant is guilty. This error, which has been referred to both as the Ultimate Issue Error<sup>45</sup> and, more famously, as the Prosecutor's Fallacy,<sup>46</sup> is committed when the RMP is subtracted from 1 and that value is offered to the jury as the probability that the matchee is guilty as charged. Here is what the prosecutor said:

Mr. Smith [Prosecutor]: Consider the fact that, what is the percentage that Troy Brown didn't commit this crime? Was it 75 percent? Are you 75 percent sure? Based on the DNA? 90 percent, 99, sometimes people use the phrase, I'm 99 percent sure about that. Well, in this case the evidence shows—how sure can you be? 99.999967 percent sure.<sup>47</sup>

As noted above, the probability that a defendant is innocent or guilty cannot be determined from the RMP alone. If this were true, then no other evidence in the case would be relevant, including evidence pertaining to the defendant's opportunity and ability to commit the crime in question. At best, an extreme DNA RMP can provide strong proof that a particular person is among the small group of people who might be the source of the DNA evidence. But, it does not address the possibility that a person may be the source of the recovered DNA evidence yet not be responsible for the crime charged. However, when jurors are expressly told that the scientific evidence enables jurors to be "99.999967% sure" that the defendant committed a crime, jurors need only determine whether this percentage is sufficiently high

---

DNA evidence the weight that it is deserves.

<sup>45</sup> AITKEN & TARONI, *supra* note 11, at 82; Koehler, *supra* note 23, at 31–32.

<sup>46</sup> See generally Thompson & Schumann, *supra* note 21.

<sup>47</sup> Jury Trial Transcript Day 4, September 30, 1994, State v. Brown, No. 5833 (Nev. Dist. Ct. 1994), reprinted in 2 Joint Appendix, *supra* note 28, at 588, 730.

to overcome any reasonable doubt they might have about the defendant's guilt. Because few things in life are more than 99.99% certain, some jurors may believe that the statistical evidence in itself provides near certain (and hence sufficient) proof of guilt.

III. THE "NEUTRAL" PRIOR PROBABILITY ASSUMPTION:  
*GRIFFITH V. STATE*

The inverse errors that arose in *McDaniel* are not unusual in cases involving DNA evidence. A similar set of statistical errors were identified and documented in DNA cases in the early 1990s.<sup>48</sup> Perhaps the most worrisome part about these errors is that they are often expressly defended by experts and courts as mathematically proper. Consider *Griffith v. State*.<sup>49</sup> In *Griffith*, the defendant was charged with raping a profoundly retarded patient at a state psychiatric hospital after the sexually inactive patient became pregnant and had a baby.<sup>50</sup> In support of its charge, the prosecution called the director of a Fort Worth-based DNA laboratory to testify about the statistical significance of a paternity DNA match.<sup>51</sup> The DNA expert planned to present a LR of 14,961 (described as a "paternity index") to describe the significance of the DNA match.<sup>52</sup> He also planned to testify that, by using what he referred to as a "neutral" 0.5 prior probability of paternity, the probability that the putative father was the father of the baby in question was greater than 99.99%.<sup>53</sup>

The defense objected to the introduction of the 0.5 prior probability as well as the computation of a probability of paternity that relied on this prior. I was the defense expert in this case. In a preliminary hearing, I testified that the use of the 0.5 prior probability was neither neutral nor an appropriate matter of speculation for the forensic scientist. Instead, I argued that the

---

<sup>48</sup> Koehler, *supra* note 23, at 28–31.

<sup>49</sup> *Griffith v. State*, 976 S.W.2d 241 (Tex. Ct. App. 1998).

<sup>50</sup> *Id.* at 242.

<sup>51</sup> *Id.*

<sup>52</sup> *Id.* at 243–44.

<sup>53</sup> *Id.* at 245.

prior probability should reflect the strength of the nongenetic evidence in the case as determined by the finders of fact. I suggested that the academic literature strongly supported my position and that the use of 0.5 as a “neutral assumption” was not generally accepted in the knowledgeable scientific community. Relatedly, I argued that it was inappropriate for the forensic scientist to offer a “probability of paternity” by using Bayes’ theorem to combine a 0.5 prior probability with a LR of 14,961. I suggested that the method of using a 0.5 prior probability amounts to an attempt to legitimate an inverse fallacy by turning the LR into a posterior odds ratio. Finally, I suggested that a posterior probability of paternity that is computed in this manner could mislead the jury about the strength and meaning of the genetic evidence.

The trial judge rejected my arguments, admitted the DNA expert’s testimony in full, and the defendant was convicted of sexual assault. The verdict was appealed to the Texas Seventh Court of Appeals on the grounds that the 0.5 prior probability violated the defendant’s right to be presumed innocent until proven otherwise.<sup>54</sup>

The defense called the court’s attention to a 1994 Connecticut Supreme Court opinion, *State v. Skipper*, in which the court rejected Bayesian computations in paternity cases that relied on a 0.5 prior probability.<sup>55</sup> In rejecting *Skipper*, the Texas appellate court mischaracterized *Skipper* as having argued that that the probability of paternity statistic assumes that the putative father did, in fact, have sex with the mother rather than *may have had* sex with the mother. *Skipper* did not rely on this argument.<sup>56</sup> Instead, *Skipper* argued that the introduction of an

---

<sup>54</sup> *Id.* at 242.

<sup>55</sup> *State v. Skipper*, 637 A.2d 1101, 1107–08 (Conn. 1994).

<sup>56</sup> In fact, the court in *Skipper* noted that:

[The probability of paternity computation was] predicated on an assumption that there was a fifty–fifty chance that sexual intercourse had occurred in order to prove that sexual intercourse had in fact occurred. The fifty–fifty assumption that sexual intercourse had occurred was not predicated on the evidence in the case but was simply an assumption made by the expert.

*Id.* at 1106 (citations omitted).

arbitrary 50% prior probability of paternity violated the presumption of innocence.<sup>57</sup> The Texas appellate court ultimately defended the 0.5 prior probability assumption because it is frequently used<sup>58</sup> and “neutral.”<sup>59</sup>

The views of the Texas appellate court on the legitimacy of using Bayes’ theorem to convert a LR into a posterior odds ratio by assuming a prior of 0.5 are not unique. Earlier this year, another appellate court cited the *Griffith* court’s arguments favorably.<sup>60</sup> However, it is far from clear that either of these courts understood the underlying math. Both courts claim that Bayes’ theorem is “required” to convert probabilities into percentages.<sup>61</sup> This is not true. As noted earlier, one in 3,000,000 may be described as a probability (.00000033) or as a percentage (.000033%). The conversion of a probability into a percentage is accomplished simply by multiplying the probability by 100 and then placing a “%” at the end of the result. Bayes’ theorem has nothing to do with it. Bayes’ theorem is a formula that tells decision makers how their prior beliefs about, say, a putative father’s paternity, should change in response to new evidence (such as a particular DNA result). It tells decision makers how to move from the probability that a hypothesis is true, to the probability that a hypothesis is true *given new information*.

---

<sup>57</sup> “[W]hen the probability of paternity statistic is introduced, an assumption is required to be made by the jury before it has heard all of the evidence—that there is a quantifiable probability that the defendant committed the crime.” *Id.* at 1107–08.

<sup>58</sup> “[M]illions of HLA and DNA tests around the nation reported paternity results using Bayes’ Theorem and the probability of paternity invoking a .5 prior probability.” *Griffith*, 976 S.W.2d at 246.

<sup>59</sup> “The use of a prior probability of .5 is a neutral assumption. The statistic merely reflects the application of a scientifically accepted mathematical theorem which in turn is an expression of the expert’s opinion testimony.” *Id.* at 247.

<sup>60</sup> *Jessop v. State*, 368 S.W.3d 653, 669 n.19, 674 (Tex. Ct. App. 2012).

<sup>61</sup> *Id.* at 669 n.19 (“Bayes’ Theorem uses a mathematical formula to determine conditional probabilities and is necessary to convert probabilities into percentages.”); *Griffith*, 976 S.W.2d at 243 (“Bayes’ Theorem is necessary to convert probabilities into percentages.”).

## IV. ERROR RATES

The previous two sections documented statistical errors associated with DNA evidence (inverse errors and flawed prior probability assumptions, respectively). One possible response to these errors is to claim that they are inconsequential. According to this argument, even if it is improper to translate a DNA RMP of one in 3,000,000 in *McDaniel* to a source probability of 99.999967%, the extremely small RMP still justifies a strong belief that the matching defendant is, in fact, the source of the recovered DNA evidence. Similarly, even if paternity experts are not justified in assuming a 0.5 prior probability of paternity for all putative fathers, the extreme LR that is commonly observed in paternity cases should give the fact finder confidence that the putative father is indeed the father of the child in question.

It is true that source probability errors and unjustifiable assumptions about prior probabilities are less significant when RMPs and corresponding LR are extreme. However, the Prosecutor's Fallacy, wherein the RMP is equated with  $P(\text{Not Guilty} \mid \text{Match})$ , remains a significant concern when the RMP is extremely small. Even if one infers, from an extremely small RMP, that the matchee is the source of the evidence, this inference should not prompt the additional inference that the matchee must have committed the crime in question. The matchee may be the source of the trace evidence in question, but he or she may not have committed the crime. The trace evidence may have been deposited by the matchee either before or after the crime was committed. Alternatively, the matchee's DNA may have been deposited by the perpetrator himself, either intentionally (as part of a frame up effort) or unintentionally (through inadvertent transfer). In short, those who commit the Prosecutor's Fallacy in cases that include very small RMPs may be relying on weak or irrelevant evidence to justify belief in a defendant's guilt.

In DNA match cases that include very small RMPs, a different consideration should take center stage when gauging the probative value of the evidence: the risk of false positive error.<sup>62</sup>

---

<sup>62</sup> Depending on the facts of the case, the risk that the true source is a



Simply put, the probative value of a DNA match is capped by the frequency with which false positive errors occur.<sup>63</sup> It makes no difference if the RMP is one in millions, billions, or even septillions<sup>64</sup>: if the probability that an analyst will erroneously report a match on two nonmatching DNA samples is 1 in 500, then the corresponding LR is, at best, 500:1.<sup>65</sup> In other words, the false positive error rate—rather than the RMP—tells us most of what we need to know about the probative value of a DNA match. With this in mind, we must ask ourselves whether it pays to risk confusion and various inverse errors by providing fact finders with the RMP at all. Elsewhere I have suggested that in cases where the RMP is several orders of magnitude smaller than the false positive error rate (e.g., RMP = 1 in 1,000,000; false positive error rate = 1 in 500), that the answer is no.<sup>66</sup> There is no need to provide the RMP in such cases because it does not contribute anything beyond the false positive error rate in terms of helping jurors understand a fact in evidence.<sup>67</sup>

What should jurors be told in cases like the one described above? They should be told something like this:

*The suspect reportedly matches the DNA evidence found at the crime scene. The chance that we would report such a match on nonmatching samples, either because of a coincidence or because of an error, is approximately one in 500.*

---

close relative of the matchee may also be an important consideration. However, as DNA matches are based on more and more loci (currently, about thirteen loci), this risk fades considerably. See generally NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., *supra* note 5, at 3-12.

<sup>63</sup> See, e.g., Jonathan J. Koehler, *Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter*, 59 HASTINGS L.J. 1077, 1079 (2008) (“[T]he false positive error rate limits and controls the probative value of the match report.”).

<sup>64</sup> *People v. Odom*, No. B225910, 2011 WL 5121175, at \*5 (Cal. Ct. App. Oct. 31, 2011) (“[The state’s DNA expert] testified that two in 24 septillion people . . . would be expected to match that profile.”).

<sup>65</sup> Recall that the LR is approximately the inverse of the RMP (i.e., 1/RMP). See *supra* text accompanying note 14.

<sup>66</sup> Koehler et al., *supra* note 11, at 210.

<sup>67</sup> The Federal Rules of Evidence require that expert testimony be helpful to the trier of fact. FED. R. EVID. 702(a).

Again, the RMP of 1 in 1,000,000 contributes nothing of value beyond this. Indeed, this RMP might actually promote confusion by inviting jurors to commit one of the fallacies described previously. Or it might tempt them to commit other errors such as averaging the RMP with the error rate, or ignoring the error rate altogether based on a mistaken belief that the RMP is the more relevant statistic. Empirical data showing that fact finders are improperly influenced by RMPs in these situations<sup>68</sup> support the argument that introduction of RMPs can be more harmful than beneficial.

At this point, one might wonder whether forensic science statistics of all sorts should simply be hidden from fact finders altogether. Perhaps we should let the forensic scientists handle the numbers in their laboratories but then have those same experts offer more qualitative opinions sans numerical data at trial. The truth is that forensic science testimony rarely includes a quantitative component outside of the DNA context. Non-DNA forensic scientists commonly offer their opinions about who or what is the source of the forensic science evidence (e.g., a hair, a shoeprint, a tire track, a bite mark, a fingerprint, a fiber, etc.). In some domains, forensic scientists use vague terms such as “consistent with,” “match,” and “could have come from” to explain their failure to find critical differences between two hairs, two fingerprints, etc. The central problem with such terms is that they lack consensus meaning. Two hairs may be “consistent with” one another because they are both brown and thick. Or they may be consistent with one another because they share a large collection of rare features. Without more information about the size of the set of included and excluded features, fact finders may find it hard to assign weight to qualitative terms.

#### V. LINGUISTIC MESS: PRELIMINARY HEARING

In some forensic areas (e.g., fingerprints and shoeprints), forensic scientists resort to strong language to report their opinions, referring to matches as “identifications” and

---

<sup>68</sup> Koehler et al., *supra* note 11, at 210–11.

“individualizations.”<sup>69</sup> Some experts use those two words interchangeably to indicate that the matching person or object is the one and only possible source of the marking to the exclusion of all others in the world. Indeed, phrases such as “to the exclusion of all others in the world” are commonly used by forensic scientists in many non-DNA disciplines to declare their opinions about who or what is the source of an evidentiary item.<sup>70</sup>

Recently, some examiners have tried to distinguish between the words “individualization” and “identification” by suggesting that individualization is a factual state of the world whereas an identification is merely the opinion of the examiner. Consider the following cross-examination of a respected fingerprint examiner in a 2008 preliminary hearing on the admissibility of fingerprint evidence:

Q: Okay. And by comparing the unknown prints to the known prints, you hope to either declare an individualization or an exclusion between the unknown and the known, correct?

A: Well, when you say individualization and it’s kind of a—when I come to my result, I’m actually referring to that as an identification. Individualization, the scientific community, kind of the international, it’s ah, more along the lines of excluding it to the possibility of all others on the face of the earth. But when we say an identification . . . I am telling you that I am confident that that latent print was made by this particular person.

Q: And that is, meaning that particular individual?

A: Yes.

Q: So that would be an individualization; you’d be saying that this individual left that print?

A: Ah, no. . . . [W]hen I say identification, it is my opinion and that I am confident in my result that this

---

<sup>69</sup> MOENSSENS ET AL., *supra* note 4, at 454.

<sup>70</sup> See Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199, 206 (2008) (quoting *United States v. Green*, 405 F. Supp. 2d 104, 107 (D. Mass. 2005)).

latent print and the known prints that I am comparing to were made by the same source.

Q: Meaning that individual?

A: No.<sup>71</sup>

At the beginning of this exchange, the expert distinguishes an individualization from an identification, suggesting that his own identification conclusion is a mere statement about who he believes is the source of the prints, rather than a statement that excludes the possibility that anyone else on earth could be the source. But, in drawing this distinction, the expert appears to directly contradict himself. At first, he says “yes” in response to the attorney’s question about whether an identification means that a print was made by “that particular individual?” But seconds later, when the attorney repeats his question (“meaning that individual?”), the expert says “no.”<sup>72</sup>

Trial transcripts are littered with confusing exchanges between attorneys and witnesses. Despite this, the exchange above is noteworthy both because the content is important and difficult and because this expert is so highly regarded. Although cross-examination has been referred to as “the greatest legal engine ever invented for the discovery of truth,”<sup>73</sup> cross-examination often does not afford experts the opportunity to expand and clarify answers to complex issues. One can only imagine what the judge (or jury) would take away from the exchange above.

## CONCLUSION

What does all of this mean for the burgeoning fields of forensic linguistics and authorship attribution? First, it means that these communities would be wise to set up clear and unambiguous standards for examining materials, documenting their findings, and reporting those findings in court. Doing so

---

<sup>71</sup> Transcript of Proceedings at 48–49, *State v. Hull*, 788 N.W. 2d 91 (Minn. 2010) (No. 48-CR-07-2336).

<sup>72</sup> *Id.* at 49.

<sup>73</sup> 5 JOHN HENRY WIGMORE, *EVIDENCE IN TRIALS AT COMMON LAW* § 1367, at 32 (1974).

will help forensic linguists persuade courts that their evidence is based on reliable methods and will be helpful to jurors.<sup>74</sup> At the very least, expert witnesses should examine materials in a common way, use agreed-upon standards for identifying and recording consistencies and inconsistencies in evidentiary materials, and use a common language to describe findings and conclusions to triers of fact. To facilitate these goals, the forensic linguistics community should establish a professional body that not only promotes these goals but also certifies experts and, where applicable, accredits training programs and laboratories.

As indicated earlier, the forensics linguistics community appears to be divided on the question of whether it favors qualitative versus quantitative methods. Whereas forensic stylists favor the qualitative approach, computational linguists and computer scientists in the field favor a quantitative approach. Regardless of which approach prevails, the field will likely succeed or fail as a function of the scientific quality of its methods. This metric favors the quantitative approach, though the field will need to do a better job developing the requisite databases and transparent methodologies. In an analogous manner, some of the more traditional forensic sciences, such as fingerprinting and voiceprint analysis, are beginning to explore quantitative approaches.<sup>75</sup>

Of course, dangers await. As the field moves toward more probabilistic analyses and outputs, inverse errors may be committed both in and out of the courtroom. It is therefore imperative that the forensic linguistics community identify clear and consistent standards for reporting and testifying about results

---

<sup>74</sup> Scientific evidence must be reliable according to the U.S. Supreme Court. *Daubert v. Merrell Dow Pharm., Inc.* 509 U.S. 579, 589 (1993). As noted previously, the Federal Rules of Evidence further require that expert testimony be helpful to the trier of fact. FED. R. EVID. 702(a).

<sup>75</sup> Christophe Champod & Ian W. Evett, *A Probabilistic Approach to Fingerprint Evidence*, 51 J. FORENSIC IDENTIFICATION 101, 117–18 (2001); Geoffrey S. Morrison, *Measuring the Validity and Reliability of Forensic Likelihood-Ratio Systems*, 51 SCI. & JUST. 91 (2011) (quantifying the accuracy of forensic voice prints); Cedric Neumann et al., *Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae*, 52 J. FORENSIC SCI. 54, 54–64 (2007).

and include training in elementary statistics and probability for its members. Regarding the latter recommendation, it is not enough that a forensic field has good scientific intentions and embraces rigorous scientific principles: expert witnesses who provide quantitative testimony must understand enough about statistics and probability to avoid, explain, and correct statistical misstatements when they arise.

The forensic linguistics community should also support a rigorous proficiency-testing program, using realistic evidentiary items, for all techniques and experts. Participation in the program, which should be conducted by an external agency that does not have an interest in demonstrating positive outcomes, should be mandatory for courtroom testimony.<sup>76</sup> Such tests can alert the field and the courts to strengths and weaknesses associated with various techniques and can provide reasonable first-pass estimates for relevant error rates.

Finally, forensic linguistics can learn from the recent battles waged over the individualization claims made by fingerprint examiners.<sup>77</sup> As the exchange in *State v. Hull* documented in Section V indicates, some examiners recognize that individualization claims reach beyond the available data in most

---

<sup>76</sup> Jonathan J. Koehler, *Proficiency Tests to Estimate Error Rates in the Forensic Sciences*, 12 LAW PROBABILITY & RISK 89 (2013); Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCIENCE 892, 893–94 (2005). Some of the traditional non-DNA forensic sciences appear to be moving in this direction. A recent Expert Working Group report on latent print examination recommended a similar testing program for fingerprint examiners. EXPERT WORKING GROUP ON HUMAN FACTORS IN LATENT PRINT ANALYSIS, LATENT PRINT EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH 187–88 (David H. Kaye ed., 2012).

<sup>77</sup> See generally Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 LAW PROBABILITY & RISK 233 (2009); Simon A. Cole, *Who Speaks for Science? A Response to the National Academy of Sciences Report on Forensic Science*, 9 LAW PROBABILITY & RISK 25 (2010); Jonathan J. Koehler & Michael J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted*, 75 BROOK. L. REV. 1187 (2010); Saks & Koehler, *supra* note 70. But see David H. Kaye, *Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies*, 75 BROOK. L. REV. 1163 (2010).

(if not all) forensic sciences. Forensic linguistics would do well to *offer conservative, descriptive claims and to support those claims with empirical data*. Source claims (e.g., “In my opinion, this text was written by the defendant”) should be avoided. Such a modest approach will not only help forensic linguistics gain a place in the courtroom, but it will also reduce the risk that jurors will overweigh this potentially important, but as yet untested, evidence.

## BEING PRAGMATIC ABOUT FORENSIC LINGUISTICS

*Edward K. Cheng\**

If my late colleague Margaret Berger taught me anything about evidence, it was that the field seldom yields easy answers. After all, law is necessarily a pragmatic discipline, especially when it comes to matters of proof. Courts must make their best decisions given the available evidence. They have neither the luxury of waiting for better, nor the ability to conjure up, evidence (or new technologies) that they wished they had.

Scholars, by contrast, are naturally attracted to the ideal, sometimes like moths to a flame. Ideals reflect the values and commitments of our society, and they provide the goals that inspire and guide research. But when assessing a new field like forensic linguistics as a legal academic, one needs to carefully separate the ideal from the pragmatic. For when it comes to real cases, evidence law can ill afford to allow the perfect to be the enemy of the good.

Bearing this admonition firmly in mind, this article aims to provide some legal context to the Authorship Attribution Workshop (“conference”). In particular, I want to offer some *pragmatic* observations on what courts will likely demand of forensic linguistics experts<sup>1</sup> and tentatively suggest what the field should aspire to in both the short and long run.

---

\* Professor of Law, Vanderbilt Law School; Ph.D. Candidate, Department of Statistics, Columbia University. Thanks to Larry Solan for organizing this remarkably interdisciplinary conference and to Dashiell Renaud for research assistance.

<sup>1</sup> While “forensic linguistics” may encompass a broader set of techniques, I will use the term synonymously with the use of linguistic methods for purposes of attributing authorship, the focus of the conference.



I. *DAUBERT*

No discussion of scientific evidence—at least no discussion of scientific evidence in the United States—can begin without referencing *Daubert v. Merrell Dow Pharmaceuticals*.<sup>2</sup> *Daubert* establishes a five-factor test for the admissibility of scientific evidence: i) falsifiability and testing; ii) publication and peer review; iii) error rates; iv) standards; and v) general acceptance.<sup>3</sup> Unfortunately, applying these factors to many of the forensic linguistic methods presented at this conference immediately raises concerns. The methods do not have rigid procedures that have been tested or have known error rates. Excepting the contributions in this issue of the *Journal of Law and Policy*, few have ever been published. And, almost by definition, since forensic linguistics is an emerging field, many techniques lack general acceptance.

The principal issue is not that forensic linguistic methods are junk. Rather, the problem is that forensic linguistic methods often change from one case to another to account for case-specific contours: Malcolm Coulthard's case study involved selecting certain misspellings and word choices made over e-mail,<sup>4</sup> while Tim Grant's study explored the peculiar grammar of text messaging.<sup>5</sup> The result is a "moving target," and while moving targets are not necessarily bad as a theoretical matter, they are a big problem for the *Daubert* test, which envisions standardized, broadly applicable (and broadly applied) techniques.

Does this mismatch spell doom for the field? Will forensic linguists thus inevitably face widespread opposition and exclusion by judges? Emphatically no. As many in the scientific evidence community have long observed, *Daubert* in practice fundamentally differs from *Daubert* in theory. In real life, courts often treat the *Daubert* factors more as incantation than as actual

---

<sup>2</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

<sup>3</sup> *See id.* at 593–94.

<sup>4</sup> Malcolm Coulthard, *On Admissible Linguistic Evidence*, 21 J.L. & POL'Y 441 (2013).

<sup>5</sup> Tim Grant, *TXT 4N6: Method, Consistency, and Distinctiveness in the Analysis of SMS Text Messages*, 21 J.L. & POL'Y 467 (2013).

requirements.<sup>6</sup> What they really impose is an impressionistic type of scrutiny, giving the expert testimony a “hard look” for intellectual rigor, but nothing more.

Courts have gravitated toward hard-look scrutiny not out of laziness or ignorance<sup>7</sup> but out of pragmatism. The *Daubert* case itself arose in the pharmaceutical context, where large datasets, standardized treatments, and statistical studies reign. What the *Daubert* test demands is thus perfectly reasonable in that context. In other contexts, however, useful expertise exists in the absence of such data. For example, like forensic linguists, accident reconstruction experts also customize their analyses based on case specifics. This customization again means little standardization or statistical justification. Yet, courts have regularly admitted reconstruction experts under hard-look review.<sup>8</sup>

The contours of this hard-look test seem to boil down to three somewhat related inquiries. First, is the expert overselling the power of his technique? Courts display little patience with expert grandstanding, strongly preferring ones who carefully delineate what their techniques can and cannot do.<sup>9</sup> Second, does the expert provide a rational explanation for how the technique works? *Daubert* is in many ways an emphatic rejection of *ipse dixit* or say-so testimony.<sup>10</sup> Even though jurors lack technical expertise, *Daubert* tasks them with engaged, reasoned, critical decision making. Blind deference to the authority of a well-

---

<sup>6</sup> Cf. 5 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 43:10, at 782 (2012) (“As a result, the Daubert factors have become something akin to incantation in the structural engineering context, rather than a roadmap for rigorous inquiry.”).

<sup>7</sup> But see Sophia I. Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-Daubert World*, 25 LAW & HUM. BEHAV. 433, 454–55 (2001) (suggesting that many state court judges may not fully understand the *Daubert* factors).

<sup>8</sup> See FAIGMAN ET AL., *supra* note 6, § 44:10, at 810 (“[C]ourts take a pragmatic view, admitting [accident reconstruction] testimony even when testing is absent or is otherwise imperfect or flawed.”).

<sup>9</sup> See *id.* §§ 45:4–7 (discussing flaws in expert economic analyses).

<sup>10</sup> Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) (“[N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.”).

credentialed expert simply will not do.<sup>11</sup> Finally, is the expert willing to acknowledge and address criticisms of his technique? Overdefensiveness or blithely ignoring well-founded objections often betrays a certain lack of understanding, another worthy ground for exclusion.<sup>12</sup>

Viewed in this light, there is little surprise that courts have generally permitted the linguists at this conference to testify in court,<sup>13</sup> and this trend will likely continue. At least within this hand-picked subpopulation, the experts do not oversell their wares and carefully circumscribe the conditions under which their methods apply. They provide reasoned explanations, and I suppose the mere fact of their attendance at this conference demonstrates a profound commitment to taking objections seriously.

## II. A (LONG-TERM) WISH LIST

As argued above, courts are likely to admit forensic linguistics as it currently stands. But presumably, this conference's focus is not merely this basic doctrinal question. Rather, Larry Solan's vision was to consider what forensic linguistics might become and how the field might best aid the legal system.<sup>14</sup> In this aspirational vein, let us therefore consider

---

<sup>11</sup> See generally Ronald J. Allen & Joseph S. Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 NW. U. L. REV. 1131 (1993) (discussing whether the role of experts is to educate the jury or to arrive at conclusions to which a jury defers).

<sup>12</sup> Cf. FAIGMAN ET AL., *supra* note 6, § 43:14, at 786–87 (discussing the courts' use of "robustness tests," which test how well an expert addresses alternative theories or contrary evidence, in the structural engineering context).

<sup>13</sup> Perhaps the most striking example is Carole Chaski, who reports having been allowed to testify in a *Frye* state even after noting repeatedly that her method was experimental and still under development, a condition clearly at odds with her methods being "generally accepted"—the sole criterion for admissibility under a *Frye* test. See Carole Chaski, *Best Practices and Admissibility of Forensic Author Identification*, 21 J.L. & POL'Y 333, 358 (2013). The suspicion, naturally, is that even in *Frye* jurisdictions, what matters to courts is not the headcount associated with a method but the intellectual rigor of the method as probed by the hard-look test.

<sup>14</sup> Lawrence Solan, *Intuition Versus Algorithm: The Case of Forensic*

a “wish list” of attributes that the law might want from the field. In an ideal world, we would probably like forensic linguistic analysis to have:

- a widely adopted, predefined algorithm (preferably automated);
- a large, random sample of known exemplars (preferably subclassified by topic and genre); and
- a well-understood theoretical underpinning.

These goals are not my brainchild but have been implicit in many comments, criticisms, caveats, and apologies heard throughout this conference. We all seem to wish that forensic linguistics had fewer ad hoc, case-specific methods so that we could have more rigorous testing and known error rates. We wish that we had a larger and more detailed set of training data so that we could be more confident about external validity. And finally, the linguists, although perhaps not the computational ones, would feel more comfortable if the methods and results were better rooted in linguistic theory.

A moment’s reflection suggests the loftiness of these goals. Only one forensic method arguably satisfies them all—DNA. DNA has a widely adopted, predefined, largely automated algorithm; a large, random sample of known exemplars; and a well-understood theoretical underpinning. That is not to say that its history and development were without controversy,<sup>15</sup> but that is where matters stand today. No other forensic field can make such claims.

Juxtaposed to DNA, forensic linguistics clearly has a long way to go. Nearly all of the procedures and algorithms presented at this conference involve some degree of ad hoc expert tweaking and customization, particularly those used for short writing samples. The computational procedures that

---

*Authorship Attribution*, 21 J.L. & POL’Y 551 (2013).

<sup>15</sup> For example, forensic DNA evidence generated two National Academy of Sciences reports in rapid succession. The first, published in 1992, failed to resolve controversies that were later largely put to rest in the second, published in 1996. See NATIONAL RESEARCH COUNCIL, *THE EVALUATION OF FORENSIC DNA EVIDENCE* 10–11 (1996) (“[W]e agree with many recommendations of the earlier [report] but disagree with others.”).

involve less tweaking ideally require a large, random sample of exemplars that currently does not exist. And in almost all cases, the theoretical underpinning for the results is opaque. For example, participants offered some off-the-cuff rationales for why *n*-grams<sup>16</sup> or the other machine learning methods work, but no one *really* understands what is going on.

These ultimate goals are surely daunting, but we should be encouraged that the leaders in the forensic linguistics community have set their sights correctly on the prize.

### III. SHORT-TERM ASPIRATIONS

With the long-term goals set, let us consider what courts might demand from forensic linguistics in the short term. As I mentioned in the introduction, the legal system must be more pragmatic in the short term, so what exactly should it demand? In this context, *Daubert* hard-look review in conjunction with the other evidentiary rules provides a convenient short-term checklist for forensic linguists.

1. *The testimony must add value.* This requirement is at the heart of the relevance standard established by Rule 401<sup>17</sup> and the “help the trier of fact” standard governing experts under Rule 702.<sup>18</sup> At the very minimum, forensic linguists should be more than highly credentialed window dressing on common sense. They must add substantive value.

This requirement appears easily met, especially when the expert moves beyond obvious identifying features such as misspellings or unusual word choices. For example, techniques exploiting syntactic structure, choice of function words or grammar, or *n*-grams clearly represent ideas beyond the ken of the average (or even sophisticated) juror.

---

<sup>16</sup> An *n*-gram is a sequence of *n* adjacent items—words, phrases, or characters—from a given text, forming the basis for analysis.

<sup>17</sup> FED. R. EVID. 401.

<sup>18</sup> FED. R. EVID. 702(a) (“A witness who is qualified as an expert . . . may testify in the form of an opinion or otherwise if: (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue . . .”).

2. *The testimony must enlighten more than distort or confuse.*

This second requirement has both evidentiary and statistical inspirations. Evidentiarily speaking, Rule 403 requires that the probative value of evidence not be substantially outweighed by its potential for unfair prejudice, confusion of the issues, or misleading the jury.<sup>19</sup> Statistically speaking, George Box's well-known maxim sums up the problem in a nutshell: "[A]ll models are wrong, but some are useful."<sup>20</sup>

On this score, somewhat counterintuitively, the trend toward quantitative and statistical measures may be more worrisome than more traditional, off-the-cuff qualitative methods. To be sure, as Jay Koehler notes, qualitative methods present hazards through loaded and vague words like "match" and "consistent."<sup>21</sup> But jurors are at least more comfortable weighing that kind of evidence, and attorneys educated about these issues can effectively attack them.

Statistical measures of linguistic similarity are another matter. Statistical methods always have underlying assumptions and potential problems, and asking jurors (or even opposing counsel) to ferret out the distortions created by flawed models is unrealistic. Unless the method is so well-trodden and well-accepted that a jury can essentially use its results uncritically, I worry that statistical models in this context may distort more than illuminate.

3. *The testimony must be sufficiently transparent to permit reasoned decision making.* This third requirement originates from *Daubert's* hard-look test, as well as Rule 702's demand that a conclusion not rest solely on the *ipse dixit* of an expert.<sup>22</sup>

All of the experts at this conference would presumably meet this criterion with ease, since they have all cogently explained and defended their methods. I can envision two instances, however, in which forensic linguistic testimony could run afoul

---

<sup>19</sup> FED. R. EVID. 403.

<sup>20</sup> GEORGE E.P. BOX & NORMAN R. DRAPER, EMPIRICAL MODEL-BUILDING AND RESPONSE SURFACES 424 (1987).

<sup>21</sup> Jonathan J. Koehler, *Linguistic Confusion in Court: Evidence from the Forensic Sciences*, 21 J.L. & POL'Y 515, 534 (2013).

<sup>22</sup> FED. R. EVID. 702; *see also* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997).

of this requirement. The first is the purely impressionistic linguist, who relies solely on his or her “training and experience.” Lest this example seem like a straw man, let me note that authentication attempts in other fields frequently proceed along these lines. For example, art experts studying the Getty kouros reported feeling an inexplicable revulsion upon first seeing the statue, and these gut feelings often provided a foundation for their assessment that the statue was a fake.<sup>23</sup> Such intuitions are surely not nonsense, and arguably the legal system should prefer an art expert’s opinion over the average juror’s, but *Daubert* makes clear that *ipse dixit*, “blink”-type testimony does not make the cut.<sup>24</sup>

The second potentially problematic instance is where a machine-learning algorithm arrives at an empirically successful identification rule (i.e., high accuracy), but researchers have little idea why it works as a matter of substantive theory.<sup>25</sup> With its emphasis on predictive accuracy over interpretability, machine learning tends toward such black boxes, and while I personally sympathize with the approach, the legal system with its emphasis on reasoned decision making typically does not.

4. *The method must have some proven empirical validity.* This final requirement is based again on the text of Rule 702<sup>26</sup> but may be the most difficult short-term aspiration for the field. The *sine qua non* of empirical validity is testing. For some of the data-intensive, quantitative methods presented at this conference, a focus on testing is practically inherent. But

---

<sup>23</sup> Georgios Dontas, *The Getty Kouros: A Look at Its Artistic Defects and Incongruities*, in THE GETTY KOUROS COLLOQUIUM 37, 37 (Angeliki Kokkou ed., Alex Doulas trans., 1993) (“In the controversy regarding the authenticity of the Getty kouros a factor that must be taken into account is, in my opinion, the unfavourable feeling it arouses at the very first glance.”); see also MALCOLM GLADWELL, BLINK 3–8 (2005) (discussing the Getty kouros).

<sup>24</sup> *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 588–89 (1993).

<sup>25</sup> See generally Leo Breiman, *Statistical Modeling: The Two Cultures*, 16 STAT. SCI. 199 (2001) (discussing the two cultures of statistics: one focused on explanation, and the other on prediction).

<sup>26</sup> FED. R. EVID. 702 (“A witness who is qualified as an expert . . . may testify in the form of an opinion or otherwise if: . . . (c) the testimony is the product of reliable principles and methods . . .”).

methods such as those proposed by Coulthard<sup>27</sup> or Grant,<sup>28</sup> which are more qualitative, subjective, or case-specific, will require experts to embrace proficiency testing and out-of-sample testing more affirmatively.

For qualitative linguistic experts, courts should demand proficiency testing—tests of ability involving known problems given under blinded conditions.<sup>29</sup> Such testing is undoubtedly no fun for the experts involved. The experts open themselves up to attack if the testing turns out badly, and the risk of endangering a lucrative line of business creates substantial disincentives to participate. Experts will thus require judicial prodding, for without such information about accuracy rates, jurors cannot assess the probative value of an expert's conclusions.

For case-customized models, any reported accuracy rates must be out-of-sample accuracy rates. Constructing models that merely fit the data on hand is one thing; successfully predicting future data is an entirely different matter. Tailoring methods or models to a specific case is a time-honored recipe for creating overfitted models, which explain the current dataset well but handle future datasets poorly. To get proper accuracy rates, researchers must divide their dataset into training and testing sets. Models should be developed only with the training set, and validation should be done only with the separate testing set. Some of the conference papers used out-of-sample testing, while others either did not or were unclear.<sup>30</sup>

Finally, part and parcel of testing is the establishment of standardized procedures. As the forensic linguistics field matures, it will have to sacrifice some of its flexibility for

---

<sup>27</sup> Coulthard, *supra* note 4.

<sup>28</sup> Grant, *supra* note 5.

<sup>29</sup> Proficiency testing has been proposed as the solution to *Daubert* in other contexts involving subjective, expert-dependent determinations, such as fingerprints. *E.g.*, Jennifer L. Mnookin, *The Courts, the NAS, and the Future of Forensic Science*, 75 BROOK. L. REV. 1209, 1217–33 (2009).

<sup>30</sup> *E.g.*, Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL'Y 299, 313 tbl.1 (2013) (uses cross-validation); Chaski, *supra* note 13, at 353 tbl.3 (uses cross-validation); Coulthard, *supra* note 4 (does not use cross validation); Grant, *supra* note 5 (does not use cross-validation).



standardization, both across cases and ultimately across experts. Standardization of the feature set used in forensic linguistic analysis is imperative if we are to have established error rates. It is also the only way to avoid confirmation bias. Without a predefined algorithm, an expert runs the significant risk of preferencing aspects that confirm her initial hypothesis over those that disprove it.<sup>31</sup>

Going forward, the challenge for forensic linguists will be to develop a method that relies less on the expertise of the individual linguist—at least on an everyday basis. The heavy lifting in developing an authorship attribution technique should occur in the lab, long before it is applied in a legal case. By the time it is applied for legal consequence, the application of the method should be largely mechanical.

#### CONCLUSION

Ours is an extremely exciting time for forensic linguistics. The field faces profound challenges in its attempt to meet the ideals and goals set by *Daubert*, and much work remains to be done. Yet, with so many motivated and intellectually engaged scholars and researchers, we can be very hopeful that progress will be steadily made.

More broadly, as a legal observer, I am curious to see how the field of forensic linguistics ultimately develops. Unlike most forensic fields, which arose long before the invention of DNA typing and the decision in *Daubert*, forensic linguistics will blossom within a modern scientific evidence framework. It will thus provide a unique opportunity to observe how the various actors and modern incentives interact. More importantly, it will help evidence scholars determine whether all the trouble collectively known as *Daubert* is really worth the candle.

---

<sup>31</sup> In this context, I am reminded of the *modus operandi* arguments made by the prosecution in *United States v. Trenkler*, 61 F.3d 45 (1st Cir. 1995), a case involving the purported “signature” of a bomber. The prosecution pointed to several common bomb parts in its argument that two bombs were constructed by the defendant. The dissent rightly wondered why one should emphasize the similarities between the two bombs rather than several significant dissimilarities. *Id.* at 64 (Torruella, J., dissenting).

# INTUITION VERSUS ALGORITHM: THE CASE OF FORENSIC AUTHORSHIP ATTRIBUTION

*Lawrence M. Solan\**

## I. INTRODUCTION

On November 6, 2012, Barack Obama was reelected President of the United States, having defeated his Republican opponent, Mitt Romney. The vote in the Electoral College—the official body that votes on a state-by-state basis—was decisive: 332–206.<sup>1</sup> Obama also won the national popular vote by a margin of about 4,850,000 votes (50.9% to 47.1%).<sup>2</sup> But Obama’s winning by a comfortable margin is not what many of the pundits on television were predicting. Some announced that Romney would win the election, including predictions that he

---

\* Don Forchelli Professor of Law, and Director, Center for the Study of Law, Language and Cognition, Brooklyn Law School. My thanks to Silvia Dahmen, Lorna Fadden, Roger Shuy and Ben Zimmer for valuable comments on earlier drafts, and to the participants in the Authorship Attribution Workshop held at Brooklyn Law School in October 2012, sponsored by the National Science Foundation. This article is a substantially expanded and revised version of my commentary, *Ethics and Method in Forensic Linguistics*, PROC. INT’L ASS’N FORENSIC LINGUISTS’ TENTH BIENNIAL CONF., 2012, at 362, available at <http://www.forensiclinguistics.net/iafl-10-proceedings.pdf>. Research on this article was sponsored by a Dean’s Summer Research Stipend from Brooklyn Law School.

<sup>1</sup> *2012 Electoral College Results*, NARA, <http://www.archives.gov/federal-register/electoral-college/2012/election-results.html> (last visited Apr. 6, 2013).

<sup>2</sup> *2012 Presidential Election*, NARA, <http://www.archives.gov/federal-register/electoral-college/2012/popular-vote.html> (last visited Apr. 6, 2013).

would win by a landslide.<sup>3</sup> Most guessed that the election would be much closer than it turned out to be.

If the television pundits were all over the lot and mostly wrong, those who used sophisticated computational techniques to draw inferences from polls fared much better. An article in the *New York Times* shortly after the election put it this way:

It was not on any ballot, but one of the biggest election contests this week pitted pundits against pollsters. It was a pitched battle between two self-assured rivals: those who relied on an unscientific mixture of experience, anecdotal details and “Spidey sense,” and those who stuck to cold, hard numbers.

When the results were tabulated, it became clear that data had bested divination.<sup>4</sup>

Perhaps most prominent among the pollsters was *New York Times* blogger Nate Silver. As of the morning of the election, his “FiveThirtyEight” blog predicted that Obama would receive 313 electoral votes to Romney’s 225, and that Obama had a 90.9% chance of winning the election.<sup>5</sup> Silver also predicted that Obama would win the popular vote by 2.5 percentage points.<sup>6</sup> He underestimated the margin of victory slightly in each measure, but not by much, and did dramatically better than did the pundits. Silver’s success made him a celebrity of sorts, including an appearance on *The Daily Show with Jon Stewart*.<sup>7</sup>

---

<sup>3</sup> See Benny Johnson, *Romney Landslide: Here Are the Biggest Names Predicting It*, THEBLAZE.COM (Nov. 4, 2012, 3:37 PM), <http://www.theblaze.com/stories/romney-landslide-here-are-the-biggest-names-predicting-it-how-it-will-happen/>. Among such predictors were Dick Morris, Karl Rove, Larry Kudlow, Joe Scarborough, and George Will. For quotes from these pundits, see *id.*

<sup>4</sup> Michael Cooper, *Election Result Proves a Victory for Pollsters and Other Data Devotees*, N.Y. TIMES, Nov. 8, 2012, at P8.

<sup>5</sup> Nate Silver, *FiveThirtyEight: When Internal Polls Misperform, a Whole Campaign May Be to Blame*, N.Y. TIMES (Dec. 1, 2012), <http://fivethirtyeight.blogs.nytimes.com/2012/12/01/when-internal-polls-misperform-a-whole-campaign-may-be-to-blame/>.

<sup>6</sup> *Id.*

<sup>7</sup> *The Daily Show with Jon Stewart* (Comedy Central television broadcast Nov. 7, 2012), available at <http://www.thedailyshow.com/watch/wed->

His book *The Signal and the Noise* is a tribute to the triumph of algorithm over intuition.<sup>8</sup>

It would be a mistake, however, to dismiss intuitive expert judgment generally, just because it fails at predicting the results of presidential elections. Not all expert opinion based upon experience can be reduced to “divination.” Through repeated experience, people develop expertise of all kinds, ranging from chess playing<sup>9</sup> to medical diagnosis.<sup>10</sup> No one accuses the best of such people of using a divining rod simply because they do not rely on computer algorithms. Moreover, we make judgments all the time about what is likely to happen next, including, for example, the judgment that it is safe to cross the street when the traffic signal favors us and the cars are all stopped. Most of the time, there is no computer algorithm with which we can compare our rate of success, but we have a good sense—confirmed by repeated experience—that we are making the right decision.

On the other hand, the use of algorithms seems to neutralize some obvious biasing factors that plague the pundits routinely. Why is it that experts paid by Fox News (a Republican-oriented network) predicted a Romney victory, whereas those paid by MSNBC (a Democrat-oriented network) predicted that Obama would win? One possibility is that most of the pundits are sufficiently corrupt to misstate their actual views if they are paid enough to do so. More likely, though, their prior commitments contribute to what information they regard as significant and color their analyses, which are sincere. This phenomenon, called confirmation bias, is well studied by psychologists. It is an

---

november-7-2012/nate-silver.

<sup>8</sup> See generally NATE SILVER, *THE SIGNAL AND THE NOISE* (2012) (investigating how statisticians distinguish meaningful indicators in ever-increasing amounts of data in order to make accurate predictions).

<sup>9</sup> Fernand Gobet & Neil Charness, *Expertise in Chess*, in *THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE* 523, 532–34 (K. Anders Ericsson et al. eds., 2006).

<sup>10</sup> Geoff Norman et al., *Expertise in Medicine and Surgery*, in *THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE*, *supra* note 9, at 339, 350.

“unwitting selectivity in the acquisition and use of evidence.”<sup>11</sup> The networks engage just those experts whose views are most likely to reinforce the views of their audience. Similarly, why is it that the television networks so routinely predicted a close election? Could it be relevant that these privately-owned media outlets make their living selling audiences to advertisers, and it is in their interest to maintain electoral drama for as long as is feasible? No doubt confirmation bias plays a role here as well.

With the election in mind, let us move to forensic authorship attribution. In his essay on the current state of the field, Professor Ronald Butters reminds us, with insight and candor, that forensic linguists, like practitioners in most areas of forensic science, have done more to advance their field substantively than they have done to advance it ethically.<sup>12</sup> The program he suggests is an ambitious one. Butters complains that forensic authorship attribution lacks not only a set of agreed understandings about methodology but also lacks, and is in need of, standards<sup>13</sup> sufficient to ensure the exclusion of bogus conclusions based on inadequate data.<sup>14</sup> In this regard, Butters places methodology beyond mere practice and elevates it to the realm of the ethical: it is simply wrong for a profession to go about its business without some verification that it is doing a good job. Professor Joseph Sanders raises similar points in an essay on the ethical duties of expert witnesses more generally.<sup>15</sup> What could be more important than making sure that those academics, whose “day jobs” are to seek the truth, do more

---

<sup>11</sup> Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 175 (1998).

<sup>12</sup> Ronald R. Butters, *Retiring President's Closing Address: Ethics, Best Practices, and Standards*, PROC. INT'L ASS'N FORENSIC LINGUISTS' TENTH BIENNIAL CONF., 2012, at 351–52. The essay is the text of the Presidential Address delivered by Professor Butters at the meetings of the International Association of Forensic Linguists, Aston University, Birmingham, U.K., July 2011.

<sup>13</sup> *Id.* at 352–53, 356.

<sup>14</sup> *See id.* at 356.

<sup>15</sup> Joseph Sanders, *Expert Witness Ethics*, 76 FORDHAM L. REV. 1539, 1583 (2007) (calling for codes of ethics for individual fields to guide experts as to their responsibility in taking an appropriate epistemological stand toward their testimony).

good than harm when they enter the courtroom with the express task of presenting analysis that will affect the lives of others in profound ways?

This essay explores some of the issues that Butters raises in the context of forensic authorship attribution analysis and that others have raised for some time in the context of other forensic sciences that rely on trace evidence.<sup>16</sup> My first point is that the conflict of interest inherent in expert forensic testimony—especially by those who make their livings, or at least a significant part of their livings, as consulting experts—can indeed best be remedied by the development of methods that are demonstrably both diagnostic and replicable. For those who rely upon judgments of coauthorship based on their knowledge of linguistic features and upon a sense that a large cluster of differences or commonalities in a particular case cannot be a matter of accident, research into methodology should be a top priority. Proficiency testing may take the place of the development of replicable methods in the short run, but the best direction for the field is to demonstrate that methods work and are not highly dependent on the skill of the practitioner alone.

My second point is that work in computer science and computational linguistics is moving toward answering many of the specific questions that Butters raises about particular standards in the field. Such matters as how much data are needed for valid conclusions to be drawn are commonplace in statistics and modeling, and can easily enough be transported to forensic linguistic application. I end this essay with some brief conjecture about why the field does not appear to have moved ahead quickly with respect to some of these questions and what it might do to adjust its course.

## II. LUCY AND LACY: TWO STYLES OF EXPERT ANALYSIS

Those who engage regularly in expert consultation, and especially in expert testimony, have an inherent conflict of

---

<sup>16</sup> See, e.g., D. Michael Risinger et al., *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1, 27–42 (2002).

interest. It is sometimes referred to as the “hired gun” syndrome, and it stems from the fact that testifying experts are encouraged to render opinions useful to the party that hires them and are subject to confirmation bias in any event. Consider the following vignette about expert witness Lucy:

Lucy is a professor of computational linguistics and currently has a grant-funded project on authorship identification, which she hopes will have practical application eventually. Last month, a lawyer phoned Lucy, saying he had heard of her work, and asked her if she would be willing to apply it to a legal case and possibly testify as an expert. Lucy was intrigued. She took the case, analyzed it according to the methods that she had developed, and concluded, by virtue of applying her algorithm, that the questioned document the lawyer presented was very unlikely to have been written by the person to whose known writings she had compared it. In her lab, Lucy was correct 88% of the time when she conducted this kind of analysis this way. She told the lawyer that she would be happy to testify to all of this, as she continues to work in her lab to improve the 88% rate of accurate rejection of authorship.

Now compare Lucy to Lacy:

Lacy is a forensic linguistic consultant. From time to time she takes authorship attribution cases. Lacy does not conduct her work computationally. Rather, she has a set of thirty-six stylistic markers by which she analyzes all documents that come to her. She has found from past experience that when the documents are long enough for comparison, some of these thirty-six markers will tend either to co-occur between a questioned document and a reference set or be noticeably different between them. There is sometimes controversy about whether her testimony will be permitted, but when she is allowed to testify, her testimony is generally convincing.

At first glance, we might prefer Lucy. After all, we know how good her methods are, making it less likely that she is a hired gun. With Lacy, in contrast, we must rely on her

persuasive rhetoric and the intuitive appeal of the data she presents.

Herein lies the problem: we have no idea which expert does a better job. It may well be that Lacy limits herself to the kinds of problems that she is certain to get right and that her success rate exceeds Lucy's not unimpressive 88%. On the other hand, it may be that Lacy gets a lot of slack from her charisma and the intuitive appeal of her analyses and that her success rate is far lower than Lucy's.

Over the past two decades, forensic linguistics, I believe, has developed as a field with more Lacys than Lucys, and this has led to some of the problems that Butters observes. Many involved in the field—especially authorship attribution specialists who rely on stylistic markers—conduct little or no laboratory work. This is true both of independent consultants and of academics who self-identify as forensic linguists. The result is a dearth of serious research, provoking reasonable questions about the legitimacy of the conclusions reached. As noted below, proficiency testing may be at least a partial solution to this problem, but no such testing currently takes place. At the same time, somewhat disconnectedly, computer scientists and computational linguists have been developing algorithms that more and more successfully predict authorship, but much of this has not yet made its way to forensic application.<sup>17</sup>

This tension was not always so pronounced. The history of “voiceprint” analysis provides quite a different story. During the 1960s, an employee of Bell Labs, which invented the sound

---

<sup>17</sup> For the state of current research, see Shlomo Argamon & Moshe Koppel, *A Systemic Functional Approach to Automated Authorship Analysis*, 21 J.L. & POL'Y 299 (2013); Moshe Koppel et al., *Authorship Attribution: What's Easy and What's Hard?*, 21 J.L. & POL'Y 317 (2013); Efstathios Stamatatos, *On the Robustness of Authorship Attribution Based on Character N-Gram Features*, 21 J.L. & POL'Y 421 (2013). For an overview, see Carole E. Chaski, *Author Identification in the Forensic Setting*, in THE OXFORD HANDBOOK OF LANGUAGE AND LAW 489 (Peter M. Tiersma & Lawrence M. Solan eds., 2012); Patrick Juola, *Authorship Attribution*, 1 FOUND. & TRENDS IN INFO. RETRIEVAL 233 (2008). For general discussion, see Lawrence M. Solan, *The Expert Witness Meets the Adversarial System*, in THE ROUTLEDGE HANDBOOK OF FORENSIC LINGUISTICS 395 (Malcolm Coulthard & Alison Johnson eds., 2010).



spectrograph, began to make extravagant claims about the ability of such devices to distinguish one voice from another, much the way fingerprints were (and still largely are) seen as distinguishable from one person to the next.<sup>18</sup> Police laboratories received training in the use of the new technology, about which there was considerable excitement.<sup>19</sup> Then, in 1979, the National Research Council issued a devastating report, pointing out that there had not been adequate testing to determine how well spectrography can be used to distinguish one voice from the other in forensic settings.<sup>20</sup> The report noted:

The degree of accuracy, and the corresponding error rates, of aural-visual voice identification vary widely from case to case, depending upon several conditions including the properties of the voices involved, the conditions under which the voice samples were made, the characteristics of the equipment used, the skill of the examiner making the judgments, and the examiner's knowledge about the case. Estimates of error rates now available pertain to only a few of the many combinations of conditions in real-life situations. These estimates do not constitute a generally adequate basis for a judicial or legislative body to use in making judgments concerning the reliability and acceptability of aural-visual voice identification in forensic applications.<sup>21</sup>

The leaders in the effort to make sure that linguistic science, if used in law enforcement efforts, would meet the high standards of science itself were chiefly academic linguists, with

---

<sup>18</sup> See Lawrence G. Kersta, *Voiceprint Identification*, 196 NATURE 1253, 1257 (1962). For discussion of this history, see LAWRENCE M. SOLAN & PETER M. TIERSMA, SPEAKING OF CRIME: THE LANGUAGE OF CRIMINAL JUSTICE 140, 140–46 (2005) and Lawrence Solan & Peter Tiersma, *Hearing Voices: Speaker Identification in Court*, 54 HASTINGS L.J. 373, 416–26 (2003).

<sup>19</sup> See SOLAN & TIERSMA, *supra* note 18, at 140.

<sup>20</sup> NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION 58 (1979).

<sup>21</sup> *Id.* at 60.

special efforts by the distinguished phonetician, Peter Ladefoged.<sup>22</sup>

Fast-forward to 2009, when the National Research Council came out with another devastating report, this time concerning forensic identification science in the United States more generally. The report decried the absence of scientific integrity in forensic identification procedures, much as the earlier report did with respect to speaker identification technology used at the time:

A body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias. Such research is sorely needed, but it seems to be lacking in most of the forensic disciplines that rely on subjective assessments of matching characteristics. These disciplines need to develop rigorous protocols to guide these subjective interpretations and pursue equally rigorous research and evaluation programs.<sup>23</sup>

Yet now, it is not the linguistic academic community taking the lead in remedying this situation on behalf of linguists who do not want to see the legal system making excessive claims about the forensic application of the language sciences. Rather, it is chiefly legal academics with expertise in the area of scientific evidence taking the lead, with the focus not on linguistics in particular but on the forensic identification sciences generally.<sup>24</sup> What has happened between 1979 and 2009? In 1979, there was no field of forensic linguistics, or at least not much of one. Linguists were occasionally called to testify as experts in court, but they did so because their academic expertise serendipitously crossed paths with a legal issue, much the way an academic physicist or engineer might be called upon to provide expert testimony. Linguists were in the business of being linguists, and

---

<sup>22</sup> See SOLAN & TIERSMA, *supra* note 18, at 140–41.

<sup>23</sup> NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 8 (2009).

<sup>24</sup> See, e.g., Jonathan J. Koehler, *If the Shoe Fits They Might Acquit: The Value of Forensic Science Testimony*, 8 J. EMPIRICAL LEGAL STUD. 21 (2011); Risinger et al., *supra* note 16; Sanders, *supra* note 15.

the legal system was as much a novelty to them as they were a novelty to the legal system. As late as 1994, Judith Levi, one of the founders of the field of forensic linguistics, wrote the inaugural article in the then-brand-new journal *Forensic Linguistics* (now, *The International Journal of Speech, Language and the Law*), explaining to the linguistics community at large the growing acceptance of linguistic experts in court.<sup>25</sup>

Since that time, things have developed considerably. With the growth of undergraduate and graduate programs in forensic linguistics,<sup>26</sup> many academics devote much of their time to applying linguistic knowledge in legal settings, as do consulting linguists without academic affiliations. The gap between the academic community that once policed its field for abuse by the legal system and practitioners within the legal system has shrunk considerably.

Some in the language and law community have recognized this gap between theory and practice to be a healthy one and have attempted to maintain it in their description of how linguists engaged in courtroom testimony should view their work. Peter Tiersma, a law professor and linguist (with whom I frequently write), suggests that the field is at its best when the reluctant scholar is asked to share her expertise with the court for the benefit of reaching a proper result in a dispute in which the expert has no interest, either intellectual or otherwise.<sup>27</sup> Roger Shuy, a distinguished scholar in applied linguistics who has been a prolific writer in the field, states the forensic linguist's ideal role more fully:

[*Forensic linguistics*] has become a useful way to refer to the use of linguistics knowledge where there are data that

---

<sup>25</sup> See Judith N. Levi, *Language as Evidence: The Linguist as Expert Witness in North American Courts*, 1 FORENSIC LINGUISTICS 1 (1994). For subsequent history, see Peter M. Tiersma & Lawrence M. Solan, *The Linguist on the Witness Stand: Forensic Linguistics in American Courts*, 78 LANGUAGE 221 (2002).

<sup>26</sup> Such programs exist at Cardiff University, Aston University (Birmingham, UK), Universitat Pompeu Fabra (Barcelona), and Hofstra University.

<sup>27</sup> Peter M. Tiersma, *Linguistic Issues in the Law*, 69 LANGUAGE 113, 122 (1993).

serve as evidence. But I have some concerns about the term itself, because it seems that when one does “forensic linguistics” one is simply doing linguistics, a type of applied linguistics, in fact.<sup>28</sup>

These views are attractive. To the extent that they are descriptively accurate views of the field, they explain how it was that the academic phoneticians were the ones who shut down the voiceprint craze of the 1960s and ‘70s.

Now, however, expert testimony on questions of language goes beyond ordinary research in linguistics into areas developed by those interested in forensic linguistics as its own discipline.<sup>29</sup> Individuals, sometimes without a great deal of training in linguistics, generally become skilled in areas of language analysis developed particularly for consumption by the legal community.<sup>30</sup> For reasons stated earlier, many practitioners have little motivation to police their own methodologies along the conventional scientific criteria of validity and reliability and typically do not engage in analysis of methods used by others, apart from disagreement within a particular case in which they have taken opposing positions. The result of this development is that it is not always possible to judge forensic testimony against ordinary practices among linguists, because linguists do not ordinarily engage in the activities that generate the expert testimony.

Does this amount to an ethical issue? It probably does if neither methodological testing nor proven individual proficiency forms a sufficient scientific basis to accept some of what passes for linguistic expertise. Moreover, to the extent that this lack of foundation results from the dearth of research that is itself a product of this conflict, then it is the fruit of a conflict of interest and is an ethical issue for this reason. Law professor and philosopher Susan Haack puts it this way:

---

<sup>28</sup> ROGER W. SHUY, *LINGUISTICS IN THE COURTROOM: A PRACTICAL GUIDE* 3 (2006).

<sup>29</sup> See Ronald R. Butters, *The Forensic Linguist's Professional Credentials*, 16 *INT'L J. SPEECH LANGUAGE & L.* 237 (2009).

<sup>30</sup> *Contra* SHUY, *supra* note 28, at 3.

Distinguishing genuine inquiry, the real thing, from pseudo-inquiry or “sham reasoning,” C.S. Peirce—a working scientist as well as the greatest of American philosophers—wrote that “the spirit . . . is the most essential thing—the motive”; that genuine inquiry consists in “actually drawing the bow upon truth with intentness in the eye, with energy in the arm.” For the same reason, I am tempted to write of advocacy “research” (in scare quotes); for it is something of a stretch to call advocacy research “research” at all. Advocacy “research” is like inquiry insofar as it involves seeking out evidence. But it is part of an advocacy project insofar as it involves seeking out evidence favoring a predetermined conclusion; and it is undertaken in the spirit, from the motive, of an advocate. In short, it is a kind of pseudo-inquiry.<sup>31</sup>

At the same time, as noted previously, there is no reason to conclude that intuitive expertise based on experience and insight fares any better or worse than does algorithmic expertise. Intuitive expertise is not necessarily unreliable. On the contrary, it is clear that at least in some settings, people are able to form sophisticated mental models of situations about which they are experts and to weigh relevant factors with great accuracy, notwithstanding that they are unable to describe how they did it. For example, Malcolm Gladwell, in his book *Blink*, made famous the story of the J. Paul Getty Museum’s acquisition of a 2,000-year-old Greek sculpture—a kouros—which is a rare thing to acquire.<sup>32</sup> The museum did its due diligence carefully, investigating the sculpture’s provenance over the centuries, engaging experts to examine the marble with microscopes, and so on.<sup>33</sup> But the day of reckoning came when the museum’s curator began inviting various experts in classical sculpture, none of whom felt that the sculpture was authentic, and one of whom remarked that seeing it caused in him a wave of “intuitive

---

<sup>31</sup> Susan Haack, *What’s Wrong with Litigation-Driven Science? An Essay in Legal Epistemology*, 38 SETON HALL L. REV. 1053, 1071 (2008).

<sup>32</sup> MALCOLM GLADWELL, *BLINK* 3 (2005).

<sup>33</sup> *Id.* at 3–4.

repulsion.”<sup>34</sup> For the most part, the experts could not put into words exactly what was bothering them, but deep within themselves, they knew.<sup>35</sup> Later, all agreed that the kouros was a kouros copy.<sup>36</sup>

The kouros story is particularly relevant here because it is, in a sense, a matter of authorship attribution: who sculpted the kouros that the museum had bought? Thus, we may wish to ask about the nature of the knowledge that the experts brought with them to the task that led to their negative reactions. As Daniel Kahneman points out,<sup>37</sup> celebrations of gut-reaction decision making, such as Malcolm Gladwell’s *Blink*, also contain stories of terrible diagnostic failure, including the misdiagnosis of heart attacks.<sup>38</sup> And we’ve already seen how well political experts do at predicting election results. Many in the business of predicting the future of the economy are on a par with the political pundits. It thus appears that intuitive expertise is neither all good nor all bad as a diagnostic tool.<sup>39</sup>

Psychologists have devoted a great deal of study to the question of expert intuition, in areas as diverse as the thinking of chess masters, medical diagnosis and treatment by physicians, neonatal intensive care nursing, and decision making about firefighting.<sup>40</sup> What appears to separate the intuitive experts from the rest of us is a deep base of knowledge that has enabled them to build mental models<sup>41</sup> that are so robust that they can be both accessed and expanded with ease. Chess masters do far better

---

<sup>34</sup> *Id.* at 5–6.

<sup>35</sup> *Id.*

<sup>36</sup> *Id.* at 7.

<sup>37</sup> DANIEL KAHNEMAN, THINKING, FAST AND SLOW 235–36 (2011).

<sup>38</sup> GLADWELL, *supra* note 32, at 130.

<sup>39</sup> For a discussion of how scientific analysis and subjective analysis blend in diagnosis in many domains, see John A. Swets et al., *Psychological Science Can Improve Diagnostic Decisions*, 1 PSYCHOL. SCI. IN PUB. INTEREST 1 (2000).

<sup>40</sup> For discussion of the circumstances in which intuitive expertise is most likely to prevail, see Daniel Kahneman & Gary Klein, *Conditions of Intuitive Expertise: A Failure to Disagree*, 64 AM. PSYCHOLOGIST 515, 522 (2009).

<sup>41</sup> For an overview of what constitutes a mental model, see P.N. JOHNSON-LAIRD, MENTAL MODELS 10–12 (1983).

than do novices in recalling the pieces in familiar chess configurations but generally do not do significantly better in recalling the pieces placed at random on a chess board.<sup>42</sup> Moreover, the more such configurations are familiar to them—because they devote their lives to studying these configurations—the broader the range of improved recall.<sup>43</sup> By the same token, experienced doctors, when asked the basis of a diagnosis, do not resort to basic science, even though they were taught and know the mechanisms underlying disease. In a hard case they may resort to their basic training, but as a general matter, they have developed, based on huge numbers of observations, a wealth of knowledge of what combinations of symptoms and conditions are indicative of disease and can do this with great accuracy, even when they are not able to articulate how it is that they assigned particular weights to particular symptoms in a particular setting.<sup>44</sup> And the same holds true for the impressive record of neonatal intensive care nurses, fire commanders, and other such experts.<sup>45</sup> Each group of experts develops mental models capable of distinguishing the successful from the unsuccessful based upon repeated exposures to a wealth of scenarios.

Thus, while Lacy relies upon intuitive judgment more than does Lucy, depending upon how conservative Lacy is in her willingness to offer expert opinions, her rate of error may be lower than that of Lucy. Yet, when given a choice, the legal system typically opts for Lucy-like expertise, not because we know that she is more likely to be correct, but because we at least know how likely she is to be right, reducing the probability

---

<sup>42</sup> Gobet & Charness, *supra* note 9, at 526–27.

<sup>43</sup> *Id.* at 526. Nonetheless, the machines have caught up to and passed the chess masters. For an interesting account, see SILVER, *supra* note 8, at 262–93 (describing Garry Kasparov's games with computer chess programs during the 1980s and 90s and his eventual loss to Deep Blue in 1997).

<sup>44</sup> Norman et al., *supra* note 10, at 346.

<sup>45</sup> For a description of some of this research and a theory of what makes intuitive expertise successful in these circumstances, see Karol G. Ross et al., *Professional Judgments and "Naturalistic Decision Making,"* in THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE, *supra* note 9, at 403, 403–15.

that cognitive biases or a witness's compelling personality will play too great a role in the outcome of a case.<sup>46</sup> Making the case for algorithmic expertise more compelling, people are much better at recognizing the biases of others than they are at recognizing their own biases.<sup>47</sup> Thus, encouraging experts to recognize and stave off the temptation of becoming too much a team player is not likely to be an adequate solution to the problem of bias.

