Big tobacco focuses on the facts to hide the truth: an algorithmic exploration of courtroom tropes and taboos

Stephan Risi,1,2 Robert N Proctor1

ABSTRACT

Objective To use methods from computational linguistics to identify differences in the rhetorical strategies deployed by defence versus plaintiffs’ lawyers in cigarette litigation.

Methods From 318 closing arguments in 159 Engle progeny trials (2008–2016) archived in the Truth Tobacco Industry Documents, we calculated frequency scores and Mann-Whitney Rho scores of plaintiffs versus defence corpora to discover ‘tropes’ (frequent terms that one side uses disproportionately) and ‘taboos’ (rare terms that one side avoids scrupulously), identifying heretofore hidden rhetorical strategies of the industry while also casting light on strategies used by plaintiffs.11 13 While cigarette makers use words or phrases that tend to focus agency on the individual smoker, attorneys working for the plaintiffs (ie, injured smokers) tend to use words that refocus agency onto the industry.

To identify terms that are distinctive for plaintiffs or defendants, we used corpus comparison methods that, while originally developed in computational linguistics, have recently become popular in the field of digital humanities.14–16 Conducting a ‘distant’, quantitative reading of a corpus of texts can be used for many different purposes. Connelly, for example, has used statistical methods to identify patterns of document destruction in State Department communications, while Underwood has used ‘distant reading’ to explore how time elapses in novels and how literary prestige leaves linguistic traces.17–19

Scholars have shown how cigarette makers use rhetorics of freedom, choice and personal responsibility to blame smokers for their injuries.1 3–7 20–24 A broad scholarly literature also details how cigarette makers falsely claim that smoking’s harm and addictiveness have long been ‘common knowledge’ 3–5 7 25–27 Computational methods offer an important complement to this literature, allowing us to show, for example, that it is family members—the husband who warned his wife about smoking, the daughter who asked her father to give up cigarettes—who put the common in the ‘common knowledge’ defence. Another strength of these new methods is that they allow us to investigate what is not said: the verbal taboos that only become visible by comparing large bodies of defence rhetoric against arguments deployed by plaintiffs. Tobacco defence teams will not talk about the companies’ ‘customers’, for example, but rather only about ‘smokers.’ They may acknowledge that someone has ‘passed away’ but will never use the word ‘killed’. Computational techniques allow detection of broad, sometimes subtle, patterns of language that might otherwise escape notice—like divergent usage of pronouns—patterns that help us better understand the industry’s courtroom strategy.

INTRODUCTION

The language used by attorneys in tobacco litigation reveals key elements of the strategies deployed by cigarette makers and their courtroom opponents.1 3–5 According to industry lawyers, for example, smokers ‘passed away’ but were never ‘killed’; they always had the ‘ability to quit’ but were not ‘addicted’. Jurors, tobacco attorneys claim, should focus on the individual ‘facts’ of the case but not on the larger ‘truth’ about the industry. Language is, per Bolinger, ‘a loaded weapon,’ which means that words are not innocent conveyers of meaning.9 There is a subtle micropolitics in human speech, expressed in the kinds of words chosen by one side or another to deploy or to avoid.7–11

To explore this divergent use of words and phrases, we analysed closing arguments in 159 Engle progeny trials from 2008 through 2016. Using methods from corpus linguistics, we constructed tables of ‘tropes’ (frequent terms that one side uses disproportionately) and ‘taboos’ (rare terms that one side avoids scrupulously), identifying heretofore hidden rhetorical strategies of the industry while also casting light on strategies used by plaintiffs.11 13

METHODS AND DATA

Howard Engle versus RJ Reynolds et al was filed in Florida in 1994, following discovery of internal
industry documents demonstrating a decades-long conspiracy to hide the hazards of cigarettes and to deny manipulation of nicotine to create and sustain addiction.24 28 The case went to trial in Miami in 1998, and in 2000 the jury reached a verdict, awarding $145 billion in damages to Florida smokers. Cigarette makers appealed to the Florida Supreme Court, however, and in 2003 managed to have the Engle class decertified, meaning that nothing would be awarded to smokers as a group.38 Instead, individual smokers would have to petition to come before the court and plead for justice. As of 2019, only about 250 Engle cases have been brought to court, with a number of others settled out of court. The industry has effectively reduced its legal exposure from 800 000 claims to only a handful.

The closing arguments examined here are from those preserved on the Truth Tobacco Industry Documents website, which collects trial testimony as part of a broader effort to preserve records from cigarette litigation.30 For our study, we included only trials for which both plaintiff and defendant closings were available and which did not end in a mistrial. Three hundred eighteen transcribed closings from 159 individual trials were obtained as full-text searchable documents, yielding a plaintiffs’ corpus consisting of 10 577 683 words and a defence corpus of 10 715 122 words.

We used an ‘open-vocabulary’ approach, analysing individual terms and phrases instead of predefined psychometric categories tracking, say, specific emotions or the temporal orientation of speech.33 Our interest here was not so much in broader affective characteristics of our corpora but rather in how courtroom adversaries deploy different rhetorical strategies. For those interested in other distinguishing linguistic features, we provide a supplementary analysis using Linguistic Inquiry and Word Count (LIWC) categories (see online supplementary file 1, ‘LIWC Categories in Engle Trial Closing Arguments’).32 33

Vectorisation
To facilitate numerical analysis, we stripped out formatting details, including line numbers, timestamps and names of stenographic services (eg, Veritext) to retain only the transcribed verbal proceedings of the actual closings. We also removed discussions between lawyers and the judges (at sidebars, for example). We then turned this collection into a document-term matrix of the 50 000 most frequent 1- to 10-grams (unique word strings), including only those terms or phrases that appear at least 33 times in the dataset. We did not use a stemmer or lemmatiser during preprocessing, but we did lower case all terms and split contractions. We also added several synthetic tokens, combining terms of identical cognate form into singular terms (‘decision’ and ‘decisions’ became ‘decision/s,’ for example). We also combined ‘he’ and ‘she’ into a single term (‘s/he’) to avoid the idiosyncrasies of any particular trial—whether the smoker was male or female, for example.