The literature on the nature of intuitive expertise raises another concern with respect to authorship attribution. Expert opinion testimony is admissible only if the expert's scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue.<sup>48</sup> It is not clear how much of the expert opinion of intuitive experts on authorship attribution is a matter of expertise. To see why this is the case, let us look at a Lacy-like analysis taken from an article by Australian linguist Robert Eagleson.<sup>49</sup> Eagleson describes a case in which a woman supposedly left a farewell letter to her husband, typed on the family typewriter, when she ran off with another man.<sup>50</sup> The police believed, however, that the husband had written the letter, and had done away with the wife.<sup>51</sup> A linguist was called in to compare the style of the farewell letter with the style of writing of documents known to be written by the husband and documents known to be written by the wife.<sup>52</sup>

---

<sup>46</sup> See Joseph Sanders, *Kumho and How We Know*, 64 LAW & CONTEMP. PROBS. 373, 374–75, 393 (2001). The propensity to overstate the role of character and to understate the circumstances in which an individual acts is called “the correspondence bias” in the psychological literature. See Daniel T. Gilbert & Patrick S. Malone, *The Correspondence Bias*, 117 PSYCHOL. BULL. 21 (1995); see also SOLAN & TIERSMA, *supra* note 18, at 29–32 (discussing evidentiary standards in the *Daubert* age).

<sup>47</sup> Emily Pronin et al., *The Bias Blind Spot: Perception of Bias in Self and Others*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 369, 369–81 (2002).

<sup>48</sup> FED. R. EVID. 702.

<sup>49</sup> Robert Eagleson, *Forensic Analysis of Personal Written Texts: A Case Study*, in LANGUAGE AND THE LAW 362 (John Gibbons ed., 1994).

<sup>50</sup> *Id.*

<sup>51</sup> *Id.*

<sup>52</sup> *Id.*



The results were dramatic. The husband's known writings and the farewell letter shared a number of nonstandard spellings (individual words, capitalization of common nouns, lower case proper nouns, use of apostrophe), syntactic structures, word structures, and punctuation nuances.<sup>53</sup> The wife's writings had none of these features.<sup>54</sup> To take one example, both the husband's writings and the farewell letter used the present tense when it would have been appropriate to use the past tense ("He threaten me.").<sup>55</sup> After the results of this analysis were presented to the husband, he confessed.<sup>56</sup>

But there is a problem here. The grouping of similarities and differences indeed requires some sophistication in the analysis of language. It is unlikely that someone not trained in linguistics would have come up with this array. Once the linguistic expert did so, however, there was no particular science behind the inference that the husband was more likely than the wife to have written the farewell letter. It only makes sense given the array of similarities with the husband's style and differences from the wife's, but it makes sense because of what our common sense notions tell us about how likely people are to be consistent about such aspects of their writing. Missing is the kind of experience that the doctors, neonatal nurses, chess players, and others describe in which the similar patterns are presented to them over and over again with the results known quickly. Other Lacy-like examples show the same characteristics—a substantial, and often intuitively convincing, number of similarities between a questioned document and the writings of a suspect, with no serious science underlying the inference of authorship identification.<sup>57</sup>

This leaves the legal system with three choices: it can accept the expert testimony, opinion and all; it can reject the expert testimony on similarities and differences entirely; or it can admit

---

<sup>53</sup> *Id.* at 371–72.

<sup>54</sup> *Id.*

<sup>55</sup> *Id.* at 367–68.

<sup>56</sup> *Id.* at 373.

<sup>57</sup> See, for example, the discussion of the Unabomber case in SOLAN & TIERSMA, *supra* note 18, at 159–64.

the testimony, but permit the jury to draw the inferences of identity for which there is not an adequate scientific basis. Each of these options has its own difficulties.

The problem with the first option is that there is no basis for crediting the opinion of the expert. In the next section, I recommend proficiency testing to legitimize expert testimony in these circumstances.

The second option, while seemingly attractive, especially to those who favor an algorithmic approach, also comes with a price. It takes from the jury the ability to evaluate evidence with more sophistication. Take, for example, a point raised by Malcolm Coulthard<sup>58</sup>: the government offers evidence in a fraud case that a letter whose authorship is in question contains the same kinds of linguistic nuances (spelling errors, punctuation, and so on) as do the known writings of the defendant.<sup>59</sup> No expert is needed.<sup>60</sup> Now, the defense offers an expert to demonstrate that these same features are present in the writings of other possible authors.<sup>61</sup> Only the expert conducted the evaluation of the documents, so without his testimony, they will not be before the jury.<sup>62</sup> It would seem that justice is better served if the expert is permitted to point out similarities with other potential authors, regardless of there being no algorithm that can demonstrate a likelihood of actual authorship. After all, that is exactly what the prosecution has done without an expert in its efforts to implicate the defendant.

The third option—permitting the expert to point out similarities and differences, but not to issue an opinion on attribution—also has its problems. The approach has initial appeal. On the one hand, the comparison seems to be relevant to the outcome. On the other hand, we have no way of knowing how good the expert's lay intuitions are likely to be. No studies have been conducted to tell us. They may be right most of the

---

<sup>58</sup> Malcolm Coulthard, *On Admissible Linguistic Evidence*, 21 J.L. & POL'Y 441 (2013).

<sup>59</sup> *Id.*

<sup>60</sup> *Id.*

<sup>61</sup> *Id.*

<sup>62</sup> *Id.*

time when so many features are either shared or differentiated, but we simply do not know. In a different context, I recommended that courts accept a “tour guide” approach to expert linguists testifying about meaning.<sup>63</sup> When each side appears to have proposed a reasonable interpretation of legally relevant language, a linguist may point out the various plausible interpretations and explain how they derive from ordinary linguistic processes.<sup>64</sup> This kind of testimony poses little danger, since the judge and jury are perfectly capable, based on their intuitions as speakers of English, to determine whether the linguist’s testimony accurately reflects their own judgments about the range of possible meanings.<sup>65</sup> Moreover, once the range of plausible interpretations is brought out, the linguist’s expert opinion about meaning is largely superfluous, since the expert will have put the jury on an equal footing with him by virtue of the testimony.<sup>66</sup>

Authorship attribution is different, however. The goal of the expert is not to make jurors sensitive to the full range of their intuitions about authorship but rather to determine who wrote the questioned document. We do not know, however, to what extent the expert testimony on similarities and differences is helpful and how much it leads jurors to intuitive judgment without adequate basis to determine whether the similarities and differences that appear so telling have any real predictive force.

Thus, as Edward Cheng points out, “[t]he heavy-lifting in developing an authorship attribution technique should occur in the lab, long before it is applied in a legal case.”<sup>67</sup> While that is happening, however, courts are faced with the uncomfortable dilemmas described above. Let us now turn to how the American legal system has reacted to these issues and how the field might develop to increase its efficacy in court.

---

<sup>63</sup> Lawrence M. Solan, *Linguistic Experts as Semantic Tour Guides*, 5 FORENSIC LINGUISTICS 87 (1998).

<sup>64</sup> *Id.* at 94–95.

<sup>65</sup> *Id.* at 95.

<sup>66</sup> *Id.* at 92.

<sup>67</sup> Edward K. Cheng, *Being Pragmatic About Forensic Linguistics*, 21 J.L. & POL’Y 541, 550 (2013).

### III. JUDICIAL REACTIONS TO AUTHORSHIP ATTRIBUTION EVIDENCE

This section comments on what the American courts have been saying about authorship identification expertise in the United States. However, as Peter Tiersma and I have pointed out,<sup>68</sup> relying upon published opinions to draw conclusions about evidentiary decisions by American courts is not likely to present a fair sense of what actually happens in trial courts. The case law provides a very deferential approach to appeals of decisions on the admissibility of expert opinion evidence,<sup>69</sup> and most (but not all) of the appeals will occur in the context of the losing party having been denied in their application to have an expert testify. The result is that most evidentiary decisions in published opinions by appellate courts are affirmances of the decision of the trial court to exclude an expert. An academic, or for that matter, a lawyer or judge, who relies on these opinions will not have any idea in how many cases experts have indeed testified at trial. Yet such testimony will occur when both sides call experts on the same issue, when one side calls an expert without objection from the other side, or when the offer of an expert survives a motion to exclude, but the case does not result in a published opinion, at least not on that issue.

Prominent examples have appeared in the press in 2011 and 2012. For example, Robert Leonard and Ronald Butters each testified in the Chicago murder case, *People v. Coleman*.<sup>70</sup> Coleman was accused of killing his wife.<sup>71</sup> Part of the prosecution's story was that he had written various threat letters, one of them painted as graffiti on a wall.<sup>72</sup> Leonard, testifying as part of the prosecution's case, was permitted to opine that the

---

<sup>68</sup> Solan & Tiersma, *supra* note 18.

<sup>69</sup> Gen. Elec. Co. v. Joiner, 522 U.S. 136, 139 (1997).

<sup>70</sup> People v. Coleman, No. 09-CF-50 (Ill. Cir. Ct. Monroe Cnty. May 5, 2011).

<sup>71</sup> Nicholas Pistor, *Jurors Shown Images of Coleman's Dead Family as Trial Starts*, ST. LOUIS POST-DISPATCH (Apr. 25, 2011, 2:15 PM), [http://www.stltoday.com/news/local/Illinois/jurors-shown-images-of-coleman-s-dead-family-as-trial-article\\_c293b985-a54c-5be1-8540-5c4b7a98432d.html](http://www.stltoday.com/news/local/Illinois/jurors-shown-images-of-coleman-s-dead-family-as-trial-article_c293b985-a54c-5be1-8540-5c4b7a98432d.html).

<sup>72</sup> *Id.*

stylistic features of the threat letters had enough in common to suggest that all were written by the same individual and that the letters bore enough similarity to the known writings of Coleman that it was a reasonable—but not proven—hypothesis that Coleman wrote them.<sup>73</sup> Butters argued that there was not enough evidence to permit one to draw legitimate inferences.<sup>74</sup> Coleman was convicted.<sup>75</sup> As an aside, the circumstantial evidence was strong: Coleman had bought the paint that was used to write the threat on the wall, and the other letters were found on his computer.<sup>76</sup> Thus, the linguistic debate was offered only to dispel the possibility that someone other than Coleman had used his computer. Nonetheless, the Coleman case demonstrates an instance in which courts permitted forensic stylistic analysis after ruling it admissible in an evidentiary hearing.

Also in the press were stories about an expert declaration by Gerald McMenemy, a linguist who specializes in forensic stylistic analysis. McMenemy testified in a case brought by Paul Ceglia against Mark Zuckerberg, claiming that Zuckerberg did not own the Facebook idea.<sup>77</sup> McMenemy opined that certain emails allegedly written by Zuckerberg were in fact not written by Zuckerberg. The methods he used have drawn criticism from Professor Butters<sup>78</sup> and some controversy in the press.<sup>79</sup> Nonetheless, in an opinion issued but not published as of the

---

<sup>73</sup> Report of Proceedings, May 3, 2011, *Coleman*, No. 09-CF-50 (testimony of Robert A. Leonard).

<sup>74</sup> Report of Proceedings, Apr. 4, 2011, *Coleman*, No. 09-CF-50 (testimony of Ronald Butters, opposing the admission of Leonard as an expert witness).

<sup>75</sup> Jim Suhr, *Christopher Coleman, Former Marine, Convicted of Killing Wife, 2 Sons*, HUFFINGTON POST (May 6, 2011, 11:15 AM), [http://www.huffingtonpost.com/2011/05/06/christopher-coleman-former\\_n\\_858460.html](http://www.huffingtonpost.com/2011/05/06/christopher-coleman-former_n_858460.html).

<sup>76</sup> *Id.*

<sup>77</sup> Decl. of Gerald R. McMenemy in Support of Motion for Expedited Discovery, *Ceglia v. Zuckerberg*, No. 1:2010-cv-00569 (W.D.N.Y. July 9, 2010), available at <http://docs.justia.com/cases/federal/district-courts/new-york/nywdce/1:2010cv00569/79861/50/0.pdf?1307100528>.

<sup>78</sup> Butters, *supra* note 12, at 354–56.

<sup>79</sup> See Ben Zimmer, *Decoding Your E-Mail Personality*, N.Y. TIMES, July 24, 2011, at SR12.

time of this writing, a federal Magistrate Judge has credited McMenamín's position<sup>80</sup> and recommended that the district court dismiss the case against Zuckerberg and Facebook.<sup>81</sup>

McMenamin has testified in authorship cases on many occasions, and in fact has written a book that describes his stylistic approach.<sup>82</sup> But, unless and until the Magistrate Judge's opinion (or another opinion discussing McMenamín's contribution) in the Facebook litigation is published, he shows up in the federal and state reporters only indirectly, through a former FBI agent and forensic linguist named James Fitzgerald. In a 2010 federal case decided by the District Court in Utah,<sup>83</sup> and a case decided by the federal court in New Jersey in 2000,<sup>84</sup> Fitzgerald testified in a *Daubert* hearing that he was employing McMenamín's methods in an authorship attribution case.<sup>85</sup> In both cases the court permitted him to testify as to similarities and differences between the defendant's known writings and those that were in question in the case but not to offer an opinion as to authorship because the methods did not meet the scientific standard required by the Federal Rules of Evidence.<sup>86</sup>

In contrast, an appellate opinion of the New Jersey state courts affirmed a conviction in a case in which Fitzgerald testified for the prosecution that the defendant had written certain anonymous documents, and Carole Chaski testified that there were significant differences between the anonymous documents and those known to be written by the defendant.<sup>87</sup> The defendant was actually acquitted of having created the

---

<sup>80</sup> Ceglia v. Zuckerberg, No. 1:2010-cv-00569, at 118–19 (W.D.N.Y. July 9, 2010), available at <http://assets.sbnation.com/assets/2383813/ceglia-v-zuckerberg-3-26-13.pdf>.

<sup>81</sup> *Id.* at 151.

<sup>82</sup> GERALD R. MCMENAMIN, *FORENSIC STYLISTICS* (1993).

<sup>83</sup> United States v. Zajac, 748 F. Supp. 2d 1340, 1343, 1351 (D. Utah 2010).

<sup>84</sup> United States v. Van Wyk, 83 F. Supp. 2d 515, 521 (D.N.J. 2000).

<sup>85</sup> *Id.*

<sup>86</sup> *Id.* at 523.

<sup>87</sup> State v. McGuire, 16 A.3d 411, 430 (N.J. Super. Ct. App. Div. 2011).

anonymous documents.<sup>88</sup> The matter of fact discussion by the appellate court, however, suggests no problem with trial judges admitting expert testimony of forensic linguists, including opinion as to authorship, whether they conduct their work through stylistic comparison or by virtue of algorithms that they have developed. When one party decides to deal with the other side's expert by hiring his own, there will typically be no rejection of either expert. Judges are not likely to exclude a witness absent an objection from the opposing party. Moreover, unless the case results in a published decision, there will be no publicly salient record of the entire episode. In fact, the court in this case mentioned the forensic linguistic testimony only as an aside, since the defendant had been acquitted on the count for which the testimony was offered.

My goal in this discussion is not to criticize the linguists whose methods were at issue in these cases. On the contrary, much of this essay is devoted to suggesting that stylistic analysis is not provably less reliable than the quantitative methods. My hope is that through communication among those who approach the field from different perspectives, it becomes possible to make such methods crisp enough to withstand scrutiny or at least to integrate their most acute insights into quantitative models.

#### IV. CURRENT TRENDS IN FORENSIC LINGUISTIC AUTHORSHIP ATTRIBUTION

The field appears to be developing to bring a healthy balance between theory and practice in forensic linguistic identification. The basic problem that the field must address is this: as we learned from Noam Chomsky more than a half century ago, language is a creative cognitive function.<sup>89</sup> By that, I do not mean to say that we can all be poets if we wish. What "creative" in this context means is that we can produce and understand infinitely many utterances because the rules of a recursive grammar that we have internalized in our minds, mostly as young children, combined with a rich vocabulary, give

---

<sup>88</sup> *Id.* at 420.

<sup>89</sup> NOAM CHOMSKY, LANGUAGE AND MIND 88-91 (3d ed. 2006).

us the capacity to do this. And we indeed use the capacity. As Malcolm Coulthard has pointed out, based on the work of some computational linguists,<sup>90</sup> if you highlight any ten-word string from any document, and then paste that string into a search engine (e.g., Google) or database (e.g., Lexis) window, you will find that your string is unique among the vast array of documents available.<sup>91</sup> This shows enormous intra-authorial variation in the use of language. We really do not say things the same way twice. What this all means is that the challenge facing authorship attribution research is to discover ways in which the variation in the use of language between authors is demonstrably greater than the variation within a particular author and to focus on the presence or absence of these distinguishing features.

As noted, there seems to be a methodological, and for that matter, cultural division between computational and stylistic analysts, with the former approaching authorship attribution problems in a manner more consistent with conventional views of expert testimony than the latter but with no evidence that either approach does a better job. Three things are likely to bring about a convergence. First, I expect that if the computational approaches reach some level of maturity and consensus, they will be far more often accepted by the courts than those approaches based largely on intuition—even excellent intuition based on experience and learning. Computational linguistics is an established subfield of linguistic inquiry. Thus, it really is possible to apply well-known linguistic techniques to a forensic problem and to ask whether the application in the forensic setting meets the standards of the field itself.<sup>92</sup> By the

---

<sup>90</sup> Fintan Culwin & Mike Child, *Optimising and Automating the Choice of Search Strings when Investigating Possible Plagiarism*, PROC. 4TH INT'L PLAGIARISM CONF., 2010. This article demonstrates that the procedure generally works with as few as six consecutive words. In earlier work, Coulthard had proposed ten words.

<sup>91</sup> Krzysztof Kredens & Malcolm Coulthard, *Corpus Linguistics in Authorship Identification*, in THE OXFORD HANDBOOK OF LANGUAGE AND LAW, *supra* note 17, at 504, 513–14. Of course, the string may also be a direct quotation of a unique expression. No doubt the text of the U.S. Constitution appears on the web in many sites.

<sup>92</sup> See Argamon & Koppel, *supra* note 17; Carole E. Chaski, *Best*



same token, computational linguists and computer scientists are accustomed to testing their algorithms to see how well they work and reporting the rate of error. One conference, for example, requires the submission of an answer to an authorship attribution task as a criterion for participation.<sup>93</sup> These procedures are consistent with contemporary views of acceptable scientific evidence.<sup>94</sup> If judges, at least in the United States, begin to accept authorship identification as a routine matter precisely because it is transparently algorithmic, with identifiable rates of error, then such work will become the standard.

Second, it will be incumbent upon those whose work is more intuitively stylistic to demonstrate its scientific underpinnings. This can be accomplished by incorporating stylistic features into the computational algorithms being developed by computational linguists and computer scientists. The insightful observations of stylistic analysts that take advantage of such nuances as word choice, punctuation, and spelling errors can be used to expand the range of factors that computer scientists include in their models, with the potential of adding power, even if only incrementally. In fact, this is already occurring.<sup>95</sup> Progress can also be made through the creation of models that demonstrate the efficacy of stylistic analysis as its own approach. Tim Grant, in his article in this volume, has taken a significant step in that

---

*Practices and Admissibility of Forensic Author Identification*, 21 J.L. & POL'Y 333 (2013); Patrick Juola, *Stylometry and Immigration: A Case Study*, 21 J.L. & POL'Y 287 (2013); Koppel et al., *supra* note 17; Stamatatos, *supra* note 17; Carl Vogel, *Attribution of Mutual Understanding*, 21 J.L. & POL'Y 377 (2013); *see also* Chaski, *supra* note 17, at 489 (summarizing the state of authorship attribution); Juola, *supra* note 17 (summarizing the state of authorship attribution); Antonio Rico-Sulayes, *Statistical Authorship Attribution of Mexican Drug Trafficking Online Forum Posts*, 18 J. SPEECH LANGUAGE & L. 53 (2011).

<sup>93</sup> The PAN Lab, held in conjunction with the CLEF conference, is referred to in the field as PAN/CLEF. *Plagiarism Detection*, PAN, <http://pan.webis.de/> (last visited Feb. 1, 2013).

<sup>94</sup> *See* SOLAN & TIERSMA, *supra* note 18, at 29–32.

<sup>95</sup> *See generally* Argamon & Koppel, *supra* note 17 (discussing the importance of language style in authorship attribution); Juola, *supra* note 92 (discussing statistical linguistics to analyze one's writing style against an ad hoc collection of distractor authors).

direction.<sup>96</sup> This will no doubt require far more collaboration between the forensic linguistic community and those adept at statistical modeling. Such collaboration has not been adequate to date. Many of the questions that Butters raises (such as how large an effect must be for it to merit evidentiary weight) are answered automatically in computational settings by virtue of the models employed. Similarly, for those engaged in the promising methods of using linguistic corpora as reference sets in authorship attribution,<sup>97</sup> collaboration with modelers will become a necessity.

In the short run, proficiency testing may substitute for the development of methods, although this remains a second-best solution because it does not eliminate the potential for bias in actual case work. Those linguists who demonstrate their ability to identify authorship correctly in a series of problems presented to them would be credited in court for their proven prowess irrespective of whether their methods are replicable. The PAN/CLEF conference mentioned above,<sup>98</sup> in which computer scientists test their methods against each other as a prerequisite for participation in the event, might be expanded to include those who analyze authorship cases through stylistic comparison. This is an especially appealing solution in that it may well be the case that the stylistic analysts are highly accurate in their conclusions. Personal communication with some of the organizers of this conference suggests that they would be enthusiastic about such participation. It is also important, of course, that those whose work depends on algorithms be willing to subject their models to testing of the sort described above. Self-proclaimed excellence is no more scientific when asserted by computer scientists than when asserted by stylistic analysts.

Third, and related to my second point, some computer scientists and some linguists have taken to looking at very large sets of features, largely stylistic markers, not concerning

---

<sup>96</sup> See Tim Grant, *TXT4N6: Describing and Measuring Consistency and Distinctiveness in the Analysis of SMS Text Messages*, 21 J.L. & POL'Y 467 (2013).

<sup>97</sup> See Kredens & Coulthard, *supra* note 91, at 504–05.

<sup>98</sup> See *supra* note 93.

themselves with which features appear relevant in a particular case, as long as enough do. One such research project uses cluster analysis to sort through a large corpus of emails and to at least reduce the pool of potential authors from a great many to a few with considerable accuracy.<sup>99</sup> Tim Grant's work also makes use of "a basket of features," concerning itself more with the number of features in common than with whether a particular feature or kind of feature is likely to be predictive across cases.<sup>100</sup>

Whether I am right or wrong in predicting the subsequent direction of the field, I firmly believe that far more collaboration among scholars with different areas of expertise is absolutely essential. And I would like nothing more than to see a significant increase in collaboration notwithstanding any conflicts experienced by the players and without a precipitating crisis.

---

<sup>99</sup> Farkhund Iqbal et al., *Mining Writeprints from Anonymous E-mails for Forensic Investigation*, 7 DIGITAL INVESTIGATION 56, 56 (2010).

<sup>100</sup> Grant, *supra* note 96.

# CODIFYING COMMON LAW: THE SELF-CRITICAL ANALYSIS PRIVILEGE AND THE NEW JERSEY PATIENT SAFETY ACT

*Adam Blander\**

## INTRODUCTION

In 2004, New Jersey enacted the Patient Safety Act (“the PSA” or “the Act”),<sup>1</sup> requiring hospitals to engage in the “comprehensive reporting of adverse patient events, systematic analysis of their causes, and creation of solutions.”<sup>2</sup> The Act was grounded in the belief that fostering “a non-punitive culture that focuses on improving processes rather than assigning blame”<sup>3</sup> was crucial in promoting disclosure and reporting. As such, it provided that materials developed from a process of “self-critical analysis” not be discoverable nor used as evidence in any subsequent trial or proceeding.<sup>4</sup>

---

\* J.D. Candidate and Health Law and Policy Fellow, Brooklyn Law School, 2013; B.A., McGill University, 2009. I thank Professor Frederic Bloom for his persistent critical feedback; the entire *Journal* staff for their patience and dedication; my father Stuart Blander for his helpful notes; and Ross Lewin and Drew Britcher for providing me with vital primary sources. Special thanks to my mother, Peri Rosenfeld, for inspiring my curiosity in health policy.

<sup>1</sup> See Patient Safety Act, N.J. STAT. ANN. § 26:2H-12.23–25 (West 2007 & Supp. 2012).

<sup>2</sup> *Patient Safety Reporting System*, STATE OF N.J. DEP’T OF HEALTH, <http://www.state.nj.us/health/ps/> (last visited Feb. 20, 2013).

<sup>3</sup> N.J. STAT. ANN. § 26:2H-12.24(e).

<sup>4</sup> *Id.* § 26:2H-12.25(g) (rendering self-critical patient safety documents immune from discovery and not “admissible as evidence or otherwise disclosed in any civil, criminal or administrative action or proceeding”).

That certain medical investigations, evaluations, and peer review reports<sup>5</sup> should be privileged is not a new evidentiary concept. At least since the early 1970s, both federal and state courts have relied on a self-critical analysis exception to the generally liberal rules of the American discovery system to prevent a litigant from gaining access to his adversary's candid assessments of its internal practices, however relevant they might be to that litigant's case.<sup>6</sup>

In some respects, the PSA merely codified an already existing, judge-made, self-critical analysis privilege, which protected medical peer review documents. In fact, the statute text explicitly adopted the holding of *Christy v. Salem*,<sup>7</sup> an important self-critical analysis case. This Note demonstrates, however, that while prior common law undoubtedly informed the drafters of the PSA, the Act actually created a fairly novel and more expansive self-critical analysis privilege. Quite simply, the values and policy concerns of the emergent "patient safety" movement that inspired the PSA differed from those that encouraged past courts to create and apply the privilege. As a result, these two privileges function quite differently: self-critical analysis under the common law (both in the federal system and in New Jersey) was traditionally a malleable and "qualified" privilege,<sup>8</sup> applied infrequently and on an ad hoc basis by trial judges in an attempt to balance competing public and private interests during discovery. In contrast, the PSA created a more

---

<sup>5</sup> See Susan O. Scheutzow, *State Medical Peer Review: High Cost But No Benefit—Is It Time for a Change?*, 25 AM. J.L. & MED. 7, 7 (1999) (defining peer review as "a process by which members of a hospital's medical staff review the qualifications, medical outcomes and professional conduct of other physician members and medical staff applicants to determine whether the reviewed physicians may practice in the hospital and, if so, to determine the parameters of their practice") (citations omitted).

<sup>6</sup> See, e.g., *Reid v. Lockheed Martin Aeronautics Co.*, 199 F.R.D. 379 (N.D. Ga. 2001); *In re Crazy Eddie Securities Litigation*, 792 F. Supp. 197 (E.D.N.Y. 1992); *Bredice v. Doctors Hosp., Inc.*, 50 F.R.D. 249 (D.D.C. 1970); *McClain v. College Hosp.*, 492 A.2d 991 (N.J. 1985).

<sup>7</sup> *Christy v. Salem*, 841 A.2d 937 (N.J. Super. Ct. App. Div. 2004). This case will be discussed in further detail in Parts II and III.

<sup>8</sup> See, e.g., *Bredice*, 50 F.R.D. at 251 (holding medical peer review reports are "entitled to a qualified privilege").

crystallized, unbending, and absolute privilege, which could likely produce more consistent, albeit perhaps less equitable, results in future litigation against hospitals.

Under a PSA regime, trial judges will have less discretion to shape the course of discovery because the relevant question in deciding whether to apply a privilege is no longer one which balances the equities and considers the discoverer's need for the information. Instead, courts will resolve distinctly statutory inquiries: whether a hospital "substantially complied" with the PSA's reporting scheme<sup>9</sup> or whether the allegedly privileged materials were created "exclusively" for the purpose of complying with the PSA.<sup>10</sup> As a result, there is a danger that the PSA, while well intentioned, will spawn unintended mischief during litigation and may undermine the underlying goal of the Act—to ensure patient safety in New Jersey.

Part I of this Note tracks the development of self-critical analysis doctrine in the federal courts, emphasizing the seminal 1970 case *Bredice v. Doctors Hospital*.<sup>11</sup> Part II examines the history of self-critical analysis in New Jersey and where it stood on the eve of the PSA's passage. Part III tells the story of the PSA—why it was needed, how the Legislature and competing interest groups united behind the Act, and how the drafters utilized existing self-critical analysis doctrine in order to further their goals. Part IV shows, through the recent case of *Applegrad ex rel. C.A. v. Bentolila*,<sup>12</sup> how the PSA has unleashed some unforeseen results, in large part because, like with any controversial legislation, interested parties are now asserting novel interpretations of the Act. Part V argues that these consequences are a result of the PSA's misguided attempt to apply its vision of patient safety to the incompatible common law principles of self-critical analysis. This Note proposes a more modest self-critical analysis rule, based not on the laws of privilege but rather on the Subsequent Remedial Measures

---

<sup>9</sup> See *Applegrad ex rel. C.A. v. Bentolila*, 51 A.3d 119, 135 (N.J. Super. Ct. App. Div. 2012).

<sup>10</sup> See *id.* at 139.

<sup>11</sup> 50 F.R.D. 249.

<sup>12</sup> 51 A.3d 119.

doctrine (commonly referred to as “Rule 407”<sup>13</sup>), which, in practice, would deem self-critical materials inadmissible at trial yet still discoverable. This paradigm strikes a proper balance between the patient’s right to uncover the truth—regardless of any intention to sue—and the public interest in encouraging constant and candid assessments of hospital procedures. Incidentally, this may even further the PSA’s objective of limiting adverse health outcomes.

#### I. THE DEVELOPMENT OF THE SELF-CRITICAL ANALYSIS PRIVILEGE

Privileges “reflect societal choices that certain relationships (such as those between husbands and wives) or activities (such as seeking legal or medical advice) should be valued above others.”<sup>14</sup> Understood another way, privileges are the product of a principled determination by the privilege creator (typically a legislature or court) that the public would benefit from certain information remaining confidential. As one author succinctly stated, “[S]ociety needs privileges because in their absence, individuals will be discouraged from engaging in certain socially desirable behavior.”<sup>15</sup> A privilege can thus be regarded as a type of public interest carve-out to the discovery process, which otherwise allows for the disclosure of all potentially relevant material.<sup>16</sup>

---

<sup>13</sup> FED. R. EVID. 407.

<sup>14</sup> Pam Jenoff, *The Case for Candor: Application of the Self-Critical Analysis Privilege to Corporate Diversity Initiatives*, 76 BROOK. L. REV. 569, 576 (2011).

<sup>15</sup> *Id.* at 577. For example, without an attorney-client privilege, a client may be reluctant to speak frankly with her lawyer, and without a doctor-patient privilege, a patient may be reluctant to inform her physician of crucial, yet possibly embarrassing, details of her personal health. Society should (and does) encourage these behaviors, which, respectively, promote justice and improve health outcomes.

<sup>16</sup> *See, e.g.*, FED. R. CIV. P. 26(b)(1) (“Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party’s claim or defense . . . . Relevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence.”). Most states have similarly expansive rules. *See, e.g.*, N.Y.

While many privileges are so deeply rooted in our culture that few would question their necessity—the privilege against self-incrimination or the attorney-client privilege, for instance—*all* privileges are controversial in that they prevent a party from uncovering facts likely crucial to its case. As Justice Scalia noted, “[J]ustice . . . is severely harmed by contravention of ‘the fundamental principle that “the public has a right to every man’s evidence.”’”<sup>17</sup> Privileges, the Supreme Court famously admonished, “are not lightly created nor expansively construed, for they are in derogation of the search for truth.”<sup>18</sup> In sum, privileges are unabashedly bold vehicles for policymaking.<sup>19</sup>

#### A. *The Doctrinal Roots of the Self-Critical Analysis Privilege*

The self-critical analysis privilege is rooted in the belief that in certain situations, public policy demands that institutions engage in evaluative internal investigations and discussions in order to pinpoint—and hopefully correct—recurring problems or prior mistakes.<sup>20</sup> Because such discussions likely contain embarrassing or damaging information, participants may not

---

C.P.L.R. 3101(a) (McKINNEY 2005) (“There shall be full disclosure of all matter material and necessary in the prosecution or defense of an action, regardless of the burden of proof . . .”).

<sup>17</sup> *Jaffee v. Redmond*, 518 U.S. 1, 19 (1996) (Scalia, J., dissenting) (quoting 7 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 2192 (3d ed. 1940) (quoting Lord Hardwicke)).

<sup>18</sup> *United States v. Nixon*, 418 U.S. 683, 710 (1974).

<sup>19</sup> Because of the extraordinary power that privileges afford, and because they reflect overarching and often controversial policy decisions, the secondary question of *who* has the authority to create a privilege is itself an important public policy inquiry. Privileges in New York, for example, are almost entirely a product of statute. New York courts have traditionally declined to create new privileges. See RICHARD T. FARRELL, PRINCE, RICHARDSON ON EVIDENCE § 5-101 (11th ed. 1995) (“Efforts have been made to induce the courts to create privileges in favor of additional classes of persons, but without success.”).

<sup>20</sup> See, e.g., Brad Bacon, Note, *The Privilege of Self-Critical Analysis: Encouraging Recognition of the Misunderstood Privilege*, 8 KAN J.L. & PUB. POL’Y 221, 223 (1999) (“The privilege is premised on the rationale that unimpeded self-criticism serves a social good outweighing the cost of evidentiary exclusion.”).



speak frankly if they know their own self-critical analyses could be discovered by outsiders, or worse, used as evidence against them in a future lawsuit.<sup>21</sup> Therefore, the argument goes, the contents of these discussions must remain confidential.

The belief that a party should not be compelled to disclose its self-evaluative material is not novel. Such a rationale is embedded in two well-recognized and existing protections: (1) the attorney-client privilege and (2) the work-product doctrine. The attorney-client privilege ensures the “full and frank communication between attorneys and their clients and thereby promote[s] broader public interests in the observance of law and administration of justice.”<sup>22</sup> A broad attorney-client privilege thus encourages a form of self-critical analysis.<sup>23</sup> The work-product doctrine, articulated in *Hickman v. Taylor*<sup>24</sup> and now codified in Rule 26(b)(3) of the Federal Rules of Civil Procedure, prevents a party from discovering documents that were prepared in “anticipation of litigation.”<sup>25</sup> The doctrine promotes the adversarial system, and more generally ensures fairness, by preventing a party from unjustly benefiting from the hard work of its adversary.<sup>26</sup> Importantly, the work-product

---

<sup>21</sup> See generally 1 DAVID M. GREENWALD ET AL., TESTIMONIAL PRIVILEGES § 1:120 (3d ed. 2005 & Supp. 2012).

<sup>22</sup> *Upjohn Co. v. United States*, 449 U.S. 383, 389 (1981). In *Upjohn*, the Supreme Court recognized that robust attorney-client privilege encourages corporate entities to investigate and root out possible illegal activities within their own ranks.

<sup>23</sup> See, e.g., Stuart E. Rickerson, *The Privilege of Self-Critical Analysis: How to Raise It and Use It*, 58 DEF. COUNS. J. 504, 507 (1991) (stating that *Upjohn* “could have become the cornerstone of the critical self-examination privilege”). An implicit assumption in both attorney-client and self-critical analysis privilege is what might be called the proactive “nip it in the bud” approach, where reliance on forward-looking internal compliance approaches produces higher degrees of conformity with the law and is therefore more efficient and desirable than post hoc deterrents and remedies through the imposition of civil or criminal liability.

<sup>24</sup> *Hickman v. Taylor*, 329 U.S. 495 (1947).

<sup>25</sup> FED. R. CIV. P. 26(b)(3)(A).

<sup>26</sup> Sherman L. Cohn, *The Work-Product Doctrine: Protection, Not Privilege*, 71 GEO. L.J. 917, 943 (1983). The doctrine serves a more forward-thinking goal as well: the quality of attorney work product would suffer if such material were easily obtainable by adversaries. See *id.* at 919–

doctrine is a protection, *not* a privilege: a court will order discovery if a litigant asserts a “substantial need” for the materials,<sup>27</sup> although “mental impressions, conclusions, opinions, or legal theories of a party’s attorney or other representative” are always protected.<sup>28</sup>

These “attorney-based protections,” however, do not extend to more general self-critical materials.<sup>29</sup> Most courts have interpreted the “anticipation of litigation” standard of the work-product doctrine fairly narrowly, protecting only work product prepared by an attorney in response to an actual event that could reasonably give rise to litigation.<sup>30</sup> As one commentator noted, many self-critical procedures and studies are designed to *prevent* litigation and thus would “not possess the requisite tie to litigation to invoke work-product protection.”<sup>31</sup> The attorney-client privilege, on the other hand, only protects confidential communications between an attorney and her client.<sup>32</sup> Information acquired by an attorney from other sources, including third parties, is not protected, however “confidential” it may seem in the colloquial sense of the term.<sup>33</sup> In sum, neither

---

20; *see also Hickman*, 329 U.S. at 516 (“Discovery was hardly intended to enable a learned profession to perform its functions either without wits or on wits borrowed from the adversary.”).

<sup>27</sup> FED. R. CIV. P. 26(b)(3)(A) provides that the discovering party can overcome the protection if it “shows that it has substantial need for the materials to prepare its case and cannot, without undue hardship, obtain their substantial equivalent by other means.”

<sup>28</sup> FED. R. CIV. P. 26 (b)(3)(B).

<sup>29</sup> Patricia L. Andel, *Inapplicability of the Self-Critical Analysis Privilege to the Drug and Medical Device Industry*, 34 SAN DIEGO L. REV. 93, 98–103 (1997).

<sup>30</sup> *See* FED. R. CIV. P. 26 advisory committee’s note (“Materials assembled in the ordinary course of business, or pursuant to public requirements unrelated to litigation, or for other nonlitigation purposes are not under the qualified immunity provided by this subdivision.”); *see also Janicker v. George Washington Univ.*, 94 F.R.D. 648, 650 (D.D.C. 1982) (“The fact that a defendant anticipates the contingency of litigation resulting from an accident or an event does not automatically qualify an ‘in house’ report as work product.”).

<sup>31</sup> Andel, *supra* note 29, at 103.

<sup>32</sup> *Id.* at 99.

<sup>33</sup> *Id.* at 100; *see also* FARRELL, *supra* note 19, § 5-101 (quoting 7

the attorney-client privilege nor the work-product doctrine can ensure the confidentiality of self-critical materials. What is needed is a distinct self-critical analysis privilege.

*B. Bredice v. Doctor's Hospital*

*Bredice v. Doctors Hospital*<sup>34</sup> is often acknowledged as the first case in which a court recognized a common-law self-critical analysis privilege.<sup>35</sup> Ms. Bredice, in her medical malpractice action, sought discovery of the minutes from medical board meetings convened by the defendant hospital concerning the treatment received by her late husband.<sup>36</sup> The court observed that these meetings, which evaluated the performance of medical staff, were required by the Joint Commissions on Accreditation of Hospitals and existed for the “sole” purpose of improving care.<sup>37</sup> The court continued:

[T]hese meetings are essential to the continued improvement in the care and treatment of patients. Candid and conscientious evaluation of clinical practices is a *sine qua non* of adequate hospital care. To subject these discussions and deliberations to the discovery

---

WIGMORE, *supra* note 17, § 2286) (“No pledge of privacy, nor oath of secrecy can avail against demand for the truth in a court of justice.”). Particularly within the context of internal compliance efforts, such as in *Upjohn*, disclosure to *anyone* outside the agency of the party, including to government agencies, may constitute a waiver of the privilege. See *In re Steinhardt Partners*, 9 F.3d 230, 235–36 (2d Cir. 1993) (deeming company’s voluntary submission of materials a waiver); Andel, *supra* note 29, at 100. Further, the privilege “does not apply when the in-house attorney, who regularly wears several hats, is performing work that requires management expertise rather than work that requires legal acumen.” Andel, *supra* note 29, at 101.

<sup>34</sup> *Bredice v. Doctors Hosp., Inc.*, 50 F.R.D. 249 (D.D.C. 1970).

<sup>35</sup> Jenoff, *supra* note 14, at 580 (“[I]n *Bredice*, a court recognized for the first time that there was a strong public interest in allowing the free discussion of information in socially useful critical self-examination, and that if discovery of such materials were allowed, the flow of information would halt.”).

<sup>36</sup> *Bredice*, 50 F.R.D. at 249.

<sup>37</sup> *Id.* at 250.

process, without a showing of exceptional necessity, would result in terminating such deliberations. Constructive professional criticism cannot occur in an atmosphere of apprehension that one doctor's suggestion will be used as a denunciation of a colleague's conduct in a malpractice suit.<sup>38</sup>

The court therefore reasoned that there was an "overwhelming public interest" in keeping these staff meetings confidential "so that the flow of ideas and advice [could] continue unimpeded."<sup>39</sup> The court further noted that "what someone . . . at a subsequent date thought of these acts or omissions is not relevant to the case."<sup>40</sup> For both of these reasons, the court concluded that the meetings "are entitled to a qualified privilege."<sup>41</sup>

### *C. Doctrinal Disputes: How Far Should the Privilege Extend?*

*Bredice* predated Federal Rule of Evidence 501 (promulgated in 1974), which created a new framework for federal courts to determine when to recognize new privileges.<sup>42</sup> While one could

---

<sup>38</sup> *Id.*

<sup>39</sup> *Id.* at 251.

<sup>40</sup> *Id.* (alteration in original) (quoting *Richards v. Me. Cent. R.R.*, 21 F.R.D. 590 (D. Me. 1957)) (internal quotation marks omitted). This contention is highly questionable. FED. R. CIV. P. 26(b) allows for the discovery of materials "reasonably calculated to lead to the discovery of admissible evidence." Documents compiled in the wake of an adverse patient occurrence are almost *certain* to include relevant evidence, particularly the identity of witnesses, and will likely serve, in the words of James F. Flanagan, as a crucial "'road map' of the events" for the discoverer. James F. Flanagan, *Rejecting a General Privilege for Self-Critical Analyses*, 51 GEO. WASH. L. REV. 551, 558 (1983). "Any evaluation of the self-critical report . . . must start with the fact that it is undeniably relevant and of assistance in resolving the case." *Id.*

<sup>41</sup> *Bredice*, 50 F.R.D. at 251.

<sup>42</sup> See FED. R. EVID. 501 ("The common law—as interpreted by United States courts in the light of reason and experience—governs a claim of privilege unless any of the following provides otherwise: the United States Constitution; a federal statute; or rules prescribed by the Supreme Court. But in a civil case, state law governs privilege regarding a claim or defense for

question whether a *Bredice* holding would survive under a Rule 501 regime, many courts have since relied on *Bredice* to shield “self-critical” medical peer reviews.<sup>43</sup> In fact, medical peer reports, along with internal disciplinary investigations<sup>44</sup> and certain types of equal employment opportunity reports,<sup>45</sup> constitute the three types of documents most often afforded self-critical analysis protection. The common denominator in all these cases is a court’s determination that the public interest in encouraging candid analysis outweighs the litigant’s right to that information.

Self-critical analysis has been litigated almost entirely at the trial court level.<sup>46</sup> “Rely[ing] on their inherent power to control discovery,”<sup>47</sup> trial judges have applied the privilege on an ad hoc basis, creating what one commentator has referred to as a “confusing body of case law” with inconsistent results.<sup>48</sup> The

---

which state law supplies the rule of decision.”); FED. R. EVID. 501 advisory committee’s note.

<sup>43</sup> Andel, *supra* note 29, at 105–06.

<sup>44</sup> Note, *The Privilege of Self-Critical Analysis*, 96 HARV. L. REV. 1083, 1088 (1983) (citing, as examples, investigations conducted by railroad companies following an accident in order to “discipline any culpable employees and ultimately to improve the railroad’s safety” and police department investigations “when, following an arrest or shooting, a plaintiff has either alleged a civil rights violation or asserted a wrongful death claim”).

<sup>45</sup> *Id.* at 1089–90 (describing government contractors’ obligation under Title VII of the Civil Rights Act of 1964 to file documents that “candid[ly]” evaluate their own nondiscrimination procedures).

<sup>46</sup> See Note, *Making Sense of Rules of Privilege Under the Structural (Il)logic of the Federal Rules of Evidence*, 105 HARV. L. REV. 1339, 1352 n.74 (1992) [hereinafter *Making Sense of Rules of Privilege*].

<sup>47</sup> Flanagan, *supra* note 40, at 575.

<sup>48</sup> GREENWALD ET AL., *supra* note 21, § 1:119. Adding to the confusion, federal courts often treat privileges as matters of “substance” (rather than procedure) under an *Erie* analysis and therefore defer to the privilege law in the state in which they sit. See, e.g., *Lawson v. Fisher-Price, Inc.*, 191 F.R.D. 381, 382 (D. Vt. 1999) (“[T]his Court applies state law in determining whether a privilege for self-critical analysis exists.”); see also *Spencer Sav. Bank, SLA v. Excell Mortg. Corp.*, 960 F. Supp. 835, 836 (D.N.J. 1997) (“[A] federal court may ‘resort to state law analogies for the development of a federal common law of privileges in instances where the

deferential “abuse of discretion” standard governing appeals of trial court discovery rulings,<sup>49</sup> as well as parties’ inability in many jurisdictions to appeal discovery rulings until a “final” judgment, have resulted in a dearth of guidance from appellate courts, which, in turn, has created more unpredictability.<sup>50</sup> As a result, “some jurisdictions have cases with conflicting outcomes that are barely recognizable.”<sup>51</sup> Many courts have also simply rejected the privilege outright.<sup>52</sup>

The privilege’s lack of coherence has forced observers to grapple with the basic question of whether the self-critical analysis should be an “absolute” relational privilege of the attorney-client or doctor-patient type or rather an equitable tool of trial judges to ensure fairness during discovery and thus more similar to protections like the work-product doctrine. Despite the inconsistent application of the privilege, one influential *Harvard Law Review* Note discerned three overarching principles to the application of the privilege in certain scenarios. First, the privilege seeks to prevent the “dual chilling effect” discovery would unleash: “the direct chilling effect on the institutional or individual self-analyst . . . [which] operates to discourage the analyst from investigating thoroughly and frankly or even from investigating at all,”<sup>53</sup> as well as the chilling effect upon the data-“supplier,” which “discourage[s] individuals from coming

---

federal rule is unsettled.” (quoting *Wm. T. Thompson Co. v. Gen. Nutrition Corp.*, 671 F.2d 100, 104 (3d Cir. 1982))).

<sup>49</sup> *See, e.g.*, *Dowling v. Am. Haw. Cruises*, 971 F.2d 423, 425 (9th Cir. 1992).

<sup>50</sup> *See* GREENWALD ET AL., *supra* note 21, § 1:119.

<sup>51</sup> *Id.*

<sup>52</sup> *Id.* (“The privilege is defined differently in different jurisdictions, but in most cases the courts have found that the privilege did not apply to facts before them.”).

<sup>53</sup> *The Privilege of Self-Critical Analysis*, *supra* note 44, at 1091–92.

Fear of lawsuits, however, is not the only cause for hesitancy on the part of self-analysts. If an individual self-analyst is asked by his superiors to conduct an internal analysis, the individual may temper his criticism out of a fear that reprisals will result if the analysis ultimately leads to liability or adverse publicity for the employer.

*Id.* at 1092.

forward with relevant information.”<sup>54</sup> The second principle is that evaluations and opinions in self-critical materials are protected from discovery but the underlying facts upon which these opinions are based are not,<sup>55</sup> a distinction which the Note criticized, observing that “chilling effects of disclosure often operate on facts as well as evaluations.”<sup>56</sup> The third principle is that the privilege is not “absolute,” meaning it is applied on a case-by-case basis, and, even when applied, may be overcome if a party shows “exceptional need” for the material.<sup>57</sup> The Note likewise criticized this principle, advocating for a more absolutist privilege approach, analogous to the attorney-client context, in which judges decline to “weight the equities” in determining whether to apply the self-critical analysis privilege.<sup>58</sup> For much the same reason, the Note criticized the “exceptional need” concession: “The more crucial the material is to the [discoverer’s] case, the more likely it is to be the type of material that the privilege was designed to protect.”<sup>59</sup> The Note therefore criticized courts for “fail[ing] to give the privilege sufficiently broad application to effectuate the important policies underlying it.”<sup>60</sup>

In contrast, James F. Flanagan, in *Rejecting a General Privilege for Self-Critical Analyses*, asserted that self-critical analysis is not an “evidentiary privilege” and should instead be regarded as “an exercise in discretionary protection founded in the court’s power over discovery.”<sup>61</sup> Self-critical analysis is thus

---

<sup>54</sup> *Id.* at 1092. “Without the privilege, as the risk of liability for the institution increases, the likelihood that witnesses will come forward decreases.” *Id.*

<sup>55</sup> *Id.* at 1093–94. This same distinction exists in work-product doctrine. See FED R. CIV. P. 26(b)(3)(B) (protecting “mental impressions, conclusions, opinions, or legal theories”); *Hickman v. Taylor*, 329 U.S. 495, 510 (1947) (recognizing that thoughts are “inviolable” and “outside the arena of discovery”).

<sup>56</sup> *The Privilege of Self-Critical Analysis*, *supra* note 44, at 1095.

<sup>57</sup> *Id.* at 1096–97.

<sup>58</sup> *Id.* at 1098.

<sup>59</sup> *Id.* at 1099.

<sup>60</sup> *Id.* at 1100.

<sup>61</sup> Flanagan, *supra* note 40, at 576.

similar to the work-product doctrine, which protects certain materials from discovery absent a showing of “sufficient need.”<sup>62</sup> Flanagan concedes that protecting medical peer reviews from malpractice plaintiffs is necessary, recognizing that “a failure to ensure [their] confidentiality will diminish the[ir] quality.”<sup>63</sup> Yet he also observes that many states nevertheless protect these reports in the form of “peer review statutes,” which, unlike a general self-critical analysis rule, “provide sufficient exceptions so that no litigant will be seriously prejudiced because he cannot discover who was present or what occurred during a relevant review proceeding.”<sup>64</sup> Flanagan thus concludes that while self-critical analysis may be a worthwhile public policy, it is undeserving of an unqualified privilege.<sup>65</sup>

#### *D. The Current State of Self-Critical Analysis*

While the self-critical analysis doctrine has likely informed many medical peer-review statutes,<sup>66</sup> it has certainly not gained recognition as a general privilege. Instead, the privilege has been maintained as an equitable tool for trial courts to shield documents not otherwise protected by the attorney-client privilege or work-product doctrine.<sup>67</sup> There is little agreement

---

<sup>62</sup> *Id.* at 575.

<sup>63</sup> *Id.* at 576.

<sup>64</sup> *Id.* at 577. As discussed *infra* Parts II & III, New Jersey is one of the few states in which medical peer-review protections are *not* derived from statute, but rather from decisional law (the exception being the “utilization review” privilege, see *infra* Parts II & III).

<sup>65</sup> Flanagan, *supra* note 40, at 582 (“At best there are compelling reasons for courts to consider requests for discovery of peer reviews, to weigh alternatives, and to seek the least harmful means of disclosure.”).

<sup>66</sup> GREENWALD ET AL., *supra* note 21, § 1:121 (“State law relating to privileges is often governed by statute, and many states have statutes adopting forms of a self-evaluative privilege in a very limited context. For example, most states afford some confidentiality to medical peer reviews of patient care.”).

<sup>67</sup> *Id.* (observing that “in order to provide additional protection [aside from the attorney-client and work-product protections], some courts have recognized [the self-critical analysis privilege] to protect institutional self-analysis from outside discovery”).



even in those three areas where courts typically apply the privilege.<sup>68</sup> For example, one court observed that employment discrimination cases “are all over the map on whether the self-evaluative privilege exists,” noting that “[t]he privilege is a creature of the state trial courts, and there is little uniformity of law even within particular states.”<sup>69</sup> Even courts recognizing a self-critical analysis privilege have mostly found it did not apply in the cases before them.<sup>70</sup>

The Supreme Court’s decision in *University of Pennsylvania v. EEOC*,<sup>71</sup> which declined to recognize a peer-review privilege, was a discernible setback for the self-critical analysis movement. The University of Pennsylvania, defendant in a Title VII discrimination suit, refused to turn over tenure review files, arguing that courts should embrace a “common law” peer review privilege under Federal Rule of Evidence 501.<sup>72</sup> The Court held that “although Rule 501 manifests a congressional desire . . . [to] provide the courts with flexibility to develop rules of privilege on a case-by-case basis . . . we are disinclined to exercise this authority expansively.”<sup>73</sup> The University’s peer review claim was, at its core, one of self-critical analysis. Both privileges posit that society should encourage the frank evaluations of experts in a given field on matters of public import, even at the expense of denying individual litigants access to plainly relevant materials.<sup>74</sup> The Court’s rebuff of the University’s policy rationale thus foreclosed a similar self-critical analysis defense in the future. As such, subsequent

---

<sup>68</sup> See *supra* notes 44–45 and accompanying text.

<sup>69</sup> *Walker v. Cnty. of Contra Costa*, 227 F.R.D. 529, 532 (N.D. Cal. 2005); see also *Siskonen v. Stanadyne, Inc.*, 124 F.R.D. 610, 611 (W.D. Mich. 1989) (observing self-critical analysis law in federal discrimination cases to be “in disarray”).

<sup>70</sup> See GREENWALD ET AL., *supra* note 21, § 1:119.

<sup>71</sup> *Univ. of Pa. v. EEOC*, 493 U.S. 182 (1990).

<sup>72</sup> *Id.* at 188–89.

<sup>73</sup> *Id.* at 189 (citations omitted).

<sup>74</sup> See, e.g., *Making Sense of Rules of Privilege*, *supra* note 46, at 1352 n.75 (observing that the academic peer review privilege “center[s] upon many of the same normative and empirical arguments that dominate the self-critical analysis privilege area”).

courts have relied on *University of Pennsylvania* to reject a self-critical analysis privilege.<sup>75</sup>

Today, many courts rely on the four-part test articulated by the Ninth Circuit in *Dowling v. American Hawaii Cruises*,<sup>76</sup> which held that a party asserting the privilege must show that:

“[1] the information must result from a critical self-analysis undertaken by the party seeking protection; [2] the public must have a strong interest in preserving the free flow of the type of information sought; [3] the information must be of the type whose flow would be curtailed if discovery were allowed.” . . . [And 4, that the document] was prepared with the expectation that it would be kept confidential, and has in fact been kept confidential.<sup>77</sup>

In *Dowling*, the court allowed the plaintiff, in his personal injury action under the Jones Act,<sup>78</sup> to discover the factual content of the defendant cruise ship’s preaccident safety committee meeting minutes, reasoning that “organizations have many incentives to conduct such reviews that outweigh the harm that might result from disclosure,” such as fear of other lawsuits or simply “to avoid developing a reputation for having an unsafe premises.”<sup>79</sup>

Altogether, the federal courts have generally declined to extend application of the self-critical analysis doctrine.<sup>80</sup> Despite

---

<sup>75</sup> Jenoff, *supra* note 14, at 585 (observing that the majority of lower courts have “seemed to take *University of Pennsylvania* as a broad mandate to reject the self-critical analysis privilege” in the employment context).

<sup>76</sup> *Dowling v. Am. Haw. Cruises, Inc.*, 971 F.2d 423 (9th Cir. 1992).

<sup>77</sup> *Id.* at 426 (citations omitted) (quoting *The Privilege of Self-Critical Analysis*, *supra* note 44, at 1086).

<sup>78</sup> The Jones Act, also known as the Merchant Marine Act of 1920, empowers injured seamen with a tort remedy. 46 U.S.C. § 30104 (2011) (“A seaman injured in the course of employment . . . may elect to bring a civil action at law . . .”).

<sup>79</sup> *Dowling*, 971 F.2d at 426. Remanding the case, the court did not explicitly rule on whether opinions and conclusions would still be protected. *Id.* at 427.

<sup>80</sup> See GREENWALD ET AL., *supra* note 21, § 1:119.

the best efforts of the defendants' bar and corporate counsel,<sup>81</sup> the dream of a broad and impenetrable general privilege, as articulated in the Harvard Note, has not been realized.

## II. SELF-CRITICAL ANALYSIS IN NEW JERSEY

Because the Patient Safety Act referred by name to self-critical analysis, and because the statute itself explicitly incorporated the holding of *Christy v. Salem*,<sup>82</sup> an important self-critical analysis case which itself was the culmination of two decades of common law development, it is crucial to understand the status of the doctrine in New Jersey prior to passage of the PSA. New Jersey courts, much like the federal courts, have approached the privilege with caution, recognizing it only in limited situations.

### A. *The Lead-Up to Christy*

*Christy* represents a synthesis of two separate streams of case law—those that confront the self-critical analysis privilege within the context of medical peer reviews, and those that deal with the privilege more generally. The most important case, *Payton v. New Jersey Turnpike Authority*,<sup>83</sup> was the latter type. *Christy* could be understood as an application of *Payton* in the medical context. To appreciate the relevance of *Christy*, a very brief

---

<sup>81</sup> For example, the Federation of Defense and Corporate Counsel, a trade group “dedicated to representation of insurers and corporations,” FED’N DEF. & CORP. COUNS., <http://www.thefederation.org/> (last visited Feb. 21, 2013), recently published an article in their quarterly journal advocating for formal adoption of the privilege, Kurtis B. Reeg & Mathew A. Temper, *The Self-Critical Analysis Privilege: It Is Time for Formal Adoption*, 62 FED’N DEF. & CORP. COUNS. Q. 80 (2011).

<sup>82</sup> The statute provides that “[n]othing in this act shall be construed to increase or decrease the discoverability, in accordance with *Christy* . . . of any documents, materials or information if obtained from any source or context other than those specified in this act.” N.J. STAT. ANN. § 26:2H-12.25(k) (West 2007) (citation omitted). For much more on this, see *infra* Part III.

<sup>83</sup> *Payton v. N.J. Tpk. Auth.*, 691 A.2d 321 (N.J. 1997).

historical sketch of self-critical analysis in New Jersey is in order.

The first case to grapple with the privilege was *Wylie v. Mills*,<sup>84</sup> a lawsuit arising out of an automobile accident, where a defendant utility company sought protection of a document titled “[City of] Elizabeth Electric Transmission & Distribution Committee Investigation—Automobile Accident,” which purported to “determine whether [the defendant] should alter its procedures to avoid future injuries to employees.”<sup>85</sup> While the court summarily rejected the defendant’s contention that the document was protected under a work-product or attorney-client privilege, it found defendant’s assertion of the “nascent” self-critical analysis protection to be a “more formidable and persuasive argument.”<sup>86</sup> Citing *Bredice*, the court determined that “confidentiality and the ‘public need for confidentiality’ are the *sine qua non* of effective internal self-critical analysis”<sup>87</sup> and protected the evaluative portions of the report while ordering disclosure of the factual portions.

One year later, in 1985, the New Jersey Supreme Court, in *McClain v. College Hospital*, expounded upon *Wylie*’s discussion of the privilege.<sup>88</sup> Within the context of medical peer reviews, the court ruled that the plaintiff/discoverer must show

particularized need that outweighs the public interest in confidentiality of the investigative proceedings, taking into account (1) the extent to which the information may be available from other sources, (2) the degree of harm that the litigant will suffer from its unavailability, and (3) the possible prejudice to the agency’s investigation.<sup>89</sup>

---

<sup>84</sup> *Wylie v. Mills*, 478 A.2d 1273 (N.J. Super. Ct. Law Div. 1984). A Westlaw search reveals this to be the first New Jersey case in which a court either used the phrase “self-critical analysis” or cited *Bredice*.

<sup>85</sup> *Id.* at 1275.

<sup>86</sup> *Id.* at 1276.

<sup>87</sup> *Id.* at 1277.

<sup>88</sup> *McClain v. Coll. Hosp.*, 492 A.2d 991, 997–98 (N.J. 1985).

<sup>89</sup> *Id.* at 993.

Guided by the United States Supreme Court's decision in *EPA v. Mink*,<sup>90</sup> which held that factual material could be "severed" and thus disclosed from otherwise privileged documents, the court remanded the case and ordered an in camera inspection of the documents, holding that "strictly factual" contents be disclosed but that "matters of opinion or conjecture" be entitled to a "higher degree of protection."<sup>91</sup>

In *Bundy v. Sinopoli*,<sup>92</sup> the court noted that the legislature created a privilege for a hospital's "utilization review committee" reports<sup>93</sup> and also provided broad immunity for participants' statements made during the peer review process<sup>94</sup> yet also observed that "[t]he Legislature has *not* . . . provided for a privilege regarding the information contained within the Peer Review process."<sup>95</sup> The court nonetheless held the evaluations therein were "absolutely protected" under the "common law" self-critical analysis doctrine enunciated under *Wylie* and *McClain*.<sup>96</sup>

Finally, the New Jersey Supreme Court in *Payton* scaled back *Wylie*'s and *Bundy*'s broad interpretation of the privilege. *Payton*, in the course of her sexual harassment suit, sought to discover the minutes of the "executive session" her employer convened in response to allegations of harassment.<sup>97</sup> The court was confronted with two competing public interests, both of which further the same goal of limiting incidents of sexual harassment: "disclosure to ensure that employers maintain effective sexual-harassment procedures and nondisclosure to

---

<sup>90</sup> *EPA v. Mink*, 410 U.S. 73 (1973).

<sup>91</sup> *McClain*, 492 A.2d at 1000.

<sup>92</sup> *Bundy v. Sinopoli*, 580 A.2d 1101 (N.J. Super. Ct. Law Div. 1990).

<sup>93</sup> These documents, resulting from peer review, are created as a condition of receiving federal funding under the Social Security Act. *See id.* at 1104. The privilege is embodied in N.J. STAT. ANN. § 2A:84A-22.8 (West 2011).

<sup>94</sup> *Bundy*, 580 A.2d at 1106. This immunity protects participants from defamation claims. *See, e.g.,* *Bainhauer v. Manoukian*, 520 A.2d 1154 (N.J. Super. Ct. App. Div. 1987).

<sup>95</sup> *Bundy*, 580 A.2d at 1105.

<sup>96</sup> *Id.* at 1106.

<sup>97</sup> *Payton v. N.J. Tpk. Auth.*, 691 A.2d 321, 325 (N.J. 1997).

enable employers to maintain effective procedures that encourage reporting and candid statements by all involved.”<sup>98</sup> The court concluded that self-critical analysis was “not qualitatively different from other confidential information, and thus [did] not require the protection of a broad privilege.”<sup>99</sup> *Payton* therefore clarified that self-critical analysis was, at most, an occasional bulwark against discovery.<sup>100</sup>

### B. *Christy v. Salem*

On February 17, 2004, the Appellate Division decided *Christy v. Salem*.<sup>101</sup> In *Christy*, the plaintiff, in his medical malpractice action, sought the defendant hospital’s peer review report after learning that an x-ray material to his claims went missing and following the depositions of several physicians which “resulted in [alleged] discrepancies . . . concerning [how] events unfolded at the hospital.”<sup>102</sup> The court noted that the “conditional” privilege established in *Payton* empowered trial courts to protect confidentiality “short of suppression” through techniques such as “redaction, issuance of confidentiality or gag orders, and sealing of portions of the record.”<sup>103</sup> The court recognized that “here, unlike *Payton* [which implicated the ‘public interest’ of preventing sexual harassment], we are required to balance the private interest of a patient against the public interest of a hospital”<sup>104</sup> and concluded that “plaintiff’s interest in disclosure does not the have the ‘strong . . . reflection of important public policies, to outweigh . . .

---

<sup>98</sup> *Id.* at 329.

<sup>99</sup> *Id.* at 331.

<sup>100</sup> *Id.*

<sup>101</sup> *Christy v. Salem*, 841 A.2d 937 (N.J. Super. Ct. App. Div. 2004). *Christy* was decided following the January 26, 2004 hearings before the New Jersey Senate Health, Human Services and Senior Citizens Committee on the proposed PSA legislation, but prior to the March 4, 2004 committee hearings in the General Assembly. *See infra* Part III.

<sup>102</sup> *Christy*, 841 A.2d at 938.

<sup>103</sup> *Id.* at 940 (quoting *Payton v. N.J. Tpk. Auth.*, 691 A.2d 321, 330 (N.J. 1997)).

<sup>104</sup> *Id.*

confidentiality concerns under most, if not all, circumstances.’”<sup>105</sup> On the other hand, the court cited a section from the New Jersey statute commonly referred to as the “Patient Bill of Rights”<sup>106</sup> as support for the proposition that patients have a “right to know . . . what happened to them while in a hospital.”<sup>107</sup>

Defendants and amicus curiae New Jersey Hospital Association (“NJHA”) argued that if the court allowed disclosure of factual materials, it would lead hospitals to simply exclude them in future peer review documents.<sup>108</sup> The court rejected this contention as “contrary to the reasoning in both *McClain* and *Payton*,”<sup>109</sup> questioning whether facts, which “provide the basis” for self-critical analysis, would be excluded “simply because [they are] discoverable.”<sup>110</sup> Defendants also argued that plaintiff in any event failed to show a “compelling need” for the reports, to which the court responded that “[t]he availability of relevant facts from multiple sources has never in and of itself prevented discovery . . . . It is not unusual to find subtle differences in both testimony and documented facts, which support an argument bearing on credibility.”<sup>111</sup> The court therefore held that the plaintiff need not make a showing of compelling need to access the factual material of the documents and ordered their disclosure. The court even ordered disclosure of a so-called “deliberative” portion concerning the inability of the committee to reach a resolution on an issue due to “missing information” because the court believed it could reasonably lead to discovery.<sup>112</sup>

---

<sup>105</sup> *Id.* (alteration in original) (quoting *Payton*, 691 A.2d at 333).

<sup>106</sup> See N.J. STAT. ANN. § 26:2H-12.8(c) (West 2007) (empowering a patient with the right “[t]o obtain from the physician complete, current information concerning his diagnosis, treatment, and prognosis in terms he can reasonably be expected to understand”).

<sup>107</sup> *Christy*, 841 A.2d at 940.

<sup>108</sup> *Id.* at 939.

<sup>109</sup> *Id.* at 941.

<sup>110</sup> *Id.* at 942.

<sup>111</sup> *Id.* at 941–42.

<sup>112</sup> *Id.* “We are convinced that [defendants] would not be prejudiced by disclosure, notwithstanding its deliberative nature, because the peer review

Nonetheless, the court refused to allow discovery of the committee's opinions and fact-findings. First, the court felt that the plaintiff did not demonstrate a compelling need for disclosure because he already "obtained and supplied opinions from three separate experts supporting his claim of medical malpractice."<sup>113</sup> Second, the court noted that justifications for disclosure were based on allegations of the factual discrepancies, and having inspected the documents in camera, the court was "convinced" that by allowing disclosure of the other material, "[the] plaintiff's compelling needs [had] been addressed."<sup>114</sup> The court further ruled that the committee's factual findings were "of no use to plaintiff, as such findings are within the sole province of the jury" and that "disclosure might discourage a peer review committee from making factual findings because such findings often include a determination of what is credible."<sup>115</sup>

Because the PSA explicitly referenced (without comment) the holding in *Christy*, it is worth asking: what *was* the holding? On a basic level, it reaffirmed two basic and interrelated principles of self-critical analysis: first, that facts are generally discoverable; and second, that privileged material can nonetheless be discovered upon a showing of "substantial" or "compelling" need. In short, the privilege is qualified.<sup>116</sup> Nonetheless, it is possible that *Christy* did *not* have any discernible holding but was instead a series of fact-sensitive rulings—a good faith attempt to balance the competing interests and equities of rival discovery claims. The *Christy* court happened to conclude that the plaintiff demonstrated a compelling need to discover factual materials. One cannot be sure that the *Christy* court would reach the same conclusion in only slightly different circumstances. For example, the plaintiff in *Christy* was denied discovery of the evaluative materials

---

committee has itself been unable to resolve the issue due to the missing information, the possible whereabouts of which is described in the subject sentence." *Id.* at 942.

<sup>113</sup> *Id.*

<sup>114</sup> *Id.*

<sup>115</sup> *Id.* at 942–43.

<sup>116</sup> The references to *Christy* during the Senate hearings seem to subscribe to this reading. *See infra* Part III.



because the court was “convinced” that his compelling needs were addressed through disclosure of the factual documents. How would the court rule if, next time, it was “convinced” that the *evaluations*, and *not* the factual material, were more likely to meet a plaintiff’s needs? *Christy* should have even less precedential value considering that the documents at issue were confidential, inspected in camera, and without description in the decision. One must therefore consider the possibility that *Christy* was simply an application of existing self-critical analysis doctrine, specifically the *McClain* and *Payton* rules. Perhaps the court never intended to make new law.<sup>117</sup>

### III. THE PATIENT SAFETY ACT

#### A. *The Patient Safety Movement*

The PSA is New Jersey’s response to the relatively recent healthcare discipline known as “patient safety,” which examines the institutional problems in complex healthcare systems that cause medical errors.<sup>118</sup> The discipline stresses that the vigilant

---

<sup>117</sup> Judge Raymond A. Reddin, the trial Judge in *Applegrad ex rel. C.A. v. Bentolilia*, see *infra* Part IV, raised a related point during oral arguments:

Cases are not firmly rooted in cement. They change. They are modified . . . . So, what happens to this Patient Safety Act if the Supreme Court either expands *Christy*, reduces the scope of *Christy*, overrules *Christy*? Does not the legislature then have to say, we read what the Supreme court did in this decision and notwithstanding that, okay, forget what we said about *Christy*, now we say the holding in whatever this new case is doesn’t change anything? . . . . [D]id not the legislature posit the Patient Safety Act on something that isn’t strong footed? I mean, did they anchor the boat to something that may not be there tomorrow?

Stenographic Transcript of Proceeding Hearings, September 7, 2011, at 14–16, *Applegrad ex rel. C.A. v. Bentolila*, No. PAS-L-908-08 (N.J. Super. Ct. Law Div. Sept. 26, 2011).

<sup>118</sup> See Linda Emanuel et al., *What Exactly Is Patient Safety?*, in 1 ADVANCES IN PATIENT SAFETY: NEW DIRECTIONS AND ALTERNATIVE APPROACHES 4 (Kerm Henriksen et al. eds., 2008), available at [http://www.ahrq.gov/downloads/pub/advances2/vol1/Advances-Emanuel-Berwick\\_110.pdf](http://www.ahrq.gov/downloads/pub/advances2/vol1/Advances-Emanuel-Berwick_110.pdf) (defining patient safety both as “a discipline in the health care sector that applies safety science methods toward the goal of achieving a

patrolling, reporting, and analysis of healthcare phenomena—particularly adverse incidents (such as the event giving rise to the peer review in *Bredice*) and so-called “near misses”—will allow providers and policymakers to locate, and ultimately fix, the mechanisms that allowed for the error in the first place.<sup>119</sup>

The 1999 Institute of Medicine Report, *To Err is Human: Building a Safer Health System*, which alarmingly estimated that between 44,000 and 98,000 Americans died each year as a result of preventable medical errors,<sup>120</sup> effectively launched the Patient Safety Movement in America.<sup>121</sup> Within months of publication, President Clinton called for the creation of nationwide error-reporting systems and mandated the introduction of patient safety programs for hospitals participating in Medicare.<sup>122</sup> Federal and state agencies, hospitals, and other health providers followed suit, initiating mandatory reporting systems, improved health records systems, and other policies attempting to root out errors in medicine.<sup>123</sup> In 2005, Congress passed the Patient Safety and Quality Improvement Act of 2005 (“PSQIA”),<sup>124</sup> which created Patient Safety Organizations (“PSOs”) “to collect, aggregate, and analyze confidential information reported by health care providers” on a privileged and confidential basis, for analysis of

---

trustworthy system of health care delivery” and as “an attribute of health care systems; it minimizes the incidence and impact of, and maximizes recovery from, adverse events”).

<sup>119</sup> See *id.* at 2, 5–6; see also N.J. STAT. ANN. § 26:2H-12.24(d) (West 2007 & Supp. 2012); George J. Annas, *The Patient’s Right to Safety—Improving the Quality of Care Through Litigation Against Hospitals*, 354 NEW ENG. J. MED. 2063, 2065 (2006).

<sup>120</sup> NAT’L RESEARCH COUNCIL, *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM 1* (2000).

<sup>121</sup> See Bob Wachter, *The Patient Safety Movement Turns Ten*, HEALTH CARE BLOG (Dec. 2, 2009), <http://thehealthcareblog.com/blog/2009/12/02/the-patient-safety-movement-turns-ten/>.

<sup>122</sup> Fred Charatan, *Clinton Acts to Reduce Medical Mistakes*, 320 BRIT. MED. J. 597, 597 (2000).

<sup>123</sup> See generally Lucian Leape & Don Berwick, *Five Years After To Err Is Human: What Have We Learned?*, 293 JAMA 2384 (2005).

<sup>124</sup> The Patient Safety and Quality Improvement Act of 2005, Pub. L. 109-41, 119 Stat. 424 (codified in scattered sections of 42 U.S.C.).

patient safety events.<sup>125</sup> New Jersey sought to create a similar patient safety regime.

*B. The New Jersey Patient Safety Act*

The New Jersey Patient Safety Act was, in part, a response to revelations that a New Jersey nurse named Charles Cullen had administered lethal doses of medication to over forty patients under his care over a several-year period at roughly a dozen different facilities.<sup>126</sup> The PSA's drafters believed that a more robust centralized reporting system could have sooner uncovered Mr. Cullen's crimes.<sup>127</sup>

The portion of the Act entitled "Findings, declarations relative to patient safety" accurately summed up some basic tenets of the patient safety movement.<sup>128</sup> Readers should take

---

<sup>125</sup> *The Patient Safety and Quality Improvement Act of 2005*, AGENCY FOR HEALTHCARE RESEARCH & QUALITY (June 2008), <http://www.ahrq.gov/qual/psoact.htm>. The agency specifically cites the Institute of Medicine Report as the impetus for the bill.

<sup>126</sup> Senator Joseph F. Vitale, Chairman of the Senate Health and Human Services and Senior Citizens Committee, remarked during committee hearings that "this [bill] is, in part, due to the recent revelations of Charles Cullen, who has professed to have killed at least 40 individuals under his care." *Testimony on the Patient Safety Act: Hearing on S.B. 557 Before the S. Comm. on Health, Human Servs. & Senior Citizens*, 211th Leg., 2004–2005 Sess., at 2 (N.J. 2004) [hereinafter *Senate Hearing*]. In 2006, Cullen pleaded guilty to twenty-nine counts of murder and six counts of attempted murder in New Jersey and Pennsylvania courts. Sophie Sohn & Allan Chernoff, *Killer Nurse Gets 11 Life Sentences*, CNN JUSTICE (Mar. 10, 2006, 11:11 AM), [http://articles.cnn.com/2006-03-02/justice/killer.nurse\\_1\\_cullen-names-killer-nurse-john-shanagher?\\_s=PM:LAW](http://articles.cnn.com/2006-03-02/justice/killer.nurse_1_cullen-names-killer-nurse-john-shanagher?_s=PM:LAW).

<sup>127</sup> See *Senate Hearing*, *supra* note 126, at 31 (testimony of U.S. Senator Frank Lautenberg's Office) (commending the committee "for convening this important hearing to discuss recommendations to improve the integrity and safety of our health-care system in the wake of the tragic murders carried out by Charles Cullen"); *id.* at 41 (testimony of David Knowlton, Chairman, New Jersey Health Care Quality Institute) ("[I]n the Cullen case, there were people who had concerns, but they—if they're a nurse, they [first] have to report it to a supervisor. . . . [T]he new bill that you've just approved . . . would provide immunity.")

<sup>128</sup> N.J. STAT. ANN. § 26:2H-12.24 (West 2007 & Supp. 2012).

special note of subsection (e), as it seems to address medical malpractice litigation:

The Legislature finds and declares that:

a. Adverse events, some of which are the result of preventable errors, are inherent in all systems, and . . . the great majority of medical errors result from systems problems, not individual incompetence; . . .

e. To encourage disclosure of these events so that they can be analyzed and used for improvement, it is critical to create a non-punitive culture that focuses on improving processes rather than assigning blame. Health care facilities and professionals must be held accountable for serious preventable adverse events; however, punitive environments are not particularly effective in promoting accountability and increasing patient safety, and may be a deterrent to the exchange of information required to reduce the opportunity for errors to occur in the complex systems of care delivery. Fear of sanctions induces health care professionals and organizations to be silent about adverse events, resulting in serious under-reporting; and

f. By establishing an environment that both mandates the confidential disclosure of the most serious, preventable adverse events, and also encourages the voluntary, anonymous and confidential disclosure of less serious adverse events, as well as preventable events and near misses, the State seeks to increase the amount of information on systems failures, analyze the sources of these failures and disseminate information on effective practices for reducing systems failures and improving the safety of patients.<sup>129</sup>

To further these legislative goals, the Act mandated healthcare facilities to report every “serious preventable adverse event” to the Department of Health and Senior Services<sup>130</sup> and to

---

<sup>129</sup> *Id.*

<sup>130</sup> *Id.* § 26:2H-12.25(c).

notify patients of such occurrences “in a timely fashion.”<sup>131</sup> Crucially, the Act provided that:

Any documents, materials or information developed by a health care facility *as part of a process of self-critical analysis* conducted pursuant to subsection b. of this section . . . shall not be . . . subject to discovery or admissible as evidence or otherwise disclosed in any civil, criminal or administrative action or proceeding . . . .<sup>132</sup>

As indicated earlier, the Appellate Division decided *Christy* while the Patient Safety bill was already in the midst of discussions at the committee level. The drafters of the PSA added subsection (k), in an attempt to clarify the new privilege it had just created: “Nothing in this act shall be construed to increase or decrease the discoverability, in accordance with *Christy v. Salem* . . . of any documents, materials or information if obtained from any source or context other than those specified in this act.”<sup>133</sup>

This late addition of subsection (k) was the product of extensive negotiations with Senate and Assembly members and representatives from both NJHA and the Association of Trial Lawyers of America—New Jersey (“ATLA-NJ”) immediately prior to the Assembly hearing.<sup>134</sup> Evidently, and somewhat

<sup>131</sup> *Id.* § 26:2H-12.25(d).

<sup>132</sup> *Id.* § 26:2H-12.25(g) (emphasis added).

<sup>133</sup> *Id.* § 26:2H-12.25(k).

<sup>134</sup> Drew Britcher, in the NJAJ amicus brief in *Applegrad* noted that:

[O]n the day that the General Assembly Health Committee was to entertain the bill, the hearing of testimony regarding the bill was held until certain amendments, namely the provision concerning *Christy* . . . were presented with the assistance of the Office of Legislative Services. Indeed, the discussions concerning the inclusion of a specific case reference were so lengthy that they warranted a comment by the Assembly Health Committee Chairperson, the Honorable Loretta Weinberg that “we just completed the longest recess in the history of committee meetings.”

Brief on Behalf of New Jersey Association for Justice (NJAJ) at 3, *Applegrad ex rel. C.A. v. Bentolila*, No. PAS-L-908-08 (N.J. Super. Ct. Law Div. Sept. 26, 2011).

surprisingly, both NJHA and ATLA-NJ found *Christy*'s holding acceptable and lobbied for its inclusion in the PSA before agreeing to publicly support the legislation. All sides professed satisfaction. Elizabeth Ryan, General Counsel of NJHA, referring to the bill as "landmark legislation," thanked the committee for "preserving" *Christy*, a case "very important to the provider community."<sup>135</sup> Drew Britcher of ATLA-NJ likewise applauded *Christy*'s codification.<sup>136</sup>

That the PSA secured the endorsement of two oft-adverse interest groups was not lost on the bill's drafters. Sitting between Ms. Ryan and Mr. Britcher, sponsor Senator Joseph Vitale proclaimed that "we're here together in accord over this bill."<sup>137</sup> Clifton Lacy, Commissioner of the Department of Health and Senior Services, emphasized the need for additional protections, citing to a National Association of State Health Policy report indicating that, of the twenty-one states with legislation mandating the reporting of medical errors, New Jersey was alone in not also legislating a peer-review privilege.<sup>138</sup> Commissioner Lacy asserted that "the focus on finding who did wrong rather than why things go wrong is the major obstacle in improving safety across this country,"<sup>139</sup> and commended the bill for "shield[ing] self-critical analysis from discovery but maintain[ing as] discoverable all that is now discoverable."<sup>140</sup> The PSA passed both houses soon thereafter, was signed into law by Governor Jim McGreevey on April 27, 2004, and went into effect on October 24 of that year.<sup>141</sup>

---

<sup>135</sup> *Legislative Hearing on Patient Safety Act Before the Gen. Assemb. Comm. on Health, Human Servs. & Senior Citizens*, 211th Leg., 2004–2005 Sess., at 7 (N.J. 2004) [hereinafter *Assembly Hearing*].

<sup>136</sup> *See id.* at 22–23 (praising the "preservation of the discoverability . . . reconfirmed recently by our courts," which "draw[s] an important balance between the absolutely vital aspect of trying to identify medical error . . . while at the same time recognizing . . . [that] the patient does need to know what has happened to them").

<sup>137</sup> *Id.* at 2.

<sup>138</sup> *Id.* at 19–20.

<sup>139</sup> *Id.* at 15.

<sup>140</sup> *Id.* at 19.

<sup>141</sup> *See* Press Release, N.J. Gov. Jim McGreevey, McGreevey Signs

## IV. DUELING INTERPRETATIONS, UNINTENDED MISCHIEF

A. *Applegrad ex rel. C.A. v. Bentolila*

Despite the good feelings on all sides surrounding the passage of the statute, a glaring inconsistency existed in the PSA: subsection (g) provided—without exception or ambiguity—that the materials developed as a process of self-critical analysis “shall not be subject to discovery or admissible as evidence,” while subsection (k) codified *Christy*, which provided for the disclosure of certain self-critical materials. The Appellate Division was confronted with this dilemma in *Applegrad ex rel. C.A. v. Bentolila*, where plaintiffs Esther and Gedalia Applegrad, on behalf of their infant child “C.A.,” alleged medical malpractice against Valley Hospital (“Valley”) for the brain damage and oxygen deprivation sustained by C.A. during delivery.<sup>142</sup> During discovery, Valley withheld six documents, which it asserted were absolutely privileged.<sup>143</sup> The motion judge sided with Valley, ruling that the PSA was a “legislative overruling” of *Christy* and that the materials were fully protected from disclosure.<sup>144</sup>

On appeal, the Appellate Division noted that “[a]lthough not specifically mentioned in *Christy*, several regulatory and professional standards existed before . . . adoption of the PSA

---

Landmark Law Protecting NJ’s Families (Apr. 27, 2004), [http://web.archive.org/web/20041116163012/http://www.state.nj.us/cgi-bin/governor/njnewsline/view\\_article\\_archives.pl?id=1884](http://web.archive.org/web/20041116163012/http://www.state.nj.us/cgi-bin/governor/njnewsline/view_article_archives.pl?id=1884); *Medical Errors and Patient Safety – New Jersey*, QUALITY & PATIENT SAFETY, [http://qups.org/med\\_errors.php?c=individual\\_state&s=31&t=1](http://qups.org/med_errors.php?c=individual_state&s=31&t=1) (last visited Feb. 24, 2013).

<sup>142</sup> *Applegrad ex rel. C.A. v. Bentolila (Applegrad I)*, No. L-0908-08, 2011 WL 13700, at \*1 (N.J. Super. Ct. App. Div. Jan. 5, 2011).

<sup>143</sup> *Id.* at \*2. Valley identified those documents: “Occurrence Report; Director of Patient Safety Post Incident Analysis; Department of Risk Management Request for Quality Assurance Review; Mother/Baby Quality Assurance/Performance Improvement Review; Department of OB/GYN Quality Assurance Response; and Utilization Review Committee, Quality Assessment and Improvement Subcommittee of the Department of OB/GYN.” *Id.*

<sup>144</sup> *Id.* at \*4. Initially, the judge ordered disclosure of two documents but changed course following an ex parte meeting with defense counsel, who for the first time asserted privilege under the PSA. *Id.*

. . . that pertain to the activities of hospitals in engaging in forms of internal self-assessments and reporting[,] some or all . . . apparently continu[ing] to this day.”<sup>145</sup> The question, therefore, was, how did the PSA alter existing law?

Valley and amicus NJHA argued that the PSA represented a “sweeping change in the law of privilege, . . . insulat[ing] from disclosure a wide range of documents and information that previously may have been subject to disclosure.”<sup>146</sup> On that view, subsection (k) simply clarified that documents *not* produced pursuant to the PSA would still be subject to a *Christy* analysis and remain partially discoverable.<sup>147</sup> Plaintiffs and amicus New Jersey Association for Justice (formerly ATLA-NJ)<sup>148</sup> argued that “*Christy*’s factual/evaluative distinction still applies to documents generated under the PSA”<sup>149</sup> and that, at any rate, the privilege should not apply because there was no proof that Valley actually reported the Applegrad event to state officials pursuant to the Act.<sup>150</sup>

Nonetheless, the court eschewed answering any of these “interpretative issues” due to what it felt were “especially troublesome” “uncertainties” in the record regarding why and how these withheld documents actually came into being<sup>151</sup>:

---

<sup>145</sup> *Id.* at \*7. For example, the court observed that the Legislature directed hospitals to develop “peer review quality assurance processes” but pointedly did *not* provide that such documents be privileged. Also, hospitals, in accordance with guidelines established by the Joint Commission on Accreditation of Healthcare Organizations, had already engaged in “self-critical analysis procedures” to determine the “root cause” of adverse occurrences. *Id.* at \*6.

<sup>146</sup> *Id.* at \*8. It is worth remembering that NJHA supported the Patient Safety bill *because of*, not in spite of, the “preservation” of *Christy*. *See supra* Part III.

<sup>147</sup> *Applegrad I*, 2011 WL 13700, at \*8. NJHA specifically cited the preservation of *Christy* as grounds for supporting the PSA at the General Assembly hearing. *See supra* Part III.

<sup>148</sup> NJAJ was represented by Drew Britcher, former president of ATLA-NJ, who testified at the General Assembly hearing. *See supra* Part III.

<sup>149</sup> *Applegrad I*, 2011 WL 13700, at \*8.

<sup>150</sup> *Id.*

<sup>151</sup> *Id.* at \*8, \*9.



One of the documents . . . bears a heading of “Occurrence Report,” with no further explanatory label or statutory cross-reference. Another one . . . contains a boxed legend on its first page citing the PSA, stating that “This Quality Assurance Document was created and is protected in accordance with N.J.S.A. 26:2H-12.23 et seq.” Two of the documents . . . bear a different heading with no statutory citation: “CONFIDENTIAL RISK MANAGEMENT / QUALITY ASSURANCE DOCUMENT.” Another document . . . contains no label and is on business letterhead. Lastly, the document dated September 10, 2007 bears this heading: “CONFIDENTIAL AND PRIVILEGED Pursuant to N.J.S.A. 2A:84A[-]22.8,” the utilization review statute.<sup>152</sup>

The court exclaimed that “mere labeling of a hospital document does not necessarily control its legal classification.”<sup>153</sup> The court also noted Valley’s inability to explain how self-critical “organizational structures and processes” actually changed in the wake of the PSA’s enactment.<sup>154</sup> The court therefore remanded the case, directing Valley to explain in greater detail “the internal processes within the hospital that generated each document, and how those processes relate to . . . other standards apart from the PSA.”<sup>155</sup>

Finally, and most curiously, the court suggested in a footnote that the PSA’s “restriction on evidential admissibility in the courts” may have improperly limited the judiciary’s powers in violation of the New Jersey Constitution and that the Legislature “apparent[ly] fail[ed] to follow the proscribed procedures for the adoption of evidence rules under the Evidence Act of 1960.”<sup>156</sup>

---

<sup>152</sup> *Id.* at \*9. A seemingly exasperated court exclaimed, “[W]e are unsure what to make of this hodgepodge of labels.” *Id.*

<sup>153</sup> *Id.*

<sup>154</sup> *Id.* For example, the record was silent on how the functions and authority of Valley’s “Patient Safety Director” (created pursuant to the PSA) corresponded to the other “related ongoing [peer-review] operations within the hospital.” *Id.*

<sup>155</sup> *Id.* at \*11.

<sup>156</sup> *Id.* at \*8 n.8; see N.J. CONST. art. VI, § 11, cl. 3 (providing that

However, the court declined to comment further because neither party challenged the PSA on separation of powers grounds.<sup>157</sup>

*B. Remand: Judge Reddin's Opinion*

On remand, Valley asserted that only two out of six documents were deserving of PSA protection: a summary of a roundtable discussion convened by the hospital's Director of Patient Safety to engage in a self-critical analysis of the Applegrad occurrence ("DV2") and a document which memorialized specific activities conducted following the roundtable discussion ("DV5").<sup>158</sup> Following in camera inspection, testimony of hospital officials, and several days of oral arguments, the trial judge, the Honorable Raymond A. Reddin, delivered an oral ruling.<sup>159</sup> While recognizing "some inconsistency between *Christy* and the language of the statute,"<sup>160</sup> he nonetheless ruled that the intent of the PSA was to allow individuals to "speak freely without a fear of retribution" and therefore the self-critical analysis documents created pursuant to the Act were "entitled to a full privilege and no *Christy* analysis is warranted."<sup>161</sup> Nonetheless, the judge held this absolute

---

"[t]he Supreme Court shall make rules governing the administration of all courts in the State and, subject to the law, the practice and procedure in all such courts"). The Evidence Act, N.J. STAT. ANN. §§ 2A:84A-33-34 (West 2011) mandates special procedures to ensure that rules of evidence be adopted only through the joint collaboration of all three branches of government. *See State v. Byrd*, 967 A.2d 285, 294-97 (N.J. 2009) (discussing the history of and rationale for the Rules of Evidence).

<sup>157</sup> *Applegrad I*, 2011 WL 13700, at \*8.

<sup>158</sup> *See Applegrad ex rel. C.A. v. Bentolila (Applegrad II)*, 51 A.3d 119, 129 (N.J. Super. Ct. App. Div. 2012) ("Defendants, on reflection, modified their earlier position that all of the withheld documents were privileged under the PSA, and instead limited their claims of PSA confidentiality to exhibits DV2 and DV5."). *See* subheadings "DV2" and "DV5," *id.* at 132-33, detailing the purpose of each document.

<sup>159</sup> Stenographic Transcript of Proceedings Decision, Sept. 12, 2011, *Applegrad ex rel. C.A. v. Bentolila*, No. PAS-L-908-08 (N.J. Super. Ct. Law Div. Sept. 26, 2011) [hereinafter Sept. 12 Record].

<sup>160</sup> *Id.* at 26.

<sup>161</sup> *Id.* at 31.

privilege be construed narrowly.<sup>162</sup> As such, the judge ordered the disclosure of the names of the participants and the date of the discussion in DV2.<sup>163</sup> He also ruled that the statute was not unconstitutionally vague and that because “[the] Legislature showed respect to the Supreme Court” in acknowledging *Christy*, the PSA did not violate separation of powers principles.<sup>164</sup>

The judge then turned to the documents in question. Having interpreted the PSA to create a full privilege, the judge was left to fashion an appropriate test for when to actually *apply* the privilege. The judge concluded that, notwithstanding Valley’s failure to report the incident, Valley had prepared DV2 in “good faith” and “substantial[ly] compli[ed]” with the PSA, and thus was entitled to the protections of the statute.<sup>165</sup> Nonetheless, the judge suggested that the eventual trial judge, as “gatekeeper,” should have a copy of the confidential documents to ensure that no eventual witness gives testimony contradicting the document because “the court can never function in a way [allowing the presentation of] improper testimony.”<sup>166</sup> The judge reasoned that this caveat was simply a matter of judicial “integrity.”<sup>167</sup> On the other hand, the judge found DV5 to be a “Risk Management” or

---

<sup>162</sup> *Id.* at 50–51.

<sup>163</sup> As support, the judge cited a prior New Jersey case ordering an attorney to reveal the address of a client, attorney-client privilege notwithstanding. *See id.* at 52–53 (citing *Horon Holding Corp. v. McKenzie*, 775 A.2d 111 (N.J. Super. Ct. App. Div. 2001)). That the judge felt compelled to utilize case precedent dealing with an entirely different privilege should indicate the novelty of the legal questions presented in *Applegrad*.

<sup>164</sup> *Id.* at 39.

<sup>165</sup> *Id.* at 51. If, on the other hand, the judge detected “bad faith or fraud or concealment or a cover up . . . . [The protection] could be lost and the sanction should be beyond [the hospital] paying a fine. Paying a fine does nothing to the person who had treatment and had a concealment occur . . . .” *Id.*

<sup>166</sup> *Id.* at 61–62.

<sup>167</sup> *Id.* The judge also suggested the appointment of a “discovery master” to monitor the process and to “see if there’s any problems that are later developed.” Stenographic Transcript of Proceedings Supplemental Decision at 24, Sept. 14, 2011, *Applegrad*, No. PAS-L-908-08 [hereinafter Sept. 14 Record].

“Quality Assurance document,” developed independent of the PSA and was therefore subject to a *Christy* analysis.<sup>168</sup>

Remarkably, the judge found that the PSA “has done nothing to change the pre-Patient Safety Act statutes and regulations . . . . All the Patient Safety Act does is encourage more reporting and how things are reported to create an atmosphere of trust.”<sup>169</sup> Specifically with regard to Valley, the court found no “tremendous difference in the way [it] investigated incidents before and after the [PSA].”<sup>170</sup>

Both parties appealed different aspects of Judge Reddin’s ruling.<sup>171</sup> Judge Reddin remarked that the “entire medical community” and the “lawyers associated with it” are “looking to see if this statute is going to be validated or invalidated [and] . . . if there really will be confidentiality.”<sup>172</sup>

### *C. Appellate Decision: Discarding “Substantial Compliance” for “Exclusivity”*

On August 9, 2012, in *Applegrad ex rel. C.A. v. Bentolila*, the Appellate Division held that the PSA’s “repeated emphasis on confidentiality . . . cannot be reconciled with plaintiffs’ claim

---

<sup>168</sup> Sept. 14 Record, *supra* note 167, at 16.

<sup>169</sup> *Id.* at 12.

<sup>170</sup> *Id.* at 6.

<sup>171</sup> Plaintiffs in particular argued that the trial judge’s interpretation of the PSA would render it unconstitutional—if the PSA did indeed create an absolute self-critical analysis privilege, it thus constituted a legislative overruling of *Payton*, in which the Supreme Court declined to recognize one. The trial court therefore “should have rejected . . . that the legislature could so cavalierly and vaguely create a new privilege . . . . [I]f the Legislature did [so] . . . without consultation with the Court,” it was in violation of the Constitution and the Evidence Act. Plaintiffs/Appellants’ Merit Brief at 37–38, *Applegrad ex rel. C.A. v. Bentolila (Applegrad II)*, 51 A.3d 119 (N.J. Super. Ct. App. Div. 2012). The New Jersey Supreme Court, to date, has not ruled on whether *Christy* was correctly decided. Certification in *Applegrad* was granted on December 6, 2012, and the court will hear argument later this year. One likely issue is whether the court will be “bound” by the Appellate Division’s holding in *Christy*, now that the Legislature has enshrined it in the PSA.

<sup>172</sup> Sept. 12 Record, *supra* note 159, at 62–63.

that the PSA's non-disclosure protections must yield to the exceptions set forth in *Christy*.<sup>173</sup> As for the inclusion of subsection (k), the court reasoned that

[T]he Legislature appears to have adopted a dual approach, i.e., (1) treating materials exclusively developed under the PSA as subject to the PSA's specific confidentiality terms; and (2) treating other internal materials that are not exclusively developed under the PSA pursuant to the residual common-law standards set forth in *Christy* or other law."<sup>174</sup>

The court thus held that "the PSA extends absolute confidential protection to 'all documents . . .' developed exclusively . . . through the PSA process."<sup>175</sup> The court made clear the privilege existed "regardless of a plaintiff's asserted need for disclosure and regardless of whether the documents contain factual information in addition to subjective opinions."<sup>176</sup> However, the court issued a crucial caveat:

If, however, such items have been created or developed through some other "source or context," then they are obtainable under the criteria governing such alternative situations. . . . Thus, if a participant in the PSA process obtains facts or opinions from other sources or contexts, such as peer-review material from the facility's continuous quality improvement program, those facts or opinions are not transformed into inaccessible "PSA materials." . . . .

The confidentiality of a particular fact or opinion under the PSA therefore hinges upon an *exclusivity test*, requiring the court to consider whether the item was developed solely under the procedures set forth in the PSA, or whether the item had an independent genesis.<sup>177</sup>

---

<sup>173</sup> *Applegrad II*, 51 A.3d at 138.

<sup>174</sup> *Id.* at 138-39.

<sup>175</sup> *Id.* at 139 (quoting N.J. STAT. ANN. §§ 26:2H-12.25(f)-(g) (West 2007 & Supp. 2012)).

<sup>176</sup> *Id.* at 123.

<sup>177</sup> *Id.* at 139 (emphasis added).

The court also made clear that “nothing in the PSA insulates the underlying facts relating to a patient mishap, *if those facts can be learned from an independent source.*”<sup>178</sup> Rather, the PSA protects the committee’s self-critical *communications*.<sup>179</sup> The court also warned that its “construction of the PSA is not an invitation to health care providers to shield information that was previously accessible under *Christy* or under other law by indiscriminately labeling such formerly accessible items as ‘PSA material’” or “to evade the limitations of [the Act] by giving job titles to hospital personnel such as ‘PSA officers’ when, in fact, they are performing functions that are not truly covered by the PSA.”<sup>180</sup> Echoing its statement a year earlier enshrining substance over form,<sup>181</sup> the court focused its inquiry on “the actual functions and activities involved, rather than the nomenclature adopted.”<sup>182</sup>

Applying these principles to the documents at hand, the Appellate Division proceeded to *reverse* Judge Reddin’s rulings, determining that DV2 should be made available to the plaintiffs.<sup>183</sup> The court also upheld the constitutionality of the

---

<sup>178</sup> *Id.* at 140 (emphasis added). “For example, if counsel for a medical malpractice plaintiff deposes employees within the hospital having personal knowledge about a patient’s care, those witnesses cannot refuse to answer factual questions because those same facts also had been made known to the hospital’s patient safety committee.” *Id.*

<sup>179</sup> *Id.* (“[W]hat the PSA guards against is the disclosure of *communications* made within the PSA process itself, including the self-critical and deliberative analyses that are undertaken by a patient safety committee.” (emphasis added)).

<sup>180</sup> *Id.* at 140–41.

<sup>181</sup> See *Applegrad ex rel. C.A. v. Bentolila (Applegrad I)*, No. L-0908-08, 2011 WL 13700, at \*9 (N.J. Super. Ct. App. Div. Jan. 5, 2011) (holding that the “mere labeling of a hospital document does not necessarily control its legal classification . . . just as the stamping of a document as ‘attorney-client’ or ‘work-product’ does not ensure that the privilege was appropriately invoked”).

<sup>182</sup> *Applegrad II*, 51 A.3d at 141.

<sup>183</sup> *Id.* While conceding that the hospital attempted to comply with the Act “in good faith,” the court was specifically troubled by the fact that the roundtable discussion was staffed by nonphysicians and that the committee chose not to refer the matter to the Patient Safety Committee, despite the gravity of the incident implicating physician error. *Id.* at 144.

PSA (essentially answering the question it raised a year earlier), noting that “the Legislature has codified other evidentiary privileges in the past without the Judiciary’s involvement” and that “[g]iven this backdrop of constitutional and legal history, we decline to pronounce the confidentiality provisions in the PSA an invalid exercise of legislative power.”<sup>184</sup>

Defendants filed a motion for leave to appeal to the New Jersey Supreme Court, arguing that the Appellate Division’s holding, which in their view imposed additional restrictions on providers, should not apply retroactively to the specific documents at issue in *Applegrad*.<sup>185</sup> Tellingly, and quite understandably, defendants did *not* appeal the Appellate Division’s overall interpretation of the statute.<sup>186</sup> The recognition of an “absolute” PSA privilege will remain the law of the land, for now.<sup>187</sup>

---

<sup>184</sup> *Id.* at 145–46; *see also id.* at 146 (“[T]he ultimate assessment of this constitutional question is best reserved to the Supreme Court, as the final arbiter of the boundaries among our three branches of State government.”).

<sup>185</sup> *See* Brief of Defendants-Appellants in Support of Motion for Leave to Appeal at 1–2, *Applegrad ex rel. C.A. v. Bentolila*, No. A-1261-11T1 (N.J. Dec. 4, 2012).

<sup>186</sup> *See* Alicia Gallegos, *Patient Safety Law Protects Some Documents in Court Case*, AM. ASS’N MED. NEWS (Aug. 29, 2012), <http://www.ama-assn.org/amednews/2012/08/27/prsd0829.htm> (quoting *Applegrad* defense attorney stating that she was “gratified that the court upheld the privilege” and that “[i]t was wonderful to see that what hospitals, physicians and nurses had been concerned about for decades has the ability to go forward [and] improve health care”).

<sup>187</sup> In December 2012, the New Jersey Supreme Court granted defendants’ motion for leave to appeal. *Applegrad ex rel. C.A. v. Bentolila*, 2012 N.J. LEXIS 1257 (N.J. 2012). Because the issue of “retroactivity” presented to the court is a fairly narrow one, NJHA is no longer involved as amicus in the case. E-mail from Ross Lewin, Drinker Biddle & Reath, LLP, to author (Oct. 17, 2012) (on file with author).

## V. CONCLUSIONS

*A. An Analytical Misfit in the Family of Privileges*

The drafters of the PSA should have foreseen that their self-critical analysis privilege, by its own terms, was destined to unleash trouble. As argued in Parts I and II of this Note, the self-critical analysis privilege, as applied in both the federal system and New Jersey, was always “qualified.” Hence, courts assessed self-critical analysis on a case-by-case basis. The privilege could be overcome through the showing of a litigant’s exceptional need, and even if applied in a given case, the privilege would not protect facts from disclosure. This dynamic is in stark contrast to long-established “unqualified” privileges, such as the attorney-client or spousal varieties, in which courts refuse to allow for compromise, no matter how compelling the need, correctly recognizing that allowing for equitable “exceptions” would undermine the underlying relationships the privilege was designed to protect.

This Note therefore suggests that much of the controversy surrounding self-critical analysis is rooted in its unfortunate and erroneous description as a “privilege,” when a much more accurate descriptor would be “protection.” In determining whether a piece of evidence is privileged, a court should *not* consider the hardship an opposing party may endure. A thorough balancing of the positive and negative practical outcomes of recognizing a privilege is surely a crucial undertaking, but this must occur at the initial privilege-creating stage, not during its application.<sup>188</sup> As the *Upjohn* Court succinctly stated, “an uncertain privilege, or one which purports to be certain but results in widely varying applications by the courts, is little better than no privilege at all.”<sup>189</sup>

New Jersey and federal courts simply have not approached self-critical analysis doctrine with this level of deference. The “uncertainty” endemic to any qualified privilege has been one of

---

<sup>188</sup> See *supra* Part I. See generally *The Privilege of Self-Critical Analysis*, *supra* note 44, at 1097–99.

<sup>189</sup> *Upjohn Co. v. United States*, 449 U.S. 383, 393 (1981).



the defining characteristics of self-critical analysis. For example, the *Dowling* court declined to extend self-critical analysis protection to the defendant's safety-review documents because they "will be *invaluable to a plaintiff* attempting to prove that his injury was caused by the company's negligent failure to make safe a hazardous condition."<sup>190</sup> (One could hardly imagine a court ordering disclosure of attorney-client or confidential psychiatric material because it would be "invaluable" to an adversary's case.) Even the *Bredice* court, which applied the privilege—and which, incidentally, referred to it as "qualified"—did so because it felt plaintiff failed to show "good cause" to discover the materials.<sup>191</sup> For the self-critical analyzer to confidently predict whether a hypothetical discoverer will successfully argue "good cause" in front of a randomly chosen judge or magistrate seems next to impossible.

Similarly, in New Jersey, the three-part balancing test established in *McClain* emphasized the discoverer's "particularized need" for the self-critical materials,<sup>192</sup> while *Payton* emphasized the court's role in overseeing an "exquisite weighing process" in determining whether to shield documents.<sup>193</sup> For example, in *Christy*, the court refused disclosure of opinions of the committee, noting that the plaintiff had already retained experts to support his medical malpractice claim, but allowed for discovery of other portions which it believed could reasonably lead to discovery and which would not prejudice defendant. Finally, the current approach in the District of New Jersey, which borrows from both New Jersey and federal law, employs a six-part self-critical analysis test, which emphasizes equity balancing at virtually every stage.<sup>194</sup>

---

<sup>190</sup> *Dowling v. Am. Haw. Cruises*, 971 F.2d 423, 427 (9th Cir. 1992) (emphasis added).

<sup>191</sup> *Bredice v. Doctors Hosp., Inc.*, 50 F.R.D. 249, 251 (D.D.C. 1970) ("Absent evidence of extraordinary circumstances, there is no good cause shown requiring disclosure of the minutes of these meetings.").

<sup>192</sup> *McClain v. College Hosp.*, 492 A.2d 991, 993 (N.J. 1985).

<sup>193</sup> *Payton v. N.J. Tpk. Auth.*, 691 A.2d 321, 331 (N.J. 1997) (quoting *Loigman v. Kimmelman*, 505 A.2d 958, 964 (N.J. 1986)) (internal quotation marks omitted).

<sup>194</sup> *See, e.g., Bracco Diagnostics, Inc. v. Amersham Health Inc.*, No. 04-

This sampling of several self-critical analysis approaches shows that courts do *not* treat it like a privilege. A true “privilege” subject to such a murky and fact-sensitive post hoc inquiry would seemingly deter the self-critical analyst from the outset.<sup>195</sup> Rather, self-critical analysis is better understood as analogous to the work-product doctrine, which offers *protections* against discovery of certain material but will not overcome a discoverer’s showing of “substantial” or “compelling” need.<sup>196</sup>

In legislating that the self-critical materials would not be discoverable in “any” litigation, thus creating an unbending rule that did not have built-in “substantial need” exceptions or fact/evaluation distinctions, the drafters of the PSA rested on the unfounded premise that the self-critical analysis protection could be codified like any other privilege. As such, they morphed a flexible common law rule of discovery into an inflexible statutory mandate.<sup>197</sup>

### *B. Reforming Tort Law Through Evidence Law*

The PSA drafters also acted from the well-intentioned yet mistaken belief that their vision of a patient safety regime could be reconciled with both the liberal rules of civil discovery, in which absolute privileges are disfavored, and the traditional American tort model, in which medical errors are deterred

---

6025, 2006 WL 2946469, at \*3 (D.N.J. Oct. 16, 2006) (“When analyzing whether the self critical analysis privilege is applicable, a court must balance (1) whether the information is the result of a self critical analysis undertaken by the party seeking protection, (2) the extent to which the information is available from other sources, (3) the degree of harm the litigant will suffer from the information’s unavailability, (4) the possible prejudice to the party asserting the privilege, (5) the public interest in preserving the free flow of the type of information sought, and (6) whether the information is of the type whose flow would be curtailed if discovery were allowed.”).

<sup>195</sup> *Upjohn Co. v. United States*, 449 U.S. 383, 393 (1981).

<sup>196</sup> *See* Flanagan, *supra* note 40, at 575.

<sup>197</sup> *See* 6 JAMES WM. MOORE ET AL., MOORE’S FEDERAL PRACTICE ¶ 26.48 (3d ed. 1999 & Supp. 2013) (“That judicial decisions have been uneven both in their willingness to recognize a privilege for self-critical analysis and in their ability to define its scope is a predictable consequence of its common law roots.”).

through fear that an injured patient, empowered with broad discovery rights, will bring suit. Patient safety, as a legal principle, is adverse to private tort litigation because the former seeks to improve overall health outcomes for patients at the expense of an individual patient's discovery rights, while the latter supports the belief that allowing a patient to discover the truth of what happened in *her* case will improve overall outcomes. As Commissioner Lacy testified in his endorsement of the PSA, "the focus on finding *who* did wrong rather than *why* things go wrong is the major obstacle in improving safety across this country."<sup>198</sup> One could dismiss the Commissioner's opinions during committee hearings as irrelevant in evaluating legislative intent.<sup>199</sup> But there is no escaping that the "Findings and Declarations" portion of the statute itself called for combating "punitive environments," which it contended "are not particularly effective in promoting accountability and increasing patient safety, and may be a deterrent to the exchange of information."<sup>200</sup> The inclusion of these portions in the bill reflects the underlying policy judgments of the bill's drafters to create a *complete* self-critical analysis privilege.<sup>201</sup> Supporters of the bill should have anticipated the Act would frustrate a plaintiff/patient's opportunity to access relevant evidence.

To be sure, the Legislature attempted to placate various interest groups by including the reference to *Christy* in the actual statute text, implying the privilege would maintain its common law qualified status.<sup>202</sup> But as both Judge Reddin and the Appellate Division have made clear, the principles embedded in *Christy* are simply inconsistent with the overall intent of the

---

<sup>198</sup> See *Assembly Hearing*, *supra* note 135, at 15 (emphasis added).

<sup>199</sup> Justice Scalia put the matter rather bluntly in *Zedner v. United States*, 547 U.S. 489, 511 (Scalia, J., concurring): "[T]he use of legislative history is illegitimate and ill advised in the interpretation of any statute."

<sup>200</sup> N.J. STAT. ANN. § 26:2H-12.24(e) (West 2007 & Supp. 2012).

<sup>201</sup> See, e.g., *Applegrad ex rel. C.A. v. Bentolila (Applegrad II)*, 51 A.3d 119, 138 (N.J. Super. Ct. App. Div. 2012) (observing that § 26:2H-12.24 expresses "clear policy objectives to provide greater incentives . . . to disclose mishaps and perceived risks to patient safety" and is thus incompatible with the "balancing criteria of *Christy*").

<sup>202</sup> See *supra* Part III.

PSA.<sup>203</sup> A review of both the legislative history and the plain meaning of the statute indicates that applying a *Christy*-like qualified privilege to PSA materials would thwart the Act's explicit goal to afford such documents the simple yet absolute cloak of confidentiality. The only fair conclusion is that, while individual drafters or supporters of the bill may have hoped otherwise, the PSA has limited a patient/plaintiff's "private right" of discovery<sup>204</sup> in its attempt to further the broader societal good of improving patient safety.

Understood another way, the Legislature has unleashed a subtle brand of tort reform.<sup>205</sup> Proponents of reform argue that an aggressive tort system, which in theory purports to deter unsafe practices and conduct, can actually yield perverse incentives. In the healthcare context, providers will rationally be reluctant to engage in conduct benefitting the patient's or public's interest (such as error-reporting) if it could foreseeably expose them to liability.<sup>206</sup>

---

<sup>203</sup> See *Applegrad II*, 51 A.3d at 146–47 (holding that PSA's "repeated emphasis on confidentiality cannot be reconciled with plaintiffs' claim that the PSA's non-disclosure protections must yield to the exceptions set forth in *Christy*"); Sept. 12 Record, *supra* note 159, at 31 (holding that PSA documents are "entitled to a full privilege and no *Christy* analysis is warranted").

<sup>204</sup> See *Christy v. Salem*, 841 A.2d 937, 940 (N.J. Super. Ct. App. Div. 2004) (referring to a medical-malpractice plaintiff's right to discovery as a private interest, which, unlike those of a civil rights plaintiff, "does not have the 'strong . . . reflection of important public policies'" (alteration in original) (quoting *Payton v. N.J. Tpk. Auth.*, 691 A.2d 321, 333 (N.J. 1997))).

<sup>205</sup> See BLACK'S LAW DICTIONARY (9th ed. 2009) (defining tort reform as "[a] movement to reduce the amount of tort litigation, usu. involving legislation that restricts tort remedies or that caps damages awards (esp. for punitive damages)" and noting that "[a]dvocates of tort reform argue that it lowers insurance and healthcare costs and prevents windfalls, while opponents contend that it denies plaintiffs the recovery they deserve for their injuries").

<sup>206</sup> See generally David A. Hyman & Charles Silver, *The Poor State of Health Care Quality in the U.S.: Is Malpractice Liability Part of the Problem or Part of the Solution?*, 90 CORNELL L. REV. 893, 909–14 (2005) (discussing the "conventional wisdom" that "malpractice liability impedes efforts to improve patient safety"). As one nursing expert put it, "The threat

Clearly, this same argument serves not only as a rationale for traditional tort reform, but also for strong self-critical analysis protection. While the majority of tort reform has focused on limiting a defendant's exposure to damage awards, often through the institution of a noneconomic recovery cap<sup>207</sup> or through the abolishment of common law joint and several liability,<sup>208</sup> the PSA seeks to redress the perceived excesses of the tort system through utilizing rules of evidence. Therefore, in a typical tort reform jurisdiction, a healthcare provider (and its insurer or indemnifier) could rest assured that damages would not exceed a specified sum per accident. The PSA's self-critical analysis protection, on the other hand, does not limit liability per se. Instead, it attempts to further the goal of both the tort reform and the patient safety movements—encouraging the reporting of errors—by rendering such reports immune to discovery.

### *C. What Has Changed—And What Will Change—Under the PSA?*

The PSA's privilege was predicated on the finding that self-critical analysis could not occur without complete confidentiality.<sup>209</sup> Admittedly, the belief that providers fail to

---

of medical malpractice litigation is one of the most obvious barriers to the improvement of patient safety. . . . These risks also serve as disincentives to participate in improvement strategies to reduce the risk of error." *Id.* at 911 (quoting Beverly Jones, *Nurses and the "Code of Silence,"* in *MEDICAL ERROR: WHAT DO WE KNOW? WHAT DO WE DO?* 84, 91–92 (Marilynn Rosenthal & Kathleen Sutcliffe eds., 2002)).

<sup>207</sup> See *Noneconomic Damages Reform*, AM. TORT REFORM ASS'N, <http://www.atra.org/issues/noneconomic-damages-reform> (last visited Feb. 24, 2013) (identifying statutory reforms enacted by state).

<sup>208</sup> See, e.g., *Joint and Several Liability Rule Reform*, AM. TORT REFORM ASS'N, <http://www.atra.org/issues/joint-and-several-liability-rule-reform> (last visited Feb. 24, 2013) (identifying statutory reforms enacted by state).

<sup>209</sup> See N.J. STAT. ANN. § 26.2H-12.24(2)(f) (West 2007 & Supp. 2012) (creating "confidential disclosure" processes, thus providing the State with a means to "increase the amount of information on systems failures, analyze the sources of these failures and disseminate information on effective practices for reducing systems failures"); *Applegrad ex rel. C.A. v. Bentolila*

report errors out of fear of liability, in the words of one public health scholar, “is plausible and has intuitive appeal.”<sup>210</sup> Still, there is reason to be skeptical. One 2005 study found an absence of “any rigorous evidence show[ing] that fear of malpractice lawsuits discourages error reporting” and that “contrary to the conventional wisdom, malpractice liability itself has the potential to kick-start quality improvement.”<sup>211</sup> For example, rocketing malpractice premiums in the 1980s (due to anesthesia-related deaths) impelled the American Society of Anesthesiologists to launch an aggressive “patient safety campaign” to study incidents of medical errors in the field.<sup>212</sup> “By the early 1980s, anesthesiologists recognized that something drastic had to be done if they were going to continue to be insured,” recalled the leader of this movement.<sup>213</sup> The campaign, while costly, proved remarkably successful and has caused an astonishing ten-to-twenty-fold decrease in deaths over the past few decades.<sup>214</sup>

Further, failure to report errors may simply be a deeply rooted cultural phenomenon rather than an economically rational response to fears of liability. For example, one health scholar cited to a survey showing that seventy-five percent of U.S. doctors failed to report errors to their patients, which was not markedly different from a showing of sixty-one percent in New Zealand, “a country that has had no-fault malpractice insurance for more than three decades.”<sup>215</sup>

---

(*Applegrad II*), 51 A.3d 119, 124 (N.J. Super. Ct. App. Div. 2012) (observing that “the Legislature was manifestly concerned” about the underreporting and analysis of adverse incidents in New Jersey resulting from the “inhibition” of medical staff from “reporting or criticizing unsafe practices within the institution”); *id.* at 127 (observing that “the PSA’s umbrella of confidentiality” was designed, among other reasons, “to foster internal self-critical analysis”).

<sup>210</sup> Annas, *supra* note 119, at 2065.

<sup>211</sup> Hyman & Silver, *supra* note 206, at 894.

<sup>212</sup> *Id.* at 919.

<sup>213</sup> *Id.* at 920 (quoting Ellison C. Pierce, Jr., *ASA Monitoring Guidelines: Their Origin and Development*, 66 AM. SOC’Y ANESTHESIOLOGISTS NEWSL., Sept. 2002, at 22, 22, available at [http://www.asahq.org/Newsletters/2002/9\\_02/feature7.htm](http://www.asahq.org/Newsletters/2002/9_02/feature7.htm)).

<sup>214</sup> *Id.* at 918.

<sup>215</sup> See Annas, *supra* note 119, at 2065.

Some scholars also question whether privileges actually enhance the frequency and quality of patient safety procedures. For example, Susan Scheutzow, a health law practitioner and academic, through analyzing the National Practitioner Data Bank, found that peer-review protections, contrary to perceived wisdom, do not promote the public policy of encouraging peer review and thus “risk being little more than special interest laws protecting physicians and hospitals.”<sup>216</sup> Scheutzow therefore argued for the elimination—or at the very least, reformation—of such laws.<sup>217</sup>

In light of these general claims, one must ask what the PSA has accomplished thus far. From a large-scale public policy perspective, it is too early to tell. Nonetheless, the *Applegrad* litigation has brought to light at least one useful case study—the patient safety apparatus of Valley Hospital.

This much is clear: the Appellate Division recognized that many regulatory and professional standards already existed prior to passage of the PSA, many of which called for peer review and self-evaluation procedures quite similar to those required by the PSA.<sup>218</sup> On remand, and following days of oral arguments and document inspection, Judge Reddin concluded that he could discern “no tremendous difference” between Valley’s procedures prior to and following the PSA. As the judge noted, this reality does not in itself raise any presumptions of wrongdoing; to the contrary, it might even show that Valley was ahead of the curve in patient safety. Yet perhaps even more remarkably, the judge further found that the PSA “has done nothing” to change pre-PSA regulations and that “all the Patient Safety Act does is encourage more reporting . . . to create an atmosphere of trust.”<sup>219</sup>

One must therefore consider the irony that the PSA may ultimately result in a raw deal for patients—as a result of a codified and absolute self-critical analysis privilege, they may

---

<sup>216</sup> Scheutzow, *supra* note 5, at 8–9.

<sup>217</sup> *Id.* at 8.

<sup>218</sup> See *Applegrad ex rel. C.A. v. Bentolila (Applegrad I)*, No. L-0908-08, 2011 WL 13700, at \*5 (N.J. Super. Ct. App. Div. Jan. 5, 2011); see also *supra* Part IV.

<sup>219</sup> Sept. 14 Record, *supra* note 167, at 12.

discover less as litigants. At the same time, as hospitals continue to engage in procedures virtually identical to those prior to the PSA, they will get little added safety in return.

Providers, on the other hand, may have much to gain. Initially, Valley asserted that all six of the suppressed materials were privileged under the PSA.<sup>220</sup> Following the Law Division's ruling that "the mere labeling of a hospital document does not necessarily control its legal classification," Valley changed course and instead argued that only *two* documents were "PSA" materials deserving of a full privilege, conceding that the rest could be subject to a *Christy* analysis.<sup>221</sup> This author wonders whether other providers have taken note of *Applegrad* and have, as a result, attempted to reconfigure their existing peer review and adverse occurrence procedures in order to be afforded the maximum level of protection. Thus, a hospital would prudently comply with any requisite PSA formalities—however minor and inconsequential they may be to actual patient safety—simply in order to demonstrate that it has complied with the statute. Procedures that could have been introduced for a variety of reasons could overnight become "patient safety" procedures.<sup>222</sup>

If a change like this occurs, it could alter the dynamics of medical malpractice litigation. In such cases, one side—the defendant—will inevitably possess the vast majority of evidence, both inculpatory and exculpatory.<sup>223</sup> Plaintiffs therefore rely on a

---

<sup>220</sup> *Applegrad I*, 2011 WL 13700, at \*2.

<sup>221</sup> This salient fact was not lost on the Appellate Division the second time around. *See Applegrad ex rel. C.A. v. Bentolila (Applegrad II)*, 51 A.3d 119, 143 (N.J. Super. Ct. App. Div. 2012) ("The Hospital exhibited its confusion about the PSA in this very litigation by first asserting other privileges and not invoking the PSA; then arguing, after the trial court's initial in camera review, that all of the withheld documents were protected by the PSA; and ultimately arguing on remand that only DV2 and DV5 are covered by the PSA.").

<sup>222</sup> Alternatively, the Appellate Division's "exclusivity test" could result in an unintended irony: hospitals which (laudably) enacted comprehensive patient safety procedures *prior* to enactment of the PSA will now be punished for their foresight because their initially voluntary practices, now mandated by the PSA, are not "exclusively" a PSA product and are thus not entitled to the statutory privilege.

<sup>223</sup> *See, e.g., Christy v. Salem*, 841 A.2d 937, 940 (N.J. Super. Ct.



liberal discovery system to ensure the disclosure of every possibly relevant document and the deposition of every possible witness. As the *Christy* court made clear, a plaintiff has good cause to discover nearly all hospital documentation concerning an adverse event, even if the facts are available from alternate sources, because of the possibility of uncovering “subtle differences in both testimony and documented facts, which support an argument bearing on credibility.”<sup>224</sup> Therefore, a plaintiff’s incentives could change if certain hospital documents become increasingly less available. Under an “absolute” PSA regime, an incident at a hospital that prompts many peer review and patient safety sessions will not necessarily imply that there will be an opportunity for broad discovery of these documents.<sup>225</sup> As a result, plaintiffs’ attorneys may be more reluctant to take on cases where negligence (or, for that matter, outright fraud or cover-up) must be argued circumstantially: for example, through showing inconsistencies in deposition testimony and patient safety documents.

There is, of course, a counterargument. While the plaintiffs’ bar may feel that the patient safety movement is simply tort reform in disguise,<sup>226</sup> a regime emphasizing the importance of

---

App. Div. 2004) (observing that critical evidence regarding claims of medical malpractice “would logically be expected to be in the possession of an adversary”).

<sup>224</sup> *Id.* at 941–42.

<sup>225</sup> One plaintiff’s attorney, in commenting on *Applegrad*, opined that: [T]rying to get discovery from some hospitals is like trying to find the proverbial needle in a haystack. Except you are first told that: (a) there is no haystack; (b) if there ever was a haystack, it did not have any needles; and finally (c) if there was a haystack with a needle, any discussion of it is privileged. If you persist, you are then advised that (d) all haystacks and needles were designed, manufactured, distributed, maintained and utilized by persons who were independent contractors and, furthermore, (e) the hospital is entitled to a limitation of liability.

Abbot S. Brown, *Hospital Malpractice: Finding the Needle in the Haystack*, N.J.L.J., May 30, 2011, at 1.

<sup>226</sup> See generally Peter P. Budetti, *Tort Reform and the Patient Safety Movement: Seeking Common Ground*, 293 JAMA 2660, 2618 (2005) (drawing parallels between the results of tort reform and the patient safety movement).

patient safety may in fact be more willing to entertain novel claims of relief for plaintiffs. The health law and bioethics scholar George Annas, in *The Patient's Right to Safety—Improving the Quality of Care Through Litigation Against Hospitals*,<sup>227</sup> argues that:

[J]udicial recognition of an explicit “right to safety” for hospital patients, with a correlative duty of hospitals to implement patient-safety measures, can become the primary motivator for the development of systems to improve patient safety. Hospitals that do not take specific actions to improve safety should be viewed as negligent and be subject to malpractice lawsuits when a violation of the right to safety results in injury.<sup>228</sup>

Annas suggests that physicians, patients, and the plaintiffs’ bar join forces to propose initiatives to “pressure hospitals to change their operating systems” to ensure patient safety.<sup>229</sup> Annas’s argument is important because it reaffirms an obvious principle of tort law that the PSA minimized: a robust tort system that constantly patrols for incidents of fault, a system that (to use the disapproving language of the PSA) cultivates a “punitive culture” focusing on “assigning blame,”<sup>230</sup> can actually *promote*, not hinder, the development and implementation of innovative safety procedures. Plaintiffs’ lawyers may therefore experiment with new patient-safety-oriented claims for relief in cases where more traditional negligence theories may be difficult to prove.

Finally, how will the PSA affect judges? If the statute’s absolute privilege is upheld, judges will be deprived of the opportunity to engage in the fact-sensitive and context-oriented balancing that previously existed under the common law. Instead, and as occurred in *Applegrad*, a court’s “exquisite weighing process” will give way to the more rigid, yet still complex, task of determining whether the defendant hospital has in “good faith” “substantially complied” with the Act; or,

---

<sup>227</sup> Annas, *supra* note 119.

<sup>228</sup> *Id.* at 2063.

<sup>229</sup> *Id.* at 2066.

<sup>230</sup> See N.J. STAT. ANN. § 26.2H-12.24(e) (West 2007 & Supp. 2012).

according to the Appellate Division's test, whether the alleged privileged material was created "exclusively" for PSA functions.<sup>231</sup> Decades of self-critical analysis jurisprudence and case law will now be of little use to judges confronted with assertions of a PSA privilege.

On the other hand, perhaps not much will change after all. Judges, now tasked with determining the contours of the statute, may develop their own "exquisite weighing process" in evaluating whether to apply the privilege. First, as the litigation in *Applegrad* made clear, merely *distinguishing* between PSA documents and related patient-safety/peer-review materials was itself a controversial and fact-sensitive inquiry.<sup>232</sup> Second, the Appellate Division imposed several requirements on providers, such as the "exclusivity" test, or the requirement that "competent" personnel of "various disciplines" administer the reviews.<sup>233</sup> A court will find it difficult to determine whether a hospital complied with these requirements, and if the facts of *Applegrad* are any indicator, it will require days of testimony (and cross-examination) of hospital staff.<sup>234</sup> How can hospitals be sure that *their* patient safety documents will be deemed absolutely privileged in the future, and if they cannot be sure, will that reality, in and of itself, frustrate the PSA's primary goal of encouraging frank discussions and full disclosure? As one commentator predicts, "Due to the highly fact specific analysis undertaken by the Court in [*Applegrad*] and apparently to be applied by the courts considering the application of this privilege, uncertainty will remain as to the ultimate outcome in any given scenario."<sup>235</sup> Yet, as this Note has described,

---

<sup>231</sup> See *Applegrad ex rel. C.A. v. Bentolila (Applegrad II)*, 51 A.3d 119, 139 (N.J. Super. Ct. App. Div. 2012) ("[W]hether the item had an independent genesis [aside from the PSA] . . . at times . . . may be obvious. At other times, it might not, and would require closer scrutiny of how each particular fact or opinion was created.").

<sup>232</sup> See *id.* at 128.

<sup>233</sup> *Id.* at 141–42.

<sup>234</sup> *Id.* at 129 ("On remand, the trial court heard testimony over seven days from persons at the Hospital who were involved in the development of the allegedly privileged records.").

<sup>235</sup> Sharlene Hunt, *Court Addresses Confidentiality Under the Patient*

“uncertainty” has always been the predominant dynamic of self-critical analysis. Perhaps the ambiguous patient safety dynamics that existed under the common law will continue under this new PSA regime, simply under a new name. The Appellate Division made the unassailable point that a hospital’s “mere labeling” of a document as “privileged” counts for very little.<sup>236</sup> So too, the Appellate Division’s recognition of the PSA privilege as being “absolute” will be of little significance to providers and plaintiffs in light of the significant caveats the court imposed.

*D. Potential for Future Patient Safety: Rule 407*

This Note concludes by offering an alternative model for analyzing self-critical patient safety documents. The fact/opinion distinction which governed the majority of critical analysis jurisprudence but was discarded under the PSA represents a doctrinal and practical compromise for litigants: the discoverer will be entitled to crucial pieces of evidence, while his adversary can still maintain a degree of confidence that he will not be penalized for his investigation. This distinction also recognizes a more basic reality of trial practice: sometimes opinions can be more damaging than facts.<sup>237</sup> This Note therefore suggests that the rationale for the self-critical analysis privilege may be better served through a different evidentiary paradigm, one recognized in virtually every state, as well as under the Federal Rules of Evidence: the Subsequent Remedial Measures doctrine.<sup>238</sup>

---

*Safety Act*, N.J. HEALTHCARE BLOG (Aug. 14, 2012), <http://www.njhealthcareblog.com/2012/08/court-addresses-confidentiality-under-the-patient-safety-act/>.

<sup>236</sup> *Applegrad II*, 51 A.3d at 141 (“What matters for judicial review is the actual functions and activities involved, rather than the nomenclature adopted by the health care facility.”).

<sup>237</sup> Flanagan, *supra* note 40, at 576 (“[T]he use of the conclusions of such reviews in litigation renders the peer reviewers involuntary experts for one of the parties.”); *see also* *Bredice v. Doctors Hosp., Inc.*, 50 F.R.D. 249, 250 (D.D.C. 1970) (“[C]onstructive professional criticism cannot occur in an atmosphere of apprehension that one doctor’s suggestion will be used as a denunciation of a colleague’s conduct in a malpractice suit.”).

<sup>238</sup> FED. R. EVID. 407 (“Subsequent Remedial Measures”) provides:

When measures are taken that would have made an earlier injury or

The New Jersey version of this rule provides that “[e]vidence of remedial measures taken after an event is not admissible to prove that the event was caused by negligence or culpable conduct.”<sup>239</sup> The rationale for this rule “rests on a social policy of encouraging people to take, or at least not discouraging them from taking, steps in furtherance of added safety.”<sup>240</sup> Courts and commentators have recognized the parallels between the self-critical analysis privilege and the subsequent remedial measures doctrine.<sup>241</sup> For example, one court recognized that both rules protect parties from the

Hobson’s choice of aggressively investigating accidents . . . , ascertaining the causes and results, and correcting the violations or dangerous conditions, but thereby creating a self-incriminating record that may be evidence of liability, or deliberately avoiding making a record on the subject (and possibly leaving the public exposed to danger) in order to lessen the risk of civil liability.<sup>242</sup>

---

harm less likely to occur, evidence of the subsequent measures is not admissible to prove: negligence; culpable conduct; a defect in a product or its design; or a need for a warning or instruction. But the court may admit this evidence for another purpose, such as impeachment or—if disputed—proving ownership, control, or the feasibility of precautionary measures.

<sup>239</sup> N.J. R. EVID. 407.

<sup>240</sup> FED. R. EVID. 407 advisory committee’s note. As Baron Bramwell described it over a century ago, this rule rejects the idea that “because the world gets wiser as it gets older, therefore it was foolish before.” *Hart v. Lancashire & Yorkshire Ry.*, 21 L. Times Rep. (n.s.) 261, 263 (Eng. 1869).

<sup>241</sup> See, e.g., *Making Sense of Rules of Privilege*, *supra* note 46, at 1351–55 (advocating that federal courts evaluate self-critical analysis claims utilizing the “activity-privilege” inquiry of Rule 407 rather than the broader “relational” privilege rules under Rule 501); see also *Reid v. Lockheed Martin Aeronautics Co.*, 199 F.R.D. 379, 384 n.2 (N.D. Ga. 2001); *Capallupo v. FMC Corp.*, 46 Fair Empl. Prac. Cas. (BNA) 1193 (D. Minn. 1988) (noting that a defendant’s claim of self-critical analysis “is perhaps most closely related to the philosophy of Rule 407 of the Federal Rules of Evidence”).

<sup>242</sup> *Reichhold Chems., Inc. v. Textron, Inc.*, 157 F.R.D. 522, 524 (N.D. Fla. 1994).

Even in the oft-cited *Dowling* case, the court recognized that “the difference between pre-accident safety reviews and post-accident investigations is an important one.”<sup>243</sup>

To be clear, Rule 407 governs questions of admissibility—it is *not* a privilege—and therefore evidence of subsequent remedial measures is still discoverable.<sup>244</sup> Consequently, the preparer or creator of documents attesting to subsequent remedies should not have any reasonable expectation that the documents will remain confidential. Still, allowing for the discovery, but not the admissibility, of patient safety and peer review documents is worth consideration, particularly in New Jersey. First, the state’s Patient Bill of Rights empowers patients with a “right to know” about the treatment they received.<sup>245</sup> But on an even more basic level—and as the *Christy* court put it—“the search for truth is paramount in the litigation process.”<sup>246</sup> A patient safety protection structured around Rule 407, rather than an absolute privilege, could strike the right balance between an injured patient’s right to *information* and the hospital’s confidence that its own safety procedures will not expose them to *liability*.

The PSA deviated from this “right to know” principle. While it mandated facilities to inform patients of any adverse events, it simultaneously shielded important documents concerning these events. Proponents would argue that this rule is vital: in certain circumstances, overall improvements in patient safety rely upon the knowledge that certain materials will remain inaccessible to an individual patient. The argument is sensible,

---

<sup>243</sup> *Dowling v. Am. Haw. Cruises*, 971 F.2d 423, 427 (9th Cir. 1992).

<sup>244</sup> See 23 CHARLES A. WRIGHT & KENNETH W. GRAHAM, JR., FEDERAL PRACTICE AND PROCEDURE § 5291 (2012) (“Rule 407 is a rule of admissibility, not a privilege; hence, subsequent remedial measures are discoverable.”); see also Donald P. Vandegrift, Jr., *The Privilege of Self-Critical Analysis: A Survey of the Law*, 60 ALB. L. REV. 171, 189 (“Rule 407 is not a privilege rule.”).

<sup>245</sup> See N.J. STAT. ANN. § 26:2H-12.8(c) (West 2007). The interplay between this statute and the PSA, in this author’s opinion, is ripe for litigation and worth further judicial exploration.

<sup>246</sup> *Christy v. Salem*, 841 A.2d 937, 942 (N.J. Super. Ct. App. Div. 2004).

but, in this author's opinion, contrary to goals of patient safety. If patients have a right to know every detail of their treatment, shouldn't they also have a right to know what occurred *following* their treatment?

Consider the case of Esther and Gedalia Applegrad, who, aside from their "search for truth" as litigants simply seek answers as a mother and father as to why their child's delivery could have gone so horribly wrong.<sup>247</sup> Did the hospital evaluate the incident, and if so, how rigorously? Did it determine how the incident occurred? Was a particular medical staff member to blame for the incident, and if so, was the person sanctioned? Was he or she involved in prior adverse incidents? Did the staff member apologize or concede fault? And most importantly, what procedures has the facility instituted or considered instituting to ensure that similar errors do not occur in the future? Clearly, such evidence should be inadmissible at trial. Still, an injured patient deserves to have these questions answered—at least as a way to provide a measure of emotional closure and mental clarity. Rule 407, which would deem patient safety documents discoverable, yet inadmissible, seems to strike a fair and reasonable balance.

To conclude, it is worth returning to one of the Harvard Note's final thoughts:

A court applying the privilege of self-critical analysis should also remember that syllogistic application embodies the policy choice of the institution that decided to adopt the privilege. In adopting the privilege, a determination was made that the public interest weighed in favor of confidentiality. Whether this decision was made by a higher court, the same court at an earlier time, or a legislature, *judges should give due weight to*

---

<sup>247</sup> See Mary Pat Gallagher, *Patient Safety Act Privilege Held Permeable in Malpractice Suit*, N.J.L.J., Aug. 13, 2012, at 4 ("The Applegrads' lawyer, Cynthia Walters . . . says there was almost no contemporaneous record of what happened during the crucial 20-minute delay in resuscitating the baby or what happened with the intubation.").

*the binding effect of a prior determination that the privilege furthers the public interest.*<sup>248</sup>

The New Jersey Legislature, in passing the PSA, made clear its binding public policy determination that materials of self-critical analysis remain confidential. Courts will do their best to elucidate this mandate. Whether the PSA's "syllogistic application" of the privilege will actually promote the principal goal of the Act—to improve overall patient safety—remains to be seen.

---

<sup>248</sup> *The Privilege of Self-Critical Analysis*, *supra* note 44, at 1099 (emphasis added).



# IMPLIED PREEMPTION AND ITS EFFECT ON LOCAL HYDROFRACKING BANS IN NEW YORK

*David Giller\**

## INTRODUCTION

Depending on whom you ask, hydrofracking is either the future of American energy or an ecological disaster waiting to happen. Hydrofracking, otherwise known as “Fracking,”<sup>1</sup> is a drilling process where underground rock formations are broken apart to extract natural gas.<sup>2</sup> A number of environmental groups have questioned the safety of hydrofracking, alleging that it can damage the environment and that the resulting runoff wastewater can harm drinking water.<sup>3</sup> Currently, there is a moratorium on hydrofracking in New York State<sup>4</sup> until the Department of

---

\* J.D. Candidate, Brooklyn Law School, 2014; B.A., State University of New York at Geneseo, 2006. I would like to thank my friends and family for their unwavering support and encouragement, especially my parents and my sister for their insight and guidance. I also want to thank the entire staff of the *Journal of Law and Policy* for their diligence and help throughout the editing process as well as Professor Christopher Serkin for his invaluable assistance.

<sup>1</sup> See, e.g., Erica Levine Powers, *Home Rule Meets State Regulation: Reflection on High-Volume Hydraulic Fracturing for Natural Gas*, A.B.A. ST. & LOC. L. NEWS, Winter 2012, at 1, 1 (2012).

<sup>2</sup> See *Marcellus Shale*, N.Y. STATE DEP’T OF ENVTL. CONSERVATION, <http://www.dec.ny.gov/energy/46288.html> (last visited Mar. 19, 2013) (providing overview of hydrofracking in New York State).

<sup>3</sup> See, e.g., Ian Urbina, *Regulation Lax as Gas Wells’ Tainted Water Hits Rivers*, N.Y. TIMES (Feb. 2, 2011), <http://www.nytimes.com/2011/02/27/us/27gas.html>.

<sup>4</sup> See N.Y. Exec. Order No. 41: Requiring Further Environmental Review (Dec. 13, 2010), available at <http://www.governor.ny.gov/archive/>

Environmental Conservation (“DEC”) completes an environmental impact review and creates new regulations.<sup>5</sup> While the DEC continues its review, a number of local municipalities in New York have enacted their own legal barriers to hydrofracking.<sup>6</sup> These include both zoning bans on hydrofracking<sup>7</sup> and moratoria against hydrofracking.<sup>8</sup> While both zoning bans<sup>9</sup> and moratoria<sup>10</sup> have been challenged in court, this Note only addresses a town’s use of zoning power to ban hydrofracking.

The New York State legislature<sup>11</sup> has delegated to local municipalities the ability to “adopt, amend and repeal zoning regulations.”<sup>12</sup> Local municipalities can use such zoning regulations to advance the public welfare, a power that has been “broadly construed.”<sup>13</sup> However, when a municipality acts

---

paterson/executiveorders/EO41.html.

<sup>5</sup> See Mary Esch, *New York Fracking Decision: Cuomo Under Pressure to Rule on Hydraulic Fracturing*, HUFFINGTON POST (Sept. 6, 2012), [http://www.huffingtonpost.com/2012/09/06/new-york-fracking-decision\\_n\\_1862112.html](http://www.huffingtonpost.com/2012/09/06/new-york-fracking-decision_n_1862112.html).

<sup>6</sup> *Id.*

<sup>7</sup> Mary Esch, *Driller to NY: Stop the Local Fracking Bans or We’ll Sue*, PRESS CONNECTS (Aug. 1, 2012), <http://www.pressconnects.com/viewart/20120731/NEWS10/307310030/Driller-NY-Stop-local-fracking-bans-we-ll-sue>.

<sup>8</sup> See Steve Reilly, *Judge Overturns Binghamton Gas Drilling Moratorium*, PRESS CONNECTS (Oct. 3, 2012), <http://www.pressconnects.com/article/20121002/NEWS11/310020090/Judge-overturns-Binghamton-gas-drilling-moratorium> (reporting that Binghamton’s moratorium was struck down for not meeting the necessary legal requirements).

<sup>9</sup> See *id.*; see also Dan Wiessner, *New York Judge Upholds Fracking Ban in Towns*, REUTERS (Feb. 21, 2012), <http://www.reuters.com/article/2012/02/22/us-newyork-fracking-idUSTRE81L05820120222>.

<sup>10</sup> Reilly, *supra* note 8.

<sup>11</sup> The term legislature when used in the remainder of the Note will refer to the New York State legislature. A reference to a local government will be expressly indicated.

<sup>12</sup> N.Y. STAT. LOCAL GOV’TS § 10(6) (McKinney 1994).

<sup>13</sup> Andrew Meyer, “*Get the Frack Out of Town: Preemption Challenges to Local Fracking Bans in New York*,” 37 COLUM. J. ENVTL. L. FIELD REPORTS (Feb. 20, 2012), <http://www.columbiaenvironmentallaw.org/articles/get-the-frack-out-of-town-preemption-challenges-to-local-fracking->

outside of these delegated powers or “intrud[es] into an area of state authority,” such action will be considered preempted by state law either expressly or impliedly.<sup>14</sup> Express preemption exists when the state, through specific language in legislation, reserves power for itself, superseding local municipal control.<sup>15</sup> Implied preemption, on the other hand, occurs where legislation does not explicitly give the state control over a local issue but insinuates that such control was intended by legislature.<sup>16</sup> To find implied preemption, courts often examine “the nature of the subject matter regulated, the purpose and scope of the state legislative scheme, and the need for statewide uniformity.”<sup>17</sup> This usually involves examining the legislature’s intent at the time the law was created.<sup>18</sup> However, such inquiries are problematic because courts are often reluctant to judge legislative intent.<sup>19</sup>

New York case law is unclear regarding the criteria necessary for a finding of implied preemption. While the New York Court of Appeals has indicated that implied preemption can be inferred from state legislative policy or a comprehensive

---

bans-in-new-york.

<sup>14</sup> Shaun Goho, *Municipalities and Hydraulic Fracturing: Trends in State Preemption*, PLAN. & ENVTL. L., July 2012, at 3, 5 (2012); see also Michael E. Kenneally & Todd M. Mathes, *Natural Gas Production and Municipal Home Rule in New York*, N.Y. ZONING L. & PRAC. REP., Jan./Feb. 2010, at 1, 3 (2010).

<sup>15</sup> See Goho, *supra* note 14, at 5; see also N.Y. COMM’N ON LOCAL GOV’T EFFICIENCY & COMPETITIVENESS, STRENGTHENING HOME RULE (2008), [http://www.nyslocalgov.org/pdf/Strengthening\\_Home\\_Rule.pdf](http://www.nyslocalgov.org/pdf/Strengthening_Home_Rule.pdf).

<sup>16</sup> See Kenneally & Mathes, *supra* note 14; see also Paul Weiland, *Preemption of Local Efforts to Protect the Environment*, 18 VA. ENVTL. L.J. 467, 470 (1999).

<sup>17</sup> Kenneally & Mathes, *supra* note 14 (citing *Albany Area Builders Ass’n v. Town of Guilderland*, 546 N.E.2d 920 (N.Y. 1989)).

<sup>18</sup> See, e.g., Goho, *supra* note 14, at 5; Kenneally & Mathes, *supra* note 14, at 3; Weiland, *supra* note 16, at 470.

<sup>19</sup> See Kenneally & Mathes, *supra* note 14, at 5 (“[S]uch curtailment should only occur under a circumstance in which the legislature’s preemptive intent is absolutely clear.”); see also *Gernatt Asphalt Prods. v. Town of Sardinia*, 664 N.E.2d 1226, 1234–35 (N.Y. 1996).

and detailed regulatory scheme,<sup>20</sup> subsequent Court of Appeals decisions have retreated from such reasoning.<sup>21</sup> This appears to be particularly true when courts examine a town's use of zoning power.<sup>22</sup> For example, in two recent trial court decisions, the trial courts upheld the town's use of zoning power to ban hydrofracking.<sup>23</sup> As part of those decisions, the courts found that the towns were not impliedly preempted<sup>24</sup> under the Oil, Gas and Solution Mining Law ("OGSML").<sup>25</sup> These two decisions are the most recent illustrations of the current difficulty in showing implied preemption without an actual statement of intent by the legislature, especially with regard to zoning.

This Note will examine the intersection of implied preemption in New York with local zoning laws and the hesitancy of New York courts to find such implied preemption. Despite the existence of implied preemption as a doctrine in New York jurisprudence, courts are unlikely to find it in fact.

---

<sup>20</sup> See *Consol. Edison Co. of N.Y. v. Town of Red Hook*, 456 N.E.2d 487, 490 (N.Y. 1983) (holding that the local zoning laws could not prohibit a power plant because the legislature had pre-empted local regulation through its "comprehensive and detailed" regulatory scheme, Article VIII of the Public Service Law (now Article X of the Public Service Law)).

<sup>21</sup> See *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903 (N.Y. 1987) (holding that the county could enact a law prohibiting sale of cesspool additives without approval by Suffolk County Commissioner since the legislature did not show a desire to preclude local regulation and the local legislation had the same motive as state legislation, safe drinking water); see also *Vatore v. Comm'r of Consumer Affairs*, 634 N.E.2d 958 (N.Y. 1994) (holding that a state statute regulating cigarette vending machines did not implicitly preempt New York City from creating more restrictive regulations).

<sup>22</sup> See *Inc. Vill. of Nyack v. Daytop Vill.*, 583 N.E.2d 928 (N.Y. 1991) (holding that New York State Mental Hygiene Law did not implicitly preempt local zoning laws even though the state law included a detailed regulatory scheme).

<sup>23</sup> These cases are *Cooperstown Holstein Corp. v. Town of Middlefield*, 943 N.Y.S.2d 722 (Sup. Ct. 2012); *Anschutz Exploration Corp. v. Town of Dryden*, 940 N.Y.S.2d 458 (Sup. Ct. 2012). Both cases are being appealed to the Appellate Division and will be described in more detail later in the Note.

<sup>24</sup> *Cooperstown*, 943 N.Y.S.2d at 730; *Anschutz*, 940 N.Y.S.2d at 474.

<sup>25</sup> Oil, Gas and Solution Mining Law, N.Y. ENVTL. CONSERV. § 23-0303(2) (McKinney 2007).

Furthermore, because the incredibly high burden for finding implied preemption cannot be met in the current cases involving hydrofracking, the New York Court of Appeals<sup>26</sup> should uphold the hydrofracking bans as a proper use of zoning power. Additionally, the Court of Appeals should recognize the reality of implied preemption and its intersection with local zoning power. Namely, with regard to zoning, implied preemption should only be found when there is an explicit indication of legislative intent. A narrow approach to implied preemption with regard to zoning power is a better policy for New York because it eliminates the ambiguity of attempting to discern intent and forces the state legislature to consider the appropriate role of local zoning power.

Part I of this Note describes the process of hydrofracking and the current controversy surrounding its use in both New York and other states. Part II examines the history of zoning and preemption in New York State with an analysis of previous New York cases involving mining and hydrofracking. Part III focuses on the narrow interpretation of what constitutes implied preemption by the New York Court of Appeals and how such an interpretation requires the Court of Appeals to uphold local hydrofracking bans. Ultimately, a narrow view of implied preemption with regard to zoning is the best policy to control hydrofracking in New York State and to promote deliberation and accountability in the state legislature.

---

<sup>26</sup> At the time of publication, the Appellate Division has unanimously upheld the hydrofracking bans. *Norse Energy Corp. USA v. Town of Dryden*, No. 515227 (N.Y. App. Div. May 2, 2013). However, hydrofracking ban opponents have indicated that they intend to seek leave to appeal to the Court of Appeals. Adam Briggie, *Cities in New York Just Got a Big Stick in the Fracking Fight*, SLATE (May 3, 2013), [http://www.slate.com/blogs/future\\_tense/2013/05/03/norse\\_energy\\_corp\\_v\\_to\\_wn\\_of\\_dryden\\_court\\_upholds\\_new\\_york\\_town\\_s\\_fracking.html](http://www.slate.com/blogs/future_tense/2013/05/03/norse_energy_corp_v_to_wn_of_dryden_court_upholds_new_york_town_s_fracking.html).

## I. HYDROFRACKING: ECONOMIC BOON OR TICKING TIME BOMB?

*A. Hydrofracking Background*

Although hydrofracking has been used by the natural gas industry for the past fifty years, it has only recently become popular.<sup>27</sup> Its increased use is attributable to the growing desirability of natural gas for environmental and economic reasons, the discovery of large gas reserves within the United States, a desire to create homegrown energy opportunities, and new advancements in the process of hydrofracking.<sup>28</sup> While scientists have known for years that certain shale formations possessed high quantities of natural gas, it is recent technological advancements that have opened up these shale formations to drilling.<sup>29</sup> One such shale formation is the Marcellus Shale, which runs underground from Ohio through northeast Virginia into Pennsylvania and southern New York.<sup>30</sup> Although it is unclear how much natural gas is recoverable from the New York portion, some estimate as much as 489 trillion cubic feet (“TCF”) of natural gas exist throughout the entire shale.<sup>31</sup> To put this into perspective, the United States’ current annual rate of gas consumption is only 25.5 TCF.<sup>32</sup> Gas from shale production alone could provide for practically all domestic natural gas demand with surplus gas that could be exported.<sup>33</sup>

---

<sup>27</sup> Goho, *supra* note 14, at 3.

<sup>28</sup> *Id.*

<sup>29</sup> Marianne Lavelle, *Forcing Gas Out of Rock with Water*, NAT’L GEOGRAPHIC NEWS (Oct. 17, 2010), <http://news.nationalgeographic.com/news/2010/10/101022-energy-marcellus-shale-gas-science-technology-water/>.

<sup>30</sup> *Marcellus Shale*, *supra* note 2.

<sup>31</sup> *Id.*

<sup>32</sup> *Natural Gas Consumption by End Use*, U.S. ENERGY INFO. ADMIN., [http://tonto.eia.doe.gov/dnav/ng/ng\\_cons\\_sum\\_dcu\\_nus\\_a.htm](http://tonto.eia.doe.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm) (last visited Mar. 19, 2013).

<sup>33</sup> *What is Shale Gas and Why Is It Important?*, U.S. ENERGY INFO. ADMIN., [http://www.eia.gov/energy\\_in\\_brief/article/about\\_shale\\_gas.cfm](http://www.eia.gov/energy_in_brief/article/about_shale_gas.cfm) (last updated Dec. 5, 2012).

For years, scientists knew of the Marcellus Shale's potential but were unable to harness the natural gas that lay underneath.<sup>34</sup> However, that changed with new technological improvements in the process of hydrofracking.<sup>35</sup> In early 2003, a geologist working for a gas company in Pennsylvania learned of a new "fracking" process pioneered by oilmen in Texas.<sup>36</sup> It relied more on water, and, while originally developed to save money, it had the added benefit of being able to fracture shale more effectively.<sup>37</sup> Larger companies saw the advantage of this new hydrofracking technique and began to combine it with another method known as horizontal drilling.<sup>38</sup> In horizontal drilling, a well is drilled from the surface to just above the gas reservoir where it is "curve[d] to intersect the reservoir . . . with a near-horizontal inclination" maximizing the amount of natural gas available.<sup>39</sup> These advancements gave companies the ability to drill and extract natural gas from areas such as the Marcellus Shale, once considered unreachable.<sup>40</sup>

The process of hydrofracking consists of "pumping an engineered fluid system and a propping agent (proppant) such as sand"<sup>41</sup> along with other chemicals into a well to break up underground rock formations to allow for the easier extraction of natural gas.<sup>42</sup> The fluid involved in hydrofracking often contains compounds such as biocide<sup>43</sup> to prevent bacteria growth

---

<sup>34</sup> Lavelle, *supra* note 29.

<sup>35</sup> *Id.*

<sup>36</sup> *Id.*

<sup>37</sup> *Id.*

<sup>38</sup> *See id.*

<sup>39</sup> Lynn Helms, *Horizontal Drilling*, 35 DMR NEWSL., no. 1, at 1, 1 (Jan. 2008), available at <https://www.dmr.nd.gov/ndgs/newsletter/NL0308/pdfs/Horizontal.pdf>.

<sup>40</sup> *See Marcellus Shale*, *supra* note 2.

<sup>41</sup> N.Y. STATE DEP'T OF ENVTL. CONSERV., REVISED DRAFT SUPPLEMENTAL GENERIC ENVIRONMENTAL IMPACT STATEMENT ON THE OIL, GAS, AND SOLUTION MINING REGULATORY PROGRAM 1-1 (2011) [hereinafter RDSGIS], available at <http://www.dec.ny.gov/data/dmn/rdsgeisfull0911.pdf>.

<sup>42</sup> *See Marcellus Shale*, *supra* note 2.

<sup>43</sup> RDSGIS, *supra* note 41, at 5-50 tbl.5.6 (explaining that biocide is an additive that "[i]nhibits growth of organisms that could produce gases (particularly hydrogen sulfide) that could contaminate methane gas [and]

and other agents to ensure the proppant remains in the fracture of the shale instead of corroding the pipes carrying the water into the fracture.<sup>44</sup> Hydrofracking is also accompanied by a drilling rig<sup>45</sup> and requires the instillation of storage and processing centers nearby.<sup>46</sup> Once the hydrofracking process is completed, the remaining fluid, known as “flowback,” returns to the surface.<sup>47</sup> If the “flowback” is not reused, then it is considered “industrial wastewater” and must be disposed of in a concentrated and safe manner.<sup>48</sup>

Supporters and opponents of hydrofracking dispute whether the benefits outweigh the risks. One benefit of hydrofracking, its supporters argue, is increased revenue and jobs. Proponents point to Pennsylvania, where more than 5,000 hydrofracking wells have been created since 2005.<sup>49</sup> According to the Pennsylvania State Department of Labor and Industry, in 2010 almost 19,000 people were employed in the hydrofracking industry with another 140,000 working in related or supporting jobs.<sup>50</sup> Additionally, the Marcellus Shale Coalition estimates that hydrofracking generated \$11.2 billion in economic activity and \$1.1 billion in state and local tax revenue for Pennsylvania in 2010 alone.<sup>51</sup> An industry study indicated that this could be just the beginning and that gas companies could generate as much as

---

prevents the growth of bacteria which can reduce the ability of the fluid to carry proppant into the fracture”).

<sup>44</sup> See *Marcellus Shale*, *supra* note 2.

<sup>45</sup> RDSGIS, *supra* note 41, at 5-135-36 tbl.5.29 (explaining that the drilling rig consists of a drill pad, drill rig, drilling fluid and materials, road construction equipment and drilling equipment such as the casing and drill pipe).

<sup>46</sup> *Id.* at 5-80-82.

<sup>47</sup> *Id.* at 5-99-117.

<sup>48</sup> *Id.* at 5-130.

<sup>49</sup> Goho, *supra* note 14, at 5.

<sup>50</sup> Laura Legere, *Industry Study: Marcellus Economic Impact Dramatic*, CITIZENVOICE.COM (July 11, 2011), <http://citizensvoice.com/news/drilling/industry-study-marcellus-economic-impact-dramatic-1.1178179>.

<sup>51</sup> TIMOTHY J. CONSIDINE ET AL., *THE PENNSYLVANIA MARCELLUS NATURAL GAS INDUSTRY: STATUS, ECONOMIC IMPACTS AND FUTURE POTENTIAL* iv (2011), available at <http://marcelluscoalition.org/wp-content/uploads/2011/07/Final-2011-PA-Marcellus-Economic-Impacts.pdf>.



\$2.6 billion in additional state and local tax revenue in 2011 and 2012.<sup>52</sup> However, it is unclear how many of the new jobs being created are going to Pennsylvania residents.<sup>53</sup> In 2008, the Pennsylvania College of Technology indicated that between seventy to eighty percent of the actual drill workers were not from Pennsylvania.<sup>54</sup> Such reports have led to doubts about whether hydrofracking is actually an effective source of revenue or jobs.<sup>55</sup>

The economic benefit for New York in particular remains unclear. Some economists estimate that hydrofracking would bring over 17,000 new construction jobs and almost 30,000 indirect jobs to New York.<sup>56</sup> Furthermore, it is predicted that hydrofracking would cause New York's personal income tax revenue to increase anywhere from \$31 million to \$125 million a year.<sup>57</sup> Landowners willing to lease or sell their land would also benefit economically. In Pennsylvania, gas companies are paying over \$1,000 per acre, plus royalties, to landowners to drill on their land.<sup>58</sup> Both the jobs and the drilling leases would benefit some of the poorest areas of New York State where jobs have been hard to find.<sup>59</sup>

Opponents of hydrofracking challenge the reliability of reports promoting the economic benefits, the prospect of viable

---

<sup>52</sup> *Id.*

<sup>53</sup> See TOM WILBER, *UNDER THE SURFACE: FRACKING FORTUNES AND THE FATE OF THE MARCELLUS SHALE* 102 (2012) (noting the trend in out-of-state workers employed at Pennsylvania drilling rigs).

<sup>54</sup> *Id.*

<sup>55</sup> See Jannette M. Barth, *Hydrofracking Offers Short-Term Boom, Long-Term Bust*, ENR NEW YORK (Mar. 7, 2011), [http://newyork.construction.com/opinions/viewpoint/2011/0307\\_HydrofrackingOffers.asp](http://newyork.construction.com/opinions/viewpoint/2011/0307_HydrofrackingOffers.asp) (discussing studies that have found that any positive economic impact from hydrofracking only occurs in the short-term).

<sup>56</sup> Joan Gralla, *Economists Clash on Jobs Fracking Brings to NY*, REUTERS (Apr. 30, 2012), <http://www.reuters.com/article/2012/04/30/us-natgas-fracking-newyork-idUSBRE83T0EH20120430>.

<sup>57</sup> *Id.*

<sup>58</sup> Steven Kastenbaum, *Fracking in New York: Risk vs. Reward*, CNN (May 2, 2012, 2:39 PM), <http://www.cnn.com/2012/03/09/us/new-york-fracking/index.html>.

<sup>59</sup> *See id.*

long term growth from hydrofracking, and the danger posed to tourism and agriculture. Some academics and economists have disputed recent reports about the economic benefits of hydrofracking.<sup>60</sup> Specifically, the accuracy of a recent Pennsylvania State University study in favor of hydrofracking has been called into question by reports that its funding came from oil and gas companies.<sup>61</sup> Other experts and scholars dispute the number of jobs that would actually be created due to the “capital intensive” nature of hydrofracking.<sup>62</sup> There are also concerns over whether any job creation would be sustainable over the long term.<sup>63</sup> In addition, many landowners are nervous about hydrofracking’s effect on New York’s large agricultural<sup>64</sup> and wine businesses.<sup>65</sup> Damage to farmland could lead to an increase in milk prices.<sup>66</sup> Furthermore, increased ozone emissions from hydrofracking could negatively affect soy and grape production.<sup>67</sup> Vineyard owners, some of whom are on the northern fringe of the Marcellus Shale, are concerned about

---

<sup>60</sup> Jim Efstahiou Jr., *Penn State Faculty Snub of Fracking Study Ends Research*, BLOOMBERG (Oct. 3, 2012), <http://www.bloomberg.com/news/print/2012-10-03/penn-state-faculty-snub-of-fracking-study-ends-research.html> (reporting that a recent hydrofracking study at Pennsylvania State University study was canceled after criticism from faculty members that the report was biased in favor of the hydrofracking industry).

<sup>61</sup> *Id.*

<sup>62</sup> Carolyn Krupski, *Experts Debate Effects of Fracking on New York State Economy, Environment*, CORNELL DAILY SUN (Nov. 16, 2012), <http://cornellsun.com/node/54307> (noting that since hydrofracking is capital-intensive, jobs are often only associated with the construction of the wells, and once the wells are complete there is often less need for labor).

<sup>63</sup> *See id.*

<sup>64</sup> *See id.* (describing the danger posed to New York’s agricultural commodities from hydrofracking based on the effect of hydrofracking in Pennsylvania and possible increased ozone emissions).

<sup>65</sup> *See* Michael Hill, *Wine and Fracking Don’t Mix, Say Vineyard Owners*, ASSOCIATED PRESS (Oct. 23, 2012), <http://bigstory.ap.org/article/wine-and-fracking-dont-mix-say-vineyard-owners> (noting the fears of upstate N.Y. vineyard owners that hydrofracking will negatively impact their businesses).

<sup>66</sup> *See* Krupski, *supra* note 62 (noting the negative impact of hydrofracking on agriculture and milk prices in Pennsylvania).

<sup>67</sup> *Id.*

possible damage to their vineyard and the perception of damage by customers.<sup>68</sup> Additionally, any damage to New York's landscape from hydrofracking could negatively affect tourism, which in 2010 was a \$6.5 billion engine for New York State.<sup>69</sup>

There is also considerable fear that hydrofracking will cause serious environmental damage. This fear is shared by a diverse group of residents and environmentalists from all over New York State.<sup>70</sup> Opponents of hydrofracking point to the environmental issues currently facing Pennsylvania.<sup>71</sup> For example, there are reports in Pennsylvania that natural gas drillers are disposing of wastewater in rivers that supply drinking water.<sup>72</sup> Environmentalists are afraid that the chemicals used in creating the hydrofracking fluid and which are present in the wastewater could be dangerous if added to drinking water.<sup>73</sup> There is apprehension about the specific nature of the chemicals used in hydrofracking, since they are currently not disclosed to the public.<sup>74</sup>

---

<sup>68</sup> See Hill, *supra* note 65 (discussing the possible damage to vineyards from hydrofracking and the “public relations nightmare” of having hydrofracking near vineyards).

<sup>69</sup> Gralla, *supra* note 56 (noting that hydrofracking could lead to “unsightly rigs and possibly scarred landscapes”).

<sup>70</sup> See *Members, NEW YORKERS AGAINST FRACKING*, <http://nyagainstfracking.org/members/> (last visited Feb. 5 2013) (listing members of Advisory Committee); *Groups Rally to Prevent Fracking in NY*, CRAIN'S N.Y. BUS. (Jan. 11, 2013, 11:50 AM), <http://www.crainsnewyork.com/article/20130111/ECONOMY/130119976> (describing efforts of antifracking groups from “New York City to Buffalo”).

<sup>71</sup> See generally David B. Caruso, ‘Fracking’ Wastewater Still a Problem in Pennsylvania, NBC NEWS (Mar. 1, 2011, 6:35 PM), [http://www.msnbc.msn.com/id/41858136/ns/us\\_news-environment/t/fracking-wastewater-still-problem-pennsylvania/](http://www.msnbc.msn.com/id/41858136/ns/us_news-environment/t/fracking-wastewater-still-problem-pennsylvania/) (detailing the harm caused in Pennsylvania by hydrofracking wastewater); Urbina, *supra* note 3 (noting the environmental concerns surrounding Pennsylvania hydrofracking).

<sup>72</sup> Caruso, *supra* note 71.

<sup>73</sup> See Caruso, *supra* note 71; see also Urbina, *supra* note 3.

<sup>74</sup> See Kate Galbraith, *Seeking Disclosure on Fracking*, N.Y. TIMES (May 30, 2012), <http://www.nytimes.com/2012/05/31/business/energy-environment/seeking-disclosure-on-fracking.html> (pointing out that while individual states have different disclosure requirements they generally contain a “trade secrets” provision that prevents public disclosure of certain

Although natural gas executives often claim that hydrofracking is not responsible for contaminated underground drinking water,<sup>75</sup> recent reports have linked tainted water wells in Pennsylvania to hydrofracking from the Marcellus Shale.<sup>76</sup> These reports indicate that some of the tainted water contained high amounts of methane, double the Pennsylvania state safety level.<sup>77</sup> Methane is dangerous because while it does not affect the smell or taste of the water, it can render the water explosive.<sup>78</sup> Methane can also migrate from a faulty well to an enclosed area where it is difficult to notice.<sup>79</sup> Pennsylvania residents nearby hydrofracking operations have reported exploding wells and homes being destroyed from methane buildup.<sup>80</sup> Additionally, residents who live nearby such operations contend that their well water has become undrinkable.<sup>81</sup> Contaminated well water could result from hydrofracking itself,<sup>82</sup> “shoddy drilling practices, accidents and poor oversight,”<sup>83</sup> or natural migration.<sup>84</sup> Environmentalists in New York State echo the concerns of Pennsylvania residents.<sup>85</sup> New York environmentalists worry that

---

chemicals that fracking companies consider proprietary material).

<sup>75</sup> Ian Urbina, *A Tainted Water Well, and Concern There May Be More*, N.Y. TIMES (Aug. 3, 2011), <http://www.nytimes.com/2011/08/04/us/04natgas.html>.

<sup>76</sup> Mark Drakem & Jim Efstahiou Jr., *Cabot's Methodology Links Tainted Water Wells to Gas Fracking*, BLOOMBERG (Oct. 2, 2012 12:01 AM), <http://www.bloomberg.com/news/2012-10-02/cabot-s-methodology-links-tainted-water-wells-to-gas-fracking.html>.

<sup>77</sup> *Id.*

<sup>78</sup> *Id.*; see also Mark Drajem, *High Methane in Pennsylvania Water Deemed Safe by EPA*, BLOOMBERG (Mar. 30, 2012), <http://www.bloomberg.com/news/2012-03-29/high-methane-in-pennsylvania-water-deemed-safe-by-epa.html> (noting that high amounts of Methane in water can become explosive, even when the water itself is not unsafe to drink according to the EPA).

<sup>79</sup> WILBER, *supra* note 53, at 89–92.

<sup>80</sup> *Id.*

<sup>81</sup> *Id.* at 133–38.

<sup>82</sup> See Drakem & Efstahiou Jr., *supra* note 76.

<sup>83</sup> Kastenbaum, *supra* note 58.

<sup>84</sup> See *id.*

<sup>85</sup> *Id.*

hydrofracking could not only affect local landowners' drinking water but also New York City drinking water.<sup>86</sup> Hydrofracking could negatively affect the watersheds in the Catskills, an area that provides much of New York City's drinking water.<sup>87</sup>

The environmental dangers from hydrofracking combined with the economic potential have galvanized both supporters and detractors in New York State.<sup>88</sup> What was once an unremarked and unknown drilling technique has become a statewide issue.<sup>89</sup> A recent protest against hydrofracking had 3,000 individuals in attendance<sup>90</sup> and over 200,000 comments have been submitted to the DEC both in support and against hydrofracking.<sup>91</sup>

### *B. Fracking in New York State*

In December of 2010, Governor David Paterson introduced a moratorium on hydrofracking in New York State.<sup>92</sup> The moratorium will continue until the DEC completes an environmental review, including a public comment period,<sup>93</sup> and

---

<sup>86</sup> *Id.*

<sup>87</sup> *See id.*

<sup>88</sup> *See* Thomas Kaplan, *Millions Spent in Albany Fight to Drill for Gas*, N.Y. TIMES (Nov. 25, 2011), <http://www.nytimes.com/2011/11/26/nyregion/hydrofracking-debate-spurs-huge-spending-by-industry.html>; *see also* Alan Chartock, *Anti-Frackers Get Gov. Andrew Cuomo's Attention*, DAILY FREEMAN (Sept. 30, 2012), <http://www.dailyfreeman.com/articles/2012/09/30/opinion/doc506061d0150a2528440730.txt>.

<sup>89</sup> *See* Goho, *supra* note 14, at 3 ("Fracking is not a new process; it has been in use for more than 50 years. But the scale and scope have expanded significantly in the last decade."); *see also* *Celebrities Lead Crowd of 3,000 in Albany Protesting Hydraulic Fracking*, CBS NEW YORK (Jan. 23, 2013), <http://newyork.cbslocal.com/2013/01/23/celebrities-lead-crowd-of-3000-in-albany-protesting-hydraulic-fracturing/>.

<sup>90</sup> *Celebrities Lead Crowd of 3,000 in Albany Protesting Hydraulic Fracking*, *supra* note 89.

<sup>91</sup> *Groups Rally to Prevent Fracking in NY*, *supra* note 70; *New Yorkers Deliver Unprecedented 200k+ Comments on Cuomo's Fracking Rules*, ECOWATCH (Jan. 11, 2013), <http://ecowatch.com/2013/comments-ny-fracking-rules/>.

<sup>92</sup> *See* N.Y. Exec. Order No. 41, *supra* note 4.

<sup>93</sup> *Groups Rally to Prevent Fracking in NY*, *supra* note 70.

crafts regulations regarding hydrofracking.<sup>94</sup> The DEC introduced a preliminary impact statement in 2011, but it is unclear when a final plan will be complete.<sup>95</sup> During the first comment period, the DEC received over 65,000 comments on the proposed regulations,<sup>96</sup> a previously record-setting number.<sup>97</sup> The DEC then had until December of 2012 to incorporate those comments and complete its proposed hydrofracking regulations.<sup>98</sup> However, before the proposed regulations were due, the DEC directed the state Health Department to begin a health assessment of hydrofracking, delaying the final decision.<sup>99</sup> The DEC then filed for a ninety day extension by submitting a revised set of DEC regulations and opening up the process for another thirty days of comment ending January 11, 2013.<sup>100</sup> This recent comment period elicited an “unprecedented” number of comments, over 200,000.<sup>101</sup> The DEC missed their recent March deadline for promulgating hydrofracking regulations, and now any new regulations will be subject to another forty-five-day comment period and additional public hearings.<sup>102</sup>

---

<sup>94</sup> See Esch, *supra* note 5.

<sup>95</sup> Rick Karlin, *Is Trial Balloon Full of Shale Gas?*, ALBANY TIMES UNION (June 13, 2012), <http://www.timesunion.com/local/article/Is-trial-balloon-full-of-shale-gas-3632650.php>.

<sup>96</sup> Kastenbaum, *supra* note 58.

<sup>97</sup> *New Yorkers Deliver Unprecedented 200k+ Comments on Cuomo's Fracking Rules*, *supra* note 91.

<sup>98</sup> Jon Campbell, *New Hydrofracking Comment Period Begins Dec. 12*, ITHACA J. (Nov. 29, 2012, 9:56 PM), <http://www.theithacajournal.com/article/20121129/NEWS11/311290097/New-hydrofracking-comment-period-begins-Dec-12>.

<sup>99</sup> Jacob Gersman, *Pressure Rises for Decision on Drilling*, WALL ST. J. (Sept. 28, 2012), <http://online.wsj.com/article/SB10000872396390443389604578024851639481218.html>.

<sup>100</sup> Campbell, *supra* note 98.

<sup>101</sup> *Groups Rally to Prevent Fracking in NY*, *supra* note 70; *New Yorkers Deliver Unprecedented 200k+ Comments on Cuomo's Fracking Rules*, *supra* note 91.

<sup>102</sup> *Fracking Deadline Passes; Health Impact Study Sought*, STEUBENCOURIER.COM (Mar. 5, 2013, 10:00 AM), <http://www.steubencourier.com/news/x846064007/Fracking-deadline-passes-health-impact-study-sought>.

While the final plan is still being developed by the DEC, an unofficial report from the DEC's office indicated that hydrofracking would be limited to Chemung, Chenango, Steuben, Tioga and Broome counties.<sup>103</sup> Additionally, development would be limited to willing communities with an initial cap of fifty wells statewide.<sup>104</sup> The Governor neither confirmed nor denied the report.<sup>105</sup> However, the Governor did say that he believed that home rule should be taken into consideration.<sup>106</sup> Such reports have been described as a "trial balloon" to possibly appease both hydrofracking proponents and critics.<sup>107</sup>

Hydrofracking has both powerful supporters and opponents. Supporters of hydrofracking include some of the largest gas and energy companies. For instance, Exxon Mobile plans to invest \$185 billion over five years to develop new sources of oil and gas.<sup>108</sup> Pro-fracking advocates also employ an army of lobbyists and industry spokespeople with the goal of bringing hydrofracking to New York.<sup>109</sup> Those opposed to hydrofracking consist of grass roots activists, conservation groups and notable celebrities.<sup>110</sup> While hydrofracking opponents aim to protect the

---

<sup>103</sup> Danny Hakim, *Cuomo Proposal Would Restrict Gas Drilling to a Struggling Area*, N.Y. TIMES (June 13, 2012), <http://www.nytimes.com/2012/06/14/nyregion/hydrofracking-under-cuomo-plan-would-be-restricted-to-a-few-counties.html>

<sup>104</sup> *Id.*

<sup>105</sup> Karlin, *supra* note 95.

<sup>106</sup> Jorteh Senah & Karen DeWitt, *Cuomo Offers Some Support for Limited Fracking*, WNYC (July 10, 2012), <http://www.wnyc.org/articles/wnyc-news/2012/jul/10/calls-independent-review-fracking-state-set-unveil-plan-summer/>.

<sup>107</sup> Karlin, *supra* note 95; *see also* Senah & DeWitt, *supra* note 106.

<sup>108</sup> Brian O'Keefe, *Exxon's Big Bet on Shale Gas*, CNNMONEY (Apr. 16, 2012, 5:00 AM), <http://tech.fortune.cnn.com/2012/04/16/exxon-shale-gas-fracking/>.

<sup>109</sup> *See* Kaplan, *supra* note 88 (noting that in 2011 companies that drill for natural gas spent more than \$3.2 million lobbying the state).

<sup>110</sup> *See* Peter Applebome, *Drilling Critics Face a Divide Over the Goal of Their Fight*, N.Y. TIMES (Jan. 9, 2012), <http://www.nytimes.com/2012/01/10/nyregion/gas-drilling-critics-in-new-york-face-a-divide-over-their-goal.html>; Chartock, *supra* note 88.

environment, there is disagreement over the best way to do that, such as a statewide ban or stringent hydrofracking regulations.<sup>111</sup>

Hydrofracking has also become an important political issue. Elected officials from both parties and different levels of government have taken a position on hydrofracking.<sup>112</sup> In recent New York State elections, both local and federal candidates have focused on the role of hydrofracking.<sup>113</sup> These have been hard fought campaigns with resources and volunteers on both sides.<sup>114</sup> Although in the last few years antifracking activists have become more pronounced in New York State,<sup>115</sup> recent election results included notable victories for pro-fracking candidates.<sup>116</sup> One such victory was Debbie Preston's successful campaign for Broome County executive against an outspoken antifracking activist.<sup>117</sup>

In the meantime, towns have been taking their own steps, with some passing resolutions in favor of hydrofracking<sup>118</sup> and others amending their laws to ban hydrofracking within their borders.<sup>119</sup> Currently, over fifty towns have passed resolutions in favor of hydrofracking.<sup>120</sup> Those towns in favor are mostly

---

<sup>111</sup> See Applebome, *supra* note 110.

<sup>112</sup> See Steve Reilly, *Hydrofracking: It Has Polarized Voters and May Decide the Election*, PRESSCONNECTS.COM (Sept. 29, 2012, 9:06 PM), <http://www.pressconnects.com/article/20120929/NEWS11/309290033/Hydrofracking-has-polarized-voters-may-decide-election>.

<sup>113</sup> See *id.*

<sup>114</sup> *Id.*

<sup>115</sup> Ellen Cantarow, *The Secret War of Anti-Fracking Activists*, MOTHERJONES (Nov. 19, 2012), <http://www.motherjones.com/environment/2012/11/fracking-new-york-cuomo>.

<sup>116</sup> Mary Esch, *NY Anti-Fracking Candidates Fared Poorly at Polls*, BUS. WK. (Nov. 8, 2012), <http://www.businessweek.com/ap/2012-11-08/ny-anti-fracking-candidates-fared-poorly-at-polls>.

<sup>117</sup> *Id.*

<sup>118</sup> See Memorandum from the Joint Landowners Coal. of N.Y., Inc. to N.Y. Local Officials (June 28, 2012) *available at* <http://www.jlcny.org/site/attachments/article/1348/JLC%20-20Resolution%20Cover%20Memo.pdf>.

<sup>119</sup> See David Slottje & Helen Holden Slottje, *A Legal Plan to Control Drilling*, SIERRA ATLANTIC (Sierra Club Atl. Chapter, Albany, N.Y.), Spring 2011, *available at* [http://newyork.sierraclub.org/SA/Vol41/Legal\\_plan.htm](http://newyork.sierraclub.org/SA/Vol41/Legal_plan.htm).

<sup>120</sup> *Map of Town Resolutions in Support of Hydrofracking*, JOINT



located in the southern tier near the Pennsylvania border, the richest area of the Marcellus Shale.<sup>121</sup> Due to the state moratorium, there is currently no hydrofracking in New York;<sup>122</sup> therefore the pro-fracking resolutions have no legal authority. However, they are a symbolic indication of support for hydrofracking.<sup>123</sup> Sometimes the resolutions specify their support for the DEC to have the final say on hydrofracking, rather than local municipalities.<sup>124</sup> These resolutions are intended to combat local hydrofracking bans and illustrate that there is substantial support for bringing hydrofracking to New York.<sup>125</sup>

Municipalities who oppose hydrofracking have used a variety of legal tactics to ban hydrofracking either in part or entirely.<sup>126</sup> So far, over fifty upstate municipalities have used their zoning power to ban hydrofracking and over one hundred have enacted their own moratoria.<sup>127</sup> Most of the municipalities that have passed bans are in central and western New York.<sup>128</sup> These areas tend to possess less natural gas than those areas closer to Pennsylvania, leading some hydrofracking supporters to question their motives.<sup>129</sup> However, some of the hydrofracking bans are in areas along the natural gas rich area of the Marcellus Shale.<sup>130</sup>

---

LANDOWNERS COAL. OF N.Y., INC., <http://www.jlcnny.org/site/attachments/article/1349/JLCNY%20NYS%20Map%208.7.12.pdf> (last visited Dec. 5, 2012).

<sup>121</sup> *Id.*; see also Matt Richmond, *Resolutions Supporting DEC's Fracking Decision Spread*, INNOVATION TRAIL (July 13, 2012), <http://innovationtrail.org/post/resolutions-supporting-decs-fracking-decision-spread>.

<sup>122</sup> See N.Y. Exec. Order No. 41, *supra* note 4.

<sup>123</sup> See Joint Landowners Coal. of N.Y., Inc., *supra* note 118.

<sup>124</sup> See Richmond, *supra* note 121.

<sup>125</sup> See *id.*

<sup>126</sup> Goho, *supra* note 14, at 4.

<sup>127</sup> *Current High Volume Horizontal Hydraulic Fracturing Drilling Bans and Moratoria in NY State*, FRAC TRACKER (Mar. 16, 2013), <http://www.fractracker.org/maps/ny-moratoria/>.

<sup>128</sup> See Joseph de Avila, *Fracking' Goes Local*, WALL ST. J. (Aug. 29, 2012, 12:01 PM), <http://online.wsj.com/article/SB10000872396390444327204577617793552508470.html>; see also Joint Landowners Coal. of N.Y., Inc., *supra* note 118.

<sup>129</sup> See Richmond, *supra* note 121; see also de Avila, *supra* note 128.

<sup>130</sup> *Current High Volume Horizontal Hydraulic Fracturing Drilling Bans*

Two local hydrofracking bans have been challenged in court.<sup>131</sup> Both were upheld at the trial court level and both were heard on appeal before the Appellate Division, Third Department on March 21, 2013.<sup>132</sup> The Appellate Division unanimously upheld the hydrofracking bans as a proper use of town zoning power, although hydrofracking proponents have indicated that they plan to appeal.<sup>133</sup>

With the moratorium against hydrofracking still in place and an ever-changing deadline for the DEC,<sup>134</sup> passions run high for both supporters and opponents of hydrofracking. Their battle has taken place in the street,<sup>135</sup> over the airwaves<sup>136</sup> and at the ballot box.<sup>137</sup> Now with the advent of hydrofracking bans all over New York State, it appears that the courts are the next major battle ground.

### *C. Fracking Legal Regulatory Structure in Other States*

While hydrofracking is still in its infancy in New York, it has been employed for some time in a number of surrounding states with legal battles already underway.<sup>138</sup> Pennsylvania was

---

*and Moratoria in NY State, supra* note 127.

<sup>131</sup> Dan Wiessner, *New York Fracking Ban in Towns Upheld by Second Judge*, HUFFINGTON POST (Feb. 24 2012, 7:55 PM), [http://www.huffingtonpost.com/2012/02/25/new-york-fracking-ban\\_n\\_1300600.html](http://www.huffingtonpost.com/2012/02/25/new-york-fracking-ban_n_1300600.html).

<sup>132</sup> Jon Campbell, *Appeals Heard on Dryden, Middlefield Fracking Bans*, ITHACAJOURNAL.COM (Mar. 21, 2013, 7:24 PM), <http://www.theithacajournal.com/article/20130321/NEWS01/303210042/Appeals-heard-Dryden-Middlefield-fracking-bans>.

<sup>133</sup> Jon Campbell, *Appeals Court: N.Y. Towns Can Ban Hydrofracking*, LOHUD.COM (May 22, 2013), <http://www.lohud.com/article/20130502/NEWS/305020063/Appeals-court-N-Y-towns-can-ban-hydrofracking>.

<sup>134</sup> See Esch, *supra* note 5.

<sup>135</sup> *Celebrities Lead Crowd of 3,000 in Albany Protesting Hydraulic Fracking*, *supra* note 89.

<sup>136</sup> See Kaplan, *supra* note 88.

<sup>137</sup> Reilly, *supra* note 112.

<sup>138</sup> See generally Francis Grandijan, *State Regulations, Litigation, and Hydraulic Fracturing*, 7 ENVTL. & ENERGY L. & POL'Y J. 47 (2012) (detailing the regulatory structure and history of hydraulic fracturing); see also Goho, *supra* note 14, at 6.

one of the first states to be part of the gas rush with companies leasing land from landowners for hydrofracking as early as 2007.<sup>139</sup> From 2008 to 2010 the number of permit applications increased from 478 to 3,314.<sup>140</sup> The permit application is supposed to involve a detailed evaluation of water intake and the process for discharging wastewater for that specific drilling site.<sup>141</sup> However, due to the overwhelming number of permits, Pennsylvania Department of Environmental Protection (“DEP”) officials have not been able to properly screen them.<sup>142</sup> This has led to an approval rate of over 99.5%.<sup>143</sup> While Pennsylvania does have general legislation to protect water supplies,<sup>144</sup> many citizens are concerned that there is no appropriate oversight of the hydrofracking industry.<sup>145</sup> Reports of exploding wells, contaminated groundwater, and destruction of nearby property have only increased those fears.<sup>146</sup>

Concerns with the state regulatory process have led a number of Pennsylvania towns to enact their own laws controlling where hydrofracking may take place.<sup>147</sup> In 2009, the Pennsylvania Supreme Court ruled that local municipalities have the ability to “control the location of wells consistent with established zoning principles.”<sup>148</sup> Such authority was pursuant to the Pennsylvania Oil and Gas Act which expressly preempted any laws regarding the specific operation of hydrofracking.<sup>149</sup> The Pennsylvania Oil and Gas Act did, however, allow

---

<sup>139</sup> See WILBER, *supra* note 53, at 17.

<sup>140</sup> *Id.* at 80.

<sup>141</sup> Grandijan, *supra* note 138, at 74.

<sup>142</sup> See WILBER, *supra* note 53, at 81.

<sup>143</sup> *Id.*

<sup>144</sup> 58 PA. CONS. STAT. § 3218 (2012) (mandating protection of well water and holding companies liable for replacing any water that they damage).

<sup>145</sup> See WILBER, *supra* note 53, at 80–82.

<sup>146</sup> See *id.* at 89–142.

<sup>147</sup> See Goho, *supra* note 14, at 6.

<sup>148</sup> Range Res.-Appalachia LLC v. Salem Twp., 964 A.2d 869, 876 (Pa. 2009).

<sup>149</sup> 58 PA. CONS. STAT. § 3302 (2012).

municipalities to ban drilling in residential areas.<sup>150</sup> Some municipalities in Pennsylvania though have gone further and banned hydrofracking entirely.<sup>151</sup> While Pennsylvania courts have ruled that towns can control the location of hydrofracking drilling sites, the legality of zoning bans under the Pennsylvania Oil and Gas Act are uncertain.<sup>152</sup>

In response to the court's support of local zoning power to control the location of hydrofracking sites, the Pennsylvania legislature enacted Act 13, amending the Oil and Gas Act, to allow hydrofracking in all zoning districts, even residential ones.<sup>153</sup> Act 13 also invalidated all existing ordinances involving hydrofracking.<sup>154</sup> However, a Pennsylvania Appellate Court recently struck down Act 13.<sup>155</sup> The court ruled that its provisions were unconstitutional in that they took too much power from local government to regulate their own communities.<sup>156</sup> That ruling is being appealed to the Pennsylvania Supreme Court.<sup>157</sup> In addition, the Public Utility Commission determined that Pittsburgh's ban on hydrofracking was not allowed under state law.<sup>158</sup> However, this is only a

---

<sup>150</sup> *Id.*

<sup>151</sup> See, e.g., *Pittsburgh Bans Hydrofracking*, BAY RIDGE JOURNAL (Dec. 1, 2010), <http://www.bayridgejournal.blogspot.com/2010/12/pittsburgh-bans-hydrofracking.html>; *Hydrofracking Bans and Moratoria*, SIERRA CLUB ACTIVIST NETWORK (May 26, 2011), <http://bit.ly/14fIwV9> (listing bans and moratoria in Pennsylvania).

<sup>152</sup> See Goho, *supra* note 14, at 6.

<sup>153</sup> *Id.*

<sup>154</sup> *Id.*

<sup>155</sup> Editorial, *Towns Win a Round in Court Against Unfettered Fracking*, PHILLY.COM (Aug. 1, 2012), [http://articles.philly.com/2012-08-01/news/32962407\\_1\\_buildings-and-water-wells-impact-fees-environmental-hazards](http://articles.philly.com/2012-08-01/news/32962407_1_buildings-and-water-wells-impact-fees-environmental-hazards).

<sup>156</sup> Marc Levy, *Pennsylvania Act 13 Provisions Struck Down by Appellate Court*, HUFFINGTON POST (July 26, 2012, 2:19 PM), [http://www.huffingtonpost.com/2012/07/26/pennsylvania-act-13-natural-gas\\_n\\_1706822.html](http://www.huffingtonpost.com/2012/07/26/pennsylvania-act-13-natural-gas_n_1706822.html).

<sup>157</sup> Scott Detrow, *Corbett Administration Files Act 13 Appeal With State Supreme Court*, STATEIMPACT (July 27, 2012, 4:53 PM), <http://stateimpact.npr.org/pennsylvania/2012/07/27/corbett-administration-files-act-13-appeal-with-state-supreme-court/>.

<sup>158</sup> Laura Olson & Joe Smydo, *PUC Says Pittsburgh's Ban on Natural*

recommendation and relies on Act 13.<sup>159</sup> It is unclear whether Pittsburgh will revise its hydrofracking ban.<sup>160</sup> As challenges to Act 13 continue to move through Pennsylvania courts, it remains unsettled whether towns in Pennsylvania will ultimately be able to control the location of hydrofracking through their zoning power.

West Virginia was also confronted with the issue of preemption with regard to hydrofracking when a number of its local municipalities passed zoning laws banning hydrofracking.<sup>161</sup> However, West Virginia's Monongalia County Circuit Court struck down a ban passed by Morgantown<sup>162</sup> that prohibited "[d]rilling a well for the purpose of extracting or storing oil or gas using horizontal drilling with fracturing or fracking methods."<sup>163</sup> The court ruled that the West Virginia Oil and Gas Act<sup>164</sup> fully "occupied the field," rendering the local ban invalid.<sup>165</sup> The court further found that the Oil and Gas Act indicated an intention for regulatory authority to be at the state

---

*Gas Extraction Conflicts with State Law*, PITTSBURGH POST-GAZETTE (Sept. 11, 2012), <http://www.post-gazette.com/stories/local/neighborhoods-city/puc-says-pittsburghs-ban-on-natural-gas-extraction-conflicts-with-state-law-652858/>.

<sup>159</sup> *Id.*

<sup>160</sup> Levy, *supra* note 156; *see also* Abby W. Schachter, *Pittsburgh Rethinks Fracking Ban*, N.Y. POST (Sept. 20, 2012), [http://www.nypost.com/p/blogs/capitol/pittsburgh\\_rethinks\\_fracking\\_ban\\_QoyPPTO8iYQNEs5BTQpt eO](http://www.nypost.com/p/blogs/capitol/pittsburgh_rethinks_fracking_ban_QoyPPTO8iYQNEs5BTQpt eO) (reporting that a Pittsburgh councilman has "proposed legislation to eliminate the current ban and replace it instead with strict zoning regulations for gas extraction").

<sup>161</sup> Goho, *supra* note 14, at 6.

<sup>162</sup> Adam Orford, *Local Bans on Hydraulic Fracturing Upheld in New York State, Struck Down in West Virginia*, MARTEN L. NEWS (Apr. 10, 2012), <http://www.martenlaw.com/newsletter/20120410-local-hydraulic-fracturing-bans>.

<sup>163</sup> Goho, *supra* note 14, at 6. (quoting An Ordinance Repealing Article 721 of the City of Morgantown's Business and Taxation Code (Morgantown, W. Va. 2011), *available at* [http://documents.foodandwaterwatch.org/doc/Frack\\_Actions\\_MorgantownWV-ban.pdf](http://documents.foodandwaterwatch.org/doc/Frack_Actions_MorgantownWV-ban.pdf)).

<sup>164</sup> West Virginia Oil and Gas Act, W. VA. CODE § 22-6-1 to -41 (West 2011).

<sup>165</sup> Orford, *supra* note 162.

level.<sup>166</sup> The court discerned such an intention by looking to the language and rules promulgated by the West Virginia DEP which gave the state ultimate responsibility for protecting the environment and indicated a “comprehensive framework.”<sup>167</sup> Additionally, the court held that West Virginia’s municipality’s powers are “narrowly proscribed” and that if there is a question as to whether a municipality has certain legislative power, the court should find that the municipality does not possess such power.<sup>168</sup> Morgantown did not appeal and other municipalities have since repealed their hydrofracking bans.<sup>169</sup> Recently, Morgantown considered limited zoning laws, controlling the location of hydrofracking rather than an outright ban, although it is unclear if even such a limited ban would be allowed.<sup>170</sup> Until appellate courts in West Virginia address the level of power local municipalities possess through their zoning power, it seems unlikely that any type of hydrofracking ban will be allowed.

The states surrounding New York, where hydrofracking already exists, have all taken different approaches to local zoning power and hydrofracking bans. Generally the courts and legislature have been more restrictive of local power with greater control given to the state.<sup>171</sup> However, the law in both West Virginia and Pennsylvania is still unsettled, with the validity of Act 13 pending before the Pennsylvania Supreme Court<sup>172</sup> and the West Virginia bans only being struck down at the trial level.<sup>173</sup>

---

<sup>166</sup> *Id.*

<sup>167</sup> *Id.*

<sup>168</sup> *Id.*

<sup>169</sup> Goho, *supra* note 14, at 6–7.

<sup>170</sup> *See* Goho, *supra* note 14, at 6–7; Orford, *supra* note 162.

<sup>171</sup> Goho, *supra* note 14, at 5–7.

<sup>172</sup> Detrow, *supra* note 157.

<sup>173</sup> Orford, *supra* note 162.

## II. ZONING AND PREEMPTION IN NEW YORK STATE

A. *History of Local Government and Zoning*

New York State consists of a myriad of different levels of local government, some existing for hundreds of years and tracing their existence to the establishment of the New York State Constitution in 1777.<sup>174</sup> The different levels of local government include county, city, town, and village governments.<sup>175</sup> The New York Constitution only confers legislative power to the New York State legislature as opposed to individual municipalities.<sup>176</sup> This gives the state the authority to “enact laws which regulate, prohibit, or require certain conduct, provided that such laws have some reasonable relation to the public health, safety, morals or welfare.”<sup>177</sup> Such broad power gives state legislatures the initial authority to impose land use restrictions.<sup>178</sup> While there are some statewide land use ordinances, such as fire laws, land use regulation is often left to local municipalities.<sup>179</sup> The rationale, as expressed by the Court of Appeals, is that towns are in the best position to evaluate community needs and use their zoning power accordingly.<sup>180</sup>

---

<sup>174</sup> See N.Y. DEP’T OF STATE, LOCAL GOVERNMENT HANDBOOK 59 (2011) [hereinafter LOCAL GOVERNMENT HANDBOOK], available at [http://www.dos.ny.gov/lg/publications/Local\\_Government\\_Handbook.pdf](http://www.dos.ny.gov/lg/publications/Local_Government_Handbook.pdf).

<sup>175</sup> *Id.* at 29.

<sup>176</sup> N.Y. CONST. art. III, § 1 (“The legislative power of this state shall be vested in the senate and assembly.”).

<sup>177</sup> PATRICIA SALKIN, NEW YORK ZONING LAW AND PRACTICE § 2:01 (4th ed. 2012).

<sup>178</sup> *Id.* (noting that a state legislature’s power to regulate land use is derived from the state’s general police power).

<sup>179</sup> *Id.*

<sup>180</sup> See *id.* § 2:01 n.3 (“A zoning resolution in many of its features is distinctively a city affair, a concern of the locality, affecting as it does the density of population, the growth of city life, and the court of city values.” (quoting *Adler v. Deegan*, 167 N.E. 705, 711 (N.Y. 1929) (Cardozo, C.J., concurring))); see also LOCAL GOVERNMENT HANDBOOK, *supra* note 174, at 147–56.

In New York, local governments<sup>181</sup> do not have any inherent law making authority; instead, such authority comes from state legislation and Article IX of the New York State Constitution.<sup>182</sup> Article IX, often referred to as the “Home Rule” article,<sup>183</sup> delegates both broad and limited powers to local government.<sup>184</sup> This includes the power to create laws that relate to the municipality’s “property, affairs or government.”<sup>185</sup> However, the ability of local governments to exercise zoning authority is not explicit in the New York Constitution.<sup>186</sup> Instead courts have held that such zoning power comes from enabling statutes such as the Statute of Local Governments and the Municipal Home Rule Law.<sup>187</sup> The Statute of Local Governments includes the power for cities, villages, and towns to “adopt, amend and repeal zoning regulations”<sup>188</sup> but allows for restriction by the state legislature.<sup>189</sup> Counties are excluded and do not have the power to enact zoning regulations.<sup>190</sup> The Municipal Home Rule Law, enacted by the Legislature, allows local governments to “have the power to adopt and amend local laws where and to the extent that its legislative body has the power to act by ordinance, resolution, rule or regulation.”<sup>191</sup> This allows for local governments to enact ordinances or zoning laws within the

---

<sup>181</sup> Local government is defined as “a county, city, town or village.” N.Y. CONST. art. IX, § 3(d)(2).

<sup>182</sup> N.Y. CONST. art. IX (defining the powers and rights of local governments).

<sup>183</sup> LOCAL GOVERNMENT HANDBOOK, *supra* note 174, at 30.

<sup>184</sup> *See id.* at 30–34.

<sup>185</sup> N.Y. CONST. art IX, § 2(c).

<sup>186</sup> *See SALKIN, supra* note 177, § 2:03.

<sup>187</sup> *See id.* §§ 2:03–04 (stating that although delegated to local government these powers are “quasi-constitutional” and can only be changed through legislation action at regular session in two calendar years).

<sup>188</sup> N.Y. STAT. LOCAL GOV'TS § 10(6) (McKinney 1994).

<sup>189</sup> *Id.* § 10 (“Grant[s of power] . . . to local governments . . . shall at all times be subject to such purposes, standards and procedures as the legislature may have heretofore prescribed or may hereafter prescribe.”).

<sup>190</sup> *See SALKIN, supra* note 177, § 2:09.

<sup>191</sup> *Id.* § 2:05 (citing N.Y. MUN. HOME RULE LAW § 10 (McKinney 1994)).



purview of their legislative power.<sup>192</sup> Though cities, towns, and villages all have similar zoning authority,<sup>193</sup> this Note will focus on the zoning power of towns.

### *B. Zoning and Preemption*

Local governments can use their police power to create laws for the “protection, order, conduct, safety, health and well-being of persons or property.”<sup>194</sup> Such police power also includes advancing the general welfare.<sup>195</sup> Under both the Statute of Local Governments and the Municipal Home Rule Law, local governments can zone under their police power.<sup>196</sup> Local government’s police power covers a broad array of activities from aesthetic concerns to preserving the character of the community.<sup>197</sup> While the zoning power of local governments is quite broad, courts have limited their authority in some areas.<sup>198</sup> For instance, the Court of Appeals in New York has generally held that local governments cannot use their zoning power to create regulations that have the effect of excluding minorities or the poor.<sup>199</sup> Another common area of contention is whether

---

<sup>192</sup> Zoning ordinances and zoning laws are interchangeable and this Note will refer to both as zoning laws. There are some procedural differences between enacting a zoning ordinance or zoning law but they are not relevant for a discussion of preemption. *See* SALKIN, *supra* note 177, §§ 3:01–03, 3:13–40; *see also* *Anschutz Exploration Corp. v. Town of Dryden*, 940 N.Y.S.2d 458, 467–68 (Sup. Ct. 2012) (citing *Gernatt Asphalt Prods. v. Town of Sardinia*, 664 N.E.2d 1226, 1234–35 (N.Y. 1996) (referring to zoning ordinances as land use laws)).

<sup>193</sup> *See* SALKIN, *supra* note 177, §§ 2:06–08 (stating that villages and towns have similar zoning authority since all of their authority comes through the Municipal Home Rule Law).

<sup>194</sup> N.Y. CONST. art IX, § 2(c)(10); *see also* MUN. HOME RULE § 10.

<sup>195</sup> SALKIN, *supra* note 177, § 6:01.

<sup>196</sup> *See* N.Y. STAT. LOCAL GOV’TS § 10(6) (McKinney 1994); MUN. HOME RULE § 10.

<sup>197</sup> *See* SALKIN, *supra* note 177, §§ 6:01–25.

<sup>198</sup> *See id.* §§ 6:02–03.

<sup>199</sup> *See id.* § 20:11 (citing *Asian Am. for Equal. v. Koch*, 527 N.E.2d 265 (N.Y. 1988)) (“The enabling acts of cities, towns and villages in New York do not authorize zoning to exclude from the enacting municipality

zoning rules can be used to ban or regulate specific uses of the land.<sup>200</sup> These disputes often involve an analysis of the extent of a town's police power and what constitutes the general welfare of a town.<sup>201</sup>

The legislature retains the ability to impose restrictions on local zoning power.<sup>202</sup> One such restriction is that zoning regulations must be part of a comprehensive plan.<sup>203</sup> Another is that they cannot be part of "spot zoning," singling out a small piece of land for a different use for the exclusive "benefit of the owner of such property and to the detriment of other owners."<sup>204</sup> This is to ensure that zoning is used to build a better community and is a "means rather than [an] end."<sup>205</sup>

State law may preempt local zoning power either expressly or impliedly.<sup>206</sup> With "express preemption," the state explicitly prevents local municipalities from addressing an issue.<sup>207</sup> Express preemption is found in the statutory text itself and clearly illustrates that the state and not a local town is responsible for handling a specific issue.<sup>208</sup> When there is "implied preemption," the legislature has evidenced an intent to supersede a local municipality in a particular area.<sup>209</sup> Implied preemption generally

---

persons of low or moderate income, and if the party attacking the ordinance establishes that it has either of an exclusory purpose or effect, the ordinance will be annulled.").

<sup>200</sup> See *id.* §§ 11:01–38.

<sup>201</sup> See *id.* § 6:01.

<sup>202</sup> See *id.* § 4:02; see also Goho, *supra* note 14, at 5.

<sup>203</sup> See SALKIN, *supra* note 177, § 4:03.

<sup>204</sup> *Id.* § 4:10 (quoting *Rodgers v. Vill. of Tarrytown*, 96 N.E.2d 731, 734 (N.Y. 1951)).

<sup>205</sup> See *id.* § 4:03 (citing *Asian Am. for Equal. v. Koch*, 527 N.E.2d 265 (N.Y. 1988)).

<sup>206</sup> See *id.* § 4:22; see also Weiland, *supra* note 16, at 470; Kenneally & Mathes, *supra* note 14.

<sup>207</sup> See, e.g., N.Y. COMM'N ON LOCAL GOV'T EFFICIENCY & COMPETITIVENESS, *supra* note 15; Goho, *supra* note 14, at 5; Weiland, *supra* note 16, at 470.

<sup>208</sup> See Weiland, *supra* note 16, at 470; Goho, *supra* note 14, at 5; see also N.Y. COMM'N ON LOCAL GOV'T EFFICIENCY & COMPETITIVENESS, *supra* note 15.

<sup>209</sup> Weiland, *supra* note 16, at 470–71.

appears in two forms. One form is “conflict preemption,” where the local law is “found to conflict with or frustrate the purpose” of the state law.<sup>210</sup> The other is “field preemption,” which occurs if state law concerning a particular issue is so broad that it “occupies the field,” leaving no ability for local discretion<sup>211</sup> or creates a “comprehensive and detailed regulatory scheme in a particular area.”<sup>212</sup>

Conflicts often arise in determining whether there is implied preemption. Unlike express preemption, which is often easily resolved based on the plain meaning of the statute,<sup>213</sup> implied preemption is more difficult to discern.<sup>214</sup> The courts often examine “the nature of the subject matter regulated, the purpose and scope of the state legislative scheme, and the need for statewide uniformity.”<sup>215</sup> Additionally, a local law is not preempted simply because it prohibits an activity that is allowed under state law.<sup>216</sup> If this were the case, the power of local governments would be “illusory.”<sup>217</sup> Furthermore, implied preemption does not require an express statement by the legislature.<sup>218</sup> Instead the court tries to discern legislative

---

<sup>210</sup> *Id.*

<sup>211</sup> *Id.*

<sup>212</sup> *Consol. Edison Co. of N.Y. v. Town of Red Hook*, 456 N.E.2d 487, 490 (N.Y. 1983).

<sup>213</sup> *See, e.g., Inc. Vill. of Lloyd Harbor v. Town of Huntington*, 149 N.E.2d 851, 854 (N.Y. 1958) (holding that a local village cannot zone out a park that a state law specifically authorizes).

<sup>214</sup> *See Kenneally & Mathes, supra* note 14, at 3.

<sup>215</sup> *Id.*

<sup>216</sup> *See, e.g., N.Y. State Club Ass’n v. City of New York*, 505 N.E.2d 915, 919–20 (N.Y. 1987) (holding that the city was not preempted, either expressly or implicitly, by the New York State Human Rights Law when it prohibited discrimination in clubs even though the city was banning an activity allowed under state law).

<sup>217</sup> *Id.* at 920.

<sup>218</sup> *See Consol. Edison Co.*, 456 N.E.2d at 489 (holding that Red Hook’s Local Law 2, which required a license for power plants that the town could deny due to zoning rules, was invalid because it was preempted by Article VIII). The Legislature made it clear that the purpose of Article VIII was to expedite the process and create a “unified procedure.” *Id.* Additionally, article VIII had a detailed regulatory scheme, which the court said was

intent.<sup>219</sup> Courts judge legislative intent by investigating the state's public policy, the language of the statute, and whether state law has created a "comprehensive and detailed regulatory scheme."<sup>220</sup> Issues commonly arise as to what type of statement by the legislature or what level of detail in a regulatory scheme is needed to show intent.<sup>221</sup> Resolving those issues often requires a fact intensive search into the statute itself or the legislative purpose and history.<sup>222</sup>

*C. Mining in New York—The Precursor to the Hydrofracking Debate*

The Court of Appeals has never addressed the issue of whether a town can use its zoning power to ban hydrofracking. However, the Court of Appeals has addressed the extent to which towns can use their zoning power to control and ban mining.<sup>223</sup> The issue in mining, similar to that of hydrofracking, is whether local zoning power is preempted by a state statute regulating that industry. In mining, the focus was on the Mined Land Reclamation Act ("MLRA"),<sup>224</sup> which bears many similarities to the OGSML.<sup>225</sup> The Court of Appeals addressed this issue in *Frew Run Gravel Products, Inc. v. Town of Carroll*<sup>226</sup> and *Gernatt Asphalt Products, Inc. v. Town of*

---

evidence of the legislature's intent to preempt. *Id.*

<sup>219</sup> See, e.g., *id.*

<sup>220</sup> See *id.*; see also *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903, 904–05 (N.Y. 1987) (upholding local law because there was no indication that state law preempted the local regulatory scheme).

<sup>221</sup> See *Jancyn Mfg. Corp.*, 518 N.E.2d at 907; see also *N.Y. State Club Ass'n*, 505 N.E.2d at 917.

<sup>222</sup> See *Consol. Edison Co.*, 456 N.E.2d at 490 (looking at the statute to discern intent); see also *Jancyn*, 518 N.E.2d at 906 (looking at the purpose of the statute, here to protect the environment).

<sup>223</sup> See, e.g., *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920 (N.Y. 1987); see also *Gernatt Asphalt Prods., Inc. v. Town of Sardinia*, 664 N.E.2d 1226 (N.Y. 1996).

<sup>224</sup> Mined Land Reclamation Act, N.Y. ENVTL. CONSERV. § 27-0030 (McKinney 2007).

<sup>225</sup> See SALKIN, *supra* note 177, § 11:23.50.

<sup>226</sup> *Frew Run*, 518 N.E.2d at 921.

*Sardinia*.<sup>227</sup> These cases dealt specifically with whether a town could use its zoning power to limit where mining could occur and if a town could ban mining entirely.<sup>228</sup>

In *Frew Run*, the Court of Appeals held that a town was permitted to use its zoning power to regulate the location of a mine. In that case, the town of Carroll passed a zoning ordinance that regulated the placement of mines within the town, although the ordinance did not ban them entirely.<sup>229</sup> A mining company sued the town claiming that the town's zoning powers were preempted by a state statute, the MLRL.<sup>230</sup> The court reviewed the text of the statute and found that the town's zoning regulations were not superseded by the MLRL because the zoning regulations did not "relat[e] to the extractive mining industry."<sup>231</sup> Local laws would be superseded only if they detailed the specific operations and practice of how the mining could occur.<sup>232</sup> Towns had the power to regulate the land itself and thereby could control the locations of the mines.<sup>233</sup> Additionally, the court held that there was no evidence of intention by the legislature to preempt local zoning power.<sup>234</sup> The legislature's intent, concern for the environment, was consistent with the aim of the zoning ordinances.<sup>235</sup>

In *Gernatt*, the Court of Appeals affirmed a town's use of its zoning power to ban mining entirely. In this case, the town of Sardinia passed a zoning law which banned the construction of any new mines in town.<sup>236</sup> The law did not affect previously constructed mines.<sup>237</sup> The town claimed this was an extension of

---

<sup>227</sup> *Gernatt*, 664 N.E.2d at 1234.

<sup>228</sup> See *Frew Run*, 518 N.E.2d at 921; see also *Gernatt*, 664 N.E.2d at 1230.

<sup>229</sup> See *Frew Run*, 518 N.E.2d at 921.

<sup>230</sup> *Id.* at 921–22.

<sup>231</sup> *Id.* at 922.

<sup>232</sup> *Id.* at 923.

<sup>233</sup> *Id.* at 923–24.

<sup>234</sup> *Id.* at 923.

<sup>235</sup> *Id.*

<sup>236</sup> See *Gernatt Asphalt Prods., Inc. v. Town of Sardinia*, 664 N.E.2d 1226, 1230–31 (N.Y. 1996).

<sup>237</sup> *Id.* at 1231

the zoning power that the Court of Appeals approved in *Frew Run*, where the mines were allowed but only in certain areas.<sup>238</sup> In *Gernatt*, the court held that the town's use of its zoning power to ban all mining within the town did not violate the MLRL.<sup>239</sup> The court noted that without a "clear expression of legislative intent to preempt local control over land use" the local zoning laws were not preempted.<sup>240</sup> The court also found that towns are not "obligated to permit the exploitation of any and all natural resources within th[at] town."<sup>241</sup>

These two cases established an important baseline for how towns may use their zoning power. However, both cases dealt only with mining and the zoning power of towns in relation to the MLRL.<sup>242</sup> Therefore, a number of oil and gas companies claim the decisions in *Frew Run* and *Gernatt* are not applicable to hydrofracking.<sup>243</sup>

#### *D. The Legal Journey of Hydrofracking in New York*

Supporters and opponents of hydrofracking hold divergent opinions as to whether zoning bans on hydrofracking are preempted by state law. Gas companies argue that hydrofracking, as a type of gas drilling, can only be controlled by state law, specifically the OGSML.<sup>244</sup> They further argue that

---

<sup>238</sup> See *Frew Run*, 518 N.E.2d at 923–24.

<sup>239</sup> See *Gernatt*, 664 N.E.2d at 1235–37.

<sup>240</sup> *Id.* at 1234.

<sup>241</sup> *Id.* at 1235 ("A municipality is not obligated to permit the exploitation of any and all natural resources within the town as a permitted use if limiting that use is a reasonable exercise of its police powers to prevent damage to the rights of others and to promote the interest of the community as a whole.").

<sup>242</sup> See *Frew Run*, 518 N.E.2d at 921; see also *Gernatt*, 664 N.E.2d at 1230.

<sup>243</sup> See Charles Gottlieb, *Regulating Natural Gas Development Through Local Planning and Land Use Controls*, N.Y. ZONING L. & PRAC. REP., May/June 2012, at 1, 3; Campbell, *supra* note 132 ("West, the Norse attorney, warned the appellate justices against falling into the 'trap' of judging based on past decisions on sand and gravel, which are regulated under a separate portion of state law.")

<sup>244</sup> Oil, Gas and Solution Mining Law, N.Y. ENVTL. CONSERV. § 23-0303(2) (McKinney 2007) ("The provisions of this article shall supersede all

the OGSML preempts local zoning laws through both express language in the statute and implicitly through state occupation of gas mining regulation and legislative intent.<sup>245</sup> Opponents of hydrofracking disagree and claim that the zoning bans are a proper exercise of the zoning power of towns.<sup>246</sup> Furthermore, antifracking advocates argue that they are following precedent set by the New York Court of Appeals<sup>247</sup> regarding the ability of towns to use their zoning power to ban mining activity within their town.<sup>248</sup> Hydrofracking opponents focus on previous Court of Appeals rulings, where the court did not find express or implied preemption in the MLRL, and cite the similar language between the OGSML and the MLRL.<sup>249</sup>

Gas companies have challenged the hydrofracking bans in two cases—*Anschutz Exploration Corp. v. Town of Dryden*<sup>250</sup> and *Cooperstown Holstein Corp. v. Town of Middlefield*.<sup>251</sup> In both cases the hydrofracking bans were upheld by the trial courts and

---

local laws or ordinances relating to the regulation of the oil, gas and solution mining industries; but shall not supersede local government jurisdiction over local roads or the rights of local governments under the real property tax law.”).

<sup>245</sup> See Gottleib, *supra* note 243, at 3.

<sup>246</sup> *Id.* at 2; see also Slottje & Slottje, *supra* note 119.

<sup>247</sup> The Court of Appeals upheld selective zoning regarding mining in *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920 (N.Y. 1987), and a town’s use of zoning power to exclude mines in *Gernatt Asphalt Prods., Inc. v. Town of Sardinia*, 664 N.E.2d 1226 (N.Y. 1996).

<sup>248</sup> See Gottleib, *supra* note 243, at 2.

<sup>249</sup> *Id.*; see also Mined Land Reclamation Law, N.Y. ENVTL. CONSERV. § 23-2703(2) (McKinney 2007) (“[F]or the purposes stated herein, this title shall supersede all other state and local laws relating to the extractive mining industry.”) (emphasis added); ENVTL. CONSERV. § 23-0303(2) (“The provisions of this article shall supersede all local laws or ordinances relating to the regulation of the oil, gas and solution mining industries; but shall not supersede local government jurisdiction over local roads or the rights of local governments under the real property tax law.”) (emphasis added).

<sup>250</sup> *Anschutz Exploration Corp. v. Town of Dryden*, 940 N.Y.S.2d 458 (Sup. Ct. 2012), *aff’d sub nom.* Norse Energy Corp. USA v. Town of Dryden, No. 515227 (App. Div. May 2, 2013).

<sup>251</sup> *Cooperstown Holstein Corp. v. Town of Middlefield*, 943 N.Y.S.2d 722 (Sup. Ct. 2012), *aff’d*, No. 515498 (App. Div. May 2, 2013).

by the Appellate Division.<sup>252</sup> However, the issue is far from settled, as hydrofracking ban opponents are currently seeking leave to appeal to the Court of Appeals.<sup>253</sup> These cases concern the extent of a town's zoning power and whether hydrofracking bans are a proper use of that power.

In *Anschutz Exploration*, the trial court found that the New York legislature did not intend to preempt local control over land use and zoning when it passed the OGSML.<sup>254</sup> Due to the similar language between the OGSML and the MLRL, the trial court based its decision largely on the precedent set by the Court of Appeals in *Frew Run*.<sup>255</sup> The court found that the OGSML's language, superseding those laws *regulating* oil and gas drilling, indicated only laws that dealt with the actual operation of drilling. The OGSML did not prevent local governments from determining where within their borders the drilling should take place.<sup>256</sup> It was within the town's land use power to ban the location of hydrofracking drilling sites if the town thought that it would negatively affect the community.<sup>257</sup> Such a ban did not rise to the level of regulation.<sup>258</sup> In effect, only the state can regulate the "how" of mining but local municipalities can regulate the "where."<sup>259</sup>

Additionally, the court in *Anschutz* found that there was no "clear expression of legislative intent" in the OGSML to preempt zoning laws, language that had been included in other state statutes.<sup>260</sup> While another trial court had interpreted the

---

<sup>252</sup> Campbell, *supra* note 133.

<sup>253</sup> *Id.*

<sup>254</sup> *Anschutz Exploration*, 940 N.Y.S.2d at 471.

<sup>255</sup> *Id.* at 471-73.

<sup>256</sup> *Id.*

<sup>257</sup> *Id.*

<sup>258</sup> *See id.* at 470-73.

<sup>259</sup> SALKIN, *supra* note 177, § 11:23.50.

<sup>260</sup> *Anschutz Exploration*, 940 N.Y.S.2d at 470. New York has clearly expressed its intent to preempt local zoning ordinances in other state statutes. *See, e.g.*, N.Y. ENVTL. CONSERV. LAW § 27-1107 (McKinney 2007) ("[N]o municipality may, except as expressly authorized by this article or the board, require any approval, consent, permit, certificate or other condition including conformity with local zoning or land use laws and ordinances" (emphasis



OGSML to preempt local fees being charged, that court had not examined the bill's language with regard to zoning.<sup>261</sup> Furthermore, the bill's language and legislative history show no indication that the legislature believed that maximizing the drilling for natural gas at the cost of local sovereignty was in the best interests of New York State.<sup>262</sup> Additionally, the OGSML only touched on technical concerns,<sup>263</sup> and it did not address common zoning problems such as traffic, noise, and protecting the character of a community.<sup>264</sup> Lastly, the court found that, as in *Gernatt*, the town did not engage in exclusionary zoning, as there is no obligation to permit the exploitation of a town's natural resources.<sup>265</sup> *Anschutz* was a clear victory for hydrofracking opponents, finding that towns could use their zoning power to ban hydrofracking.<sup>266</sup> Shortly after *Anchutz*, other trial courts would weigh in on the legality of hydrofracking bans.<sup>267</sup>

In *Cooperstown Holstein*, a different trial court upheld the local municipality's power to use their zoning power to ban hydrofracking.<sup>268</sup> The court found that the purpose and intent of

---

added)); N.Y. MENTAL HYG. LAW § 41.34(e) (McKinney 2011) ("A community residence established pursuant to this section and family care homes shall be deemed a family unit, *for the purposes of locals laws and ordinances.*" (emphasis added)).

<sup>261</sup> *Envirogas, Inc. v. Town of Kiantone*, 447 N.Y.S.2d 221 (Sup. Ct. 1982), *aff'd*, 454 N.Y.S.2d 694 (App. Div.).

<sup>262</sup> *See Anschutz Exploration*, 940 N.Y.S.2d at 469–70.

<sup>263</sup> The technical concerns in the OGSML include "where operations may be conducted, such as those governing delineation of pools, well spacing, and integration of unit" and the distance between wells to "comport with geological features of the underlying pool[s]." *Id.* at 470.

<sup>264</sup> *Id.*

<sup>265</sup> *Id.*; *see also Gernatt Asphalt Prods., Inc., v. Town of Sardinia*, 664 N.E.2d 1226, 1236 (N.Y. 1996).

<sup>266</sup> *See Anschutz Exploration*, 940 N.Y.S.2d at 471–72.

<sup>267</sup> *See Lena Groeger, Decision on Dryden Fracking Ban Could Set a National Precedent*, SYRACUSE.COM (Feb. 23, 2012, 12:50 PM), [http://www.syracuse.com/news/index.ssf/2012/02/decision\\_on\\_drydens\\_fracking\\_b.html](http://www.syracuse.com/news/index.ssf/2012/02/decision_on_drydens_fracking_b.html).

<sup>268</sup> Jinjoo Lee, *Another Court Upholds Fracking Ban*, CORNELL DAILY SUN (Feb. 27, 2012), <http://cornellsun.com/node/50051>.

the OGSML was to regulate the industry and not to preempt local land use authority.<sup>269</sup> The court relied heavily on *Frew Run* and *Gernatt* but utilized a more in-depth historical analysis than in *Anschutz*, to ascertain legislative intent.<sup>270</sup> The court, looking at previous state statutes and legislative memoranda,<sup>271</sup> found that the legislative intent was to minimize waste.<sup>272</sup> Additionally, amendments in 1978 replaced the phrase “foster, encourage and promote” regarding the state role in gas production with the word “regulate.”<sup>273</sup> The court found that this did not show clear legislative intent for state law to supersede local zoning control.<sup>274</sup>

*Anschutz* and *Cooperstown Holstein* were recently upheld by the Appellate Division, but attorneys for the hydrofracking industry have indicated that they intend to appeal.<sup>275</sup> While the Court of Appeals only grants leave to a fraction of the cases that request it,<sup>276</sup> there is a strong chance that the court will grant such leave here since it is a matter of first impression that has repercussions across the state. If the decisions are upheld by the Court of Appeals, towns will be able to ban hydrofracking through their zoning powers limiting where hydrofracking will occur in New York State. Moreover, these cases also provide the Court of Appeals an opportunity to clarify their own opaque jurisprudence on implied preemption and its appropriate application with regard to zoning.

---

<sup>269</sup> *Cooperstown Holstein Corp. v. Town of Middlefield*, 943 N.Y.S.2d 722, 730 (Sup. Ct. 2012).

<sup>270</sup> See SALKIN, *supra* note 177, § 11:23.50.

<sup>271</sup> *Cooperstown Holstein*, 943 N.Y.S.2d at 723–29 (examining Article 3-A of the Conservation Law, amendments in 1978, amendments in 1981, and the Legislative Memorandum).

<sup>272</sup> *Id.* at 728–29.

<sup>273</sup> *Id.* at 726.

<sup>274</sup> See *id.* at 729.

<sup>275</sup> Campbell, *supra* note 133.

<sup>276</sup> *Id.*

## III. IMPLIED PREEMPTION AND HYDROFRACKING

## A. Zoning and Implied Preemption

A number of New York Court of Appeals cases have addressed when zoning laws are implicitly preempted by state laws.<sup>277</sup> The issue of preemption most commonly arises in regard to exclusionary zoning<sup>278</sup> or prevention of specific uses of land.<sup>279</sup> Both are a form of “NIMBYism.” NIMBY, which stands for “not in my backyard,” refers to objections by the community about the placement of certain activities or structures in their particular neighborhood.<sup>280</sup> Such NIMBY problems often arise from projects that generate extensive benefits but impose a facility or project that negatively affects the local residents.<sup>281</sup> Examples include when communities use their zoning power to restrict housing for the low income or mentally disabled<sup>282</sup> and the placement of waste disposal facilities.<sup>283</sup> Issues arise when

---

<sup>277</sup> See, e.g., *Inc. Vill. of Nyack v. Daytop Vill., Inc.*, 583 N.E.2d 928 (N.Y. 1991); *Kamhi v. Town of Yorktown*, 547 N.E.2d 346 (N.Y. 1989); *Consol. Edison Co. of N.Y. v. Town of Red Hook*, 456 N.E.2d 487 (N.Y. 1983).

<sup>278</sup> Exclusionary zoning is often employed to describe land use laws which exclude certain people or projects from a certain community. The focus is often on individuals rather than uses. For more information see SALKIN, *supra* note 177, §§ 20:01–02.

<sup>279</sup> Often the problem arises when the specific uses of land have a relation to the public welfare. For more information see *id.* §§ 11:01–06.

<sup>280</sup> *Nimby Definition*, OXFORD DICTIONARIES, [http://oxforddictionaries.com/definition/american\\_english/Nimby](http://oxforddictionaries.com/definition/american_english/Nimby) (last visited Dec. 15, 2012) (“[A] person who objects to the siting of something perceived as unpleasant or potentially dangerous in their own neighborhood, such as a landfill or hazardous waste facility, especially while raising no such objections to similar developments elsewhere.”).

<sup>281</sup> See Barak D. Richman, *Mandating Negotiations to Solve the NIMBY Problem*, 20 UCLA J. ENVTL. L. & POL’Y 223, 223 (2001–02) (“NIMBY conflicts arise from projects that typically generate widespread dispersed benefits while imposing concentrated costs, such as homeless shelters, prisons, airports, sports stadiums, and waste disposal sites.”).

<sup>282</sup> SALKIN, *supra* note 177, §§ 20:01–02.

<sup>283</sup> Richman, *supra* note 281, at 223.

the placement of the project, while perhaps undesirable for the neighborhood, is essential for the community as a whole.<sup>284</sup>

One means to address NIMBYism is through legislation controlling placement. Since local municipalities derive their authority from the state legislature,<sup>285</sup> municipalities cannot pass zoning plans that are preempted by state law.<sup>286</sup> The local law is expressly preempted if the state law reserves control over the zoning procedure for a specific industry for itself.<sup>287</sup> However, even if the state law does not specifically reserve control over zoning, the local law could still be impliedly preempted.<sup>288</sup> In both forms of implied preemption (conflict and field), the key is to analyze the intent of the legislature.<sup>289</sup> The language in some Court of Appeals decisions seems to indicate a broad reading for what constitutes implied preemption with regard to zoning but actual decisions have created an almost impossibly narrow application.

#### *B. (Trying) To Find Implied Preemption*

The Court of Appeals has found that the intent to preempt does not have to be expressly stated and it is “enough that the Legislature has impliedly evinced its desire to do so.”<sup>290</sup> It is also not enough “that the state and local laws touch upon the same area.”<sup>291</sup> Instead, the court can look to declared state policy to infer whether the legislature intended to preempt local laws.<sup>292</sup>

---

<sup>284</sup> *Id.* at 223–24.

<sup>285</sup> *E.g.*, N.Y. CONST. art. IX, § 2.

<sup>286</sup> *See* SALKIN, *supra* note 177, § 4:22.

<sup>287</sup> *See* Weiland, *supra* note 16, at 472; Goho, *supra* note 14, at 5.

<sup>288</sup> N.Y. COMM’N ON LOCAL GOV’T EFFICIENCY & COMPETITIVENESS, *supra* note 15.

<sup>289</sup> *See id.*

<sup>290</sup> *Consol. Edison Co. of N.Y. v. Town of Red Hook*, 456 N.E.2d 487, 487 (N.Y. 1983) (citations omitted).

<sup>291</sup> *Inc. Vill. of Nyack v. Daytop Vill., Inc.*, 583 N.E.2d 928, 930 (N.Y. 1991) (quoting *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903, 907 (N.Y. 1987)).

<sup>292</sup> *Consol. Edison Co.*, 456 N.E.2d at 490 (citing *Robin v. Inc. Vill. of Hempstead*, 285 N.E.2d 285 (N.Y. 1972)).

However, in actuality, the Court of Appeals has applied a very narrow test and has been loath to find implied preemption by the state with regard to zoning without an express statement of intent.<sup>293</sup> The apparent necessity of such a clear and unequivocal statement of intent by the state raises the question of whether in the absence of such a statement any zoning act could be considered impliedly preempted.

For example, in *Incorporated Village of Nyack v. Daytop Village Inc.*,<sup>294</sup> the Court of Appeals held that “separate levels of regulatory oversight can coexist”<sup>295</sup> without preemption and that the detailed regulatory structure alone did not “evidence[] a desire” to preempt local zoning power.<sup>296</sup> The court held that the Mental Hygiene Law,<sup>297</sup> a very detailed regulatory scheme, did not preempt local zoning law since there was no clear indication of legislative intent to preempt.<sup>298</sup> Although not specifically stated, the court’s failure to find implied preemption in this case establishes an incredibly high burden for what constitutes implied preemption. *DJL Restaurant Corp. v. City of New York*

---

<sup>293</sup> See *Daytop Vill.*, 583 N.E.2d at 928–32; see also *Jancyn*, 518 N.E.2d at 906; *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920 (N.Y. 1987).

<sup>294</sup> *Daytop Vill., Inc.*, 583 N.E.2d at 929 (holding that local zoning regulations for substance abuse treatments were not preempted by state law, even though article 19 of the Mental Hygiene Law created a detailed regulatory structure because there is no evidence of legislative intent to preempt local zoning laws).

<sup>295</sup> *Id.* at 931.

<sup>296</sup> *Id.* (quoting *People v. Cook*, 312 N.E.2d 452, 457 (N.Y. 1974)).

<sup>297</sup> “DSAS [Division of Substance Abuse Services] is charged with the responsibility for establishing procedures and setting standards for the approval of substance abuse programs.” *Daytop Vill.*, 583 N.E.2d at 930 (citing N.Y. MENTAL HYG. LAW § 23.01 (McKinney 1991) (repealed 1999)). DSAS also is to “cooperate with and assist local agencies and community service boards in the development and periodic review of local comprehensive plans and programs for substance abuse services and approve such plans and programs . . . .” *Id.* (citing MENTAL HYG. § 19.07(b)(4) (McKinney 2011)). DSAS also must “inspect and approve or disapprove the facilities of and the services provided by substance abuse programs . . . .” *Id.* (citing MENTAL HYG. § 19.07(b)(5)).

<sup>298</sup> *Id.* at 931.

also demonstrates this high burden.<sup>299</sup> In that case, the Court of Appeals held that even though adult establishments were regulated by state law, local zoning rules were not impliedly preempted because state law did not address the “secondary effects” of these establishments.<sup>300</sup> Zoning laws are purposefully designed for local communities to address such concerns and protect their quality of life.<sup>301</sup> Additionally, the court held that there was no statement of legislative intent in the Alcoholic Beverage Control Law indicating that the state intended to preempt local zoning laws.<sup>302</sup>

*Frew Run* and *Gernatt* are additional examples of the reluctance of New York courts to find preemption without a specific statement of legislative intent.<sup>303</sup> In both cases, the court read the MLRL as not limiting zoning in large part because there was no explicit language of legislative intent and the local town ordinances were “consistent with the statute’s overall aim of protecting the environment.”<sup>304</sup>

The Court of Appeals has also applied this narrow view of implied preemption to questions of local power outside of zoning. In the case of *Jancyn Mfg. Corp. v. County of Suffolk*,<sup>305</sup> the Court of Appeals refused to find that a state law that prohibited the sale and use of certain sewage system cleaning additives was implicitly preempted by local laws, which set

---

<sup>299</sup> *DJL Rest. Corp. v. City of New York*, 749 N.E.2d 186, 188, 191–92 (N.Y. 2001) (holding that local zoning rules regulating adult industry locations were not preempted even though the venues served alcohol, which is regulated by the Alcoholic Beverage Control Law).

<sup>300</sup> *Id.* at 191–92 (defining “secondary effects” as “increased crime rates, reduced property values, neighborhood deterioration and inappropriate exposure of children to sexually oriented environments”).

<sup>301</sup> *See id.* at 188–89.

<sup>302</sup> *See id.* at 191.

<sup>303</sup> *See* discussion *supra* Part II.C.

<sup>304</sup> *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920, 923 (N.Y. 1987); *see also Gernatt Asphalt Prods., Inc. v. Town of Sardinia*, 664 N.E.2d 1226, 1235–36 (N.Y. 1996).

<sup>305</sup> *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903 (N.Y. 1987) (finding no preemption where plaintiff’s sewage additives, were approved for sale by state law but were not allowed to be sold according to a more stringent local standard).

stricter standards for the sale of sewage cleaning additives.<sup>306</sup> The court looked to the legislature's intent and to the statutory scheme.<sup>307</sup> Although the statutory regulatory scheme was very detailed,<sup>308</sup> the court held that it was not thorough or extensive enough to have superseded all possible future local regulation.<sup>309</sup> A key reason that the court did not find implied preemption involved the absence of an express statement from the state of its intent to preempt.<sup>310</sup> The court also held that implied preemption could not be found merely because both pieces of legislation had the same goal.<sup>311</sup> In other cases, the Court of Appeals has also held that local laws that expand a definition in state law are not preempted as long as the legislature has not "evidenced a desire" to preempt.<sup>312</sup>

When the Court of Appeals has held local zoning laws are impliedly preempted, there is often specific language in the bill itself indicating a desire for preemption.<sup>313</sup> For example, in *Consolidated Edison Co. of New York v. Town of Red Hook*,<sup>314</sup>

---

<sup>306</sup> *See id.* at 906.

<sup>307</sup> *Id.* at 905–07.

<sup>308</sup> The State law prohibited the sale and use of certain sewage system cleaning additives in Long Island. It also empowered the State Commissioner of Environmental Conservation to create regulations forcing manufacturers to disclose their chemical components and restrict sale of products with restricted chemical material after investigation and hearing. *See id.* at 903–04.

<sup>309</sup> *Id.* at 907.

<sup>310</sup> *Id.* ("Although an express statement of preemption is not required it is significant that no such statement appears in the statute . . .").

<sup>311</sup> *Id.* (finding that both the local law and state law shared the same goal, protection of the Long Island water supply).

<sup>312</sup> *N.Y. State Club Ass'n v. City of New York*, 505 N.E.2d 915, 920 (N.Y. 1987) (holding that New York City's narrower definition of what constituted a private club was not preempted by state antidiscrimination laws).

<sup>313</sup> However courts have been more likely to find implied preemption of local laws not connected to zoning. *See Albany Area Builders Ass'n v. Town of Guilderland*, 546 N.E.2d 920 (N.Y. 1989) (holding that a local law setting up a Transportation Impact Fee was impliedly preempted by the state regulatory structure regulating highway funds).

<sup>314</sup> *Consol. Edison Co. of N.Y. v. Town of Red Hook*, 456 N.E.2d 487 (N.Y. 1983).

the court held that Red Hook's Local Law 2, which allowed the town to refuse Consolidated Edison Company a permit under its zoning law, was preempted by Article VIII of the Public Service Law.<sup>315</sup> The court cited the legislature's purpose, clearly expressed in Article VIII, and the detailed regulatory structure.<sup>316</sup> Article VIII plainly indicated that the legislature intended "to provide for the expeditious resolution of all matters concerning the location of major steam electric generating facilities."<sup>317</sup> This was reaffirmed when the Legislature reenacted Article VIII in 1978, asserting "its purpose was to have the Siting Board balance all interests, including local interests, on a State-wide basis."<sup>318</sup> Although there is language indicating the role of the regulatory structure in the decision, it is clear that the holding was based primarily on the very express legislative intent indicated in Article VIII.

Together, these cases illustrate that when the Court of Appeals examines whether state law impliedly preempts local law, especially with respect to zoning, it rarely finds such preemption without an explicit statement from the state legislature. Although previous Court of Appeals decisions include language that an explicit expression of legislative intent is not required,<sup>319</sup> the reality appears to be otherwise. If the previously mentioned cases are any indication, it does not appear that any comprehensive regulatory scheme, absent a declared intention to preempt local power, will be sufficient for the Court of Appeals to find implied preemption.<sup>320</sup>

---

<sup>315</sup> See *id.* at 489-90.

<sup>316</sup> *Id.* at 490-91.

<sup>317</sup> *Id.* at 490 (quoting L. 1972, ch. 385, § 1).

<sup>318</sup> *Id.* (quoting L. 1978, ch. 708, § 1).

<sup>319</sup> See, e.g., *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903, 906 (N.Y. 1987).

<sup>320</sup> See, e.g., *Consol. Edison Co.*, 456 N.E.2d at 490 (holding that Article XIII contained an express statement about the legislature's intent to preempt local zoning rules).



*C. Hydrofracking Bans Are Not Implicitly Preempted*

If the Court of Appeals follows the exacting standard it has thus far used for determining implied preemption, it is likely to uphold the town's hydrofracking bans.<sup>321</sup> There are two aspects of implied preemption that need to be analyzed: conflict preemption and field preemption.<sup>322</sup> Either is sufficient for a law to be preempted and both are controlled by the intent of the legislature.<sup>323</sup> Since there is no explicit statement in the OGSML indicating unequivocal intent by the legislature to preempt local land use control over gas drilling, the Court of Appeals will likely find that local hydrofracking bans are not preempted.

There is no conflict preemption between the OGSML and local hydrofracking bans because the bans do not frustrate the purpose of the OGSML. There is no inherent conflict simply because the local zoning laws prohibit what state law allows, otherwise local power would be meaningless.<sup>324</sup> Instead, the court looks to legislative intent in the statute itself.<sup>325</sup> The OGSML indicates that its main purpose is not to ensure that drilling occurs anywhere that it is possible but to prevent waste and protect the rights of the general public.<sup>326</sup> While the OGSML

---

<sup>321</sup> This Note does not examine whether or not the Court of Appeals will find express preemption in the OGSML.

<sup>322</sup> See Goho, *supra* note 14, at 5; N.Y. STATE COMM'N ON LOCAL GOV'T EFFICACY & COMPETITIVENESS, *supra* note 15.

<sup>323</sup> See Goho, *supra* note 14; see also N.Y. STATE COMM'N ON LOCAL GOV'T EFFICACY & COMPETITIVENESS, *supra* note 15.

<sup>324</sup> See *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903, 907 (N.Y. 1987); see also *N.Y. State Club Ass'n v. City of New York*, 505 N.E.2d 915, 920 (N.Y. 1987).

<sup>325</sup> See, e.g., *Consol. Edison Co.*, 456 N.E.2d 487; *N.Y. State Club Ass'n*, 505 N.E.2d at 915; see *Jancyn*, 518 N.E.2d at 906 ("No preemptive intent is evident from either the Legislature's declaration of State policy . . . or the statutory scheme which has been enacted.").

<sup>326</sup> N.Y. ENVTL. CONSERV. § 23-0301 (McKinney 2007) ("It is hereby declared to be in the public interest to regulate the development, production and utilization of natural resources of oil and gas in this state in such a manner as will prevent waste; to authorize and to provide for the operation and development of oil and gas properties in such a manner that a greater ultimate recovery of oil and gas may be had and that correlative right of all

does mention the “greater ultimate recovery of oil and gas,” that is in reference to the operation and development of the sites, not where such sites should occur.<sup>327</sup> The only language which explicitly restricts local power refers to the “regulation” of mining.<sup>328</sup> Such a term though has never been interpreted to restrict all interaction with that activity.<sup>329</sup> The OGSML makes no mention of noise, traffic, and neighborhood character, all of which are responsibilities normally left to local government.<sup>330</sup> As the Court of Appeals held in *DJL Restaurant*, these are the types of concerns that are specifically meant to be addressed by zoning.<sup>331</sup> Local zoning laws that address these issues are not “regulating” hydrofracking but only affecting where hydrofracking can take place.<sup>332</sup> In addition, two levels of regulatory oversight, one stricter than the other, have been allowed<sup>333</sup> even when local law prohibits an activity allowed under state law.<sup>334</sup>

Additionally, there is no field preemption because under the Court of Appeals’ narrow view of implied preemption, the regulatory structure of the OGSML is not sufficiently detailed or comprehensive enough to eliminate local discretion.<sup>335</sup> Even in

---

owners and the rights of all persons including landowners and the general public may be fully protected.”).

<sup>327</sup> *Id.*

<sup>328</sup> *Id.* § 23-0303(2).

<sup>329</sup> See *Gernatt Asphalt Prods., Inc. v. Town of Sardinia*, 664 N.E.2d 1226, 1235 (N.Y. 1996); *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920, 923 (N.Y. 1987).

<sup>330</sup> SALKIN, *supra* note 177, § 11:23.50.

<sup>331</sup> *DJL Rest. Corp. v. City of New York*, 749 N.E.2d 186, 188, 191–92 (N.Y. 2001).

<sup>332</sup> See *Frew Run*, 518 N.E.2d at 923–24; *Gernatt*, 664 N.E.2d at 1235–36.

<sup>333</sup> See *Jancyn Mfg. Corp. v. Cnty. of Suffolk*, 518 N.E.2d 903, 907 (N.Y. 1987).

<sup>334</sup> See, e.g., *id.* (holding that a local regulation is not preempted by a state law that also addresses the same issue); *N.Y. State Club Ass’n v. City of New York*, 505 N.E.2d 915 (N.Y. 1987) (holding that local law can have a narrower definition of what constitutes a private club than state antidiscrimination laws).

<sup>335</sup> See *Inc. Vill. of Nyack v. Daytop Vill., Inc.*, 583 N.E.2d 928 (N.Y.

cases where the state legislature had a more comprehensive regulatory scheme, such as for substance abuse housing,<sup>336</sup> discrimination,<sup>337</sup> and mining,<sup>338</sup> local zoning laws were not considered preempted. The OGSML regulatory structure is not as detailed as that of the cases above, focusing only on reserving power for the state to control the *regulation* of the gas mining rather than its placement.<sup>339</sup> There is nothing to indicate that the purpose of the OGSML is to ensure hydrofracking happens anywhere that it can.<sup>340</sup> It is telling that the Court of Appeals held in *Garnett* that there is no explicit requirement that towns permit mining just because they have such resources.<sup>341</sup> Additionally, the current regulatory structure does not create a system where a single town's decision to ban hydrofracking would affect another town's ability to allow hydrofracking. While some commenters claim that natural gas production is only feasible over many municipalities,<sup>342</sup> that claim is unlikely as towns are often separated by many miles and the hydrofracking bans would only affect drilling sites within that specific town. It is also unlikely that the hydrofracking bans would be adopted by all towns due to the victory of

---

1991) (holding that local zoning regulations for substance abuse treatments were not preempted by a detailed state regulatory structure because there was no evidence of legislative intent to preempt local zoning laws); *see also Jancyn*, 518 N.E.2d at 907 (upholding a local law banning cleaning additives even though cleaning additives were also regulated through state scheme).

<sup>336</sup> *See Daytop Vill., Inc.*, 583 N.E.2d at 928–29.

<sup>337</sup> *See N.Y. State Club Ass'n*, 505 N.E.2d at 916.

<sup>338</sup> *See Gernatt Asphalt Prods., Inc., v. Town of Sardina*, 664 N.E.2d 1226, 1227 (N.Y. 1996); *Frew Run Gravel Prods., Inc. v. Town of Carroll*, 518 N.E.2d 920, 921 (N.Y. 1987).

<sup>339</sup> N.Y. ENVTL. CONSERV. § 23-0303(2) (McKinney 2007) (“The provisions of this article shall supersede all local laws or ordinances relating to the regulation of the oil, gas and solution mining industries; but shall not supersede local government jurisdiction over local roads or the rights of local governments under the real property tax law.”).

<sup>340</sup> *See Anschutz Exploration Corp. v. Town of Dryden*, 940 N.Y.S.2d 458, 464–66 (Sup. Ct. 2012).

<sup>341</sup> *Gernatt*, 664 N.E.2d at 1235.

<sup>342</sup> Gregory R. Nearpass & Robert J. Brenner, *High Volume Hydraulic Fracturing and Home Rule: The Struggle for Control*, 76 ALB. L. REV. 167, 188–89 (2013).

hydrofracking proponents in recent elections<sup>343</sup> and the presence of prohydrofracking resolutions in over fifty towns.<sup>344</sup> Even if all towns did enact hydrofracking bans, that would not alter the preemption argument since the language in the OGSML speaks to regulation of drilling where it occurs,<sup>345</sup> not the maximization of gas drilling everywhere. The limited regulatory structure created by the OGSML is not comprehensive enough to imply that the legislature intended to occupy the field and preempt all local zoning laws.

The Court of Appeals is not likely to find that towns' hydrofracking bans are impliedly preempted due to their own narrow interpretation of what constitutes implied preemption. Court of Appeals jurisprudence appears to indicate that only an explicit statement of legislative intent will preempt even the most exacting of state regulations. While the OGSML does specifically discuss control over the regulation of gas drilling, there is no explicit statement indicating that the state intended to reserve power over the placement of gas drilling locations.<sup>346</sup> Without such an explicit statement, the Court of Appeals is unlikely to find local zoning concerns impliedly preempted.

#### CONCLUSION

The Court of Appeals' narrow interpretation of implied preemption is appropriate public policy for New York in general and specifically with regard to hydrofracking. Although hydrofracking has been conducted for many years in other states, there are still a number of questions as to its effect on the local environment, including tainted water and methane explosions.<sup>347</sup> These environmental concerns are important as they could affect the drinking water of local towns and New

---

<sup>343</sup> Esch, *supra* note 116.

<sup>344</sup> See *Map of Positive Resolutions for Hydrofracking*, *supra* note 120 (showing specifically that the towns in favor of hydrofracking are also along the Marcellus Shale, the most lucrative area for hydrofracking).

<sup>345</sup> ENVTL. CONSERV. § 23-0303(2).

<sup>346</sup> See SALKIN, *supra* note 177, § 11:23.50.

<sup>347</sup> See WILBER, *supra* note 53, at 89-92; Caruso, *supra* note 71.

York City,<sup>348</sup> as well as impact tourism and local agriculture.<sup>349</sup> Ensuring that local governments are able to ban hydrofracking within their communities will provide another level of protection against any possible dangers from hydrofracking. It will also allow those communities eager for jobs and economic benefits to permit hydrofracking. An open debate about the pros and cons of hydrofracking will increase residents' knowledge and through the local political process, individuals will be able to have their voices heard.

Furthermore, town hydrofracking bans do not present a NIMBY problem. Unlike a waste reactor, which is often necessary for the community and needs to be placed somewhere, hydrofracking does not need to occur. Hydrofracking is not necessary for a community and while it may bring economic benefits, those benefits also come with risks. Individual towns should have the ability to decide for themselves if the costs outweigh the benefits. Additionally, even if towns are able to enact hydrofracking bans, it is unlikely that would end hydrofracking in New York State. Over fifty towns have already enacted resolutions supporting hydrofracking.<sup>350</sup> The ability of local governments to ban hydrofracking also appears to be in line with recent unofficial reports from the DEC indicating that hydrofracking would only occur in those areas that desire it.<sup>351</sup>

Additionally, allowing local governments to ban hydrofracking will not negatively affect other energy producers in New York State. The recent Power NY Act of 2011<sup>352</sup> includes express preemption language that creates a "one stop approval process for new and expanded power plans" including wind farms.<sup>353</sup> Since the Power New York Act expressly gives

---

<sup>348</sup> See Kastenbaum, *supra* note 58.

<sup>349</sup> See Gralla, *supra* note 56.

<sup>350</sup> *Map of Positive Resolutions for Hydrofracking*, *supra* note 120.

<sup>351</sup> Karlin, *supra* note 95 (citing recent reports, which the Governor refuses to deny, indicating that his administration is considering a plan where hydrofracking would only be permitted in willing communities).

<sup>352</sup> Power NY Act of 2011, S. 5844, Assemb. 8510, Reg. Sess. (codified in scattered sections of N.Y. PUB. SERV. LAW).

<sup>353</sup> *New York Legislature Enacts Power Plant Siting Law*, Bryan Cave Bulletin (Bryan Cave, LLP, St. Louis, MO), Aug. 9, 2011, at 1, *available at*

control over zoning to the state, there is no need to look for implied preemption. This is different from the OGSML, which only has specific language preempting regulation and does not have any explicit language regarding zoning or the placement of drilling sites.

Permitting New York towns to ban hydrofracking does go further than other states but that is a positive development. While West Virginia courts have struck down hydrofracking bans, their reasoning focused on the “narrow” power held by municipalities and West Virginia DEP’s primary authority to protect the environment.<sup>354</sup> For the New York Court of Appeals to analyze the OGSML in a similar manner would upend years of jurisprudence that allowed local municipalities greater control through zoning. It is also not clear in Pennsylvania what level of control local municipalities will have over hydrofracking.<sup>355</sup> It is possible that the Pennsylvania Supreme Court will take a similar position to that of the New York Court of Appeals and adopt a broad view of zoning power. Regardless, the environmental issues that Pennsylvania has encountered in its quick embrace of hydrofracking<sup>356</sup> are additional evidence that the best path forward is greater local control.

The New York Court of Appeals should also take this opportunity to clarify that implied preemption should only be found with regard to zoning if there is an explicit statement of intent from the legislature. Such a statement would simply codify what is already effectively unstated law. This would have a number of policy benefits for New York State. It would create a clear bright line rule that would give local municipalities a greater sense of what they are able to do and would decrease the number of lawsuits challenging their authority.

---

[http://www.bryancave.com/files/Publication/b03ff613-d188-440a-bd0b-01f498b4e1bc/Presentation/PublicationAttachment/f8e28986-f958-4912-91e0-119d5dccd6e0/New%20York%20Power%20Plant%20Siting%20Article%20X%20Client%20Alert\\_v7%20\(AP\).pdf](http://www.bryancave.com/files/Publication/b03ff613-d188-440a-bd0b-01f498b4e1bc/Presentation/PublicationAttachment/f8e28986-f958-4912-91e0-119d5dccd6e0/New%20York%20Power%20Plant%20Siting%20Article%20X%20Client%20Alert_v7%20(AP).pdf)

<sup>354</sup> Orford, *supra* note 162.

<sup>355</sup> Detrow, *supra* 157.

<sup>356</sup> See Caruso, *supra* note 71; see also Drakem & Efstahiou Jr., *supra* note 76.

Greater control for local municipalities is especially important with regard to zoning. A municipality's zoning power is its most effective weapon to protect their community. As Judge Cardozo commented, "a zoning resolution in many of its features is distinctively a city affair, a concern of the locality, affecting as it does the density of population, the growth of city life, and the court of city values."<sup>357</sup> Due to the unique importance of zoning, it is proper for the Court of Appeals to adopt such a bright line rule that forces the legislature to explicitly state if they intend to remove a municipality's zoning power.

In addition, a requirement of express intent for preemption would help the judiciary and the legislature. The judiciary will no longer have to struggle to discern unclear legislature intent. Instead, courts could look at the legislation itself for an explicit statement to determine if the state reserved zoning power for itself, otherwise local municipalities would retain that authority. Government, both on the state and the local level, would also benefit. State legislatures going forward would have to truly contemplate if the laws they are enacting would be better served through local involvement or through laws controlling zoning power. This would create an environment conducive to better lawmaking. Local governments would also be spared the threat of constant litigation based on the intended thoughts of the legislature.

The legality of hydrofracking bans will likely remain precarious until the Court of Appeals clarifies the limits of implied preemption. In the interim, local municipalities will continue to use their zoning power to decide for themselves whether the risks of hydrofracking outweigh its rewards.

---

<sup>357</sup> See SALKIN, *supra* note 177, § 2:01 n.3 (citing *Adler v. Deegan*, 167 N.E. 705 (N.Y. 1929)).

# IS LITIGATION YOUR FINAL ANSWER? WHY THE HEALTHY WORKPLACE BILL SHOULD INCLUDE AN ADR PROVISION

*Florence Z. Mao\**

## I. INTRODUCTION

On the morning of May 2, 2005, Marlene Braun wrote in a two-page e-mail to a coworker, “I cannot bear the thought of coming into the office this morning or ever again . . . I cannot take any more abuse . . . and any more of the humiliation I have had to endure for the past year.”<sup>1</sup> Moments later, Marlene used a .38 blue steel revolver to shoot and kill her dogs before turning the gun to her head and pulling the trigger.<sup>2</sup>

Marlene had served as monument manager at the Carrizo Plain National Monument in Bakersfield, California and had been a federal employee at the Bureau of Land Management (“BLM”) for nineteen years.<sup>3</sup> One year before Marlene’s death, the BLM office in Bakersfield acquired a new director who

---

\* J.D. Candidate, Brooklyn Law School, 2014; B.A., Barnard College, Columbia University, 2008. I thank my family and friends for their unconditional support and encouragement, and members of the *Journal of Law and Policy* for their time, patience, and thoughtful feedback. Special thanks to my mother, Christina Gong, for inspiring my love of learning.

<sup>1</sup> OFFICE OF INSPECTOR GEN., U.S. DEP’T OF INTERIOR, REPORT OF INVESTIGATION: CARRIZO PLAIN INCIDENT 15 (2006), *available at* [http://www.workplacebullying.org/multi/pdf/braun\\_oig\\_report.pdf](http://www.workplacebullying.org/multi/pdf/braun_oig_report.pdf).

<sup>2</sup> *Id.*; see also KAMuston, *Fault Lines*, DAILY KOS (May 11, 2007, 3:34 AM), <http://www.dailykos.com/story/2007/05/11/333443/-Fault-Lines>.

<sup>3</sup> *An Act Concerning State Employees and Violence and Bullying in the Workplace: Hearing on Substitute H.B. 5464 Before the H. Comm. on Labor & Pub. Emps.*, 2010 Leg. (Conn. 2010) [hereinafter *Hearing 2010*] (statement of Katherine Hermes, Conn. Healthy Workplace Advocates); OFFICE OF INSPECTOR GEN., *supra* note 1, at 3.



strongly disagreed with Marlene's plan to regulate land use and cattle grazing in an effort to preserve native plant species at the Monument.<sup>4</sup> During that year, Marlene and her new supervisor engaged in a series of heated interactions.<sup>5</sup> In one incident, when she attempted to explain her position to him, he repeatedly yelled, "Did you hear what I said?"<sup>6</sup> The next day he continued to shout at her in front of other employees.<sup>7</sup> Marlene later wrote that she "felt like a bully had just beaten [her] up," and she was so upset that she vomited.<sup>8</sup> Another time, her supervisor threatened her when he blocked her on a narrow road with his truck, exited the vehicle, and told her that she had "brought this on herself."<sup>9</sup> Despite having a previously spotless employment record, after another conflict, her supervisor suspended her for five days without pay.<sup>10</sup> Even worse, Marlene appealed the suspension, but was denied.<sup>11</sup> In less than a year, Marlene received five written reprimands.<sup>12</sup>

While Marlene was once a healthy individual,<sup>13</sup> during the last year of her life she lost forty pounds, grew anxious and depressed,<sup>14</sup> and took prescription tranquilizers and sleeping pills.<sup>15</sup> When Marlene requested medical leave for the first time

---

<sup>4</sup> Julia Cart & Maria L. La Ganga, *Suicide Casts a Shadow on Conservation Battle*, L.A. TIMES (Aug. 20, 2005), <http://articles.latimes.com/print/2005/aug/20/local/me-carrizo20>.

<sup>5</sup> OFFICE OF INSPECTOR GEN., *supra* note 1, at 6.

<sup>6</sup> KAMuston, *supra* note 2. Marlene privately documented conflicts with her supervisor in a thirty-five-page chronology, including instances of being "yelled at." See Cart & La Ganga, *supra* note 4.

<sup>7</sup> KAMuston, *supra* note 2.

<sup>8</sup> *Id.*

<sup>9</sup> *An Act Concerning Bullying in the Workplace: Hearing on S.B. 60 Before the S. Comm. on Labor & Pub. Emps.*, 2008 Leg. (Conn. 2008) [hereinafter *Hearing 2008*] (statement of Katherine Hermes, Workplace Bullying Inst.).

<sup>10</sup> KAMuston, *supra* note 2; Cart & La Ganga, *supra* note 4.

<sup>11</sup> KAMuston, *supra* note 2.

<sup>12</sup> *Id.*

<sup>13</sup> *Id.*

<sup>14</sup> See *Hearing 2008*, *supra* note 9.

<sup>15</sup> KAMuston, *supra* note 2.

in her career, her supervisor denied the request.<sup>16</sup> In a suicide note to her best friend, Marlene wrote that the new director had made her life “utterly unbearable,”<sup>17</sup> and she could no longer suffer the abuse and humiliation.<sup>18</sup>

Unfortunately, Marlene’s experience with her supervisor is not atypical of “Targets”<sup>19</sup> of workplace bullying. Targets often suffer psychological, emotional, and physical harm as a result of the abuse.<sup>20</sup> For example, Targets can suffer from severe psychological harm akin to posttraumatic stress disorder<sup>21</sup> and, in a number of cases, may even resort to suicide.<sup>22</sup> A survey conducted by the Workplace Bullying Institute (“WBI”) in 2012 found that eighty percent of respondents experienced anxiety from workplace bullying, and forty-nine percent reported being diagnosed with clinical depression.<sup>23</sup> Moreover, twenty-nine percent of respondents considered suicide, and sixteen percent had an actual plan to commit it.<sup>24</sup> Despite the psychological and physical toll that workplace bullying has on American workers, currently there is no state or federal law that adequately addresses the phenomenon and protects workers against it.<sup>25</sup> The

---

<sup>16</sup> See *Hearing 2008*, *supra* note 9.

<sup>17</sup> *Hearing 2010*, *supra* note 3; Cart & La Ganga, *supra* note 4.

<sup>18</sup> OFFICE OF INSPECTOR GEN., *supra* note 1, at 1.

<sup>19</sup> Gary and Ruth Namie, the leading researchers of workplace bullying and author of several books on the subject, use the word “Target” instead of “victim” out of respect for the affected individuals. See GARY NAMIE & RUTH NAMIE, *BULLYPROOF YOURSELF AT WORK!* 10–12 (1999) [hereinafter NAMIE & NAMIE, *BULLYPROOF YOURSELF*].

<sup>20</sup> For further discussion regarding the harmful effects of workplace bullying, see *infra* Part II.

<sup>21</sup> See Heinz Leymann & Annelie Gustafsson, *Mobbing at Work and the Development of Post-Traumatic Stress Disorders*, 5 *EUR. J. WORK & ORGANIZATIONAL PSYCHOL.* 251, 252–54 (1996).

<sup>22</sup> *The Toll of Workplace Bullying on Employee Health: WBI Survey*, WORKPLACE BULLYING INST. (Aug. 12, 2009), <http://www.workplacebullying.org/2012/08/09/2012-d/>.

<sup>23</sup> *Id.*

<sup>24</sup> *Id.*

<sup>25</sup> Many European countries have passed anti-workplace-bullying legislation. See Susan Harthill, *Bullying in the Workplace: Lessons from the United Kingdom*, 17 *MINN. J. INT’L L.* 247, 263–66 (2008) (discussing anti-

only plausible avenues of relief available to Targets are to file a claim under either the common law tort of intentional infliction of emotional distress (“IIED”) or under Title VII of the Civil Rights Act of 1964 for a hostile work environment. These legal options, however, may be insufficient to address the often subtle nature of workplace bullying.<sup>26</sup>

As a result, Professor David Yamada of Suffolk Law School, a leading proponent of creating a distinct cause of action for workplace bullying and the architect of the Healthy Workplace Bill, has written extensively on the need for status-blind legislation to address workplace bullying.<sup>27</sup> Since workplace bullying, like sexual harassment, is subtle in nature, Yamada modeled the bill largely after the hostile work environment doctrine under sexual harassment case law.<sup>28</sup> The Healthy Workplace Bill would create a civil cause of action for victims of workplace bullying and hold employers liable for creating or maintaining an abusive work environment.<sup>29</sup> As of March 2013, twenty-four states—New York, Illinois, and

---

workplace-bullying laws in Europe and Canada); Amanda E. Lueders, Note, *You'll Need More Than a Voltage Converter: Plugging European Workplace Bullying Laws into the American Jurisprudential Outlet*, 25 ARIZ. J. INT'L & COMP. L. 197, 207–11 (2008).

<sup>26</sup> See *infra* Part II.C.

<sup>27</sup> See, e.g., David C. Yamada, *Crafting a Legislative Response to Workplace Bullying*, 8 EMP. RTS. & EMP. POL'Y J. 475 (2004) [hereinafter Yamada, *Crafting a Legislative Response*]; David C. Yamada, *Employment Law as if People Mattered: Bringing Therapeutic Jurisprudence into the Workplace*, 11 FLA. COASTAL L. REV. 257 (2010) [hereinafter Yamada, *Therapeutic Jurisprudence*]; David C. Yamada, *Human Dignity and American Employment Law*, 43 U. RICH. L. REV. 523 (2009) [hereinafter Yamada, *Human Dignity*]; David C. Yamada, *The Phenomenon of “Workplace Bullying” and the Need for Status-Blind Hostile Work Environment Protection*, 88 GEO. L.J. 475 (2000) [hereinafter Yamada, *Status-Blind Hostile Work Environment*]; David C. Yamada, *United States: Workplace Bullying and American Employment Law: A Ten-Year Progress Report and Assessment*, 32 COMP. LAB. L. & POL'Y J. 251 (2010) [hereinafter Yamada, *Ten-Year Progress Report*].

<sup>28</sup> See Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 524–25.

<sup>29</sup> See Yamada, *Crafting a Legislative Response*, *supra* note 27, at 517–20.

Oregon among them—have introduced the bill in state legislatures, but none has succeeded in passing it into law.<sup>30</sup> Opponents of the Healthy Workplace Bill primarily argue that such legislation would negatively impact the economy and flood the courts with frivolous lawsuits.<sup>31</sup>

A provision in the Healthy Workplace Bill encouraging the alternative dispute resolution (“ADR”) processes of mediation and arbitration as potential alternatives to litigation will increase the likelihood of its passage through state legislatures and provide Targets with more efficient and cost-effective solutions. Part II defines workplace bullying, examines its effects on employers and employees, and discusses possible avenues of relief currently available. Part III explores the text of the Healthy Workplace Bill and its progress in various state legislatures. Part IV discusses the use of mediation in employment disputes and the Equal Employment Opportunity Commission’s (“EEOC”) mediation program. Part V looks at existing court-connected and nonprofit arbitration programs. Reflecting upon the success of mediation and arbitration in other forms of employment disputes, Part VI examines the potential for inclusion of such programs in the Healthy Workplace Bill. Adopting these measures would address many concerns of the bill’s opponents—namely, crowding of court dockets—and encourage its swift passage.

## II. WORKPLACE BULLYING

### *A. Definition and Prevalence of Workplace Bullying*

Bullying has received increased media attention in recent years; however, because of its subtle nature, many have struggled to establish a uniform definition that accurately captures the phenomenon.<sup>32</sup> Nonetheless, the growing awareness

---

<sup>30</sup> HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/> (last visited Apr. 5, 2013).

<sup>31</sup> See Yamada, *Ten-Year Progress Report*, *supra* note 27, at 269–70.

<sup>32</sup> See, e.g., Emily Bazelon, Op-Ed., *Defining Bullying Down*, N.Y. TIMES (Mar. 11, 2013), <http://www.nytimes.com/2013/03/12/opinion/defining-bullying-down.html>.

of workplace bullying has prompted the development of various definitions in an attempt to address the issue. For example, the WBI defines workplace bullying as “the repeated, health-harming mistreatment of one or more persons (Target) by one or more perpetrators (supervisors or coworkers)” that can consist of verbal abuse, threatening or humiliating conduct, interference with work-related tasks, and even sabotage.<sup>33</sup> Similarly, Professor Yamada defines workplace bullying as “the intentional infliction of a hostile work environment upon an employee by a coworker or coworkers, typically through a combination of verbal and nonverbal behaviors.”<sup>34</sup> Others have characterized workplace bullying as “persistent,”<sup>35</sup> “unreasonable,”<sup>36</sup> and “malicious.”<sup>37</sup> These definitions have three unifying themes: (1) the bullying activity is persistent and intentional; (2) the Target suffers a combination of psychological, physical, and economic harm as a result; and (3) the bullying activity creates an overall hostile work environment.

Workplace bullying is distinguishable from general incivility and status-based harassment.<sup>38</sup> Unlike general aggression or incivility, which involve isolated instances of rudeness or crass

---

<sup>33</sup> *The WBI Definition of Workplace Bullying*, WORKPLACE BULLYING INST., <http://www.workplacebullying.org/individuals/problem/definition/> (last visited Apr. 5, 2013).

<sup>34</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 481.

<sup>35</sup> Pamela Lutgen-Sandvik, *Water Smoothing Stones: Subordinate Resistance to Workplace Bullying 1* (Aug. 2005) (unpublished Ph.D. dissertation, Ariz. State Univ.), *available at* <http://www.unm.edu/~plutgen/Resistance%20to%20Workplace%20Bullying%20Lutgen-Sandvik%20Dissertation%202005.pdf>.

<sup>36</sup> WASH. STATE DEP'T OF LABOR & INDUS., *WORKPLACE BULLYING AND DISRUPTIVE BEHAVIOR: WHAT EVERYONE NEEDS TO KNOW*, SHARP Report # 87-2-2011, at 1 (Apr. 2011) [hereinafter *WHAT EVERYONE NEEDS TO KNOW*], *available at* <http://www.lni.wa.gov/Safety/Research/Files/Bullying.pdf>.

<sup>37</sup> Lutgen-Sandvik, *supra* note 35, at 1.

<sup>38</sup> See Lynn M. Andersson & Christine M. Pearson, *Tit for Tat? The Spiraling Effect of Incivility in the Workplace*, 24 ACAD. MGMT. REV. 453, 466 (1999). Status-based harassment refers to harassment based on a person's protected status under Title VII of the Civil Rights Act of 1964, such as gender, race, or age. See 42 U.S.C. § 2000e-2 (2011).

behavior, workplace bullying involves *repetition*, *duration*, and *escalation*, creating an ongoing pattern of abusive behavior.<sup>39</sup> Workplace bullying is also unique in that it can consist of both covert and overt tactics.<sup>40</sup> Examples include excessive monitoring or micromanaging, being sworn at, unwarranted or invalid criticism, being humiliated and yelled at in front of others, exclusion from important meetings, social isolation, and being given unrealistic deadlines.<sup>41</sup> While an uncivil worker may be rude and boorish, this behavior is generally not targeted at anyone and is not personalized.<sup>42</sup> Conversely, bullying is a “laser-focused, systematic campaign of interpersonal destruction” that “escalate[s] in abusiveness.”<sup>43</sup> Therefore, workplace bullying goes far beyond general incivility and rudeness; it is the repeated and targeted abuse of an individual that has devastating consequences for that person.

Bullies may take the form of either a supervisor or coworker.<sup>44</sup> One study by the National Institute for Occupational Safety and Health found that bullying by coworkers was more common than bullying by bosses.<sup>45</sup> Another survey found that coworkers were bullies in forty-three percent of cases, compared to supervisor involvement in thirty-six percent of cases.<sup>46</sup> In

---

<sup>39</sup> See WHAT EVERYONE NEEDS TO KNOW, *supra* note 36, at 1; Lutgen-Sandvik, *supra* note 35, at 24.

<sup>40</sup> See WHAT EVERYONE NEEDS TO KNOW, *supra* note 36, at 3.

<sup>41</sup> See *id.* at 1; *Early Signs of Bullying*, WORKPLACE BULLYING INST., <http://www.workplacebullying.org/individuals/problem/early-signs/> (last visited Apr. 5, 2013).

<sup>42</sup> See Gary Namie, *Workplace Bullying: Escalated Incivility*, IVEY BUS. J., Nov.–Dec. 2003, at 1 [hereinafter Namie, *Escalated Incivility*].

<sup>43</sup> GARY NAMIE & RUTH NAMIE, *THE BULLY-FREE WORKPLACE: STOP JERKS, WEASELS, AND SNAKES FROM KILLING YOUR ORGANIZATION* 6 (2011) [hereinafter NAMIE & NAMIE, *BULLY-FREE WORKPLACE*]; *The WBI Definition of Workplace Bullying*, *supra* note 33.

<sup>44</sup> Press Release, Nat'l Inst. for Occupational Safety & Health, CDC, *Most Workplace Bullying Is Worker to Worker, Early Findings from NIOSH Study Suggest* (July 28, 2004), available at <http://www.cdc.gov/niosh/updates/upd-07-28-04.html>.

<sup>45</sup> *Id.*

<sup>46</sup> Loreleigh Keashly & Joel H. Neuman, *Bullying in the Workplace: Its Impact and Management*, 8 EMP. RTS. & EMP. POL'Y J. 335, 344 (2004).

2010, the WBI and Zogby International (“WBI-Zogby”) released a comprehensive survey measuring the prevalence of workplace bullying in the United States.<sup>47</sup> Based on this online survey of 2,092 adults, approximately 53.5 million Americans, or thirty-five percent of the workforce, have been bullied at work, and fifty percent have been affected by workplace bullying either as a Target or a witness to the behavior.<sup>48</sup> The WBI-Zogby survey results reveal that workplace bullying is a pervasive phenomenon with harmful effects that are widely felt by a large portion of the American workforce.<sup>49</sup>

*B. Negative Consequences of Workplace Bullying for Employees and Employers*

When bullying exists in the workplace, it can have serious economic, psychological, and emotional consequences for both the employee and employer. Targets experience psychological effects such as stress, depression, loss of sleep, and low self-esteem, as well as feelings of shame, guilt, and embarrassment.<sup>50</sup> In more severe instances, they may develop posttraumatic stress disorder, which, if left untreated, may cause an individual to react violently against either the bully or another coworker.<sup>51</sup> Targets may also manifest physical symptoms, such as stress headaches, high blood pressure, digestive problems, and even reduced immunity to infection.<sup>52</sup>

---

<sup>47</sup> The survey asked respondents, “At work, what is your experience with any or all of the following types of repeated mistreatment: sabotage by others that prevented work from getting done, verbal abuse, threatening conduct, intimidation or humiliation?” *U.S. Workplace Bullying Survey*, WORKPLACE BULLYING INST. 2 (2010), [http://workplacebullying.org/multi/pdf/WBI\\_2010\\_Natl\\_Survey.pdf](http://workplacebullying.org/multi/pdf/WBI_2010_Natl_Survey.pdf).

<sup>48</sup> *Id.* The survey had a margin of error of +/- 2.2 percentage points. *Id.* at 1.

<sup>49</sup> *See generally id.*

<sup>50</sup> NAMIE & NAMIE, BULLYPROOF YOURSELF, *supra* note 19, at 69.

<sup>51</sup> *See id.* at 69–70; *see also* Leymann & Gustafsson, *supra* note 21, at 252–54 (discussing the diagnostic criteria and symptoms of posttraumatic stress disorder).

<sup>52</sup> NAMIE & NAMIE, BULLYPROOF YOURSELF, *supra* note 19, at 70.

An employee suffering from the stress and fatigue of workplace bullying will generally become less productive and efficient.<sup>53</sup> In turn, the employer could suffer direct, indirect, and opportunity costs.<sup>54</sup> For example, employers may see a significant increase in medical and workers' compensation claims due to work-related stress as well as increased legal fees and settlement costs.<sup>55</sup> Other direct costs include hiring temporary staff to fill in for those who call in sick or eventually quit, the loss of ex-workers who take valuable company knowledge with them, and additional expenses in recruitment and training.<sup>56</sup> Moreover, an abusive environment may lead to indirect costs for the employer by creating a general atmosphere filled with "fear and mistrust, resentment, hostility, feelings of humiliation, withdrawal, play-it-safe strategies, and hiding mistakes."<sup>57</sup> Other indirect costs could include high turnover rates, poor customer service, frequent absenteeism, and acts of sabotage and revenge.<sup>58</sup> Finally, an employer may incur opportunity costs resulting from a worker's disengagement and disinterest.<sup>59</sup> For instance, a 2002 survey of 9,000 federal employees revealed that workplace harassment over a two-year period cost the U.S. government more than \$180 million in lost time and productivity.<sup>60</sup> Therefore, the effects of workplace

---

<sup>53</sup> See Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 483.

<sup>54</sup> *Id.*

<sup>55</sup> *Id.*

<sup>56</sup> See Judith Lynn Fisher-Blando, *Workplace Bullying: Aggressive Behavior and Its Effect on Job Satisfaction and Productivity* 8 (Feb. 2008) (unpublished Ph.D. dissertation, Univ. of Phx.) (citing A. NEEDHAM, *WORKPLACE BULLYING: THE COSTLY BUSINESS SECRET* (2003)) (on file with author).

<sup>57</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 483-84 (quoting EMILY S. BASSMAN, *ABUSE IN THE WORKPLACE: MANAGEMENT REMEDIES AND BOTTOM LINE IMPACT* 141 (1992)).

<sup>58</sup> See *id.* at 484 (citing BASSMAN, *supra* note 57, at 142-44).

<sup>59</sup> See *id.*

<sup>60</sup> *Workplace Bullying's High Cost: \$180 Million in Lost Time, Productivity*, ORLANDO BUS. J. (Mar. 18, 2002), <http://www.bizjournals.com/orlando/stories/2002/03/18/focus1.html?page=all>. Similarly, another survey showed that workplace bullying led to reduced employee



bullying can impose long- and short-term costs on employers resulting from lost productivity and low morale in the organization.

### *C. Current Legal Remedies to Address Workplace Bullying*

Currently, American common law does not recognize a tort of workplace bullying, and no state or federal statute directly addresses the issue either.<sup>61</sup> Existing common law and statutory remedies are insufficient to address the particular nature of workplace bullying.<sup>62</sup> For example, under Title VII of the Civil Rights Act of 1964, it is unlawful for an employer to discriminate against or harass any individual because of his or her protected status, such as race, religion, sex, or national origin.<sup>63</sup> Targets of workplace bullying, however, could be subjected to a status-blind, “equal-opportunity abusive work environment.”<sup>64</sup> In fact, workplace bullying frequently includes same-sex and same-race harassment.<sup>65</sup> Research has shown that

---

productivity and increased employee attrition. The survey revealed that “[t]wenty-eight percent lost work time avoiding the [bully], fifty-three percent lost work time worrying about [a past] incident or future interactions with the [bully] . . . forty-three percent contemplated changing jobs to avoid the [bully], and twelve percent actually changed jobs. . . .” Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 484 (quoting Christine M. Pearson, *Incivility and Aggression at Work: Executive Summary* (July 1998) (unpublished ms.)). Similarly, another survey showed that twenty-two percent “lost work time avoiding the [bully],” twenty-four percent “lost work time worrying about incidents and future interactions,” and thirty-five percent “changed jobs to avoid the [bully].” *See* Fisher-Blando, *supra* note 56, at 132.

<sup>61</sup> *See* Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 484.

<sup>62</sup> Other scholars have also argued that statutory and common law remedies are inadequate to address workplace bullying. *See, e.g.*, Michael E. Chaplin, *Workplace Bullying: The Problem and the Cure*, 12 U. PA. J. BUS. L. 437 (2010); Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27.

<sup>63</sup> *See* 42 U.S.C. § 2000e-2(a) (2011).

<sup>64</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 508.

<sup>65</sup> *See* Namie, *Escalated Incivility*, *supra* note 42, at 2.

at least half of all bullying is woman-on-woman.<sup>66</sup> Even though the Supreme Court has ruled that same-sex harassment is actionable under Title VII,<sup>67</sup> it may be difficult to prove. In addition to showing that same-sex harassment was “because of” a plaintiff’s sex,<sup>68</sup> he or she must establish that the sexual conduct was both overt and unwelcomed.<sup>69</sup> Moreover, nonsexual conduct may be “too remotely related to a tangible job benefit” to bring a prima facie case.<sup>70</sup> Therefore, unless a Target can prove that the bullying conduct was overtly sexual in nature and “because of” his or her sex, he or she cannot bring a claim for hostile work environment<sup>71</sup> and is left with no legal redress.

Professor Susan Harthill of Florida Coastal School of Law has suggested expanding the Occupational Safety and Health Act (“OSHA”) to cover bullying as a recognized workplace health and safety hazard.<sup>72</sup> OSHA requires employers to maintain a workplace free from physically harmful hazards and to “comply with occupational safety and health standards.”<sup>73</sup> At the same time, Harthill acknowledges that OSHA, in its current form, is

---

<sup>66</sup> *Id.*

<sup>67</sup> *Oncale v. Sundowner Offshore Servs., Inc.*, 523 U.S. 75, 79 (1998).

<sup>68</sup> 42 U.S.C. § 2000e-2(a) (2011). The hostile work environment doctrine mostly revolves around questions of sexual harassment. Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 511. Although the Supreme Court has not reviewed the hostile work environment doctrine in the context of racial or same-race harassment, in *Harris v. Int’l Paper Co.*, 765 F. Supp. 1509, 1512–13 (D. Me. 1991), the court recognized a racial harassment claim based on unwanted racially discriminatory conduct that created a hostile work environment. *Id.*

<sup>69</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 511 (citing Vicki Schultz, *Reconceptualizing Sexual Harassment*, 107 *YALE L.J.* 1683, 1713).

<sup>70</sup> *Id.* (quoting Schultz, *supra* note 69, at 1721).

<sup>71</sup> *See* 42 U.S.C. § 2000e-2(a) (2011). *See generally* Yamada, *supra* note 27 (arguing for status-blind legislation to combat workplace bullying).

<sup>72</sup> *See* Susan Harthill, *The Need for a Revitalized Regulatory Scheme to Address Workplace Bullying in the United States: Harnessing the Federal Occupational Safety and Health Act*, 78 *U. CIN. L. REV.* 1250, 1298–99 (2010) (arguing that OSHA should reflect hazards like workplace bullying because it is “likely to cause serious physical harm” under the general duty clause of the Act).

<sup>73</sup> 29 U.S.C. § 654(a)(1)–(2) (2011).

ineffective because (1) its monetary sanctions are not heavy enough to compel employers to prevent or combat workplace bullying in their organizations, and (2) it would be impossible for OSHA inspectors to conduct adequate investigations of every instance of workplace bullying.<sup>74</sup> Furthermore, as Yamada points out, Targets do not have a private cause of action under OSHA.<sup>75</sup>

Similarly, the common law tort of intentional infliction of emotional distress (“IIED”) does not provide an adequate response to workplace bullying. In particular, the subtle nature of workplace bullying usually does not rise to the level of “extreme and outrageous conduct” required by the tort.<sup>76</sup> Professor Michael Chaplin of California State University has suggested that courts consider tailoring IIED to bullying in the workplace because Targets suffer undeniable emotional harm.<sup>77</sup>

---

<sup>74</sup> See Harthill, *supra* note 72, at 1297. Under OSHA, the maximum fine that an employer can incur for a “willful” violation is \$70,000. *Id.*; see also Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 522; Stephen J. Beaver, Comment, *Beyond the Exclusivity Rule: Employer’s Liability for Workplace Violence*, 81 MARQ. L. REV. 103, 127–30 (1997) (arguing that OSHA is inadequate to address the issue of workplace violence).

<sup>75</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 522.

<sup>76</sup> Most courts rely upon the definition of IIED as outlined in the RESTATEMENT (SECOND) OF TORTS, which reads:

One who by extreme and outrageous conduct intentionally or recklessly causes severe emotional distress to another is subject to liability for such emotional distress, and if bodily harm to the other results from it, for such bodily harm.

Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 493 (quoting RESTATEMENT (SECOND) OF TORTS § 46(1) (1965)); see also Howell v. N.Y. Post Co., 612 N.E.2d 699, 702 (N.Y. 1993) (dismissing all IIED claims because plaintiff failed to allege conduct that was “sufficiently outrageous”); Magidson v. Wachovia Bank, NA, No. 1:07CV505, 2007 WL 4592230, at \*4 (M.D.N.C. Dec. 27, 2007) (“North Carolina courts have been reluctant to extend intentional infliction of emotional distress liability in the workplace.”); Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 493–508 (examining cases in which workplace bullying claims failed because the plaintiff failed to show “extreme and outrageous conduct”).

<sup>77</sup> See Chaplin, *supra* note 62, at 461–62.

Under the proposed modified tort of Intentional Infliction of Workplace Abuse (“IIWA”), the Target would only need to show that he or she was exposed to bullying conduct that was “intentional or reckless” consisting of “two or more negative acts on a weekly basis for at least six months,” which resulted in “mental or physical harm.”<sup>78</sup> Chaplin has argued that since tort law may readily evolve to address different claims in changing circumstances, IIWA is a more appropriate solution to workplace bullying.<sup>79</sup> Unfortunately, he admits that “courts are not inclined to adopt new causes of action.”<sup>80</sup> Moreover, Chaplin suggests IIWA as a solution partly due to the Healthy Workplace Bill’s failure to be passed.<sup>81</sup> Though IIWA is a creative solution, it may be unnecessary for courts to wait for the right factual situation to adopt the modified tort if the Healthy Workplace Bill is passed into law.

A recent case may signify a willingness among courts to utilize common law civil assault to address workplace bullying. In *Raess v. Doescher*,<sup>82</sup> the Indiana Supreme Court affirmed a jury award for civil assault<sup>83</sup> for a Target of workplace bullying. In that case, the plaintiff Doescher was a cardiovascular perfusionist during a medical procedure performed by the defendant Dr. Raess.<sup>84</sup> When the two men entered into a work-related argument, Dr. Raess’s face turned red, and with his fists balled at his side, he angrily walked towards Doescher, who

---

<sup>78</sup> *Id.* at 462–63.

<sup>79</sup> *Id.* at 466.

<sup>80</sup> *Id.*

<sup>81</sup> *Id.* at 453.

<sup>82</sup> *Raess v. Doescher*, 883 N.E.2d 790 (Ind. 2008).

<sup>83</sup> Civil assault is an act intended to cause harmful or offensive contact, causing the victim to fear such contact. *See* RESTATEMENT (SECOND) OF TORTS § 21 (1965).

<sup>84</sup> *Raess v. Doescher*, 858 N.E.2d 119, 121 (Ind. Ct. App. 2006), *vacated*, 883 N.E.2d 790 (Ind. 2008). During open heart surgeries or other heart procedures, a cardiovascular perfusionist operates circulation equipment when a patient’s circulatory or respiratory functions need to be artificially supported or temporarily replaced. *See Cardiovascular Perfusionist*, MAYO SCH. OF HEALTH SCI., <http://www.mayo.edu/mshs/careers/cardiovascular-perfusionist> (last visited Apr. 5, 2013).

backed up against the wall.<sup>85</sup> Believing that the surgeon intended to “smack the [shit] out of [him],” Doescher raised his hands in defense.<sup>86</sup> Instead, Dr. Raess walked past Doescher and exited the room after screaming, “You’re over. You’re history. You’re finished.”<sup>87</sup> At trial, Doescher testified that he “felt assaulted” by Dr. Raess’s behavior<sup>88</sup> and retained Dr. Gary Namie of the WBI to testify as an expert witness.<sup>89</sup> The Indiana Supreme Court affirmed the trial court’s decision to admit Dr. Namie’s testimony<sup>90</sup> and the jury verdict that Dr. Raess was liable for civil assault.<sup>91</sup>

Dissenting Justice Boehm, however, believed that the trial court erred in admitting Dr. Namie’s testimony.<sup>92</sup> Before trial, Dr. Raess had moved to exclude Dr. Namie’s testimony because “workplace bullying” was not a recognized tort and had no legal definition, but the trial court denied the motion without explanation.<sup>93</sup> Justice Boehm believed that without a legal context for workplace bullying, Dr. Namie’s testimony—that Dr. Raess was a “workplace abuser” and the incident was “an episode of workplace bullying”—amounted to “highly prejudicial name-calling.”<sup>94</sup> Although some commentators are hopeful that the result of *Raess* will help protect Targets from workplace bullying,<sup>95</sup> Justice Boehm’s dissent and the trial court’s decision to exclude the term “workplace bullying” from jury instructions demonstrate judicial reluctance to adopt tort relief directly addressing the phenomenon and highlight the need for workplace bullying to be legally recognized and statutorily defined.

---

<sup>85</sup> *Raess*, 858 N.E.2d at 121.

<sup>86</sup> *Id.*

<sup>87</sup> *Id.*

<sup>88</sup> *Id.*

<sup>89</sup> *Raess*, 883 N.E.2d at 801 (Boehm, J., dissenting).

<sup>90</sup> *Id.* at 797 (majority opinion).

<sup>91</sup> *Id.* at 799.

<sup>92</sup> *Id.* at 800 (Boehm, J., dissenting).

<sup>93</sup> *Id.*

<sup>94</sup> *Id.* at 801.

<sup>95</sup> See Jordan F. Kaplan, Comment, *Help Is on the Way: A Recent Case Sheds Light on Workplace Bullying*, 47 HOUS. L. REV. 141, 172–73 (2010).

## III. THE HEALTHY WORKPLACE BILL

A. *The Model Act*

In 2000, Professor Yamada proposed a model act to address workplace bullying under a theory called “Intentional Infliction of a Hostile Work Environment (IIHW).”<sup>96</sup> The new cause of action would advance the important policy goals of prevention, self-help, compensation, and punishment.<sup>97</sup> He suggested drawing upon the statutory text and case law under Title VII of the Civil Rights Act of 1964 and the elements of common law IIED as guidance for crafting a statute to address workplace bullying.<sup>98</sup> Furthermore, he argued that the hostile work environment doctrine could extend to all workers regardless of any protected status.<sup>99</sup> To address criticisms that plaintiffs might rush the courthouse with frivolous claims, Yamada argued limiting IIHW to a private cause of action because the plaintiffs’ bar would serve an effective gatekeeping function.<sup>100</sup> Presumably if a plaintiffs’ attorney represents his client on a contingency fee basis, he is less likely to bring a weak case.<sup>101</sup>

In 2004, Yamada crafted the model legislation for the Healthy Workplace Bill.<sup>102</sup> The model act’s primary policy objectives are to promote prevention and compensation while discouraging frivolous and marginal claims.<sup>103</sup> The cause of action, definitions of terms, and affirmative defenses are mostly drawn from hostile work environment doctrine and common law IIED.<sup>104</sup> The model act creates a private right of action<sup>105</sup> and

---

<sup>96</sup> Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 524–25.

<sup>97</sup> *Id.* at 524.

<sup>98</sup> *Id.*

<sup>99</sup> *Id.* at 523–24.

<sup>100</sup> *See id.*

<sup>101</sup> *See id.*

<sup>102</sup> *See* Yamada, *Crafting a Legislative Response*, *supra* note 27, at 498.

<sup>103</sup> *See id.*

<sup>104</sup> *See* Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 524–25.

<sup>105</sup> *See* Yamada, *Crafting a Legislative Response*, *supra* note 27, at 521.

makes it unlawful for an employer to subject an employee to an “abusive work environment,” defined as “when the defendant, acting with malice,<sup>106</sup> subjects the complainant to abusive conduct<sup>107</sup> so severe that it causes tangible harm to the complainant.”<sup>108</sup> The model act explicitly states that a single act would not constitute “abusive conduct,” unless it is “especially severe and egregious.”<sup>109</sup>

Furthermore, under the proposed legislation, liability is not limited to the bully as an individual.<sup>110</sup> The employer can be held vicariously liable for both an employee’s abusive conduct and bullying between coworkers.<sup>111</sup> However, employers are provided two affirmative defenses.<sup>112</sup> The first affirmative defense is available when the employer “exercised reasonable care to prevent and correctly prompt any actionable behavior,” and the employee “unreasonably failed to take advantage of appropriate preventative or corrective opportunities provided by the employer.”<sup>113</sup> The second affirmative defense is available when the employee’s “complaint is grounded primarily upon a negative employment decision made consistent with an

---

<sup>106</sup> “[M]alice” is defined as “the desire to see another person suffer psychological, physical, or economic harm without legitimate case or justification” and may be inferred from the bully’s conduct, including “outward expressions of hostility” and “harmful conduct inconsistent with an employer’s legitimate business interests,” among others. *Id.* at 518.

<sup>107</sup> “Abusive conduct” is “conduct that a reasonable person would find hostile, offensive, and unrelated to an employer’s legitimate business interests.” *Id.* In considering whether conduct is “abusive,” the trier should “weigh the severity, nature, and frequency” of the bully’s conduct, such as intimidation, humiliation, and repeated verbal abuse.” *Id.*

<sup>108</sup> *Id.* For the full text of the model act as proposed by Yamada in 2004, see *id.* at 517–21.

<sup>109</sup> *Id.* at 519.

<sup>110</sup> *Id.*

<sup>111</sup> *Id.*

<sup>112</sup> Affirmative defenses under the model act are similar to those provided to employers in sexual harassment cases. For more information, see U.S. EQUAL EMP’T OPPORTUNITY COMM’N, NO. 915.002, ENFORCEMENT GUIDANCE ON VICARIOUS EMPLOYER LIABILITY FOR UNLAWFUL HARASSMENT BY SUPERVISORS (1999), available at <http://www.eeoc.gov/policy/docs/harassment.html>.

<sup>113</sup> Yamada, *Crafting a Legislative Response*, *supra* note 27, at 520.

employer's legitimate business interests, such as a termination or demotion based on [the] employee's poor performance; or . . . [the employer's] reasonable investigation about potentially illegal or unethical activity."<sup>114</sup>

A Target has several avenues for redress under the model act.<sup>115</sup> These include reinstatement, injunctive relief, and/or monetary compensation for back pay, front pay, medical expenses, emotional distress, and attorney's fees.<sup>116</sup> Where an employer is vicariously liable for the actionable conduct of a Target's coworker, and there was no negative employment decision, emotional distress damages are capped at \$25,000.<sup>117</sup> The statute of limitations is only one year, and the Target would not be able to file for workers' compensation benefits.<sup>118</sup>

Since drafting the model act, Yamada has written extensively on theories that support legal redress for workplace bullying. In particular, he has advocated for a humanistic approach to the law that promotes dignity in the workplace.<sup>119</sup> Indeed, the model act could fill the void in existing law to reduce instances of workplace bullying and improve the health of American workers.

### *B. Current State of Anti-Workplace-Bullying Legislation*

While Yamada's model act has served as the basis for anti-workplace-bullying legislation in twenty-four states, such laws have yet to pass.<sup>120</sup> In 2003, California became the first state to introduce a version of the Healthy Workplace Bill in its state legislature, but it subsequently died in committee.<sup>121</sup> Shortly

---

<sup>114</sup> *Id.*

<sup>115</sup> *See id.* at 520–21.

<sup>116</sup> *Id.*

<sup>117</sup> *See id.* at 521.

<sup>118</sup> *Id.*

<sup>119</sup> *See Yamada, Human Dignity, supra* note 27, at 524.

<sup>120</sup> As of March 2013, twenty-four states have introduced anti-workplace-bullying legislation. HEALTHY WORKPLACE BILL, *supra* note 30.

<sup>121</sup> *See California, HEALTHY WORKPLACE BILL*, <http://www.healthyworkplacebill.org/states/ca/california.php> (last visited Apr. 5, 2013).



after, legislatures in Oklahoma,<sup>122</sup> Hawaii,<sup>123</sup> Massachusetts,<sup>124</sup> Oregon,<sup>125</sup> and Washington<sup>126</sup> followed suit. Of the twenty-four states, some have proposed bills calling for the state to study the problem of workplace bullying,<sup>127</sup> while others have limited the

---

<sup>122</sup> See *Oklahoma*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/ok/oklahoma.php> (last visited Apr. 5, 2013). The bill (H.B. 2467) was first introduced in 2004 but died in committee. The bill was reintroduced in 2007 (H.B. 1467) and 2009 (H.B. 1685) but suffered the same result. *Id.*

<sup>123</sup> See *Hawaii*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/hi/hawaii.php> (last visited Apr. 5, 2013). In 2005, 2007, and 2012, both the House and Senate of Hawaii introduced versions of the bill, and all have died in committee. *Id.*

<sup>124</sup> See *Massachusetts*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/ma/massachusetts.php> (last visited Apr. 5, 2013). Since 2005, House Representative Ellen Story has repeatedly introduced a petition for the state to study and develop a mandated program for employers to combat workplace bullying. A Joint Committee on Labor and Workforce Development held a public hearing on January 27, 2010, but no further action has been taken. On February 13, 2013, Representative Story reintroduced the bill (H.B. 1766) yet again. *Id.*

<sup>125</sup> See *Oregon*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/or/oregon.php> (last visited Apr. 5, 2013). House Representatives Jackie Dingfelder and Diane Rosenbaum introduced two versions of the Healthy Workplace Bill, which both died in committee. *Id.* In 2007, Senator Avel Gordly introduced a version of the bill, and a public hearing was held before the Senate Commerce Committee, but the bill failed to advance. In 2009, Senator Ginny Burdick introduced another bill, but it died in committee. *Id.*

<sup>126</sup> See *Washington*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/wa/washington.php> (last visited Apr. 5, 2013). In the 2005–06 legislative session, a version of the bill passed the Commerce and Labor Committee but died after never being heard by Appropriations. *Id.* In 2007, House Representatives introduced a version of the Healthy Workplace Bill designed to protect only state workers. In 2009, another version of the bill was introduced that only required policies aimed at state employees of the three regional universities to be written. In 2011, the House (H.B. 1928) and Senate (S.B. 5789) each introduced a version of the bill. *Id.*

<sup>127</sup> See, e.g., *Massachusetts*, *supra* note 124; *Connecticut*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/ct/connecticut.php> (last visited Apr. 5, 2013); *New York*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/ny/newyork.php> (last visited Apr. 5, 2013).

scope of employees who would be protected under the law.<sup>128</sup> Although not every bill introduced is identical to Yamada's model legislation, each is intended to eliminate the problem of workplace bullying.

To date, only Illinois and New York have successfully passed the bill through one chamber of their respective state legislatures.<sup>129</sup> In 2010, the Illinois Senate passed a version of the Healthy Workplace Bill; however, it died in the House Rules Committee in 2012.<sup>130</sup> In 2010, the New York Senate passed a version of the Healthy Workplace Bill, but it stalled in the State Assembly.<sup>131</sup> Recently, in February 2013, New York Assemblyman Steve Englebright and Senator Diane Savino, along with seventy-four sponsors, reintroduced the bill in their respective chambers.<sup>132</sup> Even though almost half of the states in America have introduced anti-workplace-bullying legislation, Targets remain without legal redress.

### *C. Criticisms of the Healthy Workplace Bill*

Although workplace bullying is a problem that affects approximately half of the American workforce, the Healthy Workplace Bill has failed to pass in every state legislature in which it has been introduced.<sup>133</sup> Critics have argued that the legislation's definition of bullying conduct is too vague and exposes employers to potentially unlimited liability. For example, Suzanne Lucas, author of the blog *Evil HR Lady*, opposes the restrictions on, and interference with, an employer's

---

<sup>128</sup> See *Connecticut*, *supra* note 127.

<sup>129</sup> See *Illinois*, HEALTHY WORKPLACE BILL, <http://www.healthyworkplacebill.org/states/il/illinois.php> (last visited Apr. 5, 2013); *New York*, *supra* note 127.

<sup>130</sup> See *Illinois*, *supra* note 129.

<sup>131</sup> See *New York*, *supra* note 127.

<sup>132</sup> Assemblyman Englebright introduced A.B. 4965 on February 13, 2013, and Senator Savino introduced S.B. 3863 on February 25, 2013. The Senate version of the bill was referred to the Senate Labor Committee, of which Senator Savino serves as chair. *Id.*

<sup>133</sup> See *supra* Part III.B.

business decisions.<sup>134</sup> First, she argues that anti-workplace-bullying legislation will make employers hesitant to hire employees when a claim could too easily be made for a boss “being mean,” especially because bullying behavior may be difficult to clearly define.<sup>135</sup> Second, she claims that such legislation will not provide sufficient incentive for supervisors and coworkers to stop their bullying behavior.<sup>136</sup> Finally, she argues that anti-workplace-bullying legislation would interfere with the freedom of employers and human resources managers to run their organizations without having to constantly fear that their employment decisions could lead to legal action.<sup>137</sup>

Other critics have echoed the concern about employer liability. Small business owners argue that the model act’s vague language would place them at risk of costly lawsuits.<sup>138</sup> Additionally, two Manhattan Institute researchers, Edmund McMahon and James Copland, believe that anti-workplace-bullying legislation would strike at the heart of the “employment at-will” doctrine.<sup>139</sup> The “at-will” doctrine means that “an employer is free to discharge individuals ‘for good cause, or bad cause, or no cause at all,’ and the employee is equally free to quit, strike or otherwise cease work.”<sup>140</sup> Thus, the argument goes that an anti-workplace-bullying law would essentially allow every discharged employee to bring suit against his or her

---

<sup>134</sup> See Suzanne Lucas, *Why Workplace Bullying Should Be Legal*, CBS NEWS (Mar. 23, 2011 9:00 AM), [http://www.cbsnews.com/8301-505125\\_162-44941976/why-workplace-bullying-should-be-legal/](http://www.cbsnews.com/8301-505125_162-44941976/why-workplace-bullying-should-be-legal/).

<sup>135</sup> *Id.*

<sup>136</sup> *See id.*

<sup>137</sup> *See id.*

<sup>138</sup> See, e.g., Workplace Bullying Inst., *2012 Biz Lobby Opposition to Healthy Workplace Bill*, YOUTUBE (Jan. 21, 2012), <http://www.youtube.com/watch?v=wMekFHpb018>.

<sup>139</sup> See E.J. McMahon & James Copland, Op-Ed., *New York’s Latest Job Killer: A New Bill Would Give Workers Broad Rights to File Suit When Fired*, N.Y. DAILY NEWS (Mar. 19, 2010), <http://www.nydailynews.com/opinion/new-york-latest-job-killer-new-bill-give-workers-broad-rights-file-suit-fired-article-1.446041>.

<sup>140</sup> *See id.*

former employer.<sup>141</sup> These critics argue that enacting the Healthy Workplace Bill would essentially be a “job killer.”<sup>142</sup>

Public officials are also concerned about the ramifications of workplace-bullying legislation. In 2012, soon after the New York Senate passed its version of the bill, New York City Mayor Michael Bloomberg’s administration sided with business owners and issued a statement opposing it.<sup>143</sup> Similarly, in January 2012, during a public hearing before the Washington Senate Labor, Commerce and Consumer Protection Committee, the Office of the Attorney General adamantly opposed passage of the bill, citing its vague definition of “abusive conduct” and arguing that workplace conflicts should not be resolved in courts.<sup>144</sup>

Indeed, passage of the Healthy Workplace Bill may have serious consequences for employers who have difficulties instituting adequate policies to avoid liability. One method to address these concerns is to incorporate more specific characterizations of workplace bullying into the bill’s definitions of “malice” and “abusive conduct.” For example, the bill could further define workplace bullying as conduct that is “intentional, repetitive, and escalates” over a specified period of time.

---

<sup>141</sup> *See id.*

<sup>142</sup> *Id.*

<sup>143</sup> *See* R.M. Schneiderman, *State Anti-Bully Law Would Let Workers Sue for Nastiness*, WALL ST. J. METROPOLIS BLOG (May 14, 2010, 6:18 PM), <http://blogs.wsj.com/metropolis/2010/05/14/state-anti-bully-law-would-let-workers-sue-for-nastiness/>.

<sup>144</sup> During that same hearing, Washington State Senator Janea Homquist Newbry voiced concerns over the bill’s vague terms and definitions and the subjective nature of allowing a plaintiff to sue an employer for refusing to promote him or her for any reason. *See 2012 Biz Lobby Opposition to Healthy Workplace Bill*, *supra* note 138. Similarly, in July 2012, lawmakers of the West Virginia Joint Judiciary Committee also voiced concerns over the “poorly-defined” terms that would “open doors to problems.” David Beard, *Lawmakers Question Legislation’s Proponents*, DOMINION POST, July 25, 2012, at 2-A, available at <http://www.workplacebullying.org/multi/pdf/dompost072512.pdf>. They were also concerned about how an employee’s preexisting mental and physical health issues would factor into the lawsuit and whether an employer should be liable for coworker bullying when the employer had no knowledge of the problem. *Id.*

Alternatively, lawmakers could preserve the model act's current definitions but create a legislative history that provides more insight or examples as to what constitutes workplace bullying and how employers may address it.<sup>145</sup> Since factual circumstances surrounding workplace bullying will vary from case to case, it is important that the Healthy Workplace Bill allows flexibility in interpretation. Moreover, both employers and employees should have access to legal redress beyond a lengthy and costly litigation process. They should be encouraged to pursue dispute resolution outside of court to efficiently resolve workplace-bullying disputes.

#### IV. STATE-LEVEL MEDIATION IN EMPLOYMENT DISPUTES

The subtle and unique nature of workplace bullying is often compared to sexual harassment.<sup>146</sup> Accordingly, Yamada crafted the Healthy Workplace Bill around theories and case law underlying the hostile work environment doctrine under Title VII.<sup>147</sup> Before sexual harassment law evolved in the 1980s and 1990s, the concept of sexual harassment in the workplace was often cast into doubt for its vague and broad definitions.<sup>148</sup> At that time, Professor Kingsley R. Browne, who specialized in

---

<sup>145</sup> When statutory text is ambiguous or unclear, courts will often look to legislative history for background context as authoritative evidence of the enacting legislature's "specific intent" behind the statute. William E. Eskridge, Jr., *Legislative History Values*, 66 CHI.-KENT. L. REV. 365, 370-71 (1990). Legislative history is generally composed of committee reports, floor debates, sponsor statements, and other materials. See WILLIAM N. ESKRIDGE, JR. ET AL., *CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY* 972 n.d (4th ed. 2007). For further discussion and analysis of the role of legislative history in statutory interpretation, see generally *id.*

<sup>146</sup> See Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 524-25.

<sup>147</sup> See *id.*

<sup>148</sup> See Deborah Epstein, *Can a "Dumb Ass Woman" Achieve Equality in the Workplace? Running the Gauntlet of Hostile Environment Harassing Speech*, 84 GEO. L.J. 399, 408 n.57 (1996) (outlining various statements made by men in the media fearing that they might inadvertently be liable for sexual harassment because they are confused about its definition).

employment discrimination law, argued that the law gave “little notice” of what constituted sexual harassment speech.<sup>149</sup> Indeed, much of sexual harassment case law has revolved around defining the kind of “conduct” that would create a sexually hostile environment.<sup>150</sup> Similarly, in the workplace-bullying context, contemporary scholars and commentators face the challenge of defining what constitutes “abusive conduct” that would create a status-blind hostile work environment claim.

Unfortunately, unlike sexual harassment, workplace bullying has not received federal statutory or judicial recognition.<sup>151</sup> The first necessary step towards achieving this goal is to pass state legislation and allow courts to interpret the law. In order to pass the legislation, however, drafters must make the Healthy Workplace Bill more palatable to gain sufficient support. Therefore, rather than creating a sole private right of action, drafters should include an ADR provision in the Healthy Workplace Bill, which would provide a more cost-effective and efficient alternative to an expensive and prolonged lawsuit that neither the employer nor the employee wants.

#### A. ADR: A Brief Overview

ADR processes are methods of dispute resolution that take place outside of courts.<sup>152</sup> Forms of ADR include negotiation, arbitration, mediation, summary jury trial, mini trial, and early neutral evaluation.<sup>153</sup> The goals behind ADR are to reduce court congestion, minimize cost and delay, tailor a dispute resolution process to the unique needs of each party, facilitate access to justice, and utilize a collaborative approach to dispute

---

<sup>149</sup> *Id.* at 408 (quoting Kingsley R. Browne, *Title VII as Censorship: Hostile-Environment Harassment and the First Amendment*, 52 OHIO ST. L.J. 481, 502 (1991)).

<sup>150</sup> *See id.* at 416–17.

<sup>151</sup> *See supra* Part III.

<sup>152</sup> See LEONARD L. RISKIN ET AL., *DISPUTE RESOLUTION AND LAWYERS* 1–2 (abridged 4th ed., 2009). For further discussion and analysis of ADR processes, see generally *id.*

<sup>153</sup> *See id.* at 14–16.

resolution.<sup>154</sup> In recent decades, courts and federal agencies have increasingly favored ADR processes for their efficiency and cost-effectiveness.<sup>155</sup> In fact, some ADR processes are suggested, offered, or mandated by state and federal courts.<sup>156</sup> Many commentators believe that parties obtain better quality solutions and a more satisfying outcome than they would in a trial.<sup>157</sup> This

---

<sup>154</sup> See *id.* at 11; Frank E.A. Sander, *Alternative Methods of Dispute Resolution: An Overview*, 37 U. FLA. L. REV. 1, 3 (1986).

<sup>155</sup> Known as the “ADR movement,” since the 1970s, ADR has increasingly gained support from lawyers and nonlawyers. RISKIN ET AL., *supra* note 152, at 11. In 1991, the Supreme Court in *Gilmer v. Interstate/Johnson Lane Corp.*, 500 U.S. 20 (1991), held that an employee was bound by the mandatory arbitration clause in his employment contract to arbitrate his statutory employment claim. See also *Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc.*, 473 U.S. 614, 628 (1986) (applying an arbitration agreement to antitrust claims arising under the Sherman Antitrust Act because “[b]y agreeing to arbitrate a statutory claim, a party does not forgo the substantive rights afforded by the statute”). Under the Federal Arbitration Act (“FAA”), arbitration agreements are enforceable if they are “written provision[s] in any maritime transaction or a contract evidencing a transaction involving commerce.” See 9 U.S.C. § 2 (2011). The 1996 Administrative Dispute Resolution Act (“ADRA”) permanently authorizes federal agencies to use ADR to resolve complaints filed by federal employees. 5 U.S.C. § 572 (2011). “In 2000, the EEOC required all federal agencies to establish or make available an ADR program during the pre-complaint and formal complaint stages of the EEO process.” *Federal Sector Alternative Dispute Resolution*, U.S. EQUAL EMP’T OPPORTUNITY COMM’N, <http://www.eeoc.gov/federal/adr/index.cfm> (last visited Apr. 6, 2013). Most recently, in January 2011, the United States District Court for the Southern District of New York began mandating early mediation through the court’s ADR program for all employment discrimination cases, except those filed under the Fair Labor Standards Act (“FLSA”). See Peter J. Dugan, *Southern District of New York Mandates Early Mediation in Employment Discrimination Cases*, EMP. L. ALERT (Mar. 18, 2011), <http://www.employmentlawalert.com/2011/03/articles/alternative-dispute-resolution/southern-district-of-new-york-mandates-early-mediation-in-employment-discrimination-cases/>.

<sup>156</sup> For further discussion of ADR processes in state courts, see *infra* Part V.

<sup>157</sup> See, e.g., ROBERT F. COCHRAN, JR. ET AL., *THE COUNSELOR-AT-LAW: A COLLABORATIVE APPROACH TO CLIENT INTERVIEWING AND COUNSELING* 198–202 (1999), reprinted in RISKIN ET AL., *supra* note 152, at 54–55.

Note will primarily focus on the two most commonly used ADR processes to resolve employment disputes—mediation and arbitration.<sup>158</sup>

### 1. Mediation

Actual processes can vary greatly, but generally, mediation is a voluntary, informal, and confidential process in which a neutral third party helps two or more parties resolve a dispute.<sup>159</sup> Mediators assist parties to guide the dialogue, generate options, maintain a flow of information, and agree on a resolution.<sup>160</sup> Mediation is less time-consuming than going to court because hearings often last for one day,<sup>161</sup> whereas the litigation process may not resolve a dispute for years.<sup>162</sup> Even though parties must

---

<sup>158</sup> See Susan A. FitzGibbon, *Arbitration, Mediation, and Sexual Harassment*, 5 PSYCH. PUB. POL'Y & L. 693, 697 (1999) (citing U.S. GEN. ACCOUNTING OFFICE, ALTERNATIVE DISPUTE RESOLUTION: EMPLOYERS' EXPERIENCES WITH ADR IN THE WORKPLACE, 1997 WL 709361, at \*14–15 (reporting eighty percent of employers use mediation and nineteen percent use arbitration to resolve disputes with nonunion workers)).

<sup>159</sup> RISKIN ET AL., *supra* note 152, at 14, 203–230 (discussing and analyzing models of mediation, including facilitative, transformative, and understanding-based mediation). In a facilitative mediation, the mediator guides parties to discuss the conflict and promote mutual understanding. See Briana L. Seagriff, Note, *Keep Your Lunch Money: Alleviating Workplace Bullying with Mediation*, 25 OHIO ST. J. ON DISP. RESOL. 575, 591 (2010). In a transformative mediation, mediators encourage and support the parties in improving the quality of conflict interaction to reach a positive outcome. See ROBERT A. BARUCH BUSH & JOSEPH FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT (rev. ed. 2005), reprinted in RISKIN ET AL., *supra* note 152, at 222. In understanding-based mediation, mediators help parties reach a mutually agreeable solution by encouraging parties to understand the substance of the conflict and collaboratively make decisions in the dispute resolution process. See GARY FRIEDMAN & JACK HIMMELSTEIN, CHALLENGING CONFLICT: MEDIATION THROUGH UNDERSTANDING (2008), reprinted in RISKIN ET AL., *supra* note 152, at 224–26.

<sup>160</sup> See RISKIN ET AL., *supra* note 152, at 221–22; Seagriff, *supra* note 159, at 591.

<sup>161</sup> See FitzGibbon, *supra* note 158, at 717.

<sup>162</sup> See, e.g., Gordon W. Netzorg & Tobin D. Kern, *Proportional Discovery: Making it the Norm, Rather than the Exception*, 87 DEN. U. L.



pay mediator fees, attorney's fees, and costs of acquiring a meeting room, the shorter duration of a mediation hearing leads to lower overall costs to resolve the dispute.<sup>163</sup> Moreover, many private organizations and state courts offer free mediation services.<sup>164</sup> Once the parties reach a settlement, the terms are memorialized in a signed writing and become an enforceable legal contract.<sup>165</sup>

## 2. Arbitration

Like mediation, arbitration is confidential.<sup>166</sup> Unlike mediation, however, arbitration is a more formal adjudicatory process in which an impartial third party considers evidence submitted by the parties to make a legally binding and enforceable decision.<sup>167</sup> Before an arbitration hearing, parties can jointly agree on an informal or formal discovery process.<sup>168</sup> In general, evidential and procedural rules in arbitration are more flexible than in litigation.<sup>169</sup> At the hearing, parties may present evidence as in a court of law, including witness testimony and

---

REV. 513, 522–23 (2010) (discussing how, in a federal civil case, the pretrial discovery process and motion practice can delay trial for more than two years).

<sup>163</sup> See FitzGibbon, *supra* note 158, at 717.

<sup>164</sup> For example, the New York Peace Institute is a nonprofit organization that offers free mediation services. See *Facts About Mediation*, N.Y. PEACE INST., <http://www.nypeace.org/mediation-services/> (last visited Apr. 6, 2013). The New York City Civil Court also offers free court-connected mediation. See *Resolving Your Case Through Mediation in Civil Court of the City of New York*, N.Y. STATE UNIFIED COURT SYS. (Mar. 16, 2010), <http://www.courts.state.ny.us/courts/nyc/civil/pdfs/mediation.pdf> [hereinafter *Resolving Your Case*].

<sup>165</sup> See FitzGibbon, *supra* note 158, at 702.

<sup>166</sup> Frederick L. Sullivan, *Accepting Evolution in Workplace Justice: The Need for Congress to Mandate Arbitration*, 26 W. NEW ENG. L. REV. 281, 311 (2004).

<sup>167</sup> RISKIN ET AL., *supra* note 152, at 369–70.

<sup>168</sup> See *ADR Frequently Asked Questions*, JAMS, <http://www.jamsadr.com/adr-faqs/> (last visited Apr. 6, 2013).

<sup>169</sup> See Russell D. Feingold, *Mandatory Arbitration: What Process Is Due?*, 39 HARV. J. ON LEGIS. 281, 283 (2002).

exhibits.<sup>170</sup> After the proceeding, parties may request a transcript of the hearing and file post-hearing briefs.<sup>171</sup> Public arbitration through courts is based on statutes and case law, whereas private arbitration is based on contract, either before or after the dispute has arisen.<sup>172</sup> In some instances, a party may appeal an arbitrator's decision in state court.<sup>173</sup> A party may also petition the arbitrator if he believes that a procedural mistake has been made, but typically a court will not review the merits of an arbitrator's decision.<sup>174</sup> In cases of private arbitration, an arbitrator's decision is appealable if the parties agree in advance.<sup>175</sup>

Many types of state-level mediation and arbitration programs are available. Nonprofit organizations, like JAMS, the American Arbitration Association ("AAA"), and Center for Conflict Resolution,<sup>176</sup> and many state courts offer free or low-cost mediation and arbitration services.<sup>177</sup> Additionally, many states

---

<sup>170</sup> See RISKIN ET AL., *supra* note 152, at 374.

<sup>171</sup> *Id.*

<sup>172</sup> See *id.* Most states have adopted arbitration statutes modeled after the Uniform Arbitration Act ("UAA"). *Id.* at 369.

<sup>173</sup> *Appealing Decisions*, CTR. FOR CONFLICT RESOLUTION, <https://ccr.byu.edu/content/appealing-decisions/> (last visited Apr. 6, 2013). However, time limits often apply: for example, in Utah, a party has ninety days after the arbitrator issues the decision to appeal to a state court. *Id.*

<sup>174</sup> *Id.*

<sup>175</sup> See *JAMS Optional Arbitration Appeal Procedure*, JAMS, <http://www.jamsadr.com/rules-optional-appeal-procedure/> (last visited Apr. 6, 2013).

<sup>176</sup> Many states have their own Center of Conflict Resolution, such as Minnesota and Washington, and cities, too, like Chicago. See CTR. FOR CONFLICT RESOLUTION – MINN., <http://crcminnesota.org> (last visited Apr. 6, 2013); CTR. FOR CONFLICT RESOLUTION IN CHI., <http://www.ccrchicago.org> (last visited Apr. 6, 2013); *Dispute Resolution Centers*, WASH. STATE COURTS, [http://www.courts.wa.gov/court\\_dir/?fa=court\\_dir.dispute](http://www.courts.wa.gov/court_dir/?fa=court_dir.dispute) (last visited Apr. 6, 2013).

<sup>177</sup> For example, the New York City Civil Court and the Los Angeles County Court offer free mediation programs. See *Resolving Your Case*, *supra* note 164; *Department of Consumer Affairs*, CNTY. OF L.A., <http://dca.lacounty.gov/tsMediation.html> (last visited Apr. 6, 2013). The Washington State Courts also offer Dispute Resolution Centers that provide free services or use an income-based sliding fee scale. See *Dispute Resolution*

offer court-connected arbitration programs for mandatory and voluntary arbitration hearings.<sup>178</sup> As workplace disputes are often resolved through arbitration or mediation, ADR programs may present a viable forum for addressing workplace-bullying claims.

### *B. The EEOC Mediation Program*

Since the EEOC mediation program exemplifies a government-instituted, out-of-court process that is consistent with an enacting legislature's intent to maximize ADR methods, it can provide guidance to drafters and sponsors of the Healthy Workplace Bill on how to implement a similar state-level scheme for workplace-bullying claims.<sup>179</sup> Under Title VII of the Civil Rights Act of 1964, the EEOC is obligated to investigate every charge of employment discrimination and litigate in federal court to enforce the statute.<sup>180</sup> For example, from 1997 to 2012, sexual harassment claims made up approximately thirty percent of all charges filed with the EEOC.<sup>181</sup> After the agency determines that there is reasonable cause to believe that the charge is true, it may file suit in federal court on behalf of the public interest.<sup>182</sup> However, the EEOC files less than two percent

---

*Centers, supra* note 176.

<sup>178</sup> For example, Oregon state law requires civil actions involving claims for damages of less than \$50,000 to go to arbitration. *See What Is Arbitration?*, OR. COURTS, <http://courts.oregon.gov/OJD/programs/adr/pages/whatisarbitration.aspx> (last visited Apr. 6, 2013). Arizona Superior Courts have a mandatory, nonbinding arbitration program as a component of its civil court system for disputes valued under \$50,000. *See What Is Arbitration?*, JUDICIAL BRANCH OF ARIZ., MARICOPA CNTY., <http://www.superiorcourt.maricopa.gov/SuperiorCourt/CivilDepartment/Arbitration/> (last visited Apr. 6, 2013).

<sup>179</sup> *See Mediation*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, <http://www.eeoc.gov/eeoc/mediation/> (last visited Apr. 6, 2013).

<sup>180</sup> 42 U.S.C. § 2000e-6(e) (2011); *see also* Yamada, *Status-Blind Hostile Work Environment, supra* note 27, at 529.

<sup>181</sup> *See Charge Statistics FY 1997 Through FY 2012*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, <http://www.eeoc.gov/eeoc/statistics/enforcement/charges.cfm> (last visited Apr. 6, 2013).

<sup>182</sup> *See The Charge Handling Process*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, <http://www.eeoc.gov/employers/process.cfm> (last visited Apr. 6,

of its antidiscrimination claims in federal court.<sup>183</sup> One reason that the EEOC files so few claims may be due to its incredibly successful mediation program.<sup>184</sup>

In 1991, the EEOC launched a pilot mediation program in four field offices as a response to the increasing number of charges filed with the agency.<sup>185</sup> In 1995, after the EEOC's ADR Task Force found mediation to be a successful and sustainable method of resolving employment discrimination disputes, the agency decided to fully implement the mediation program.<sup>186</sup> Since then, the mediation program has seen great success, resolving sixty to seventy-six percent of charges submitted to the EEOC each year.<sup>187</sup>

Before the EEOC investigates a discrimination charge or files suit, the agency offers parties the opportunity to participate in the mediation program to reach an out-of-court resolution.<sup>188</sup> The program is voluntary and confidential.<sup>189</sup> The program's goals are to lessen a victim's intimidation from filing a charge by providing a less expensive and contentious method for dispute resolution and to free up the EEOC's resources for investigating and litigating other employment discrimination

---

2013).

<sup>183</sup> See *EEOC v. Waffle House, Inc.*, 534 U.S. 279, 290 n.7 (2002).

<sup>184</sup> In 2012, the EEOC mediation program achieved a seventy-seven percent settlement rate. See *EEOC Mediation Statistics FY 1999 Through FY 2012*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, [http://www.eeoc.gov/eeoc/mediation/mediation\\_stats.cfm](http://www.eeoc.gov/eeoc/mediation/mediation_stats.cfm) (last visited Apr. 6, 2013).

<sup>185</sup> Matthew A. Swendiman, Note, *The EEOC Mediation Program: Panacea or Panicked Reaction?*, 16 OHIO ST. J. ON DISP. RESOL. 391, 397 (2001).

<sup>186</sup> *Id.*; see also Press Release, U.S. Equal Emp't Opportunity Comm'n, Commission Votes to Incorporate Alternative Dispute Resolution into Its Charge Processing System; Defers Decisions on State and Local Agencies (Apr. 28, 1995), available at <http://www.eeoc.gov/eeoc/newsroom/release/archive/4-28-95.html>.

<sup>187</sup> See *EEOC Mediation Statistics FY 1999 Through FY 2012*, *supra* note 184.

<sup>188</sup> See *Questions and Answers About Mediation*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, <http://www.eeoc.gov/eeoc/mediation/qanda.cfm> (last visited Apr. 6, 2013); *The Charge Handling Process*, *supra* note 182.

<sup>189</sup> See *Mediation*, *supra* note 179.

matters.<sup>190</sup> Once the parties agree to mediation, the case is assigned to a neutral mediator,<sup>191</sup> who is an internal EEOC mediator or third-party mediator contracted to mediate cases.<sup>192</sup> The mediation is free for both parties.<sup>193</sup> If the parties reach an agreement, their written and signed settlement is enforceable under contract law.<sup>194</sup> If the dispute cannot be resolved through mediation, the EEOC will resume investigation of the initial charge or file suit in federal court.<sup>195</sup>

Some scholars and commentators advocate using mediation for employment disputes,<sup>196</sup> including workplace bullying.<sup>197</sup> Because of the privacy and confidentiality in the mediation process, it is particularly suitable for resolving disputes in which parties want to preserve a long-term relationship, like an employment relationship.<sup>198</sup> When mediators encourage mutual

---

<sup>190</sup> See U.S. EQUAL EMP'T OPPORTUNITY COMM'N, NO. 915.002, EQUAL EMPLOYMENT OPPORTUNITY COMMISSION'S ALTERNATIVE DISPUTE RESOLUTION POLICY STATEMENT §§ I-II.A (2002), available at <http://www.eeoc.gov/policy/docs/adrstatement.html>.

<sup>191</sup> The Standards of Conduct for Mediators establish the professional conduct of mediation and consists of standards for mediator impartiality, confidentiality, conflict of interest, and other factors to ensure the ethical practice of mediation. See FitzGibbon, *supra* note 158, at 716 (citing John D. Feerick, *Standards of Conduct for Mediators*, 79 JUDICATURE 314 (1996)).

<sup>192</sup> See *Questions and Answers About Mediation*, *supra* note 188.

<sup>193</sup> *Id.*

<sup>194</sup> See *id.*; *The Charge Handling Process*, *supra* note 182.

<sup>195</sup> See *Facts About Mediation*, U.S. EQUAL EMP'T OPPORTUNITY COMM'N, <http://www.eeoc.gov/eeoc/mediation/facts.cfm> (last visited Apr. 6, 2013).

<sup>196</sup> See Allison Balc, *Making It Work at Work: Mediation's Impact on Employee/Employer Relationships and Mediator Neutrality*, 2 PEPP. DISP. RESOL. L.J. 241 (2002) (concluding that mediation is a cost- and time-effective alternative to litigation that employers are increasingly utilizing); FitzGibbon, *supra* note 158, at 714 (asserting that mediation offers to resolve sexual harassment disputes faster and may exert a conduct-regulating effect on the workplace).

<sup>197</sup> See Seagriff, *supra* note 159 (advocating for employers' adoption of internal mediation procedures to resolve workplace bullying disputes without going to court or losing profit and productivity).

<sup>198</sup> See FitzGibbon, *supra* note 158, at 718 (citing Lon L. Fuller, *Mediation—Its Forms and Functions*, 44 S. CAL. L. REV. 305, 307–09

understanding, provide for an open dialogue, and help generate resolution options, not only can parties resolve the current conflict situation, but they can also avoid future controversies.<sup>199</sup> Moreover, mediators are impartial third parties who have experience in the mediation process and who have special knowledge and understanding of the particular claims.<sup>200</sup> In addition, mediation does not involve credibility determinations, which could have negative consequences for future employment prospects.<sup>201</sup> Finally, parties pursuing mediation will likely expend less time and money to resolve the dispute, reducing the burden of employees with little to no resources.<sup>202</sup>

Opponents, however, argue that the mediation process could impose “undue settlement pressures” on the weaker party.<sup>203</sup> As mediation is a nonadjudicative and less formal process, the weaker party may feel intimidated and forced to accept an unfair agreement.<sup>204</sup> Moreover, the mediator may not always recognize the power imbalance between parties.<sup>205</sup> As for the EEOC mediation program, civil rights activists argued that language barriers could restrict access for some plaintiffs pursuing charges.<sup>206</sup> Since mediation is a voluntary and collaborative

---

(1971)); Seagriff, *supra* note 159, at 598.

<sup>199</sup> See RISKIN ET AL., *supra* note 152, at 222–26; Sander, *supra* note 154, at 13–14.

<sup>200</sup> See FitzGibbon, *supra* note 158, at 717 (citing Christopher A. Barreca et al., *A Due Process Protocol for Mediation and Arbitration of Statutory Disputes Arising out of the Employment Relationship*, DISP. RESOL. J., Oct.–Dec. 1995, at 37–39); Seagriff, *supra* note 159, at 596–99 (discussing how having a neutral third-party facilitator can help resolve workplace bullying disputes).

<sup>201</sup> FitzGibbon, *supra* note 158, at 718 (citing Sara Adler, *Sexual Harassment Claims Lend Themselves to Mediation*, L.A. DAILY J., Feb. 18, 1994, at 7).

<sup>202</sup> See *id.* at 717; Roselle L. Wissler, *The Effects of Mandatory Mediation: Empirical Research on the Experience of Small Claims and Common Pleas Courts*, 33 WILLAMETTE L. REV. 565, 567–68 (1997).

<sup>203</sup> See Wissler, *supra* note 202, at 573.

<sup>204</sup> *Id.*

<sup>205</sup> See *id.*

<sup>206</sup> See Swendiman, *supra* note 185, at 402–03.

process,<sup>207</sup> however, a party who feels intimidated may address these concerns with the mediator or refuse to accept the agreement altogether. Accordingly, the party may still pursue other alternatives, like litigation or arbitration, which are more formal adjudicatory processes.<sup>208</sup> Finally, concerns about language barriers impeding access to justice are valid, but they “will continue to exist regardless of the use of ADR.”<sup>209</sup>

### *C. Mediation for Workplace Bullying Claims*

Given the success of mediation in federal employment discrimination disputes,<sup>210</sup> this ADR method can also be an effective solution to workplace bullying because it emphasizes efficiency and fairness in resolving deeply personal conflicts, providing Targets closure to a very painful situation. Drs. Gary and Ruth Namie, renowned researchers of workplace bullying and authors of several books on the topic, have opposed mediation as a potential avenue for redress, arguing that there is an inherent power imbalance between the bully and his Target.<sup>211</sup> Their rejection of mediation, however, overlooks the Target’s need for a fair and efficient solution. Whereas Targets in litigation may wait years before their cases are resolved, in mediation, collaboration between the Target, bully, and employer can more quickly generate effective solutions. For example, as part of a settlement agreement, an employer could agree to discharge the bully and implement an in-house antibullying grievance procedure. This would simultaneously provide continued employment for the Target while incentivizing the employer to prevent future instances of workplace bullying.<sup>212</sup> Since settlement terms are legally enforceable, if an

---

<sup>207</sup> See *supra* Part IV.A.1.

<sup>208</sup> See also *supra* Part IV.B.

<sup>209</sup> Swendiman, *supra* note 185, at 404.

<sup>210</sup> See *supra* Part IV.B.

<sup>211</sup> NAMIE & NAMIE, BULLY-FREE WORKPLACE, *supra* note 43, at 138–39.

<sup>212</sup> See Moira Jenkins, *Practice Note: Is Mediation Suitable for Complaints of Workplace Bullying?*, 29 CONFLICT RESOL. Q. 25, 28 (2011) (“As the conflict escalates and one or more parties becomes more aggressive,

employer breaches the agreement, the Target could file suit and seek relief under contract law.

In addition, early mediation could clear up any misunderstanding or miscommunication before the bully, employer, or Target spends too much time and resources on litigation. Moreover, the collaborative nature of the mediation process, and its goal of preserving the employer-employee relationship, align with Yamada's humanitarian and dignitarian approach to crafting the Healthy Workplace Bill.<sup>213</sup> In fact, one study found that some employees supported their employers' use of in-house ADR procedures, such as mediation, to resolve workplace-bullying disputes.<sup>214</sup> Employees with fewer resources might prefer to utilize available state-level free or low-cost mediation programs. During mediation, if the Target feels "undue settlement pressures" as a weaker party, he may opt out of the process and still have the option of pursuing a more formal and adversarial process, such as arbitration or litigation.<sup>215</sup>

Since the federal government has seen immense success with ADR programs in resolving employment discrimination claims, drafters of the Healthy Workplace Bill should include language endorsing out-of-court alternatives to resolve workplace-bullying disputes. An ADR provision in the Healthy Workplace Bill could include language similar to Section 118 of the Civil Rights Act of 1991.<sup>216</sup> In Yamada's model act,<sup>217</sup> "Section 8—

---

. . . the mediator will have to be particularly aware of the power differences between the parties, . . . and the importance of follow-up built into the settlement agreement.").

<sup>213</sup> See Yamada, *Human Dignity*, *supra* note 27, at 539 (arguing that employment law focused on human dignity helps to "define both rights and responsibilities that promote healthy and productive workplaces").

<sup>214</sup> See Suzy Fox & Lamont E. Stallworth, *Employee Perceptions of Internal Conflict Management Programs and ADR Processes for Preventing and Resolving Incidents of Workplace Bullying: Ethical Challenges for Decision-Makers in Organizations*, 8 *EMPL. RTS. & EMPLOY. POL'Y J.* 375, 394-96, 398.

<sup>215</sup> See Jenkins, *supra* note 212, at 28-29.

<sup>216</sup> Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified as amended in scattered sections of 42 U.S.C.). Section 118 of the Civil Rights Act of 1991 states, "Where appropriate and to the extent



Procedures” states, “1. Private right of action. This Chapter shall be enforced solely by a private right of action.”<sup>218</sup> Drafters, however, could make the following revision:

Section 8—Procedures

1. Private right of action and alternative dispute resolution. This Chapter may be enforced by a private right of action. Where appropriate and to the extent authorized by law, the use of alternative means of dispute resolution, including settlement negotiations, mediation, and arbitration, is encouraged to resolve disputes arising under this Chapter.

It is important to recognize, however, that as state law, the Healthy Workplace Bill would not be enforced by the EEOC but rather by state fair employment agencies.<sup>219</sup> Currently, those agencies only have jurisdiction over discrimination claims,<sup>220</sup> and like the EEOC, they are overloaded and suffer substantial delays in investigating claims.<sup>221</sup> In fact, a claim investigated by the EEOC takes on average up to two years, whereas an investigation into a claim filed at some state anti-discrimination agencies may take up to twenty-two months to commence, and

---

authorized by law, *the use of alternative means of dispute resolution, including settlement negotiations, conciliation, facilitation, mediation, factfinding, minitrials, and arbitration*, is encouraged to resolve disputes arising under the Acts of provisions of Federal law amended by this title.” *Id.* (emphasis added).

<sup>217</sup> See Yamada, *Crafting a Legislative Response*, *supra* note 27, at 517–21.

<sup>218</sup> *Id.* at 521.

<sup>219</sup> Many state EEO enforcement agencies have implemented ADR programs to resolve statutory-based labor and employment disputes. Fox & Stallworth, *supra* note 214, at 383.

<sup>220</sup> Most states have a Civil Rights Commission or Department of Labor to enforce state employment discrimination laws. For more information, see *State Laws on Employment-Related Discrimination*, NAT’L CONFERENCE OF STATE LEGISLATURES (Oct. 2010), <http://www.ncsl.org/documents/employ/DiscriminationChart-III.pdf>.

<sup>221</sup> See Lamont Stallworth & Linda Stroh, *Who Is Seeking to Use ADR? Why Do They Choose to Do So?*, 51 DISP. RESOL. J. 30, 30 (1996). State anti-discrimination agencies struggle to investigate and resolve claims in a timely manner due to heavy caseloads, budget cuts, and staff reduction. *Id.*

more than four years before the agency issues a final investigatory determination.<sup>222</sup> Fortunately, many state employment agencies have readily available ADR processes to help resolve these claims at an early stage.<sup>223</sup> Accordingly, using these existing ADR procedures to resolve workplace-bullying claims is equally viable. Furthermore, state employment agencies could share this added burden with court-connected ADR programs or by contracting nonprofit ADR organizations.<sup>224</sup> Therefore, the Healthy Workplace Bill could encourage a more cost-effective and efficient resolution of workplace-bullying claims by utilizing existing state-level ADR programs. Given the lengthy investigatory procedures of state employment agencies, it is even more important that the Bill maximizes ADR procedures to resolve disputes.

#### V. STATE-LEVEL ARBITRATION IN EMPLOYMENT DISPUTES

Depending on the factual situation, Targets may prefer to use arbitration to resolve their workplace-bullying conflict if they believe that mediation would not generate a satisfactory resolution, or if the bully and employer are particularly hostile to finding a collaborative solution. While in some circumstances, Targets may prefer litigating a case, they nonetheless should be provided with the opportunity to pursue a less formal and costly process through arbitration. Currently, many workplace-related disputes are resolved through arbitration under mandatory arbitration clauses of employment contracts.<sup>225</sup> Although

---

<sup>222</sup> *Id.* at 30–31.

<sup>223</sup> For example, Indiana's Civil Rights Commission and Colorado's Civil Rights Division have Alternative Dispute Resolution Units that mediate employment discrimination claims. See *Alternative Dispute Resolution/Mediation*, COLO. DIV. OF CIVIL RIGHTS, <http://www.colorado.gov/cs/Satellite?c=Page&childpagemame=DORA-DCR%2FDORALayout&cid=1251629148334&pagemame=CBONWrapper> (last visited Apr. 6, 2013); *Alternative Dispute Resolution (ADR)*, IND. CIVIL RIGHTS COMM'N, <http://www.in.gov/icrc/2386.htm> (last visited Apr. 6, 2013).

<sup>224</sup> For a discussion of court-connected and private arbitration programs, see *infra* Part V.A.

<sup>225</sup> See Nicole Karas, Note, *EEOC v. Luce and the Mandatory Arbitration Agreement*, 53 DEPAUL L. REV. 67, 71–73 (2003); Sullivan,

arbitration is more formal and adversarial than mediation, it is less time-consuming and expensive than litigation.<sup>226</sup> Since many state-court-connected and private arbitration programs exist, Targets should have the opportunity to utilize these programs.

*A. Arbitration Programs Offered by State Courts or  
Nonprofit Organizations*

Some state courts have a court-connected ADR scheme that includes mediation and arbitration programs.<sup>227</sup> Court-connected arbitration programs vary from state to state.<sup>228</sup> Many state statutes, however, mandate court-connected arbitration for certain types of cases, such as family law matters and civil actions involving claims for damages valued at less than \$50,000.<sup>229</sup> Mandatory court-connected arbitration is intended to reduce litigation costs and mitigate docket crowding.<sup>230</sup> In some states, the judge or the parties must choose a lawyer to arbitrate the case.<sup>231</sup> After the arbitration hearing, the arbitrator will issue

---

*supra* note 166, at 296–304 (“Congress has promulgated substantive employment laws and the federal court has developed its common law, which supports arbitration as not only an alternative to litigation, but as a legitimate means of resolving employment-related disputes outside of the litigation context.”).

<sup>226</sup> For further discussion of the benefits of arbitration, see *infra* Part V.B.

<sup>227</sup> Barbara McAdoo & Nancy Welsh, *Court-Connected General Civil ADR Programs: Aiming for Institutionalization, Efficient Resolution and the Experience of Justice*, in ADR HANDBOOK FOR JUDGES 1, 1 (Donna Stienstra & Susan Yates eds., 2004), available at <http://ssrn.com/abstract=1726213>.

<sup>228</sup> *Id.*

<sup>229</sup> See, e.g., NEV. REV. STAT. § 38.250 (2012); OR. REV. STAT. § 36.400 (2011); CAL. CIV. PROC. CODE § 1141.11(a) (West 2007); 42 PA. CONS. STAT. § 7361 (2013). However, New York County’s mandatory arbitration program requires all claims seeking damages in an amount up to \$10,000 first be decided by an arbitrator. FERN A. FISHER, CIVIL COURT OF THE CITY OF N.Y., HOW TO TRY OR DEFEND A CIVIL CASE WHEN YOU DON’T HAVE A LAWYER 12 (Nov. 2005), available at <http://www.courts.state.ny.us/publications/guideforproses.pdf>.

<sup>230</sup> See *Mandatory Arbitration*, OR. STATE BAR (Aug. 2009), [http://www.osbar.org/public/legalinfo/1216\\_MandatoryArbitration.htm](http://www.osbar.org/public/legalinfo/1216_MandatoryArbitration.htm).

<sup>231</sup> See *id.* In the Washington Superior Courts, “the parties may select an

a decision and award to the prevailing party.<sup>232</sup> In some courts, a new trial can be requested.<sup>233</sup> Because many state courts impose limits on monetary awards and the types of cases eligible for mandatory court-connected arbitration,<sup>234</sup> Targets who have claims that exceed that amount or who reside in a jurisdiction that declines to hear workplace-bullying claims may opt for private arbitration.

Private arbitration procedures can be instituted by the employer<sup>235</sup> or a nonprofit organization.<sup>236</sup> Parties may choose an

---

arbitrator by stipulation”; however, if they fail to choose an arbitrator within fourteen days after the case enters the arbitration calendar, the court will select an arbitrator. WASH. SUPER. CT. MANDATORY ARB. R. 2.3. The case will then fall under the jurisdiction of the court. *See id.*; WASH. SUPER. CT. MANDATORY ARB. R. 1.3. In New York County, however, “an arbitrator is often a retired judge.” FISHER, *supra* note 229, at 12.

<sup>232</sup> *See, e.g.*, WASH. SUPER. CT. MANDATORY ARB. R. 6.1 (“The award shall be in writing and signed by the arbitrator. The arbitrator shall determine all issues raised by the pleading, including a determination of damages.”); NEV. ARB. R. 17 (“Within 7 days after the conclusion of the arbitration hearing . . . the arbitrator shall file the award with the commissioner, and also serve copies of the award on the attorneys of record, and on any unrepresented parties.”); *What Is Arbitration?*, OR. COURTS, *supra* note 178 (“The arbitrator should issue a decision within 20 days after the hearing is finished.”).

<sup>233</sup> *See, e.g.*, WASH. SUPER. CT. MANDATORY ARB. R. 7.1 (“Any aggrieved party not having waived the right to appeal may request a trial de novo in the superior [court].”); FISHER, *supra* note 229, at 12 (“If either of the parties . . . disagrees with the decision, that party has the right to demand a new trial before a Judge or jury.”); *Arbitration Guide*, JUDICIAL BRANCH OF ARIZ., MARICOPA CNTY., <http://www.superiorcourt.maricopa.gov/SuperiorCourt/CivilDepartment/Arbitration/Index.asp> (last visited Apr. 6, 2013) (“When an arbitration award is appealed, the case is sent back to the assigned judge for a new trial.”).

<sup>234</sup> *See, e.g.*, NEV. REV. STAT. § 38.250 (2012); OR. REV. STAT. § 36.400 (2011); CAL. CIV. PROC. CODE § 1141.11(a) (West 2007); 42 PA. CONS. STAT. § 7361 (2013).

<sup>235</sup> *See* Lewis Maltby, *Private Justice: Employment Arbitration and Civil Rights*, 30 COLUM. HUM. RTS. L. REV. 29, 30–32 (discussing employers’ growing use of private arbitration systems in recent decades).

<sup>236</sup> *See Arbitration Defined*, JAMS, <http://www.jamsadr.com/arbitration-defined/> (last visited Apr. 6, 2013) (“[Arbitration] is often ‘administered’ by a private organization . . .”).

impartial arbitrator who has expertise in employment disputes.<sup>237</sup> Private arbitration can be mandatory or voluntary. In a voluntary arbitration, either party may choose to initiate arbitration.<sup>238</sup> If an employment contract contains a mandatory arbitration clause, the workplace-related dispute must be resolved in private arbitration.<sup>239</sup> Moreover, arbitration can be binding or nonbinding. Most arbitration proceedings are binding and legally enforceable, and the result is appealable only in extremely limited circumstances, such as fraud or collusion.<sup>240</sup> In a nonbinding arbitration, if either party is dissatisfied with the arbitral decision, he may still file a complaint in court.<sup>241</sup>

### *B. Arguments for Arbitration to Resolve Employment Disputes*

Proponents of arbitration in workplace disputes argue that it provides a more cost-effective, timesaving, and accessible resolution. In fact, some argue that arbitration could improve the “rank-and-file employee’s” access to justice through reduced costs.<sup>242</sup> Some have even suggested that, for employees earning below \$60,000, arbitration is, unlike litigation, a “plausible dispute resolution option.”<sup>243</sup> Indeed, studies have shown that an

---

<sup>237</sup> *Id.* The AAA maintains a roster of employment arbitrators and mediators and offers arbitration services. See AM. ARBITRATION ASS’N, ALTERNATIVE DISPUTE RESOLUTION SERVICES FOR NEGOTIATED EMPLOYMENT CONTRACTS 2 (2009), available at [http://www.adr.org/aaa/ShowPDF?doc=ADRSTG\\_008048](http://www.adr.org/aaa/ShowPDF?doc=ADRSTG_008048).

<sup>238</sup> AM. ARBITRATION ASS’N, REPRESENTING YOURSELF IN EMPLOYMENT ARBITRATION: AN EMPLOYEE’S GUIDE 1 (2006), available at [http://www.adr.org/aaa/ShowPDF?doc=ADRSTG\\_004412](http://www.adr.org/aaa/ShowPDF?doc=ADRSTG_004412) [hereinafter REPRESENTING YOURSELF].

<sup>239</sup> See Feingold, *supra* note 169, at 283.

<sup>240</sup> *Arbitration Defined*, *supra* note 236; see also REPRESENTING YOURSELF, *supra* note 238, at 1, 6.

<sup>241</sup> See *Arbitration Defined*, *supra* note 236.

<sup>242</sup> Maltby, *supra* note 235, at 63. (“By reducing the costs, private arbitration holds the potential for bringing justice to many to whom it is currently denied.”).

<sup>243</sup> Alexander J.S. Colvin, *Empirical Research on Employment Arbitration: Clarity Amidst the Sound and Fury?*, 11 EMP. RTS. & EMP.

employee-plaintiff is more likely to prevail in arbitration than litigation,<sup>244</sup> and while a comprehensive study is unavailable, data suggests that legal fees in employment-related disputes can range between \$3,000 and \$14,000,<sup>245</sup> whereas litigation legal fees can cost at least \$50,000.<sup>246</sup> In addition, research by the Rand Institute showed that “the cost of employment litigation is increasing at a rate of fifteen [to] twenty-four percent per year.”<sup>247</sup> Statistics also show that arbitration proceedings usually conclude within twelve months.<sup>248</sup> By contrast, an employment-related litigation can last two-and-a-half years at the trial stage, and at least one and a half years to conclude by pretrial settlement or motion.<sup>249</sup> Moreover, arbitration protects the privacy of the parties,<sup>250</sup> which is crucial to the employee’s future prospects of employment.

Scholars have also recognized modern changes in the workplace that align with the use of private arbitration. Researchers have found that the long-term employer-employee relationship has given way to more “flexible work relations” where employees expect to have a “boundaryless career,” in which they move within and across various firms and organizations.<sup>251</sup> In arbitration, employers and employees address

---

POL’Y J. 405, 419 (2007).

<sup>244</sup> An AAA survey conducted from 1993–95 revealed that “employees who arbitrated their claims won sixty-three percent of the time,” whereas employees who litigated their claims in federal district courts prevailed only 14.9% of the time. Maltby, *supra* note 235, at 46 (citation omitted). Moreover, a survey of EEOC trials between 1974 and 1983 showed that employee-plaintiffs had a 16.8% success rate. *Id.*

<sup>245</sup> *Id.* at 54–55.

<sup>246</sup> *Id.* at 56 (“The cost of litigating an employment dispute is at least \$10,000, even if the case is resolved without trial. If a trial is required, the cost increases to at least \$50,000. Costs of this magnitude represent several years’ pay for most employees and far exceed their ability to pay under the best of circumstances.”); Sullivan, *supra* note 166, at 309.

<sup>247</sup> Maltby, *supra* note 235, at 62.

<sup>248</sup> Sullivan, *supra* note 166, at 309.

<sup>249</sup> *Id.*

<sup>250</sup> *See id.* at 311; *see also* Boyd A. Byers, *Mandatory Arbitration of Employment Disputes*, 67 J. KAN. B. ASS’N 18, 19 (1998).

<sup>251</sup> Katherine Van Wezel Stone, *Dispute Resolution in the Boundaryless*

and resolve the dispute before a third-party neutral decision-maker, who is removed from the normal chain of command.<sup>252</sup> This form of procedural justice reinforces both the employer and employee's perception of fairness and trust in the relationship.<sup>253</sup> It can also "inject an external standard of fairness" to address "abuses of hidden authority" in the workplace.<sup>254</sup>

Supporters of arbitration also contend that employers are not at an advantage in arbitration merely because they repeatedly access the service.<sup>255</sup> Known as the "repeat player effect," critics of arbitration assert this theory to demonstrate the power imbalance between an employer who routinely uses arbitration to resolve employment disputes and an employee who is accessing the service for the first time.<sup>256</sup> However, studies neither prove nor disprove the "repeat player" theory.<sup>257</sup> One study revealed that even in a highly impartial private arbitration system, employees still prevailed in forty-three percent of cases over a three-year period.<sup>258</sup> Another study found that, in mandatory private arbitration where the AAA, a nonprofit organization, did not dismiss the claim as meritless, there was no evidence of the "repeat player effect" against employees, including those of

---

*Workplace*, 16 OHIO ST. J. ON DISP. RESOL. 467, 471-72, 477 (2001).

<sup>252</sup> *Id.* at 471-72.

<sup>253</sup> *Id.* at 479-82.

<sup>254</sup> *Id.* at 487. "Hidden authority" can come in the form of cliques between coworkers or patronage networks that impose invisible authority onto the workplace. *Id.* at 486-87.

<sup>255</sup> See Sullivan, *supra* note 166, at 319-20.

<sup>256</sup> *Id.* at 319 n.235; see also Jean R. Sternlight, *Creeping Mandatory Arbitration: Is It Just?*, 57 STAN. L. REV. 1631, 1650. For further discussion of the "repeat player effect," see *infra* Part V.C.

<sup>257</sup> Professor Lisa Bingham of Indiana University conducted several studies on cases decided between 1993 and 2000, and could not empirically prove the "repeat player" theory or its causes. See Elizabeth Hill, *Due Process at Low Cost: An Empirical Study of Employment Arbitration Under the Auspices of the American Arbitration Association*, 18 OHIO ST. J. ON DISP. RESOL. 777, 785-87 (2003).

<sup>258</sup> Maltby, *supra* note 235, at 50. The study focused on a private arbitration system established by the securities industry that "has been highly criticized for its impartiality." *Id.*

middle and lower income.<sup>259</sup> Since arbitration provides a cost-effective and procedurally sound method for resolving employment disputes, Targets with valid and substantiated workplace-bullying claims should be given the opportunity to utilize it without resorting to an expensive and lengthy litigation.

### *C. Arguments Against Arbitration to Resolve Employment Disputes*

Opponents of arbitration often raise due process concerns. First, a party may only challenge an arbitral award for judicial review when arbitrators “exceed[] their powers” in interpreting law or fact, or for arbitral misconduct.<sup>260</sup> Both are rarely successful.<sup>261</sup> On the other hand, courts have found that “procedural safeguards” exist within state arbitration statutes to prevent violation of due process.<sup>262</sup> Second, opponents contend that the arbitration-related costs make the process inaccessible to employees with fewer resources.<sup>263</sup> For example, the costs of filing for arbitration and paying for the arbitrator’s hourly rates may deter them from pursuing their claim.<sup>264</sup> Since most cases

---

<sup>259</sup> Hill, *supra* note 257, at 805–09. In cases that were not deemed meritless, the win/loss ratio for employees was 0.96. *Id.* at 808.

<sup>260</sup> Stephen Wills Murphy, Note, *Judicial Review of Arbitration Awards Under State Law*, 96 VA. L. REV. 887, 898–903 (2010).

<sup>261</sup> *See id.*

<sup>262</sup> *Id.* at 903.

<sup>263</sup> Reginald Alleyne, *Arbitrators’ Fees: The Dagger in the Heart of Mandatory Arbitration for Statutory Discrimination*, 6 U. PA. J. LAB. & EMP. L. 1, 30, 30–32 (2003); Melissa G. Lamm, Comment, *Who Pays Arbitration Fees?: The Unanswered Question in Circuit City Stores, Inc. v. Adams*, 24 CAMPBELL L. REV. 93, 112 (2001)

<sup>264</sup> Employment arbitrator fees are generally based on hourly rates ranging from \$200 to \$400. Parties who use the American Arbitration Association (“AAA”) to resolve their employment disputes could pay several different kinds of fees, including a filing a fee of \$125, a postponement/cancellation fee of \$120, and administrative fees ranging from \$500 to \$13,000, depending on the scope of the claim. *See* Alleyne, *supra* note 263, at 30–31; *see also* Ryan P. Steen, Comment, *Paying for Employment Dispute Resolution: Dilemmas Confronting Arbitration Cost Allocation Throw the Arbitration Machine into Low Gear*, 7 J. SMALL & EMERGING BUS. L. 181, 182 (2003).



are disposed of in less time than in litigation, the overall expenses incurred by employee-plaintiffs would still be considerably less.<sup>265</sup> Moreover, if the claim is valid, employee-plaintiffs are more likely to recover an arbitration award in less time.<sup>266</sup> Third, opponents contend that the “inadequate” rules of discovery in arbitration may prevent the employee-plaintiff from fully uncovering evidence, since employers control pertinent information.<sup>267</sup> Equally, since arbitration discovery rules are more flexible and less well defined than federal evidence rules, an employer may also be at a disadvantage in the absence of highly relevant evidence to establish defenses.<sup>268</sup> Therefore, arbitration discovery rules pose challenges to both parties in the fact-finding process.

Critics also contend that under the “repeat player” theory, employers prevail more often because they routinely use arbitration to resolve disputes. For instance, private arbitration organizations, which have financial incentives to keep an employer’s business, are more likely to favor the employer in a proceeding.<sup>269</sup> Employers also develop continuing relationships with the same arbitrators.<sup>270</sup> However, many nonprofit arbitration

---

<sup>265</sup> See Pat K. Chew, *Arbitral and Judicial Proceedings: Indistinguishable Justice or Justice Denied?*, 46 WAKE FOREST L. REV. 185, 198 (2011) (citing a study that concluded that “arbitrations resolved disputes in a timelier manner than litigation”); see also *supra* Part V.B.

<sup>266</sup> Chew, *supra* note 265.

<sup>267</sup> Martin H. Malin, *Privatizing Justice—But by How Much? Questions Gilmer Did Not Answer*, 16 OHIO ST. J. ON DISP. RESOL. 589, 594 (2001).

<sup>268</sup> See Michael Z. Green, *Debunking the Myth of Employer Advantage from Using Mandatory Arbitration for Discrimination Claims*, 31 RUTGERS L.J. 399, 437–40 (2000) (“The right to ‘[t]ake depositions early in litigation and use the plaintiff’s own words to prove that the challenged reason [for an adverse employment decision] was nondiscriminatory’ is essential, because ‘if you know your rules of evidence, you can win a case just on evidentiary issues.’”).

<sup>269</sup> Sternlight, *supra* note 256, at 1650. Companies often enter into agreements with arbitration organizations and name them as the provider of arbitration services involving certain types of disputes. *Id.*

<sup>270</sup> Sarah Rudolph Cole, *Uniform Arbitration: “One Size Fits All” Does Not Fit*, 16 OHIO ST. J. ON DISP. RESOL. 759, 774 (2001); Malin, *supra* note 267, at 603.

organizations exist to resolve claims without consideration of future business.<sup>271</sup> And while research suggests that the “repeat player” theory does put the employee-plaintiff at some disadvantage,<sup>272</sup> the studies are inconclusive. Indeed, the “repeat player” effect also applies to litigation proceedings, as parties who are frequently in court will develop informal relations with judges.<sup>273</sup> Therefore, a continuing relationship between a party and adjudicator could equally bias litigation and arbitration proceedings.

#### *D. Arbitration for Workplace Bullying Claims*

The frequent use of arbitration to resolve labor and employment disputes<sup>274</sup> makes it susceptible to translation in the workplace-bullying context. Since the Healthy Workplace Bill is state legislation,<sup>275</sup> a Target can file for an arbitration hearing through state courts to obtain a faster and less expensive resolution.<sup>276</sup> Arbitration programs could possibly benefit short-term or part-time employees, who earn less income and cannot afford to bring suit.<sup>277</sup> If the damages are valued at less than \$50,000, many state courts could utilize their existing mandatory

---

<sup>271</sup> For example, the AAA and National Academy of Arbitrators are nonprofit arbitration organizations. See *About American Arbitration Association*, AM. ARBITRATION ASS’N., <http://www.adr.org/aaa/faces/s/about> (last visited Apr. 6, 2013); NAT’L. ACAD. OF ARBITRATORS, <http://www.naarb.org> (last visited Apr. 6, 2013).

<sup>272</sup> See, e.g., Colvin, *supra* note 243, at 428–32 (discussing studies that suggest a repeat player effect, but maintaining that more statistically significant research with larger sample sizes are necessary to prove or disprove the theory).

<sup>273</sup> See Bahaar Hamzehzadeh, Note, *Repeat Player Vs. One-Shotter: Is Victory All That Obvious?*, 6 HASTINGS BUS. L.J. 239, 243–44 (2010) (“The heightened level of familiarity with institutional actors allows repeat players to occasionally disobey court rules or obtain information that is not readily accessible to the public.”). For further discussion of the “repeat player effect” in litigation, see generally *id.*

<sup>274</sup> See FitzGibbon, *supra* note 158, at 697.

<sup>275</sup> See *supra* Part III.

<sup>276</sup> See *supra* Part V.A.

<sup>277</sup> See *supra* Part V.A.

arbitration programs to resolve the dispute.<sup>278</sup> Otherwise, if damages are valued at more than \$50,000, Targets can still elect to pursue arbitration with nonprofit organizations.<sup>279</sup> Those with fewer resources and who want to move past the conflict as quickly as possible should only choose to file suit as a last resort given the costs associated with litigation. Moreover, nonprofit arbitration organizations are widely available to provide impartial proceedings at affordable rates. Finally, procedural safeguards in state arbitration statutes exist to protect Targets from an abuse of due process in proceedings. In the event that a bully or employer refuses to utilize arbitration, Targets are still afforded a private right of action under the Healthy Workplace Bill.

## VI. CONCLUSION

Workplace bullying is a real and serious problem affecting millions of workers.<sup>280</sup> It is therefore crucial that the Healthy Workplace Bill be passed into law. Adding an ADR provision to the Bill would ease state legislators' and opponents' concerns about plaintiffs "flooding" courts with frivolous claims and exposing employers to unlimited liability. Since workplace-bullying incidents are very fact-specific,<sup>281</sup> allowing both litigation and ADR procedures as potential avenues of relief will give Targets and employers more flexibility to resolve disputes. Mediation and arbitration could resolve disputes quickly and with less expense, which is important for most employees and for small employers with fewer resources. Parties who pursue ADR would also reduce the likelihood that state court dockets become overloaded with workplace bullying claims.<sup>282</sup> To address workplace-bullying claims through ADR processes, states could utilize existing institutions and programs, such as private ADR organizations, the state labor department, or court-

---

<sup>278</sup> See *supra* note 229.

<sup>279</sup> See *supra* Part V.A.

<sup>280</sup> See *supra* Part II.

<sup>281</sup> See *supra* Part II.

<sup>282</sup> See *supra* Part IV.

connected mediation and arbitration programs.<sup>283</sup> Therefore, the Healthy Workplace Bill's dual purposes of maintaining dignity in the workplace<sup>284</sup> and preventing workplace bullying<sup>285</sup> could equally be achieved through ADR procedures and should not be limited solely to a private cause of action.

---

<sup>283</sup> See *supra* Parts IV–V.

<sup>284</sup> See Yamada, *Human Dignity*, *supra* note 27 (arguing that employment law focused on human dignity helps define both rights and responsibilities that promote healthy and productive workplaces); Yamada, *Therapeutic Jurisprudence*, *supra* note 27 (arguing that therapeutic jurisprudence, which focuses on the law's impact on emotional life and psychological well-being, should play an important role in promoting a "dignitarian" framework in shaping employment law).

<sup>285</sup> See Yamada, *Status-Blind Hostile Work Environment*, *supra* note 27, at 492–93.

# **“EYE SEE YOU”: HOW CRIMINAL DEFENDANTS HAVE UTILIZED THE NERD DEFENSE TO INFLUENCE JURORS’ PERCEPTIONS**

*Sarah Merry\**

*For the great enemy of truth is very often not the lie—deliberate, contrived and dishonest—but the myth—persistent, persuasive and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought.<sup>1</sup>*

Eyeglasses are “one of the most important artifacts used in the courtroom.”<sup>2</sup> In 2012, a defendant’s use of eyeglasses at trial went to appeal in the District of Columbia Court of Appeals in *Harris v. United States*.<sup>3</sup> “[A]t the heart of” the appeal was whether the defendant’s rights were prejudiced by the Superior Court’s issuing a change-of-appearance instruction,<sup>4</sup> prompted by

---

\* J.D. Candidate, Brooklyn Law School, 2014; B.S., Santa Clara University, 2009. I would like to thank my parents for their continued support and encouragement. I would also like to offer a special thanks to Professor Kathleen Darvil for her thoughtful suggestions and research advice, and to the staff of the *Journal of Law and Policy* for their careful edits and insightful comments.

<sup>1</sup> President John F. Kennedy, Yale University Commencement Address 2 (June 11, 1962) (transcript available in the National Archives), *available at* <http://research.archives.gov/description/193922>.

<sup>2</sup> LAWRENCE J. SMITH & LORETTA A. MALANDRO, *COURTROOM COMMUNICATION STRATEGIES* § 1.21, at 42 (1985).

<sup>3</sup> *See Harris v. United States*, No. 08-CF-1405, at 4–6 (D.C. Cir. 2012), *available at* <http://legaltimes.typepad.com/files/moj-08-cf-1405.pdf>.

<sup>4</sup> A change-of-appearance instruction is given to a jury in circumstances in which a defendant has changed his or her appearance after the commission

the defendant's use of unnecessary eyeglasses at trial.<sup>5</sup> This was the first time that the appeals court considered a defendant's instructional challenge to a change-of-appearance instruction issued solely because the defendant donned unnecessary eyewear at trial.<sup>6</sup> The court of appeals upheld the change-of-appearance instruction and determined that the evidence supported the instruction because the defendant had, among other things, donned unnecessary eyeglasses.<sup>7</sup> Importantly, the defendant's

---

of a crime and such alteration in appearance may be considered by the jury as "an indication of the defendant's awareness of guilt and fear of identification." See comments to BARBARA BERGMAN, CRIMINAL JURY INSTRUCTIONS FOR THE DISTRICT OF COLUMBIA § 2.303(B) (5th ed. 2008). The change-of-appearance instruction issued in *Harris* states:

You heard evidence that Donnell Harris attempted to change his appearance to avoid being identified. It is up to you to decide that he took these actions. If you find he did so, you may consider this evidence as tending to show his feelings of guilt which you may in turn consider as tending to show actual guilt. On the other hand, you may also consider that he may have taken these actions for reasons fully consistent with innocence in this case. If you find that Donnell Harris attempted to change his appearance to avoid being identified, you should consider such evidence along with all the other evidence in this case and give it as much weight as you think it deserves.

Trial Transcript at 87–88, *United States v. Harris*, No. CF1-18801-07 (D.C. Super. Ct. 2008), reprinted in Appellant's Limited Appendix, *Harris*, No. 08-CF-1405.

<sup>5</sup> Zoe Tillman, *Glasses an Issue in Appeal over Public Defender Intern's Slaying*, BLT: BLOG OF LEGAL TIMES (Sept. 13, 2012, 1:32 PM), <http://legaltimes.typepad.com/blt/2012/09/glasses-an-issue-in-appeal-over-public-defender-interns-slaying.html%20/> (highlighting Donnell Harris' use of eyeglasses at trial, "a seemingly innocuous detail [that was] a key issue at the heart of Harris' appeal").

<sup>6</sup> Compare Brief for Appellee at 32, *Harris*, No. 08-CF-1405 (noting that the D.C. Court of Appeals has never directly considered a defendant's challenge to a change-of-appearance instruction prompted solely by a defendant's use of eyewear at trial), with *United States v. Carr*, 373 F.3d 1350, 1353 (D.C. Cir. 2004) (rejecting a defendant's instructional challenge to a change-of-appearance instruction because the lower court considered a combination of the defendant's beard, weight, and eyeglasses to equate to "profound alterations" in appearance and therefore justifying the resulting jury instruction).

<sup>7</sup> *Harris*, No. 08-CF-1405, at 5 (quoting the trial court, which informed

identity was a key issue in the case.<sup>8</sup> However, the court of appeals left open a critical question: can a court issue a change-of-appearance instruction if a defendant wears nonprescriptive eyeglasses to trial when the defendant's identity is not specifically at issue? This tactic is known as the "nerd defense"—a persistent and unrealistic change in one's appearance aimed at persuading a jury of the defendant's low propensity to commit a crime.<sup>9</sup> The court in *Harris* highlighted the importance of the "glasses issue" by observing that an increasing number of defendants have appeared at trial wearing nonprescriptive eyeglasses.<sup>10</sup>

Evidence concerning a defendant's appearance is rarely admitted because it is often considered more prejudicial to the defendant than it is probative.<sup>11</sup> However, it is well documented

---

the jury that "there is no evidence in the record that Mr. Harris needs glasses to read or anything else" and found that Harris' explanation for the use of eyeglasses was mere "speculation"); *see also* Brief for Appellee, *supra* note 6, at 30 n.32.

<sup>8</sup> While one witness heard "'two muffled gunshots' and 'could see the gun being held and . . . most of the [shooter's] arm,'" the witness did not see the shooter's face. Another witness also heard the gunshots but did not see the shooter. *Harris*, No. 08-CF-1405, at 2.

<sup>9</sup> Keith L. Alexander, *Trendy, Non-Prescription Eyewear Latest in Criminal Defendant Strategic Attire*, WASH. POST (Mar. 27, 2012), [http://www.washingtonpost.com/local/crime/trendy-non-prescription-eyewear-latest-in-criminal-defendant-strategic-attire/2012/03/17/gIQA62xJeS\\_story.html](http://www.washingtonpost.com/local/crime/trendy-non-prescription-eyewear-latest-in-criminal-defendant-strategic-attire/2012/03/17/gIQA62xJeS_story.html) (quoting New York defense lawyer Harvey Slovis, who encourages "all his clients [to] wear glasses" to make them "appear less intimidating"); *see also* Michael J. Brown, *Is Justice Blind or Just Visually Impaired? The Effects of Eyeglasses on Mock Juror Decisions*, JURY EXPERT, Mar. 2011, at 1, 3, available at [http://www.thejuryexpert.com/wp-content/uploads/2011/03/TJEVol23Num2\\_Mar2011.pdf](http://www.thejuryexpert.com/wp-content/uploads/2011/03/TJEVol23Num2_Mar2011.pdf) (finding that defendants who wear eyeglasses appear to be more intelligent and less physically threatening).

<sup>10</sup> Tillman, *supra* note 5 (noting that Chief Judge Eric Washington found the eyeglasses issue particularly "compelling" because "a growing number of defendants had been showing up for trial wearing glasses").

<sup>11</sup> *See* FED. R. EVID. 404 advisory committee's notes ("Character evidence is of slight probative value and may be very prejudicial. It tends to distract the trier of fact from the main question of what actually happened on the particular occasion. It subtly permits the trier of fact to reward the good man and to punish the bad man because of their respective characters despite what the evidence in the case shows actually happened." (quoting CAL. LAW

that juries do consider a defendant's appearance at trial.<sup>12</sup> Lisa Wayne, President of the National Association of Criminal Defense Lawyers, stated, "the bottom line is we know people judge a book by its cover," a fact that implicates "the fundamental fairness process."<sup>13</sup> A defendant who intends to mislead the jury with respect to his or her altered appearance—for instance, by wearing nonprescriptive eyeglasses to trial—circumvents character evidence rules<sup>14</sup> by unofficially introducing into evidence positive character traits associated with eyeglasses (e.g., intelligence, honesty, decreased propensity to commit a violent crime).<sup>15</sup> Unless a defendant's identification is

---

REVISION COMM'N, TENTATIVE RECOMMENDATION AND A STUDY RELATING TO THE UNIFORM RULES OF EVIDENCE (ART. VI. EXTRINSIC POLICIES AFFECTING ADMISSIBILITY) 615 (1964)).

<sup>12</sup> See, e.g., Mark S. Brodin, *Accuracy, Efficiency, and Accountability in the Litigation Process—The Case for the Fact Verdict*, 59 U. CIN. L. REV. 15, 48 n.137 (1990) ("[P]hysical appearance of the litigant . . . can influence the verdict."). See generally Brown, *supra* note 9 (discussing a study linking juror perceptions of defendant appearance with the likelihood of guilty or not-guilty verdicts); Annie Murphy Paul, *Judging by Appearance*, PSYCHOL. TODAY (Nov. 1, 1997), <http://www.psychologytoday.com/articles/200909/judging-appearance> (commenting that the influence of a defendant's appearance inside the courtroom is so great that an "entire industry has emerged to advise lawyers, plaintiffs, and defendants on their aesthetic choices").

<sup>13</sup> Jenny Montgomery, *Dressing Defendants*, IND. LAW. (May 23, 2012), <http://www.theindianalawyer.com/article/print?articleId=28848> (quoting Lisa Wayne).

<sup>14</sup> See FED. R. EVID. 404(a)(1) (stating that "[e]vidence of a person's character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait"). A defendant's appearance serves as a "substitute for any real discussion of character during a trial." Josephine Ross, *"He Looks Guilty": Reforming Good Character Evidence to Undercut the Presumption of Guilt*, 65 U. PITT. L. REV. 227, 231 n.10 (2004).

<sup>15</sup> See SMITH & MALANDRO, *supra* note 2, § 1.21, at 42 (suggesting that defendants should wear eyeglasses with wider lenses because people with a "wide-eyed look or open-eyed look are considered to be more trustworthy, more likable, and oftentimes more innocent"); Brown, *supra* note 9, at 3 (discussing a study finding that defendants who wear eyeglasses appear more intelligent and less physically threatening); see also CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, 1 FEDERAL EVIDENCE § 4:23 (3d ed.



relevant at trial, wearing unnecessary eyeglasses to artificially alter appearance will not *officially* put a defendant's appearance at issue.<sup>16</sup> Under the Federal Rules of Evidence, when a defendant's identification is not specifically at issue, the prosecution is unable to counter this unofficial introduction of character evidence.<sup>17</sup> Allowing a defendant to purposefully falsify a vision defect by wearing nonprescriptive eyeglasses is akin to allowing a defendant to appear in a wheelchair before the jury when he or she is perfectly mobile.<sup>18</sup> Both actions fabricate

---

2009) (stating that “[c]haracter embraces qualities like honesty or dishonesty, [and] being peaceful or prone to violence”).

<sup>16</sup> Generally, *only* when a defendant's identification is a relevant trial issue may a prosecutor comment on the defendant's change in appearance from the time that the crime was committed to the time that the defendant appears at trial. GARY MULDOON, *HANDLING A CRIMINAL CASE IN NEW YORK* § 9:262 (2012–2013); *see also* *People v. Sanders*, 622 N.Y.S.2d 986, 987 (App. Div. 1995) (reasoning that the prosecutor's comment on the defendant's change in hairstyle was not prejudicial, in part, because the defendant's identification was a factor in the trial).

<sup>17</sup> A defendant must first introduce evidence of his or her pertinent character traits at trial, and only when such evidence is *officially* admitted may the prosecution offer evidence to rebut it. *See* FED. R. EVID. 404(a)(2)(A); *see also* FED. R. EVID. 405 (stating that proof of “character” at trial through instances of specific conduct is limited to situations in which “a person's character or character trait is an essential element of a charge, claim, or defense” or is otherwise admissible). However, when a defendant's identity is contentious at trial, eyeglasses may serve as a disguise, thereby hindering identification and thus making a defendant's use of eyeglasses admissible as evidence relevant to the case. *See* Steve D. Charman & Gary L. Wells, *Eyewitness Lineups: Is the Appearance-Change Instruction a Good Idea?*, 31 *LAW & HUM. BEHAV.* 3, 5 (2007), *available at* [http://www.psychology.iastate.edu/~glwells/wells\\_articles\\_pdf/charman&wells\\_appearance\\_change.pdf](http://www.psychology.iastate.edu/~glwells/wells_articles_pdf/charman&wells_appearance_change.pdf) (noting that disguises typically involve the addition of items such as hats, eyeglasses, or masks, and observing the “strong debilitating effect of disguise on accurate recognition rates”); *District of Columbia v. Carter*, No. 2010 CF1 005677 (D.C. Super. Ct. 2012) (offering an example of a prosecutor questioning a defendant's sudden use of eyeglasses at trial when the defendant's identification is at issue).

<sup>18</sup> *See* Jack Marshall, *The Perplexing “Nerd Defense,”* *ETHICS ALARMS* (Feb. 21, 2011, 8:18 AM), <http://ethicsalarms.com/2011/02/21/the-perplexing-nerd-defense/#more-4880> (“Glasses convey information about physical capabilities the same way coming into court on crutches or in a wheelchair does.”).

a physical disability in order to mislead the jury. Wearing nonprescriptive eyeglasses can be deceptive because jurors might not be able to discern whether a defendant truly requires eyeglasses,<sup>19</sup> and a defendant's unnecessary use of eyeglasses may subtly persuade the jury by playing upon one of society's most deeply rooted stereotypes: that wearing eyeglasses equates to higher intelligence.<sup>20</sup> Additionally, the jury may never even consider the motive behind the defendant's use of such a prop.<sup>21</sup> Such intentional misdirection undermines the truth-seeking principles of the judicial system.<sup>22</sup>

Part I of this Note focuses on how the wearing of eyeglasses significantly affects the way an individual is perceived and briefly examines the influence of popular culture on the deeply ingrained stereotype that wearing eyeglasses correlates to increased intelligence. Part II analyzes the unofficial role of a defendant's appearance in the courtroom and discusses cases that highlight the impact of a defendant's appearance on criminal

---

<sup>19</sup> See generally Brown, *supra* note 9, at 3 (discussing a controlled study of jurors' perceptions of eyeglasses); Alexander, *supra* note 9 (finding that eyeglasses often "escape notice").

<sup>20</sup> See Åke Hellström & Joseph Tekle, *Person Perception Through Facial Photographs: Effects of Glasses, Hair, and Beard on Judgments of Occupation and Personal Qualities*, 24 EUR. J. SOC. PSYCHOL. 693, 695 (articulating that judgments about intelligence and success can be traced back to the development of myopia caused by extensive schoolwork in childhood days); see also Brown, *supra* note 9, at 3 (finding that eyeglasses have a positive correlation to increased intelligence in juror perceptions of defendants); Francine C. Jellesma, *Do Glasses Change Children's Perceptions? Effects of Eyeglasses on Peer- and Self-Perception*, EUR. J. DEVELOPMENTAL PSYCHOL. 1, 5 (2012) (arguing that "the association between eyeglasses and intelligence is part of the nerd stereotype" because "almost 50% of the people think eyeglasses are part of the physical appearance of nerds").

<sup>21</sup> Wolfgang Manz & Helmut E. Lueck, *Influence of Wearing Glasses on Personality Ratings: Crosscultural Validation of an Old Experiment*, 27 PERCEPTUAL & MOTOR SKILLS 704, 704 (1968) (describing the wearing of eyeglasses as "an irrelevant cue," which may lead jurors to be unconsciously persuaded by a defendant's use of unnecessary eyeglasses).

<sup>22</sup> See Franklin Strier, *Making Jury Trials More Truthful*, 30 U.C. DAVIS L. REV. 95, 99 (1996) ("None of the trial's functions are more central to its legitimacy than the search for truth.").

trials despite the inadmissibility of such character evidence. Part III introduces the so-called “nerd defense” and examines the effect of a defendant’s use of eyeglasses on juror perceptions. This section further explores the use of the nerd defense by criminal defendants to purposefully mislead jurors. Part IV examines the intersection of the nerd defense and change-of-appearance instructions. Additionally, this section criticizes the *Harris* opinion for failing to address the use of eyeglasses for the purpose of jury persuasion. Part V of this Note acknowledges the need to balance a defendant’s right of free expression against the potential for jury manipulation and proposes two solutions: (1) imposing a modified change-of-appearance instruction that removes language relating to consciousness of guilt and (2) allowing for an eyeglasses inquiry when a defendant suddenly dons eyeglasses at trial.

#### I. THE PSYCHOLOGY BEHIND EYEGLASSES

The National Eye Institute reports that, as of 2008, more than 150 million Americans used corrective eyewear.<sup>23</sup> Eyeglasses are specifically designed to correct “congenital or acquired vision deficits such as myopia, presbyopia, or astigmatism.”<sup>24</sup> However, it is clear that wearing eyeglasses is no longer only for those with vision deficiencies.<sup>25</sup> As of 2011, the Vision Council estimates that approximately sixteen million Americans wear nonprescriptive eyeglasses for the purpose of changing their appearances.<sup>26</sup>

---

<sup>23</sup> PREVENT BLINDNESS AM. & NAT’L EYE INST., VISION PROBLEMS IN THE U.S. 12 (2008), available at [http://www.preventblindness.net/site/DocServer/VPUS\\_2008\\_update.pdf](http://www.preventblindness.net/site/DocServer/VPUS_2008_update.pdf).

<sup>24</sup> Helmut Leder et al., *The Glasses Stereotype Revisited: Effects of Eyeglasses on Perception, Recognition, and Impression of Faces*, 70 SWISS J. PSYCHOL. 211, 211 (2011).

<sup>25</sup> *See id.* (describing how eyeglasses not only serve to correct eyesight but also function as facial accessories that are linked to fashion demands); *see also* ESSILOR OF AM. & LUXOTTICA GRP., 20/20 OPTICIANS’ 2008 HANDBOOK 4–5 (4th ed. 2008) (observing that eyeglasses can be useful for those who wish to project their individualism or who simply desire to appear fashionable).

<sup>26</sup> Michelle Healy, *Prescription Eyeglass Frames Get Softer Look*, USA

*A. The Impact of the Eye Region on Perception of the Face*

Humans, from their earliest stages of life, are drawn to the eye region.<sup>27</sup> In fact, infants recognize eyes before they are able to recognize faces.<sup>28</sup> The eyes play a critical role in developing perceptions of the face.<sup>29</sup> The eye region is also fundamental to nonverbal communication because emotions, attention, and intentions are all perceived through observing one's eye gaze.<sup>30</sup> For example, wide-open eyes signal the emotions of surprise and fear.<sup>31</sup> A study designed specifically to measure the relative time a subject looks at the eye region during a "social impression-formation task" revealed that eyes are the facial feature that people spend the most time analyzing.<sup>32</sup> When presented with static facial displays,<sup>33</sup> subjects spent 43.4% of their visual inspection time on the eye region and only 12.6% of their visual inspection time on the mouth region.<sup>34</sup> The social impression-formation task is pertinent in a courtroom setting because a

---

TODAY (Sept. 15, 2011), [http://usatoday30.usatoday.com/LIFE/usaedition/2011-09-15-eyeglass-frames\\_ST\\_U.htm](http://usatoday30.usatoday.com/LIFE/usaedition/2011-09-15-eyeglass-frames_ST_U.htm).

<sup>27</sup> Jellesma, *supra* note 20, at 2.

<sup>28</sup> M.J. Taylor et al., *Eyes First! Eye Processing Develops Before Face Processing in Children*, 12 NEUROREPORT 1671, 1676 (2001).

<sup>29</sup> Dan Nemrodov and Roxane J. Itier, *The Role of Eyes in Early Face Processing: A Rapid Adaptation Study of the Inversion Effect*, 102 BRIT. J. PSYCHOL. 783, 793 (2011).

<sup>30</sup> R.J. Itier & M. Batty, *Neural Bases of Eye and Gaze Processing: The Core of Social Cognition*, 33 NEUROSCIENCE & BIOBEHAVIORAL REVS. 843, 844 (2009) (noting that "the human brain has developed a very complex cognitive system of gaze direction analysis based on perceptual elements of faces and eyes").

<sup>31</sup> *Id.* at 845.

<sup>32</sup> Stephen W. Janik et al., *Eyes as the Center of Focus in the Visual Examination of Human Faces*, 47 PERCEPTUAL & MOTOR SKILLS 857, 857-58 (1978); *see also* Leder et al., *supra* note 24, at 211 (noting that eyes are located in a "prominent position in the visual field").

<sup>33</sup> Static facial displays depict no movement in the facial region of the person shown in the slide. Janik et al., *supra* note 32, at 858.

<sup>34</sup> Subjects spent a greater portion of their looking time on the eye region as compared to the hair, nose, ear, or mouth regions, regardless of the facial expression or sex of the person depicted in the slide. *Id.*

defendant's facial features will be seemingly static to the jury.<sup>35</sup> Generally, if a defendant does not take the stand to speak in his or her own defense,<sup>36</sup> there will not be occasion for prolonged interaction between a defendant's eyes and mouth that may affect the viewer's primary focus on the defendant's eye region.<sup>37</sup> Because eyeglasses significantly alter the appearance of the eye region, wearing eyeglasses impacts the type of social information that is perceived through facial processing.<sup>38</sup>

*B. Studies Concerning the Effect of Eyeglasses on Judgment and Perception*

Social information about others is gleaned through facial processing, and "even the briefest of glances at a face is sufficient to furnish a wealth of knowledge about its owner."<sup>39</sup> To form judgments and perceptions of others, people rely heavily on their cognitive representations (schemata).<sup>40</sup> Collectively shared schemata can be described as widely held

---

<sup>35</sup> Compare *id.* at 857–58 (discussing a study conducted by presenting the subjects with various slides depicting static faces), with SMITH & MALANDRO, *supra* note 2, § 1.12, at 22 (noting that "[m]ost of the time jurors are watching what is referred to as 'static' facial behavior in the courtroom").

<sup>36</sup> See Stephen J. Schulhofer, *Some Kind Words for the Privilege Against Self-Incrimination*, 26 VAL. U. L. REV. 311, 329–30 (1991) (describing a sample study of Philadelphia felony defendants that revealed that only half of criminal defendants took the stand); Scott E. Sundby, *The Capital Jury and Absolution: The Intersection of Trial Strategy, Remorse, and the Death Penalty*, 83 CORNELL L. REV. 1557, 1561 (1998) (finding that "most defendants [in California capital jury trials] did not testify"). But see HARRY KALVEN, JR. & HANS ZEISEL, *THE AMERICAN JURY* 146–48 (1966) (finding that the percentage of defendants who testify depends on the defendants' prior records and the amount of evidence against the defendants).

<sup>37</sup> See Janik et al., *supra* note 32, at 858 (concluding that the study does not determine the extent to which a subject's primary focus would change due to a person's eye and mouth movements during prolonged inspection by the subject).

<sup>38</sup> Jellesma, *supra* note 20, at 2.

<sup>39</sup> K.A. Quinn & C.N. Macrae, *Categorizing Others: The Dynamics of Person Construal*, 88 J. PERSONALITY & SOC. PSYCHOL. 467, 476 (2005).

<sup>40</sup> *Id.* at 467.

social stereotypes.<sup>41</sup> Character traits are associated with schemata, and, when one schema is activated, “the associated traits are attributed to the target person in the form of a first impression.”<sup>42</sup> Eyeglasses greatly impact both the perception and recognition of others because they frame the eyes and make more distinct the facial region found to receive the most notable fixation.<sup>43</sup>

Numerous studies demonstrate that a perceived correlation between wearing eyeglasses and heightened intelligence develops in early childhood and continues to strengthen with age. This perception also exists among children who wear eyeglasses themselves, suggesting that some children, through their own experiences, might learn to associate myopia with intelligence.<sup>44</sup> Sarah Sandow, Reader in Education at the West London Institute, conducted a study revealing that children as young as eight years old draw a connection between wearing eyeglasses and possessing intelligence.<sup>45</sup> Children ages eight to ten consistently drew a “very clever” person with eyeglasses but did not do the same for stupid or nasty people.<sup>46</sup> Hannu Rätty and Leila Snellman, professors at the University of Joensuu in Finland, led a similar study that asked children to draw an “intelligent” person and found that children consistently drew eyeglasses in their images.<sup>47</sup> However, when asked to draw an

---

<sup>41</sup> Roger L. Terry & John H. Krantz, *Dimensions of Trait Attributions Associated with Eyeglasses, Men’s Facial Hair, and Women’s Hair Length*, 23 J. APPLIED SOC. PSYCHOL. 1757, 1757 (1993).

<sup>42</sup> *Id.*

<sup>43</sup> Leder et al., *supra* note 24, at 221.

<sup>44</sup> Jeffrey J. Walline et al., *What Do Kids Think About Kids in Eyeglasses?*, 28 OPHTHALMIC & PHYSIOLOGICAL OPTICS 218, 223 (2008) (noting that another origin of children’s development of the stereotype that wearing eyeglasses equates to higher intelligence could be the media’s depictions of “intelligent-nerds”).

<sup>45</sup> See Sarah Sandow, *The Good King Dagobert, or Clever, Stupid, Nice, Nasty*, 12 DISABILITY & SOC’Y 83, 86–91 (1997) (commenting that “[i]t was fascinating that the wearing of glasses has survived as a stereotype for cleverness” and that “spectacles lend an air of dignity and bookishness, and the wearers are cool and confident”).

<sup>46</sup> *Id.* at 91–92.

<sup>47</sup> Hannu Rätty & Leila Snellman, *Children’s Images of an Intelligent*

“ordinary” person, children rarely sketched a person with eyeglasses.<sup>48</sup> David Chambers, Director of the Sciences in Society Centre at Deakin University, developed the well-known Draw-a-Scientist-Test (“DAST”), designed to determine when children develop stereotypical images of a scientist (“a man of knowledge”).<sup>49</sup> Chambers’ test was administered over an eleven-year period to nearly 5,000 children and found that the association between scientists and eyeglasses continues to increase with age.<sup>50</sup> When Mark Thomas, a doctoral student in the Department of Psychology at Mississippi State University, administered a modified DAST to college-aged students (with a mean age of roughly twenty-one years), it revealed that the stereotype of eyeglasses correlating to higher intelligence does not fade with age: the drawings depicted a scientist with eyeglasses nearly seventy percent of the time.<sup>51</sup>

---

*Person*, 12 J. SOC. BEHAV. & PERSONALITY 773, 778 (1997) (noting that older children depicted eyeglasses more frequently than younger children and that “eyeglasses are an almost archetypal sign of a ‘bookworm,’ a person absorbed in mental activity”); see also Hannu Rätty & Leila Snellman, *On the Social Fabric of Intelligence*, 4 PAPERS ON SOC. REPRESENTATIONS 1, 2–3 (1995) (concluding that “children have captured some central value-bound ideas of intelligence prevalent in our culture well before being capable of understanding them conceptually”).

<sup>48</sup> Rätty & Snellman, *Children’s Images of an Intelligent Person*, *supra* note 47, at 778.

<sup>49</sup> David Wade Chambers, *Stereotypic Images of the Scientist: The Draw-a-Scientist Test*, 67 SCI. EDUC. 255, 256–58 (1983) (noting that eyeglasses are associated with eyestrain and therefore are associated with acute observation). In Chambers’ study, each drawing was analyzed for seven predetermined indicators of a scientist: lab coat, eyeglasses, growth of facial hair, symbols of research, symbols of knowledge, technology (products of science), and relevant captions. *Id.*; see also Rätty & Snellman, *Children’s Images of an Intelligent Person*, *supra* note 47, at 781 (noting significant overlap between the results of the Children’s Images of an Intelligent Person study and Chambers’ DAST results of children’s portrayals of the scientist as “a man of knowledge”).

<sup>50</sup> Chambers, *supra* note 49, at 257–58 (reporting that the number of indicators in children’s standard images of a scientist increased from fourteen in kindergarten-age children to 1,524 in fifth-grade-aged children).

<sup>51</sup> Mark D. Thomas et al., *The Draw a Scientist Test: A Different Population and a Somewhat Different Story*, 40 C. STUDENT J. 140, 144 (2006).

The following studies indicate that perceptions and judgments of those who wear eyeglasses permeate cultural, gendered, and racial divides. As far back as 1944, G.R. Thornton, a professor in the Department of Psychology at Purdue University, found that people who wear eyeglasses are judged as being more intelligent, more industrious, more honest, and more dependable than those who do not wear eyeglasses.<sup>52</sup> A subsequent cross-cultural study conducted twenty-five years later paralleled Thornton's findings.<sup>53</sup> A study led by Åke Hellström, professor in the Department of Psychology at Stockholm University, conducted a facial attributes rating analysis that directly linked the wearing of eyeglasses with professionalism and intellect.<sup>54</sup> Specifically, this study revealed a strong perceived correlation between the wearing of eyeglasses and both prestigious occupations and positive character traits.<sup>55</sup> In 1991, a gender-based study on stereotypes associated with eyeglasses found that both men and women who wear eyeglasses are perceived as

---

<sup>52</sup> G.R. Thornton, *The Effect of Wearing Glasses upon Judgments of Personality Traits of Persons Seen Briefly*, 28 J. APPLIED PSYCHOL. 203, 203 (1944). Subjects wearing eyeglasses and judged via photographic slides were rated as more intelligent, more industrious, more honest, and more dependable. *Id.* However, subjects wearing eyeglasses and judged in person were rated as more intelligent and more industrious, but not necessarily as more honest. *Id.* at 207. When judged only by photographs, the subjects' dress, demeanor, and overall appearance were excluded, supporting the proposition that a person's demeanor is also taken into account when personality traits are judged. *Id.*; see also G.R. Thornton, *The Effect upon Judgments of Personality Traits of Varying a Single Factor in a Photograph*, 18 J. SOC. PSYCHOL. 127, 127 (1943).

<sup>53</sup> Manz & Lueck, *supra* note 21, at 704 (replicating Thornton's study with German students 25 years later, with subjects in photographs wearing eyeglasses producing higher ratings than subjects not wearing eyeglasses in the categories of intelligence, industriousness, dependability, and honesty).

<sup>54</sup> Hellström & Tekle, *supra* note 20, at 694.

<sup>55</sup> Hellström and Tekle's study found that wearing eyeglasses positively correlates to the occupations of physician, lawyer, professor, engineer, pastor, politician, psychologist, and bank clerk, and positively correlates to judged character attributes of trustworthiness, helpfulness, and intelligence. *Id.* at 699. However, the study found that wearing eyeglasses negatively correlates to the occupations of factory worker, colonel, farmer, and salesman, and to the character attributes of masculinity and being suspect. *Id.*



being more intelligent, well-educated, well-read, and better employed.<sup>56</sup> In 1993, a study examining the effects of eyeglasses and gender on perceived social forcefulness and mental competence confirmed that, overall, both men and women who wear eyeglasses are judged as having decreased forcefulness and heightened mental capacity.<sup>57</sup> However, this study found that eyeglasses tend to detract from social appeal more in women than in men.<sup>58</sup> In 2011, a study found that faces depicted with eyeglasses were consistently judged to be significantly more successful, more trustworthy, and more intelligent than faces depicted without eyeglasses.<sup>59</sup> Even details such as whether the eyeglasses worn are rimless or full-rimmed can have an impact on trustworthiness and facial recognition.<sup>60</sup> An earlier study using African-American and Caucasian subjects analyzed the effects of wearing eyeglasses in a courtroom setting.<sup>61</sup> Echoing the results of previous studies, researchers found a strong link between wearing eyeglasses and perceived intelligence and a correlation between perceived intelligence of a defendant and decreased likelihood of a juror to render a guilty verdict.<sup>62</sup>

The perceived correlation between wearing eyeglasses and heightened intelligence may be the result of a “nerd stereotype” that is deeply rooted in one’s schemata, in one’s social

---

<sup>56</sup> See Mary B. Harris, *Sex Differences in Stereotypes of Spectacles*, 21 J. APPLIED SOC. PSYCHOL. 1659, 1674–75 (1991).

<sup>57</sup> Terry & Krantz, *supra* note 41, at 1757, 1765–66.

<sup>58</sup> *Id.* at 1759.

<sup>59</sup> Leder et al., *supra* note 24, at 218–19.

<sup>60</sup> *Id.* at 216–19 (noting that “faces without eyeglasses [are] judged to be less successful and less intelligent than faces with full-rim glasses or rimless glasses,” and observing that it takes longer to recognize faces with full-rim glasses than it does to recognize faces either without glasses or with rimless glasses).

<sup>61</sup> See Brown, *supra* note 9, at 3 (finding no significant difference in the number of guilty verdicts rendered against African-American defendants (forty-nine percent) and Caucasian defendants (fifty-one percent) and concluding that, overall, participants rendered “guilty” verdicts forty-four percent of the time against defendants who wore eyeglasses while rendering “guilty” verdicts fifty-six percent of the time against defendants who did not wear eyeglasses).

<sup>62</sup> *Id.*

experiences, and in the media's portrayal of intelligent people.<sup>63</sup> Stereotypes about people who wear eyeglasses abound in popular culture—in Hollywood film characters,<sup>64</sup> highly rated television series,<sup>65</sup> best-selling novels,<sup>66</sup> classic comic books,<sup>67</sup> and,

---

<sup>63</sup> Jellesma, *supra* note 20, at 2–5.

<sup>64</sup> In 1918, Harold Lloyd's "Glasses Character" became the "persona for which he would ultimately be celebrated." Maurizio Giammarco, *Harold Lloyd: Horatio Alger in Straw Hat and Horn-Rims*, in PLAYBILLS TO PHOTOPAYS: STAGE PERFORMERS WHO PIONEERED THE TALKIES 1, 143–47 (Brenda Loew ed., 2010). Lloyd's eyeglasses marked him as "more gentle, kind, and clever in nature." *Id.*; see also Annette M. D'Agostino, *Harold Lloyd: The Glasses*, SILENTS ARE GOLDEN (1998), <http://www.silentsaregolden.com/hlloydglassesarticle.html> (quoting Harold Lloyd) ("There is more magic in a pair of horn-rimmed glasses than the opticians dream of, nor did I guess the half of it when I put them on in 1917."). Later, the screwball comedy *Bringing Up Baby* portrayed David Huxley as a bespectacled paleontologist marked by horn-rimmed eyeglasses that were intended to function as the visual marker of his "nerd" persona. See Eddie Deezen, *Why Do Nerds So Often Wear Glasses?*, NEATORAMA (Jan. 11, 2012, 5:03 AM), <http://www.neatorama.com/2012/01/11/why-do-nerds-so-often-wear-glasses/>.

<sup>65</sup> Steve Urkel, a character on the popular television show *Family Matters*, was known for his nerd persona marked by thick-rimmed eyeglasses and suspenders. See Hannah Jones, "Urkel" Is Now a Verb—and a High-School Fashion Don't, TIME (Dec. 9, 2010), <http://newsfeed.time.com/2010/12/09/urkel-is-now-a-verb-and-a-high-school-fashion-dont/#ixzz2KS3ABLYn>. For a suggestion of the influence of Urkel's eyeglasses on both American culture and the American legal system, see Bruce Carton, "Change of Appearance" Instruction Upheld in Case of Defendant Wearing Eyeglasses to Court, LEGAL BLOG WATCH (Oct. 25, 2012, 4:15 PM), [http://legalblogwatch.typepad.com/legal\\_blog\\_watch/2012/10/change-of-appearance-instruction-upheld-in-case-of-defendant-wearing-hipster-glasses.html](http://legalblogwatch.typepad.com/legal_blog_watch/2012/10/change-of-appearance-instruction-upheld-in-case-of-defendant-wearing-hipster-glasses.html) ("So in D.C. courts, at least, defendants wear their Urkel glasses at their own peril.").

<sup>66</sup> In the popular *Superman* comic book, Clark Kent used eyeglasses to disguise himself as an astute reporter. See LES DANIELS, SUPERMAN: THE COMPLETE HISTORY (2004); see also *Superman Sales Figures*, COMICHRON (Oct. 25, 2009), <http://www.comichron.com/titlespotlights/superman.html> (documenting that, from 1960 to 1986, *Superman* sold an estimated 110–15 million copies).

<sup>67</sup> The *Harry Potter* novels became a wildly successful global phenomenon that influenced millions of people. See SUSAN GUNELIUS, HARRY POTTER: THE STORY OF A GLOBAL BUSINESS PHENOMENON (2008); Guy Dammann, *Harry Potter Breaks 400m in Sales*, GUARDIAN (June 18,

recently, in the media's coverage of professional athletes.<sup>68</sup> The aforementioned studies, coupled with popular culture's portrayal of intelligent people, demonstrate the significant impact of wearing eyeglasses on the development of judgments and perceptions of others.

## II. THE UNOFFICIAL ROLE OF APPEARANCE IN THE COURTROOM

Physical appearance is intimately tied to stereotypes about a person's character traits,<sup>69</sup> and the triggering of stereotypes based on appearance does not fade in a courtroom setting.<sup>70</sup>

---

2008, 7:30 AM), <http://www.guardian.co.uk/books/2008/jun/18/harrypotter.news> (noting that *Harry Potter* book sales "have topped 400m worldwide"). In the novels, Harry is marked by his thick-rimmed eyeglasses and consistently outsmarts his nemeses. See, e.g., J.K. ROWLING, *HARRY POTTER AND THE SORCERER'S STONE* 20 (1998).

<sup>68</sup> Harvard economist Roland Fryer suggests that professional athletes use eyeglasses to convey a message: a positive, but false, message that eyeglasses imbue the wearer with greater intelligence. See Stephen J. Dubner, *Playing the Nerd Card: A New Marketplace Podcast*, FREAKONOMICS (May 31, 2012, 9:26 AM), <http://www.freakonomics.com/2012/05/31/playing-the-nerd-card-a-new-marketplace-podcast/>. The "nerd" appearance is transforming the images of National Basketball Association (NBA) players, with eyeglasses intended to signify, "We're much, much smarter than you think." Sean Gregory, *NBA Nerd Alert*, TIME (May 14, 2012), <http://www.time.com/time/magazine/article/0,9171,2113816-2,00.html>; see also Matt Ufford, *NBA Finals Fashion: Shut Up, Everyone Looks Fine*, SBNATION (June 13, 2012, 9:00 AM), <http://www.sbnation.com/nba/2012/6/13/3079618/nba-finals-fashion-fake-glasses> (quoting Dwayne Wade) (stating that NBA players wear eyeglasses because "[i]t's cool to be smart [and] educated").

<sup>69</sup> SMITH & MALANDRO, *supra* note 2, §1.48, at 86.

<sup>70</sup> A study conducted on the interplay between a defendant's appearance and an evaluation of a defendant found that attractive females (long hair and cosmetics as opposed to short hair and no cosmetics) were more often given short-term imprisonment ratings rather than long-term imprisonment ratings. Angela S. Ahola et al., *Is Justice Really Blind? Effects of Crime Descriptions, Defendant Gender and Appearance, and Legal Practitioner Gender on Sentences and Defendant Evaluations in a Mock Trial*, 17 PSYCHIATRY, PSYCHOL. AND LAW 304, 319–20 (2010). This study further noted that faces often trigger stereotypes, such that "[a] baby-faced defendant will be considered less likely to have committed an offence intentionally, and

While inferences drawn about a defendant's character based on his or her appearance may not be entirely inaccurate,<sup>71</sup> such inferences are arbitrarily drawn and difficult to verify.<sup>72</sup> Thus, a jury is generally precluded by the Federal Rules of Evidence from taking into consideration a defendant's character.<sup>73</sup> However, the physical appearance of a defendant still plays a substantive role at trial.<sup>74</sup> In fact, a defendant's physical appearance is of such vital importance to a trial that an entire industry has developed for the purpose of advising a defendant on his or her aesthetic appearance at trial.<sup>75</sup>

*A. Character Evidence—Evidence the Jury Can Consider  
Versus Evidence the Jury Does Consider*

The courtroom, comprised of individuals who fill specialized and particular roles,<sup>76</sup> provides a dynamic platform for discovering the truth. The jury trial is a central component of the American adversarial system, its purpose being to sort

---

more likely to have committed an offence by negligence than a defendant with a mature face." *Id.* at 312.

<sup>71</sup> Robert Agnew, *Appearance and Delinquency*, 22 CRIMINOLOGY 421, 424, 429 (1984) (finding a positive correlation between delinquency rates and unattractive appearance and noting that unattractive people—premised on the appearance of physical traits, dress, and grooming—are perceived as having significantly less favorable characteristics than attractive people).

<sup>72</sup> See FED. R. EVID. 404 advisory committee's notes (cautioning against the use of character evidence at trial because it raises questions of relevancy and proof).

<sup>73</sup> See *id.*

<sup>74</sup> See David L. Wiley, *Beauty and the Beast: Physical Appearance Discrimination in American Criminal Trials*, 27 ST. MARY'S L.J. 193, 209 (1995); M.G. Efran, *The Effect of Physical Appearance on the Judgment of Guilt, Interpersonal Attraction, and Severity of Recommended Punishment in a Simulated Jury Task*, 8 J. RES. PERSONALITY, 45, 45 (1974) (finding that physically attractive defendants were judged with less certainty of guilt and received milder punishments than unattractive defendants).

<sup>75</sup> See Paul, *supra* note 12 (observing that jury consultants, "often trained in both psychology and law," advise defendants on what to wear and how to appear in the courtroom).

<sup>76</sup> These individuals include the judge, defendant(s), legal counsel, audience, and witness(es).

through competing positions in order to arrive at the truth.<sup>77</sup> However, jurors hinder the truth-seeking process when they consider evidence that is deemed inadmissible due to its prejudicial effects.<sup>78</sup> Though not always irrelevant,<sup>79</sup> character evidence that is used to prove that a defendant acted, in a specific instance, in conformity with a character trait is often so weakly probative of guilt that the prejudice of admitting such evidence is likely to substantially outweigh the evidence's probative value.<sup>80</sup> Character evidence is traditionally forbidden because evidence of a defendant's particular character trait does not necessarily correlate to a defendant having "acted in

---

<sup>77</sup> Strier, *supra* note 22, at 100 ("Arguably, the most compelling claim supporting the adversary system of trial court dispute resolution is that it is the best judicial system for truth-finding."); *see also* Barbara A. Babcock, *Introduction: Taking the Stand*, 35 WM. & MARY L. REV. 1, 9 n.31 (1993) (noting that the Supreme Court "has recognized that the purpose of a trial is to sort truth from untruth").

<sup>78</sup> *See* Michelle Pan, *Strategy or Stratagem: The Use of Improper Psychological Tactics by Trial Attorneys to Persuade Jurors*, 74 U. CIN. L. REV. 259, 262 (2005); *see also* Barrett J. Anderson, *Recognizing Character: A New Perspective on Character Evidence*, 121 YALE L.J. 1912, 1928-29 (2012) (stating that "[l]egal historians have commonly understood courts to have developed the law of evidence to prevent jurors' 'cognitive and decisional failings' from impacting their solemn duty to find the truth," but also noting that despite such laws, jurors are often unable to properly consider character evidence) (citing Frederick Schauer, *On the Supposed Jury-Dependence of Evidence Law*, 155 U. PA. L. REV. 165, 199 (2006)).

<sup>79</sup> A defendant's appearance may become relevant evidence if it "forms the basis for identification." Laurie L. Levenson, *Courtroom Demeanor: The Theater of the Courtroom*, 92 MINN. L. REV. 573, 577 n.19 (2008).

<sup>80</sup> Aviva A. Orenstein, *No Bad Men!: A Feminist Analysis of Character Evidence in Rape Trials*, 49 HASTINGS L.J. 663, 669-70 (1998) ("Even assuming that such [character] evidence is reliable, a proposition which is itself open to doubt, character evidence can be invasive, unfair, and prejudicial."); *see also* FED. R. EVID. 403 ("The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence."); *United States v. Baytank*, 934 F.2d 599, 614 (5th Cir. 1991) (noting that an instruction to introduce character evidence should be refused where character evidence is not "central or crucial").

conformity with that trait or characteristic.”<sup>81</sup> Nevertheless, a defendant’s physical appearance at trial, whether consciously or unconsciously acknowledged, significantly influences perception of the defendant’s character and can influence the outcome of a case.<sup>82</sup>

Behind the decision to refrain from giving jurors a proper instruction about consideration of a defendant’s appearance at trial lie two incorrect assumptions: (1) jurors are unbiased<sup>83</sup> and (2) jurors consider only relevant evidence at trial.<sup>84</sup> However, jurors tend to favor defendants whom they find more relatable, regardless of the facts.<sup>85</sup> If jurors do, in fact, follow a court’s

---

<sup>81</sup> Orenstein, *supra* note 80, at 668.

<sup>82</sup> See Steven Fein et al., *Hype and Suspicion: The Effects of Pretrial Publicity, Race, and Suspicion on Jurors’ Verdicts*, 53 J. SOC. ISSUES 487, 488–89 (1997) (observing that one factor contributing to “prejudicial effects of nonevidentiary information” is a juror’s inability to block out an image or thought of the defendant); Ross, *supra* note 14, at 227, 232 (“[T]he perceived character of an accused affects the outcome of jury trials . . .”); see also Steven Shepard, Note, *Should the Criminal Defendant Be Assigned a Seat in Court?*, 115 YALE L.J. 2203, 2208 (2006) (“A defendant’s appearance matters to the jury and can affect the outcome of a trial.”).

<sup>83</sup> See HAZEL THORNTON, HUNG JURY: THE DIARY OF A MENENDEZ JUROR 101–02 (1995); see also Brown, *supra* note 9, at 6 (quoting Tara Trask, a jury consultant with seventeen years of experience in litigation strategy, who observed that “jurors tend to assign credibility to those who fit the stereotypes they have”).

<sup>84</sup> THORNTON, *supra* note 83, at 101–02. Jurors operate as “detectives, assimilating important visual information to add to evidence,” even when the visual information is irrelevant to the facts of the case. SMITH & MALANDRO, *supra* note 2, § 1.50, at 90.

<sup>85</sup> E.H. SUTHERLAND & D.R. CRESSEY, PRINCIPLES OF CRIMINOLOGY 442 (7th ed. 1966) (“The main work of a trial lawyer is to make a jury like his client, or, at least, to feel sympathy for him; facts regarding the crime are relatively unimportant.”); see also Douglas Keene, *Tattoos: When Should You Clean Up Your Witness?*, KEENE TRIAL CONSULTING (Dec. 6, 2010), <http://keenetrial.com/blog/2010/12/06/tattoos-when-should-you-clean-up-your-witness/> (“The goal of the attorney presenting a witness is to help the jury see the witness as ‘kind of like me’ or ‘someone I can trust.’ Appearance is a part of that. If someone looks scary or unfamiliar, they are judged as less trustworthy and less believable. The goal is to help them be more ‘relatable,’ regardless of the facts.”); Melanie Tannenbaum, *Casey’s Case: What Psychology Says About Anthony’s Acquittal*, PSYSOCIETY (July 10, 2011),

cautionary instructions to consider only relevant evidence,<sup>86</sup> then why do courts,<sup>87</sup> jury consultants,<sup>88</sup> and defense teams<sup>89</sup> go to such lengths to alter a defendant's appearance at trial? Such

---

<http://psysociety.wordpress.com/2011/07/10/casey-anthony/> (“Overall, jurors are more likely to be lenient towards defendants that are similar to them in some meaningful way. For example, jurors are less likely to convict defendants if they are of the same gender or race, or if they come from a similar socioeconomic background.”); John Schwartz, *Extreme Makeover: Criminal Court Edition*, N.Y. TIMES (Dec. 5, 2010), <http://www.nytimes.com/2010/12/06/us/06tattoo.html> (“‘It’s easier to give someone who looks like you a fair shake,’ said [defense attorney] Bjorn E. Brunvand.”).

<sup>86</sup> CTJNY § 3:2 (2012) (describing to the jury what kind of evidence may be considered during deliberations by stating that “[e]vidence consists of the sworn testimony elicited both on direct examination and cross-examination, and redirect and recross, if any, plus any concessions made during the trial by counsel, and any exhibits received and marked in evidence”); *see also* FED-JI § 12:03 (6th ed. 2013) (“The evidence in this case consists of the sworn testimony of the witnesses—regardless of who may have called them—all exhibits received in evidence—regardless of who may have produced them—all facts which may have been agreed to or stipulated and all facts and events which may have been judicially noticed.”); PATTERN JURY INSTRUCTIONS, Fifth Circuit, Criminal Cases, § 1.01, 3–4 (2012) (explaining to the jury what is *not* evidence, which includes “[s]tatements, arguments, and questions by lawyers . . . [o]bjections to questions . . . [t]estimony that the court has excluded . . . [and] [a]nything [the jurors] may have seen, heard, or read outside the courtroom . . .”).

<sup>87</sup> *See, e.g.*, Neil Nagraj, *Court Pays \$150 a Day to Cover Neo-Nazi John Ditullio’s Tattoos During Murder Trial*, N.Y. DAILY NEWS (Dec. 7, 2009), [http://articles.nydailynews.com/2009-12-07/news/17940784\\_1\\_neo-nazi-compound-tattoos-extreme-makeover](http://articles.nydailynews.com/2009-12-07/news/17940784_1_neo-nazi-compound-tattoos-extreme-makeover) (describing a Florida court’s order that required the state to pay a cosmetologist \$150 for each day of trial in an effort to cover the defendant’s facial tattoos (barbwire and teardrops) and neck tattoos (large swastikas and a vulgar phrase)).

<sup>88</sup> Literature on trial practice devotes significant portions to how to alter a defendant’s appearance in accordance with known juror perceptions and stereotypes. *See generally* SMITH & MALANDRO, *supra* note 2, §§ 1.01–2.08.

<sup>89</sup> *See, e.g.*, Ivy Bigbee, *Fashioning a Defense: Casey Anthony’s Evolving Style*, CRIM. REP. DAILY (Apr. 29, 2011), [http://web.archive.org/web/20120520050256/http://blogs.discovery.com/criminal\\_report/2011/04/fashioning-a-defense-casey-anthonys-evolving-style.html](http://web.archive.org/web/20120520050256/http://blogs.discovery.com/criminal_report/2011/04/fashioning-a-defense-casey-anthonys-evolving-style.html) (noting that, due to her defense team, “Anthony’s courtroom appearances at hearings ironically have spoken volumes, without her uttering a single word”).

measures are taken because it is ingrained in American society to judge others based on physical appearance:

We live in a society where people are bombarded with so much information each day that they have learned to use shortcut techniques to make decisions. One of these shortcut techniques is to judge people based on initial perceptions of their appearance, background, and behavior. Once we have made these initial decisions about an individual, all further communication is filtered through this arrived-at perception. If we decide a person “looks like a law student,” then we will proceed to treat that perception as if it were an actual fact. We will respond to the individual as though he actually was in law school; that is, we might assume that he is an intellectually capable, academically motivated, and career-oriented person. It makes little difference whether the initial perception is correct. People treat the perception as accurate and make decisions from this base of information.<sup>90</sup>

Juries—composed of a cross-section of American society—judge in this same way.<sup>91</sup> Jurors tend to consider any artificial altering of a defendant’s physical appearance, including: eyeglasses,<sup>92</sup> clothing style,<sup>93</sup> clothing color,<sup>94</sup> makeup,<sup>95</sup>

---

<sup>90</sup> SMITH & MALANDRO, *supra* note 2, §1.48, at 86.

<sup>91</sup> Levenson, *supra* note 79, at 576–77.

<sup>92</sup> *See generally* Brown, *supra* note 9, at 1 (finding that appliances that “alter the appearance of eyes—namely eyeglasses—may influence our perceptions of an individual who uses such devices”).

<sup>93</sup> SMITH & MALANDRO, *supra* note 2, § 1.19, at 36–37 (“It is important to coach your client and witnesses with regard to personal appearance factors. Witnesses do not have a good understanding of how much their clothing can affect the total impact in the courtroom.”). Although clothing alone may not change perception, it takes only one juror to notice clothing details and to share them with others in order to have an impact on the decision-making process. *Id.* § 1.16, at 29.

<sup>94</sup> An industry has emerged, called “color consulting,” in which consultants advise defendants on what clothing and make-up to wear at trial. Black colors should not be worn by defendants on trial for murder because “the connotations associated with black tend to be consistent across cultures and are deeply embedded in our minds.” *Id.* § 1.28, at 55. Red is associated



jewelry,<sup>96</sup> and hairstyle.<sup>97</sup> Yet, unless given specific instruction, jurors might not know how to properly consider those judgments in relation to the facts of the case.<sup>98</sup>

### *B. The Impact of the Defendant's Appearance on Juror Perceptions*

As the following cases highlight, a defendant's appearance, although generally inadmissible as evidence, can still impact jurors' perceptions at trial. In *Estelle v. Williams*, the Supreme Court recognized the impact of a defendant's appearance on

---

with "passion, violence, excitement, and blood," and should not be worn by defendants. *Id.*

<sup>95</sup> W.J. McKeachie, *Lipstick as a Determiner of First Impressions of Personality: An Experiment for the General Psychology Course*, 36 J. SOC. PSYCHOL. 241, 242 (1952) (concluding from a study that women who wear lipstick are judged as being more frivolous, more conscientious, and having more overt interest in males).

<sup>96</sup> SMITH & MALANDRO, *supra* note 2, § 1.23, at 44 ("Jewelry, to the juror-detective, offers many stereotypes which will affect the total perception of the individual."). For a discussion of rings, see *id.* § 1.50, at 91 (explaining that a ring worn on the ring finger indicates a stable relationship). For a discussion of defendants wearing watches to trial, see Martha Neil, *Defendant Puts Best Face Forward, After Extreme Makeover, in Capital Murder Case*, A.B.A. J. (Dec. 6, 2010, 1:43 PM), <http://www.abajournal.com/news/article/defendant-puts-best-face-forward-sans-most-tattoos-in-capital-murder-case/> (reporting that, in the Enron Trial, consultant Douglas Keene told his clients "to be sure they didn't wear their \$10,000 watches to trial").

<sup>97</sup> See SMITH & MALANDRO, *supra* note 2, § 1.24, at 50 ("Hair that is curly will make the person appear energetic. For the perception of credibility, the hair should be short, tailored, and professional.").

<sup>98</sup> The jury may be instructed on how to specifically consider a defendant's appearance at trial when the defendant's appearance is relevant, such as when the defendant's identification is at issue. See MULDOON, *supra* note 16, § 9:262 (stating that "[w]ith the time lag between the occurrence of the crime and the trial, the defendant's appearance may change, whether with time, fashions or because of an attempt at disguise for trial" and "[t]he defendant's appearance at the time of the crime is relevant for purposes of identification"); see also *Joseph v. State*, 642 So. 2d 613, 615 (Fla. Dist. Ct. App. 1994) (noting that the trial court had the authority to instruct jurors on how to properly consider religious dress in the courtroom).

juries.<sup>99</sup> The Court found that forcing a defendant to wear prison attire before a jury infringed upon his or her Fourteenth Amendment due process rights.<sup>100</sup> A defendant clothed in an orange jumpsuit at trial can give the jury the impression that the defendant is more likely to have committed the crime, something the Court deemed inconsistent with the presumption of innocence in the American justice system.<sup>101</sup> Even though the character evidence derived from the defendant's appearance was inadmissible, the Court recognized the likely prejudicial effect of the defendant's clothing on the jury.<sup>102</sup> The Court concluded that jurors, at least in some instances, are unable to ignore a defendant's appearance.<sup>103</sup>

In 2010, a Florida judge recognized the likelihood of a defendant's appearance impacting the jury.<sup>104</sup> John Ditullio faced the death penalty for charges related to the violent stabbing and death of a teenager but ultimately received life in prison without the possibility of parole.<sup>105</sup> Ditullio's defense team successfully argued that his neo-Nazi tattoos—although acquired after his arrest—would be too distracting and too prejudicial for the

---

<sup>99</sup> *Estelle v. Williams*, 425 U.S. 501, 505 (1976) (“The defendant’s clothing is so likely to be a continuing influence throughout the trial that, not unlike placing a jury in the custody of deputy sheriffs who were also witnesses for the prosecution, an unacceptable risk is presented of impermissible factors coming into play.”).

<sup>100</sup> *Id.* at 512–13.

<sup>101</sup> *Id.* at 503 (“The presumption of innocence, although not articulated in the Constitution, is a basic component of a fair trial under our system of criminal justice.”); *see also* THORNTON, *supra* note 83, at 111–12 (supporting the proposition that a defendant who wears prison attire is more likely to be convicted); Shepard, *supra* note 82, at 2208.

<sup>102</sup> *Estelle*, 425 U.S. at 505.

<sup>103</sup> *Id.* at 518 (finding that prison attire “surely tends to brand [the defendant] in the eyes of the jurors with an unmistakable mark of guilt”).

<sup>104</sup> *State v. Ditullio*, No. CRC06-05827CFAWS (Fla. Cir. Ct. 2009); *see also* Schwartz, *supra* note 85.

<sup>105</sup> Carlin DeGuerin Miller, *John Allen Ditullio Guilty Verdict: Neo-Nazi Convicted of Murder, Sentenced to Life in Prison*, CBS NEWS (Dec. 17, 2010, 1:14 PM), [http://www.cbsnews.com/8301-504083\\_162-20026011-504083.html](http://www.cbsnews.com/8301-504083_162-20026011-504083.html).

jurors to see.<sup>106</sup> Alternatively, Alan Dershowitz, a criminal defense attorney and Harvard Law School professor, suggested that “the swastika and other tattoos [were] an extension of Ditullio’s persona, and masking the marks could be construed as misleading to a jury.”<sup>107</sup> Nonetheless, the court agreed with Ditullio’s defense attorney that unless Ditullio’s neo-Nazi tattoos were covered, his physical appearance could prejudice the jury.<sup>108</sup> Ditullio’s tattoos may have suggested to the jury that he had unfavorable characteristics—essentially, that embodied in Ditullio’s persona was an outwardly racist and hateful being.<sup>109</sup>

In one of the most publicized capital murder trials in history,<sup>110</sup> Casey Anthony’s defense team strategically selected preppy clothing to project a childish and innocent image.<sup>111</sup>

---

<sup>106</sup> Schwartz, *supra* note 85 (reporting that Ditullio’s lawyer argued for his client’s tattoos to be covered up because “[t]here’s no doubt in my mind without the makeup being used, there’s no way a jury could look at John and judge him fairly”). Ditullio’s second trial was widely discussed because of the court’s decision to have the state pay for his neo-Nazi tattoos to be covered at trial. *Id.*

<sup>107</sup> Cindy E. Rodriguez, *Neo-Nazi Accused of Hate-Crime Murder Gets Make-over*, ABC NEWS (Dec. 7, 2010), <http://abcnews.go.com/US/neo-nazi-accused-hate-crime-murder-makeover/story?id=12324409#.UIXSiUKRrww>.

<sup>108</sup> *Ditullio*, No. CRC06-05827CFAWS; *see also* K. McKinney, *Brunvand Wins Motion to Have Murder Defendant’s Tattoos Covered During Trial*, ACQUITTER.COM (Dec. 7, 2009), <http://acquitter.com/news/florida/brunvand-wins-motion-cover-tattoos/>.

<sup>109</sup> Greg Wims, the President of the Victims’ Rights Foundation, stated, “People should be able to see these tattoos. The jury should see what kind of person he is. Of course those tattoos are central to the case.” Rodriguez, *supra* note 107; *see* Ryan Lozar, *Tattoos as Evidence*, CALIFORNIA LAWYER 57–58 (2012) (explaining that snap judgments about a defendant’s character that are based on physical appearance are especially severe when the defendant has a tattoo, and noting that depending on the subject of the tattoo, the defendant may be perceived as “seedy, provocative, or downright dangerous”).

<sup>110</sup> *State v. Anthony*, No. 48-2008-CF-015606-O (Fla. Cir. Ct. 2011); *see also State of Florida v. Casey Marie Anthony*, NINTH JUD. CIRCUIT CT. FLA., <http://www.ninthcircuit.org/news/High-Profile-Cases/Anthony/> (last visited Mar. 16, 2013) (highlighting the court’s media policies).

<sup>111</sup> *See* Bigbee, *supra* note 89 (describing a change in Casey Anthony’s courtroom attire from “stylish in a sexually-suggestive way” to “a modest, plaid shirt under a drab gray cardigan sweater”). Casey Anthony was accused

Anthony's appearance evolved throughout the trial: from oversized<sup>112</sup> and pastel-colored shirts<sup>113</sup> to preppy sweaters and long hair.<sup>114</sup> Anthony's defense team crafted a story about her

---

of murdering her two-year-old daughter, Caylee Anthony, but the jury acquitted her of first-degree murder. Michael Winter, *Casey Anthony Acquitted of Murder*, USA TODAY (July 5, 2011), <http://content.usatoday.com/communities/ondeadline/post/2011/07/casey-anthony-jury-reaches-verdict/1>. In the infamous case of the Menendez brothers, the defendants used a similar preppy look. The brothers were convicted of violently murdering their parents. Throughout their trial, the brothers donned preppy sweaters. Although the jurors acknowledged that the defendants' appearance was not admissible evidence, they still "discussed the fact that the defendants wore sweaters as opposed to suits to court." THORNTON, *supra* note 83, at 111; *see also* Dominick Dunne, *The Menendez Murder Trial*, VANITY FAIR (Oct. 1993), <http://www.vanityfair.com/magazine/archive/1993/10/dunne199310> (reporting that the Menendez brothers' "Armani-type clothes [were] replaced in the courtroom by sensible shirts, slacks, and sweaters, brought freshly washed and ironed each morning for them to change into from the L.A. County Jail uniforms they [were] wearing when they arrive[d] at court").

<sup>112</sup> *Casey's Appearance Could Be Changed to Influence Jury*, (Apr. 7, 2008, 6:00 PM), <http://www.wftv.com/news/news/caseys-appearance-could-be-changed-to-influence-ju/nK99f/> (describing Anthony's clothes as "baggy and disheveled" and noting that she transformed her image to fit a "librarian look" and donned clothing that matched her defense team's attire); SMITH & MALANDRO, *supra* note 2, § 1.19, at 36 ("To create a victimized look or a look of helplessness, the individual should wear oversized clothing . . . ."). One explanation for matching attire might be the proximity between Anthony and one of her female attorneys. If there is a stark difference in dress between attorney and client, any images will be perceived as "more extreme." *Id.* at 35.

<sup>113</sup> *Compare Lillian Glass, Jose Baez and Casey Anthony's Body Language Show Delight in Jurors' Negative Reactions Towards Nancy Grace, Casey's Courtroom Tension Builds on Day 6*, DR. LILLIAN GLASS BODY LANGUAGE BLOG, (May 15, 2011), <http://drlillianglassbodylanguageblog.wordpress.com/2011/05/15/jose-baez-and-casey-anthonys-body-language-show-delight-in-jurors-negative-reactions-towards-nancy-grace-caseys-courtroom-tension-builds-on-day-6/> (depicting Casey Anthony's use of pastel-colored and ruffled clothing at trial), *with* SMITH & MALANDRO, *supra* note 2, § 1.19, at 36 ("To create a submissive look for a woman, pastels and ruffles should be worn.").

<sup>114</sup> *See* Jessica Hopper & Ashleigh Banfield, *Casey Anthony Trial: Defense Team Claims Caylee Anthony Drowned in Family Pool*, ABC NEWS, (May 24, 2011), [http://abcnews.go.com/US/casey\\_anthony\\_trial/casey-](http://abcnews.go.com/US/casey_anthony_trial/casey-)

character through her appearance that suggested that Anthony was a child-like woman “forever stuck in adolescence.”<sup>115</sup> Perhaps to the jurors, Anthony appeared as a woman incapable of having committed the brutal crime of which she was accused. The court did not consider Anthony’s appearance to be relevant admissible evidence, yet legal analysts suggested that the defense counsel attempted to “subtly influence the judge” by altering Anthony’s appearance.<sup>116</sup> It follows that what may subtly influence the judge may also influence the jury.<sup>117</sup> At the trial’s conclusion, the jury acquitted Anthony of the capital murder charge.<sup>118</sup> As these cases highlight, the Supreme Court, defense attorneys, prosecuting attorneys, and law professors all recognize that a defendant’s appearance has the potential to encourage certain biases in jurors.

### III. THE NERD DEFENSE

Since eyeglasses can drastically change one’s appearance, they are a particularly powerful tool with which to alter juror perceptions.<sup>119</sup> A defendant who wears eyeglasses to trial without any need to correct vision impairment utilizes the nerd defense

---

anthony-trial-defense-claims-caylee-anthony-drowned/story?id=13674375; *see also* SMITH & MALANDRO, *supra* note 2, § 1.24, at 48 (noting that to help portray a soft and submissive look in the courtroom, a client’s hair should be longer).

<sup>115</sup> Bigbee, *supra* note 89.

<sup>116</sup> *Casey’s Appearance Could Be Changed to Influence Jury*, *supra* note 112.

<sup>117</sup> Judges are more likely than jurors to notice defense teams’ strategies because judges are trained to examine the law and are attuned to the strategies that defense teams employ. *See* SAUL M. KASSIN & LAWRENCE S. WRIGHTSMAN, *THE AMERICAN JURY ON TRIAL: PSYCHOLOGICAL PERSPECTIVES* 126 (1988) (discussing the argument against jury trials by acknowledging that judges and lawyers are “courtroom veterans” as compared to the jury).

<sup>118</sup> Winter, *supra* note 111.

<sup>119</sup> *See* Tom Davies, *Framed! Sharon Osbourne*, OPTICIAN ONLINE (Nov. 18, 2005), <http://www.opticianonline.net/Articles/2005/11/18/14675/Framed!+Sharon+Osbourne.htm> (stating that eyeglasses can cause the ugly to “become cool” and the cool to “become intelligent”).

in the hopes of appearing more intelligent<sup>120</sup> and therefore less likely to have committed a crime.<sup>121</sup> Juries are less likely to convict defendants whom they find more “likeable,”<sup>122</sup> and studies show that wearing eyeglasses helps to make a defendant appear more likeable.<sup>123</sup> Wearing nonprescriptive eyeglasses fabricates a defect in a defendant and plays upon one of society’s most deeply rooted stereotypes: that glasses are synonymous with higher intelligence.<sup>124</sup> In this sense, wearing nonprescriptive eyewear is analogous to using crutches or a wheelchair despite lacking a physical injury.<sup>125</sup> By contrast, dressing a defendant in a suit and tie, adding accessories (e.g., watches and rings), altering hairstyle, or applying makeup can affect juror perception, but such changes do not falsely represent a handicap or a physical defect.<sup>126</sup> Similarly, a defendant who

---

<sup>120</sup> Alexander, *supra* note 9; *see also* Brown, *supra* note 9, at 1.

<sup>121</sup> Brown, *supra* note 9, at 3; *see also* Rita Handrich, *The Glasses Create a Kind of Unspoken Nerd Defense*, KEENE TRIAL CONSULTING (Mar. 7, 2011, 6:01 AM), <http://keenetrial.com/blog/2011/03/07/the-glasses-create-a-kind-of-unspoken-nerd-defense> (“The idea that the Nerd Defense might work (or help) is an extension of the fact that Nerds are evidently viewed as being less likely to commit crimes . . . . If they create an image of someone who ‘doesn’t look like they would do that sort of thing,’ it will aid in the defense.”).

<sup>122</sup> SUTHERLAND & CRESSEY, *supra* note 85, at 442. However, attractiveness is also a component of likability, and glasses are perceived as making the wearer less attractive. Terry & Krantz, *supra* note 41, at 1766 (noting that the “negative” perceptions of those who wear eyeglasses lead to “increas[ed] ratings of character, compassion, honesty, and sensitivity [and] decreas[ed] ratings of attractiveness [and forcefulness]”); *see generally* Leder et al., *supra* note 24 (connecting the wearing of eyeglasses with increased intelligence and decreased attractiveness).

<sup>123</sup> *See, e.g.*, SMITH & MALANDRO, *supra* note 2, § 1.21 at 42 (noting that wide lenses help to create an open-eyed look that is associated with traits of trustworthiness, likability, and innocence).

<sup>124</sup> Thornton, *The Effect of Wearing Glasses upon Judgments of Personality Traits of Persons Seen Briefly*, *supra* note 52, at 203; Brown, *supra* note 9, at 3.

<sup>125</sup> *See* Marshall, *supra* note 18 (“Glasses convey information about physical capabilities the same way coming into court on crutches or in a wheelchair does.”).

<sup>126</sup> *Id.* (contrasting use of fake eyewear with “haircut, a shave, a suit and

chooses to wear contacts rather than eyeglasses to trial has not falsely represented a handicap. The real problem lies in concocting a handicap that brings with it such powerful social stereotypes. Richard Waites, the Chief Executive of a jury consulting firm, observes that “[j]urors expect to see defendants wearing [nice shirts and ties],” but “[j]urors don’t expect to see defendants wearing glasses if they don’t have to.”<sup>127</sup> Eyeglasses are now one of the world’s most popular fashion accessories,<sup>128</sup> viewed as possessing the unique power to “transform you like no other accessory.”<sup>129</sup> Defense attorneys have taken note of this pop-culture trend and are increasingly employing the nerd defense as a courtroom tactic.<sup>130</sup>

*A. Studies Concerning the Effect of Eyeglasses on Jurors’ Perceptions*

In one analysis of a study conducted in 2008, psychologist Michael J. Brown examined the social-cognitive processes involved when individuals make decisions, attributions, and judgments.<sup>131</sup> In Brown’s study, 220 students were presented with a portfolio containing the vignette of a fictitious trial involving a violent crime, the defendant’s photograph and physical description, and a survey asking the reader to render a

---

shined shoes,” which are not deceptive). Although eyeglasses are being increasingly worn as fashion accessories, their original purpose was to correct for an eye defect. Leder et al., *supra* note 24, at 211 (“The primary use of eyeglasses is their ability to correct congenital or acquired vision deficits such as myopia, presbyopia, or astigmatism.”).

<sup>127</sup> Alexander, *supra* note 9.

<sup>128</sup> Leder et al., *supra* note 24, at 211; Joel Stein, *The TIME 100 Most Influential Things in the World*, TIME (Apr. 21, 2011), [http://www.time.com/time/specials/packages/article/0,28804,2066367\\_2066584\\_2066602-3,00.html](http://www.time.com/time/specials/packages/article/0,28804,2066367_2066584_2066602-3,00.html) (ranking “nerd glasses” as the 74th most influential thing in the world).

<sup>129</sup> Davies, *supra* note 119.

<sup>130</sup> Tillman, *supra* note 5; Debra Cassens Weiss, *Instruction on Slaying Defendant’s New Eyeglasses at Issue on Appeal*, A.B.A. J. (Sept. 14, 2012, 8:52 AM), [http://www.abajournal.com/news/article/instruction\\_on\\_slaying\\_defendants\\_new\\_eyeglasses\\_at\\_issue\\_on\\_appeal/](http://www.abajournal.com/news/article/instruction_on_slaying_defendants_new_eyeglasses_at_issue_on_appeal/).

<sup>131</sup> Brown, *supra* note 9, at 1.

verdict and to rate the defendant as either “more” or “less” physically threatening, intelligent, attractive, and friendly.<sup>132</sup> The portfolio included the defendant’s photograph in one of four possible combinations: male Caucasian wearing eyeglasses; male Caucasian not wearing eyeglasses; African American wearing eyeglasses; or African American not wearing eyeglasses.<sup>133</sup> In this study, participants rendered a “guilty” verdict only forty-four percent of the time against defendants who wore eyeglasses, while defendants who did not wear eyeglasses were found “guilty” fifty-six percent of the time.<sup>134</sup> The study found no significant difference between the verdicts for Caucasian defendants and the verdicts for African-American defendants.<sup>135</sup>

Brown’s follow-up study, using the same general format and method noted above, examined the effect of eyeglasses in a white-collar crime context.<sup>136</sup> Consistent with the previous study, defendants who wore eyeglasses were rated as being more intelligent.<sup>137</sup> However, increased ratings of intelligence positively correlated with an increased number of guilty verdicts. In Brown’s presentation of a white-collar crime, eyeglasses had a “detrimental indirect effect” on a defendant by making the defendant appear more intelligent.<sup>138</sup> Brown’s studies did not definitively conclude that wearing eyeglasses equates to

---

<sup>132</sup> *Id.* at 3.

<sup>133</sup> *Id.* (using models comparable in age, weight, hair color, hair length, eye size, and facial hair, and wearing the same eyeglasses in each photograph).

<sup>134</sup> *Id.*

<sup>135</sup> *Id.* (concluding that Caucasians received guilty verdicts fifty-one percent of the time, while African Americans received guilty verdicts forty-nine percent of the time). Although there was not a significant difference in verdicts based on race, “race was a significant predictor of several perceived defendant characteristics.” *Id.* When both race and eyeglasses were taken into account, African Americans were perceived as more attractive and more friendly, while Caucasians were perceived as less attractive and less friendly. Moreover, African-American defendants wearing eyeglasses were perceived as less physically threatening than Caucasian defendants wearing eyeglasses. *Id.*

<sup>136</sup> *Id.*

<sup>137</sup> *Id.* at 3–4.

<sup>138</sup> *Id.* at 3.



either a “not guilty” verdict or a “guilty” verdict.<sup>139</sup> Rather, these studies found that wearing eyeglasses has a significant indirect effect on verdict outcome.<sup>140</sup> This is because wearing eyeglasses relates to increased ratings of intelligence, and perceived intelligence positively affects jurors’ verdicts in violent-crime scenarios and negatively affects jurors’ verdicts in white-collar-crime scenarios.<sup>141</sup> While these studies merely scratch the surface of the effect of eyeglasses on juror perception, they lend support to the premise that jurors do not relinquish their biases concerning eyewear in a courtroom setting.<sup>142</sup>

Research shows that jurors discriminate on the basis of appearance,<sup>143</sup> race,<sup>144</sup> and gender.<sup>145</sup> To compensate for juror biases,<sup>146</sup> defendants are urged to appear before the court well groomed and in business-type attire.<sup>147</sup> Are unnecessary eyeglasses simply another means to offset negative juror biases? Wearing unnecessary eyeglasses, like wearing proper courtroom

---

<sup>139</sup> *Id.* at 4.

<sup>140</sup> *Id.*

<sup>141</sup> *Id.* at 2–4.

<sup>142</sup> *Id.* at 3–4, 6 (including a response from Tara Trask, a jury consultant with 17 years of experience in litigation strategy, who stated, “I have seen jurors tend to assign credibility to those who fit the stereotypes they have”).

<sup>143</sup> Efran, *supra* note 74, at 45–54.

<sup>144</sup> See Fein et al., *supra* note 82, at 491 (indicating that “research has found that a jury’s racial composition . . . can have a significant effect on the verdict that jury reaches”); see also Wiley, *supra* note 74, at 214 (noting that “it is easier for jurors to imagine themselves in the defendant’s situation when the defendant is of the same race as the juror”).

<sup>145</sup> Ahola et al., *supra* note 70, at 321 (finding that “[i]n the courtroom situation, the defendant will be judged more severely by a judge or jurors of the same gender as the defendant him/herself; being sentenced by a judge of the opposite sex will be to the advantage of the defendant”).

<sup>146</sup> See THORNTON, *supra* note 83, at 103–08 (noting that the legal system acknowledges that jurors “bring to any new experience all past experiences and attitudes,” but that “it is not always possible to recognize those biases and eliminate those jurors” through the jury selection process).

<sup>147</sup> Mark J. Sullivan, *A Defendant’s Guide to Courtroom Etiquette*, CRIME, JUST. & AM. 34, 35 (2001) (suggesting to criminal defendants that jurors should, at first impression, be unable to discern through dress who is the defendant and who is the attorney).

attire, portrays favorable characteristics to the jury.<sup>148</sup> However, by analogy, a defendant who seeks to offset juror bias might utilize a multitude of props (such as unnecessary crutches or neck braces) that are designed to manipulate the jury and elicit misplaced sympathy and favorable judgment. Wearing proper courtroom attire does not fabricate a defect in any way. However, wearing unnecessary eyeglasses to trial is akin to telling the jury a lie without consequence. Such behavior undermines a judicial system that is designed to arrive at the truth.<sup>149</sup>

### *B. Reception of the Nerd Defense in Criminal Trials*

Strong opinions abound about a defendant's use of unnecessary eyeglasses at trial.<sup>150</sup> The use of nonprescriptive eyewear by defendants is becoming increasingly popular, with inmates strategically swapping eyeglasses before hearings, friends and family delivering eyeglasses during visits to inmates, and lawyers supplying clients with eyeglasses.<sup>151</sup> The nerd defense has received significant media attention, with commentators both endorsing and criticizing the use of

---

<sup>148</sup> See *id.* (noting that formal dress for defendants in the courtroom leads jurors to believe that the defendant is serious and leads judges to believe that defendants have respect for the courtroom); see also Brown, *supra* note 9, at 2–4.

<sup>149</sup> See Anderson, *supra* note 78, at 1928 (stating that the jury has a “solemn duty to find the truth”).

<sup>150</sup> Kevin Deutsch, *Defense Lawyers Swear by Gimmick of Having Defendants Wearing Glasses at Trial*, N.Y. DAILY NEWS (Feb. 13, 2011), [http://articles.nydailynews.com/2011-02-13/news/28613008\\_1](http://articles.nydailynews.com/2011-02-13/news/28613008_1) (“If a jury thinks the defendant looks incapable of a brutal crime, then it’s certainly an advantage for the defense . . . . The glasses create a kind of unspoken nerd defense.”); Marshall, *supra* note 18 (“If glasses made a guy like Larry Davis look gentle, they can work for anybody . . . . I always tell clients to get a pair. The nerdier the better.”); Alexander, *supra* note 9 (noting that “[o]ften times it’s about perception, and glasses help with that perception” and “[eyeglasses are] masks . . . [t]hey’re designed to confuse the witness and influence the jury”); Weiss, *supra* note 130 (quoting Harvey Slovis, who stated that “I’ve tried cases where there’s been a tremendous amount of evidence, but my client wore glasses, dressed well and got acquitted”).

<sup>151</sup> Alexander, *supra* note 9.

nonprescriptive eyewear by defendants in a courtroom setting.<sup>152</sup> The following cases provide examples of the utilization of the nerd defense in criminal trials.

In the infamous case of *People v. Davis*, a twenty-one-year-old defendant was charged with the attempted murder of nine police officers as they sought to arrest him.<sup>153</sup> During a lengthy trial period, the prosecution used police testimony to portray Larry Davis, the defendant, as a “gold chain-clad thug.”<sup>154</sup> The defense countered by altering Davis’ appearance, giving him a clean-cut look and a pair of horn-rimmed eyeglasses<sup>155</sup> to make him “look like Mr. Peepers.”<sup>156</sup> In a trial that spanned seven months, Davis was acquitted of the attempted murder of the officers.<sup>157</sup>

In 2011, a Bronx jury acquitted Thomas Cordero, known as “[the] nude housekeeper,” on charges of stabbing John Conley to death.<sup>158</sup> On the advice of his lawyer, Harvey Slovis, Cordero donned eyeglasses throughout the trial.<sup>159</sup> Despite Cordero’s

---

<sup>152</sup> See Marshall, *supra* note 18 (raising concerns about whether putting nonprescriptive eyeglasses on a defendant with 20-20 eyesight rises to the level of deceit prohibited by New York Legal Ethics Rule 8.4); see also MODEL RULES OF PROF’L CONDUCT R. 8.4 (2002); John Pertzborn, *Chet Pleban: Nerd Defense in the Courtroom*, FOX2NOW ST. LOUIS (Apr. 2, 2012, 8:18 AM), <http://fox2now.com/2012/04/02/chet-pleban-nerd-defense-in-the-courtroom/>.

<sup>153</sup> *People v. Davis*, 537 N.Y.S.2d 430 (Sup. Ct. 1988).

<sup>154</sup> Deutsch, *supra* note 150.

<sup>155</sup> See *id.* (observing that at trial, Davis resembled a studious young adult “dressed like a college student, sporting horn-rimmed glasses and sweaters”).

<sup>156</sup> Marshall, *supra* note 18.

<sup>157</sup> See William G. Blair, *Jury in Bronx Acquits Larry Davis in Shooting of Six Police Officers*, N.Y. TIMES (Nov. 21, 1988), <http://www.nytimes.com/1988/11/21/nyregion/jury-in-bronx-acquits-larry-davis-in-shooting-of-six-police-officers.html>.

<sup>158</sup> Kevin Deutsch & Bill Hutchinson, “Nude Housekeeper” Acquitted of Stabbing Client to Death Despite DNA Evidence, Confession, N.Y. DAILY NEWS (Jan. 13, 2011), [http://articles.nydailynews.com/2011-01-13/news/27087260\\_1\\_dna-evidence-confession-thomas-cordero](http://articles.nydailynews.com/2011-01-13/news/27087260_1_dna-evidence-confession-thomas-cordero).

<sup>159</sup> Deutsch, *supra* note 150; see also Hannah Rand, *The “Nerd Defense”: How Violent Criminals Are Turning to Thick-Framed Hipster Glasses to Persuade Juries They Are “More Intelligent, More Honest,”* MAIL

confession and the admission of DNA evidence linking him to the murder, Cordero was acquitted.<sup>160</sup> Cordero “ditched [the eyeglasses] the moment he was free.”<sup>161</sup>

In perhaps the most publicized use of the nerd defense thus far, five young men went on trial in Washington, D.C. in 2010 for first degree murder committed during “one of the District’s deadliest outbreaks of violence.”<sup>162</sup> Each of the defendants arrived to court wearing large-framed and heavy-rimmed glasses.<sup>163</sup> This sparked the attention of the prosecution, prompting Assistant U.S. Attorney Michael Brittin to ask his key witness if he had ever seen any of the men wearing eyeglasses prior to trial.<sup>164</sup> In the dozens of hearings before the

---

ONLINE (Mar. 30, 2012, 3:52 PM), <http://www.dailymail.co.uk/news/article-2122990/The-nerd-defense-How-violent-criminals-turning-framed-hipster-glasses-persuade-juries-intelligent-honest.html> (reporting that defense attorney Harvey Slovis gives “all his defendants thick-rimmed, nonprescription glasses before court appearances to make them look less threatening”).

<sup>160</sup> Minara El-Rahman, *Nude Housekeeper Not Guilty of Murder*, FINDLAW (Jan. 21, 2011, 6:15 AM), <http://blogs.findlaw.com/blotter/2011/01/nude-housekeeper-not-guilty-of-murder.html> (reporting that in a taped statement, Cordero confessed to police that John Conley attempted to rape him, at which time he took a knife from Conley and stabbed him, but that at trial Cordero stated that he wanted to recant his confession because “he was coerced by Detective Steven Berger”).

<sup>161</sup> Deutsch, *supra* note 150.

<sup>162</sup> Keith L. Alexander, *Deliberations Begin in Five-Man District Murder Trial*, WASH. POST (Apr. 25, 2012, 4:26 PM), [http://www.washingtonpost.com/blogs/crime-scene/post/deliberations-in-five-man-district-murder-trial/2012/04/25/gIQAT15RhT\\_blog.html](http://www.washingtonpost.com/blogs/crime-scene/post/deliberations-in-five-man-district-murder-trial/2012/04/25/gIQAT15RhT_blog.html); *see also* *South Capitol Street Murders: Sanquan Carter, Orlando Carter, Jeffrey Best, Robert Bost and Lamar Williams to be Sentenced*, ABC NEWS (Sept. 11, 2012 5:47 AM), <http://www.wjla.com/articles/2012/09/south-capitol-street-murders-sanquan-carter-orlando-carter-jeffrey-best-robert-bost-and-lamar-willia.html>.

<sup>163</sup> Alexander, *supra* note 9.

<sup>164</sup> The defendants’ identities were at issue in the case, allowing the prosecution to question two witnesses. Key witness Nathaniel Simms, who had previously pled guilty in the case, responded with an unequivocal “no” to U.S. Attorney Michael Brittin’s question concerning Orlando Carter’s use of eyewear prior to trial. The answer was the “same for each of Carter’s co-defendants.” *Id.*; *see also* *District of Columbia v. Carter*, No. 2010 CF1 005677 (D.C. Super. Ct. 2012).

trial, only one defendant had donned eyeglasses.<sup>165</sup> Prosecutors took advantage of the opportunity to suggest to jurors that the defendants were being “dishonest in misrepresenting their appearance.”<sup>166</sup> All five defendants were found guilty.<sup>167</sup> After the trial, one prosecutor suggested that the defendants were putting on a “schoolboy act.”<sup>168</sup> Patricia Jefferies, grandmother of one of the victims, agreed, arguing that the defendants’ strategy was aimed at “influencing the jury, trying to make them think they’re Boy Scouts or something.”<sup>169</sup> Together, these cases demonstrate the ease with which a defendant can add eyeglasses to his or her look to influence the jury.

#### IV. THE INTERSECTION OF THE NERD DEFENSE AND THE CHANGE-OF-APPEARANCE INSTRUCTION

##### A. *Harris v. United States*

The case of *Harris v. United States* marks the first instance in which a defendant’s use of nonprescriptive eyewear at trial became an explicit issue on appeal.<sup>170</sup> In July 2008, a jury for the Superior Court of the District of Columbia found Donnell

---

<sup>165</sup> See Benjamin R. Freed, *Murder Defendants Try Wearing Hipster Glasses in Fashionable Attempt to Win Over Juries*, DCIST (Mar. 27, 2012, 4:15 PM), [http://dcist.com/2012/03/defense\\_attorneys\\_try\\_giving\\_client.php](http://dcist.com/2012/03/defense_attorneys_try_giving_client.php) (noting that Lamar Williams was the only defendant known to have ever worn eyeglasses before trial).

<sup>166</sup> *Id.*

<sup>167</sup> *Five Men Convicted of Charges for Their Roles in Crimes That Led to Five Murders, Nine Other Shootings*, U.S. DEP’T OF JUSTICE (May 7, 2012), <http://www.justice.gov/usao/dc/news/2012/may/12-161.html> (reporting that all five men were convicted of murder, conspiracy, and other charges related to violent crimes that culminated in the March 30, 2010 fatal shootings on South Capitol Street); see also Carter, No. 2010 CF1 005677; District of Columbia v. Carter, No. 2010 CF1 005176 (D.C. Super. Ct. 2012); District of Columbia v. Best, No. 2010 CF1 007370 (D.C. Super. Ct. 2012); District of Columbia v. Bost, No. 2010 CF1 007155 (D.C. Super. Ct. 2012); District of Columbia v. Williams, No. 2010 CF1 007157 (D.C. Super. Ct. 2012).

<sup>168</sup> Alexander, *supra* note 9.

<sup>169</sup> Freed, *supra* note 165.

<sup>170</sup> *Harris v. United States*, No. 08-CF-1405, at 5 (D.C. Cir. 2012).

Harris guilty of second degree murder for the fatal shooting of Michael Richardson.<sup>171</sup> Soon after sentencing, Harris filed a notice of appeal with the District of Columbia Court of Appeals.<sup>172</sup> Harris' use of eyewear, a "seemingly innocuous detail" during the trial, was a "key issue at the heart of [his] appeal."<sup>173</sup> On appeal, Chief Judge Eric Washington found Harris' use of eyeglasses to be one of the case's most compelling issues.<sup>174</sup> Throughout trial, Harris consistently donned eyeglasses despite not having worn eyeglasses prior to trial.<sup>175</sup> This prompted the prosecution to request a change-of-appearance instruction, a request that the judge granted.<sup>176</sup>

When a court issues a change-of-appearance instruction, the language used by the court can be damning to the defendant if the jury determines that the defendant has, in fact, changed his or her appearance.<sup>177</sup> This is due to the inference of a

---

<sup>171</sup> The government presented evidence that Harris entered Joe's Steak and Egg Restaurant at approximately 2:00 AM on June 29, 2007. Harris asked to use the phone, and, when his request was denied, he left the restaurant. Harris subsequently reentered the restaurant, where witnesses saw him arguing with Richardson, an intern with the D.C. Public Defender Service. Some time later, multiple gunshots were fired, fatally wounding Richardson. *Id.*

<sup>172</sup> See *Court Cases Online Database*, D.C. CTS., <https://www.dccourts.gov/cco/maincase.jsf> (last visited Dec. 1, 2012) (search for "Harris, Donnell") (reporting that Harris first filed a notice of appeal on Oct. 21, 2008).

<sup>173</sup> Tillman, *supra* note 5. On appeal, Harris also argued that the trial court erred by (1) overruling the defendant's objections to statements made during the prosecution's closing argument, (2) excluding from jury instructions the defendant's theory that someone else committed the murder, and (3) denying a motion for acquittal despite there being insufficient evidence against the defendant. See Brief of Appellant, *Harris*, No. 08-CF-1405.

<sup>174</sup> Tillman, *supra* note 5.

<sup>175</sup> *Harris*, No. 08-CF-1405, at 5.

<sup>176</sup> *Id.* at 6.

<sup>177</sup> Change-of-appearance instructions "contemplate[] some independent evidence" that the defendant is the one who actually changed his or her appearance. *United States v. Perkins*, 937 F.2d 1397, 1403 (9th Cir. 1991). For example, this occurs where a defendant, shortly after committing a crime, cuts or colors his or her hair or shaves his beard. *Id.*

defendant's consciousness of guilt, which "flows from any change of appearance" instruction that is given to the jury.<sup>178</sup> Change-of-appearance instructions generally contain language that attributes to the defendant "consciousness of guilt" or "fear of being identified."<sup>179</sup> In *Harris*, the change-of-appearance instruction, issued as a result of the defendant's sudden use of eyeglasses, raised questions as to whether the defendant's rights had been prejudiced.<sup>180</sup> Defense counsel asserted that the change-of-appearance instruction is reserved for situations that "refer[]

---

<sup>178</sup> *Id.*

<sup>179</sup> *See, e.g.,* United States v. Carr, 373 F.3d 1350, 1353 (D.C. Cir. 2004) ("A defendant's attempt to change his appearance after a crime has been committed does not create a presumption of guilt. An innocent person charged with a serious offense may resort to various means, both lawful and unlawful, to avoid prosecution. On the other hand, you may consider evidence of the defendant's attempt to change his appearance as tending to prove the defendant's *fear of being identified and therefore his consciousness of guilt*. You are not required to do so." (emphasis added)); *see also* Perkins, 937 F.2d at 1402 n.3 (discussing a defendant's appeal of a conviction for bank robbery because the district court instructed the jury that "[a] defendant's intentional change of his appearance immediately after the commission of a crime or after he is accused of a crime that has been committed, is not, of course, sufficient in itself to establish his guilt, but may be considered by the jury in the light of all other evidence in the case in determining guilt or innocence," and noting that "[w]hether or not evidence of a change of appearance shows a *consciousness of guilt* and the significance to be attached to any evidence, are matters exclusively within the province of the jury" (emphasis added) (citing Devitt & Blackmar, FEDERAL JURY PRACTICE AND INSTRUCTIONS, § 15.08)).

<sup>180</sup> Trial Transcript at 87–88, United States v. Harris, No. CF1-18801-07 (D.C. Super. Ct. 2008), *reprinted in* Appellant's Limited Appendix, *Harris*, No. 08-CF-1405 ("You heard evidence that Donnell Harris attempted to change his appearance to avoid being identified. It is up to you to decide that he took these actions. If you find he did so, you may consider this evidence as tending to show his *feelings of guilt* which you may in turn consider as *tending to show actual guilt*. On the other hand, you may also consider that he may have taken these actions for reasons fully consistent with innocence in this case." (emphasis added)); *see also* BARBARA BERGMAN, CRIMINAL JURY INSTRUCTIONS FOR THE DISTRICT OF COLUMBIA § 2.303(B) (5th ed. 2008). The defense argued that the trial court issued this instruction in error because the government did not establish that Harris was "attempting to conceal his identity by wearing glasses." Brief of Appellant, *supra* note 173, at 6.

to [a defendant] doing things like shaving his head, as opposed to having dreadlocks . . . [and] shaving his beard, as opposed to having facial hair.”<sup>181</sup> The defense further argued that Harris needed the eyeglasses in order to “read through voluminous material,” although there was no such evidence presented at trial to back this assertion.<sup>182</sup> The prosecution rebutted the defense’s stance by calling two key witnesses to testify that they had never previously seen Harris wearing eyeglasses.<sup>183</sup> At trial, the government argued that Harris’ eyewear was “an attempt at concealment” because “eyeglasses do change appearance . . . .”<sup>184</sup> The D.C. Court of Appeals agreed with the trial court that the “wearing of glasses at trial had some probative value[] and that the prejudicial effect did not outweigh its probative value.”<sup>185</sup> The appeals court affirmed the lower court’s ruling, in part, because the appeals court determined that Harris’ rights had not been prejudiced by the change-of-appearance instruction.<sup>186</sup>

---

<sup>181</sup> *Harris*, No. 08-CF-1405, at 4–5.

<sup>182</sup> *Id.* at 5. The trial court informed the jury that “there is no evidence in the record that Mr. Harris needs glasses to read or anything else,” finding Harris’ explanation for his use of eyeglasses to be mere “speculation.” *Id.*; see also Brief for Appellee, *supra* note 6, at 30 n.32.

<sup>183</sup> See Brief of Appellant, *supra* note 173, at 6 (noting that the defense argued that eyeglasses could not conceal the identity of the defendant to someone who knew him well and pointing out that one witness knew Harris his entire life and another witness encountered Harris on a regular basis as a routine customer); see also *Harris*, No. 08-CF-1405, at 4–5 (observing that Francis Iwuh knew Harris since infancy and Marion Sesay knew Harris as a regular customer at the Steak and Egg Restaurant where the shooting took place).

<sup>184</sup> Brief of Appellant, *supra* note 173, at 5; Leder et al., *supra* note 24, at 212; Tillman, *supra* note 5 (quoting U.S. Attorney John Gidez, who argued “that even if [wearing eyeglasses] was not a profound change, it could still alter Harris’ appearance enough to potentially cause a non-identification or misidentification by witnesses who didn’t know Harris well enough to recognize him with or without glasses”).

<sup>185</sup> *Harris*, No. 08-CF-1405, at 5. The trial court’s reasoning aligns with the proposition that a defendant’s appearance may become relevant evidence if it “forms the basis of identification” in the case. Levenson, *supra* note 79, at 577 n.19.

<sup>186</sup> *Harris*, No. 08-CF-1405, at 6 (affirming the lower court ruling and



The D.C. Court of Appeals premised its justification for upholding the lower court's change-of-appearance instruction on two factors: (1) a defendant's identification must be at issue and (2) a defendant must have "significant[ly]" changed his or her appearance before trial.<sup>187</sup> The fundamental problem with this opinion is that it does not define the scope of "identification" matters for purposes of issuing a change-of-appearance instruction.<sup>188</sup> Does the defendant's identification need to be specifically at issue, as it was in *Harris*, for the court to properly issue a change-of-appearance instruction? Or can the holding in *Harris* be interpreted to encompass all situations in which a witness is asked to identify the defendant simply as a procedural requirement—even when no genuine issue of identification is present?<sup>189</sup> *Harris* leaves unanswered two critical

---

holding that even if the change-of-appearance instruction was in error, it was harmless under *Kotteakos v. United States*, 328 U.S. 750, 764–65 (1946)); Zoe Tillman, *Appeals Court Rules Glasses Instruction Not Prejudicial in Murder Trial*, BLT: BLOG OF LEGAL TIMES (Oct. 24, 2012, 1:33 PM), <http://legaltimes.typepad.com/blt/2012/10/appeals-court-rules-glasses-instruction-not-prejudicial-in-murder-trial.html>.

<sup>187</sup> *Harris*, No. 08-CF-1405, at 5–6 (focusing on identification by relying heavily on *United States v. Carr*, 373 F.3d 1350, 1353 (D.C. Cir. 2004), which reasoned that a defendant's change in appearance should have been coupled with "anticipa[tion] that witnesses would be called at trial to identify him").

<sup>188</sup> *Harris*, No. 08-CF-1405, at 5 (noting that the trial court recognized that a change in the defendant's appearance must be significant to warrant a change-of-appearance instruction). The trial court touched on the scope of the change-of-appearance instruction when the trial judge stated, "I'm not sure if [wearing glasses] is [an] attempt to change his appearance so he couldn't be identified. It's not like he changed his appearance before a lineup or before some photographic identification. He's wearing glasses now." Brief of Appellant, *supra* note 173, at 5. Still, the trial court issued the instruction and the appeals court affirmed the instruction without any further clarification. *Harris*, No. 08-CF-1405, at 6.

<sup>189</sup> When the defense stipulates to a defendant's identification, there is no genuine issue as to identification. *Compare* *United States v. Alexander*, 48 F.3d 1477, 1490 (9th Cir. 1995) ("Identification of the defendant as the person who committed the charged crime is always an essential element which the government must establish beyond a reasonable doubt."), *with* *United States v. Darrell*, 629 F.2d 1089, 1091 (5th Cir. 1980) ("[A] witness need not physically point out a defendant so long as the evidence is sufficient

questions: (1) Can the change-of-appearance instruction be given in a case in which a defendant's identification is not specifically at issue but the defendant wears nonprescriptive eyeglasses to trial? and (2) How should a judge instruct the jury concerning a defendant's "significant" change of appearance when the defendant's identity is not specifically at issue?<sup>190</sup> This type of defendant is not attempting to avoid identification;<sup>191</sup> rather, he or she is attempting to misguide the jury with persistent and subtle changes in appearance that are intimately linked with society's most deeply rooted stereotypes.<sup>192</sup> To avoid such misguidance, a jury should be instructed in a manner that balances a defendant's right of expression against the jury's right to the truth.

*B. Carefully Balancing a Defendant's Constitutional Rights  
Against Potential Jury Manipulation*

A defendant's right to a fair trial is one of his or her fundamental liberties, a right protected by the Due Process Clause of the Fourteenth Amendment.<sup>193</sup> Due Process Clause principles firmly hold that the State cannot force a defendant to appear before a jury in a manner that suggests that the defendant

---

to permit the inference that the person on trial was the person who committed the crime.").

<sup>190</sup> This question is unanswered by the court in *Harris*. *Harris* is the first case focusing solely on a defendant's use of unnecessary eyeglasses to signal a "significant" change in appearance and thus justifying the resulting jury instruction. Compare *Harris*, No. 08-CF-1405, at 5, with *Carr*, 373 F.3d at 1353 (looking at a combination of the defendant's beard, weight, and glasses to signal "profound alterations" in appearance and justifying the resulting jury instruction).

<sup>191</sup> A defendant who stipulates to identification is not attempting to avoid being identified. Attempting to avoid identification through a significant change in appearance would warrant the issuing of a change-of-appearance instruction to the jury. See *Harris*, No. 08-CF-1405, at 4-5.

<sup>192</sup> See Hellström & Tekle, *supra* note 20, at 695 (articulating that judgments about intelligence and successfulness can be traced back to the development of myopia caused by extensive schoolwork in childhood days).

<sup>193</sup> U.S. CONST. amend. XIV; *Estelle v. Williams*, 425 U.S. 501, 503 (1976).

is guilty.<sup>194</sup> During trial, a defendant is “on display for the jury.”<sup>195</sup> Consequently, members of the jury might notice and take into account details of the defendant’s appearance that nonjurors might find irrelevant.<sup>196</sup> But how far may a defendant go to change his or her appearance in order to convey innocence? Courts have held that a defendant may wear to trial such clothing items as religious cult wear,<sup>197</sup> sweatshirts with religious symbols,<sup>198</sup> and official military academy dress uniforms.<sup>199</sup> A defendant has a First Amendment right to control his or her appearance at trial.<sup>200</sup> Generally, this right is subject to the judge’s discretion.<sup>201</sup> However, when a defendant’s dress in the courtroom involves religious attire, the standard for

---

<sup>194</sup> Shepard, *supra* note 82, at 2208.

<sup>195</sup> Levenson, *supra* note 79, at 575.

<sup>196</sup> THORNTON, *supra* note 83, at 112; *see also* Levenson, *supra* note 79, at 574 (noting that “the outcome of the case is affected by many factors that are technically not evidence: the quality of the lawyers’ presentations, the appearance and reaction of the defendant in the courtroom, and even the presence of the victim’s representatives”).

<sup>197</sup> *See* United States v. Yahweh, 779 F. Supp. 1342, 1345 (S.D. Fla. 1992) (holding that defendants may choose to wear to trial religious cult uniforms, including white robes and white turbans).

<sup>198</sup> Joseph v. State, 642 So. 2d 613, 613 (Fla. Dist. Ct. App. 1994) (holding that defendants may wear to trial shirts with religious symbols).

<sup>199</sup> Johnson v. Commonwealth, 449 S.E.2d 819, 820–21 (Va. Ct. App. 1994) (holding that defendants may wear to trial official military uniforms).

<sup>200</sup> U.S. CONST. amend. I; *see In re Palmer*, 386 A.2d 1112, 1115 (R.I. 1978) (recognizing the mandate in *Sherbert v. Verner*, 374 U.S. 398 (1963), to strike a balance between a defendant’s First Amendment right and the “interest of the court in maintaining decorum in its proceedings by regulating dress in the courtroom”); *see also Yahweh*, 779 F. Supp. at 1345 (stating that defendants may choose to wear “suitable clothing of their choice in the courtroom”).

<sup>201</sup> *Johnson*, 449 S.E.2d at 820–21 (“The conduct of a trial includes courtroom decorum. The trial court has the duty and the authority, in the exercise of sound discretion, to require persons attending court to dress in a manner appropriate to their functions and consistent with the publicity and dignity of the courtroom.”); *see also* Catherine Theresa Clarke, *Missed Manners in Courtroom Decorum*, 50 MD. L. REV. 945, 1001 (1991) (noting the concern that some clothing can distract or offend judges as “a breach of etiquette because it undermines the serious, professional atmosphere of the proceedings”).

regulating such dress is higher: the government must demonstrate a compelling interest.<sup>202</sup>

By wearing nonprescriptive eyeglasses to trial, a defendant attempts to cultivate an image premised on potentially misleading character traits that are associated with wearing eyeglasses.<sup>203</sup> A defendant's use of nonprescriptive eyeglasses therefore presents a unique challenge to the criminal court system: it is inconsistent with the First Amendment to prohibit a defendant's free expression through the use of nonprescriptive eyeglasses at trial, but it is also inconsistent with the truth-seeking principles of the judicial system to allow a defendant to purposefully mislead a jury. This Note proposes a modified change-of-appearance instruction that mitigates potential jury manipulation and that does not carry with it the same presumption of guilt as a standard change-of-appearance instruction concerning specific identification matters.

## V. PROPOSED SOLUTIONS

*Harris* confirms that the prosecution may inquire into a defendant's use of unnecessary eyeglasses and request a change-

---

<sup>202</sup> See *In re Palmer*, 386 A.2d at 1115 (noting the need in the courtroom to “accommodate the right to exercise the religious freedoms safeguarded by the first amendment with the right of the state to regulate these individual freedoms for the sake of societal interests”); see also *McMillan v. State*, 265 A.2d 453, 456 (Ct. App. Md. 1970) (stating that “[w]e are fully aware that the orderly administration of courts of justice requires the maintenance of dignity and decorum and for that reason rules of conduct and behavior to govern participants are essential . . . Understandably, respect for the courts is something in which the State has a compelling interest”).

<sup>203</sup> See *Brown*, *supra* note 9, at 2–6 (finding that defendants who wear eyeglasses appear more intelligent and less physically threatening); *Terry & Krantz*, *supra* note 41, at 1766 (finding that wearing eyeglasses increases ratings for character, compassion, honesty, and sensitivity—but that eyeglasses decrease ratings of attractiveness and forcefulness); *Harris*, *supra* note 56, at 1674 (finding that those who wear eyeglasses appear more timid and more intelligent than those who do not wear eyeglasses); Aylin Zafar, “Hipster” Glasses Might Get You Off the Hook in Court, *TIME* (Mar. 30, 2012), <http://newsfeed.time.com/2012/03/30/hipster-glasses-might-get-you-off-the-hook-in-court/> (noting that eyeglasses help make an individual appear “a little emasculated”).

of-appearance instruction when the defendant's identification is specifically at issue.<sup>204</sup> Presumably, this is because eyeglasses tend to cover a significant portion of the eye region and can restructure the appearance of facial features,<sup>205</sup> making it difficult to recognize a defendant who wears eyeglasses.<sup>206</sup> In *Harris*, the appeals court agreed with the lower court that the "wearing of glasses at trial [has] some probative value" which is not outweighed by its prejudicial effect.<sup>207</sup> However, when a defendant's identification is not specifically at issue, the prosecution cannot request a change-of-appearance instruction because such an instruction is designed to address changes in appearance related to potential misidentification.<sup>208</sup> This Note proposes two possible solutions to this problem.

*A. Modifying the Change-of-Appearance Instruction When the Defendant's Identification Is Not Specifically at Issue*

A defendant's use of nonprescriptive eyewear at trial generally constitutes a specific attempt to intentionally misguide the jury, and it works against the fundamental principles of a judicial system that seeks the truth in all cases. Jury awareness of this tactic will help to lessen the impact of intentional jury manipulation. As currently utilized by courts, the change-of-appearance instruction is particularly harsh because it can imply a consciousness of guilt.<sup>209</sup> Although change-of-appearance

---

<sup>204</sup> *Harris v. United States*, No. 08-CF-1405, at 5–6 (D.C. Cir. 2012).

<sup>205</sup> SMITH & MALANDRO, *supra* note 2, § 1.21, at 42.

<sup>206</sup> See Leder et al., *supra* note 24, at 216–18 (finding that "glasses impede the immediate recognition of faces" because it takes longer to recognize faces with full-rim glasses than it does to recognize faces either without glasses or with rimless glasses).

<sup>207</sup> *Harris*, No. 08-CF-1405, at 5.

<sup>208</sup> *United States v. Perkins*, 937 F.2d 1397, 1403 (9th Cir. 1991); see also MULDOON, *supra* note 16, § 9:262 ("The prosecutor may properly comment on the defendant's changed appearance at trial, as compared to the time of the crime, where identification is a trial issue.").

<sup>209</sup> Inferences drawn about a defendant's "consciousness of guilt" reasonably "flow[] from any change of appearance" instruction that is given to the jury. *Perkins*, 937 F.2d at 1403; *People v. Slutts*, 259 Cal. App. 2d

instructions typically specify that the charge does not carry a presumption of guilt, jurors might be unable to ignore the harsh language of the instruction or jurors might give the inference too much weight.<sup>210</sup> For these reasons, the standard change-of-appearance instruction should be reserved for situations in which a defendant has significantly changed his or her appearance and where his or her identification is specifically at issue in the case.

A defendant who seeks to encourage misidentification through the use of unnecessary eyewear should be distinguished from a defendant who requires a prescription for eyeglasses.<sup>211</sup> When a defendant dons unnecessary eyeglasses for purposes of persuasion<sup>212</sup>—but *not* for purposes of misidentification—the jury should be made aware through a modified change-of-appearance instruction.

This Note's proposed modification of the change-of-appearance instruction removes the language connecting a defendant's change of appearance to his or her consciousness of guilt in order to account for a defendant's right of free expression. The modification expands the scope of a standard change-of-appearance instruction to cover a defendant's use of eyeglasses as a means to unofficially introduce persuasive and

---

886, 890, 893 (1968) (noting that the defendant's significant change in appearance correlated to consciousness of guilt because "consciousness of guilt can reasonably be inferred from [defendant's] action in shaving off his beard shortly after the [crime]"); *see also Harris*, No. 08-CF-1405, at 5 (quoting *Lazo v. United States*, 930 A.2d 183, 187 (D.C. Cir. 2007)) ("When supported by the evidence, we have recognized the legitimacy of the change of appearance argument based on the attempted change of appearance jury instruction.").

<sup>210</sup> *See generally* Levenson, *supra* note 79, at 581 ("Juries are not machines and courtrooms are not laboratories. Laboratories are controlled environments in which trial and error are accepted protocol. Even with rules of evidence, trials do not assume the same type of controlled, sterile environment.").

<sup>211</sup> A defendant who "significantly changes" his or her appearance before trial in a case where his or her identity is specifically at issue has, by default, changed appearance to avoid identification. Attempting to avoid identification would warrant the issuing of a standard change-of-appearance instruction. *Harris*, No. 08-CF-1405, at 5–6.

<sup>212</sup> For example, a defendant with no history of wearing eyeglasses who then wears eyeglasses to trial *and* whose identity is not specifically at issue.

misleading character evidence and further instructs the jury on how to properly consider such a tactic. This change-of-appearance instruction should be given when: (1) a defendant dons unnecessary eyeglasses to trial and (2) identification of the defendant is not specifically at issue. The proposed modification, adapted from the instructions given in *Carr*<sup>213</sup> and *Harris*,<sup>214</sup> reads as follows:

During trial, the defendant changed his or her appearance by wearing eyeglasses that he or she does not need. This particular alteration in appearance after the commission of a crime and in preparation for trial does not create a presumption of guilt. It is entirely possible that an innocent person would resort to both lawful and unlawful means to avoid prosecution. The wearing of unnecessary eyeglasses at trial is lawful.

In this case, the defendant's wearing of eyeglasses constituted a falsification of a vision deficiency. You may consider this falsification an attempt by the defendant to gain favorable judgment based upon the positive social stereotypes associated with the wearing of eyeglasses, which can include truthfulness, intelligence, and nonaggressive demeanor.

When you consider the evidence presented in this case, you may take into account the defendant's choice to appear at trial wearing eyeglasses that he or she does not need. You are not required to do so.

A defendant's use of unnecessary eyeglasses at trial silently and unofficially introduces character evidence.<sup>215</sup> Consequently, when

---

<sup>213</sup> See *United States v. Carr*, 373 F.3d 1350, 1353 (D.C. Cir. 2004).

<sup>214</sup> See Trial Transcript at 87-88, *United States v. Harris*, No. CF1-18801-07 (D.C. Super. Ct. 2008), reprinted in Appellant's Limited Appendix, *Harris*, No. 08-CF-1405.

<sup>215</sup> See *Brown*, *supra* note 9, at 3 (using a case with "purposefully ambiguous evidence" to examine the effect of eyeglasses on juror perceptions of defendants and finding both a direct link between eyeglasses and perception of increased intelligence and a correlation between increased intelligence and fewer guilty verdicts). Everything about a defendant's appearance has an "impact in the courtroom." SMITH & MALANDRO, *supra* note 2, § 1.26, at 54.

a defendant's identity is only at issue as a procedural requirement,<sup>216</sup> the prosecution is limited to informing the jury of the defendant's use of unnecessary eyewear through relevant admitted evidence, such as photographs,<sup>217</sup> answers to juror inquiries,<sup>218</sup> or evidence first introduced by the defendant.<sup>219</sup> However, these methods are insufficient because they are unpredictable and leave jurors to consider evidence concerning the defendant's use of unnecessary eyeglasses but *without* proper instruction as to how to consider such conduct.<sup>220</sup>

Any inquiry into a defendant's unnecessary use of eyeglasses at trial will likely be aimed at attacking a defendant's truthfulness.<sup>221</sup> However, even if a defendant first introduces evidence of his or her truthfulness by taking the stand,<sup>222</sup> Federal

---

<sup>216</sup> For example, if the defense stipulates to the defendant's identity before trial.

<sup>217</sup> If a photograph presented as evidence at trial depicts a defendant without eyeglasses and the defendant subsequently wears eyeglasses at trial, the jury may acknowledge the defendant's change in appearance.

<sup>218</sup> Certain states allow jurors to pose questions to defendants during trial. During Jodi Arias' capital murder trial, the jury posed two specific questions concerning her eyeglasses: "What is your eye prescription?" and "If you are so nearsighted then how could you drive?" Graham Winch, *Arias Grilled With Questions By Jurors*, HLN LIVE BLOG (Mar. 6, 2013), <http://www.hlnv.com/article/2013/03/06/live-blog-what-will-jurors-ask-jodi-arias>.

<sup>219</sup> See FED. R. EVID. 404(a)(2)(A).

<sup>220</sup> See generally 3 CLIFFORD S. FISHMAN & ANNE T. MCKENNA, JONES ON EVIDENCE § 16:26 (7th ed. 1997) (stating that twelve federal circuits express a strong preference that when a jury is instructed on the issue of a defendant's character, "the judge instruct the jury to consider evidence relating to defendant's character together with the rest of the evidence in the case").

<sup>221</sup> "Character" embraces the quality of truthfulness, and although "character" does not include having either "good eyesight or impaired vision," a defendant's eyesight becomes linked to his or her truthfulness when determining whether the defendant truly requires eyeglasses. MUELLER & KIRKPATRICK, *supra* note 15, § 4:23.

<sup>222</sup> 2 MUELLER & KIRKPATRICK, *supra* note 15, § 4:43 (stating that "[w]hen defendants [who take the stand] describe good behavior, patterns, an honest, hardworking, nonviolent, or caring disposition, they open to prosecutors the right to cross-examine on specific acts relevant to that testimony").



Rule of Evidence 608(b) precludes the prosecution's use of extrinsic evidence for the sole purpose of attacking the defendant's truthfulness.<sup>223</sup> Subject to the court's discretion, on cross-examination a prosecutor may inquire into the defendant's use of unnecessary eyeglasses if the court deems such information to be "probative of the [defendant's] character for truthfulness or untruthfulness."<sup>224</sup> Nonetheless, this evidence may still be excluded under Federal Rule of Evidence 403 due to its potential for prejudice.<sup>225</sup> Therefore, the modified change-of-appearance instruction is necessary to adequately inform jurors of the defendant's purposeful attempt to misguide the jury and to ensure that jurors are properly instructed as to *how* to consider the defendant's actions. This Note's proposed instruction functions as a safeguard against potential jury manipulation because it provides the prosecution with a means of countering a defendant's strategic use of eyeglasses as a prop to elicit juror biases. It ensures that jurors are made aware of and know how to consider such information, while at the same time it informs jurors that the nerd defense does not correlate to a defendant's consciousness of guilt.

### *B. Making an Eyeglasses Inquiry the "Norm" at Trial*

Prosecution teams and law students should be exposed to the tactics employed by defense teams. It is important for current and future prosecutors to learn how and under what circumstances to request a change-of-appearance instruction and to learn how to ask questions about a defendant's misleading utilization of eyeglasses. This will ensure that a jury is better

---

<sup>223</sup> See FED. R. EVID. 608 advisory committee's notes (stating that Rule 608(b) "has been amended to clarify that the absolute prohibition on extrinsic evidence applies only when the sole reason for proffering that evidence is to attack or support the witness' character for truthfulness"); see also *United States v. Fusco*, 748 F.2d 996, 998 (5th Cir. 1984) (noting that the principles "embodied in Federal Rule of Evidence 608 . . . limit the use of evidence designed to show that the witness has done things, unrelated to the suit being tried, that make him more or less believable per se").

<sup>224</sup> FED. R. EVID. 608(b).

<sup>225</sup> FED. R. EVID. 403.

equipped to properly consider a defendant's strategic change of appearance.

If the prosecution is suspicious of a defendant's sudden use of eyeglasses at trial, the prosecution should be allowed to inquire, in the absence of the jury, into the defendant's need for eyeglasses. If the defendant is unable or unwilling to offer proof of his or her need for eyeglasses—for example, through a prescription, evidence of prior use of eyeglasses, or an eye exam—then the court should grant the prosecution's request for a modified change-of-appearance jury instruction. One likely objection to this rule is that indigent defendants might be unable to pay for an eye exam that is necessary to prove their need for eyeglasses. As such, any rule requiring defendants to offer proof of their need for eyeglasses needs to be accompanied by a rule requiring the state to pay for any necessary eye exams. Another objection to this rule might be that defendants should not be required to assist in their own prosecution. However, wearing unnecessary eyeglasses is a defendant's choice and such a strategic accessory serves to mislead the jury. Making an eyeglasses inquiry the norm might lead defense attorneys and defendants to think twice before employing the nerd defense—and therefore lessen the ability of defendants to hinder the truth-seeking process by purposefully eliciting deep-seated biases in jurors.

#### CONCLUSION

The Supreme Court has long recognized that the right to an impartial jury, afforded by the Sixth Amendment,<sup>226</sup> is fundamental to a fair trial.<sup>227</sup> The right to an impartial jury includes the right to exclude potentially biased jurors.<sup>228</sup> While

---

<sup>226</sup> See U.S. CONST. amend. VI (providing that “[i]n all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury”).

<sup>227</sup> See, e.g., *Taylor v. Louisiana*, 419 U.S. 522, 525 (1975) (stating that the Sixth Amendment guarantees an impartial jury trial).

<sup>228</sup> See *Turner v. Murray*, 476 U.S. 28, 36 (1986) (holding that “by refusing to question prospective jurors on racial prejudice, the trial judge failed to adequately protect petitioner's constitutional right to an impartial

the trial process offers a system to exclude jurors with potential biases, “jurors may not be willing to reveal their biases, or they simply may not recognize they have any biases.”<sup>229</sup> To counteract unconscious biases held by jurors,<sup>230</sup> the Court has held that empaneling jurors from “a cross-section of the community” is a necessary ingredient of the selection of an impartial jury.<sup>231</sup> However, when a bias is widely held,<sup>232</sup> selection of a jury in this manner is insufficient by itself to counteract such a bias.<sup>233</sup> Research shows that stereotypes about those who wear eyeglasses are so powerful as to cross cultural,<sup>234</sup> gendered,<sup>235</sup> and racial divides.<sup>236</sup> The biases

---

jury”); Wiley, *supra* note 74, at 227.

<sup>229</sup> THORNTON, *supra* note 83, at 108 (noting that “it is very difficult, if not impossible, to empanel a jury of twelve ‘blank slates’ capable of hearing evidence free of influence of past experiences”); *see also* SUTHERLAND & CRESSEY, *supra* note 85, at 442 (noting that in certain cases, “several thousand prospective jurors have been examined before twelve were secured” and “[i]n one Chicago trial 9,425 persons were summoned for jury duty and 4,821 were examined before twelve were finally selected”).

<sup>230</sup> *See* Darryl K. Brown, *The Role of Race in Jury Impartiality and Venue Transfers*, 53 MD. L. REV. 107, 122 (1994) (noting that the fair-cross-section doctrine is designed to address juror biases resulting from “deep-seated hunches and judgments about social life”).

<sup>231</sup> *See Taylor*, 419 U.S. at 530 (holding that “[w]e accept the fair-cross-section requirement as fundamental to the jury trial guaranteed by the Sixth Amendment and are convinced that the requirement has solid foundation”). For a discussion of the jury-selection process in the context of gender discrimination, *see Ballard v. United States*, 329 U.S. 187, 192 (1946). For a discussion of the jury-selection process in the context of racial discrimination, *see Smith v. State of Texas*, 311 U.S. 128, 130 (1940).

<sup>232</sup> Wiley, *supra* note 74, at 230 (arguing that discrimination based on physical appearance may be even more “menacing” in American culture than racial or gender discrimination because everyone discriminates based on appearance).

<sup>233</sup> *See People v. Wheeler*, 583 P.2d 748, 755 (1978) (“The only practical way to achieve an overall impartiality is to encourage the representation of a variety of such groups on the jury so that the respective biases of their members, to the extent they are antagonistic, will tend to cancel each other out.”).

<sup>234</sup> *See Manz & Lueck*, *supra* note 21, at 704.

<sup>235</sup> *See Harris*, *supra* note 56, at 1674–75.

<sup>236</sup> *See Brown*, *supra* note 9, at 3.

associated with those who wear eyeglasses are deeply ingrained in our minds in early stages of life.<sup>237</sup>

The United States judicial system is designed to eliminate juror biases. Purposefully eliciting *any* biases from the jury undermines the goal of the judicial system, which is to seek the truth in all cases.<sup>238</sup> While defendants have the right to control their appearance at trial, there exists a distinction between a defendant who simply presents himself or herself in “neat and clean attire” and with “good grooming” and a defendant who uses attire to present “an unrealistic suggestion of character.”<sup>239</sup> A defendant who wears *unnecessary* eyeglasses fabricates a vision handicap that is intimately tied to stereotypes of favorable characteristics and manipulates the jury into believing a lie: that the defendant truly requires eyeglasses. By providing a jury with a modified change-of-appearance instruction, a court will enable the jury to have a more complete and truthful base of knowledge when considering the facts of the case and the jury will be better equipped to consider the defendant’s change in appearance.

---

<sup>237</sup> Walline et al., *supra* note 44, at 223 (describing a study finding that children as young as six years old correlate wearing eyeglasses with character traits of intelligence and honesty).

<sup>238</sup> Strier, *supra* note 22, at 99.

<sup>239</sup> Johnson v. Commonwealth, 449 S.E.2d 819, 821 (Va. Ct. App. 1994).