Divergent terms
We used two different algorithms in our analysis—a frequency score (FS) and the Mann-Whitney Rho (MWR) score—each of which captures a different facet of how best to understand rhetorical distinctiveness (or divergence) between two different textual corpora.

The FS indicates how often a given term appears in plaintiffs’ as opposed to defence closings.31 33 The score ranges from 0 (when only defence attorneys use the term) to 1 (only plaintiffs use the term). A score of 0.8, for example, means that 80% of all instances of the term occur in plaintiffs’ closings, that is, it appears four times more often in plaintiffs’ than in defence closings. To account for the fact that the defence corpus contains slightly more words, we normalised scores by using relative frequencies rather than absolute counts. FSs are useful for identifying taboos: terms generally avoided by one side or the other. Some of these are trivial and uninteresting: names of attorneys score high by this metric (‘Cofer’ or ‘Gdanski’), for example, as do legal expressions like ‘her burden (of proof)’. We often find, however, that dramatically divergent FSs (close to 1 or 0) can reveal significant taboos.36 Cigarette industry lawyers will almost never use terms such as ‘profits’ or ‘replacement smokers’, for example, or ‘addictive drugs’, just as plaintiffs’ lawyers are not likely to say that the smoker in question ‘had the ability to quit’ or that ‘every pack’ carried a warning.

One drawback of using Frequency Scores is that they compare the entire plaintiffs’ corpus against the entire defence corpus, which means that we cannot tell how consistently a divergence appears throughout such closings—because a strongly divergent term might simply be due to idiosyncrasies of an individual trial, such as the name of a testifying witness or the verbal tic of a particular lawyer (only Steve Hammer and Alex Alvarez say ‘Objection Judge’, for example, instead of the more usual ‘Objection Your Honor’).

To obtain a measurement for consistency across all closings we used the MWR statistic, which ranks all documents by the frequency with which any given term appears in every closing.15 37 Normalised to produce scores between 0 and 1, MWR indicates whether most plaintiff or defence closings are clustered at the top or bottom of this ranking. The term ‘they’, for example, has an MWR score of 0.96, which means that, given a randomly selected plaintiff’s closing and a randomly selected closing by the defence, there is a 96% chance that ‘they’ will appear more often in the plaintiff’s closing. MWR is efficient at identifying general patterns that remain consistent across all documents, because it gives less weight to terms that appear in only a few trials. Such terms generally have nothing to do with strategy but rather only with the particulars of a given trial—such as who is trying the case or on whose behalf.

The two measures are complementary. The FS produces immediately interpretable results, telling us how much more often plaintiffs or defendants use a term. Terms scoring highest (or lowest) by this metric tell us about the taboos of each side, which one can imagine as imperatives or injunctions: “Whatever you do, don’t mention ‘profits’, or ‘replacement smokers’.” MWR scores allow us to identify subtler patterns that might otherwise escape notice: the interestingly divergent use of pronouns, for example. By FS the word ‘they’ is not particularly remarkable—plaintiffs’ lawyers use it only about twice as often as defence teams. MWR, however, reveals that this is an extremely important term for the plaintiffs, who almost invariably use it more often than defence attorneys as a way of drawing attention to misdeeds ‘they’ (the cigarette makers) have perpetrated. By contrast, smoker-focused terms such as ‘he’ or ‘she’ are deployed consistently more often by the defence. In this way, MWR helps us identify rhetorical strategies that, in the pre-algorithmic age, managed to hide in plain sight. Who would have thought that pronoun usage would be of strategic legal significance?

RESULTS
The most divergent terms by both FS and MWR capture many of the patterns we would expect to see in such trials (tables 1 and 2). Defence attorneys reference the plaintiffs, the risks taken by the smoker, the decisions he or she made and his or her ability...
Plaintiffs’ attorneys will often say:

<table>
<thead>
<tr>
<th>Term</th>
<th>Defence</th>
<th>Plaintiffs</th>
<th>MWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>plaintiff</td>
<td>4306</td>
<td>536</td>
<td>0.06</td>
</tr>
<tr>
<td>quit smoking</td>
<td>2975</td>
<td>435</td>
<td>0.07</td>
</tr>
<tr>
<td>s/he never</td>
<td>11 668</td>
<td>3922</td>
<td>0.08</td>
</tr>
<tr>
<td>Mr/Mrs/Ms</td>
<td>32 912</td>
<td>15 514</td>
<td>0.10</td>
</tr>
<tr>
<td>decision/s</td>
<td>2857</td>
<td>761</td>
<td>0.12</td>
</tr>
<tr>
<td>smoking</td>
<td>19 707</td>
<td>11 419</td>
<td>0.13</td>
</tr>
<tr>
<td>smoking is/was dangerous</td>
<td>23</td>
<td>631</td>
<td>0.13</td>
</tr>
<tr>
<td>risk/s</td>
<td>2580</td>
<td>791</td>
<td>0.14</td>
</tr>
<tr>
<td>warnings/s</td>
<td>2864</td>
<td>813</td>
<td>0.14</td>
</tr>
</tbody>
</table>

P values < 0.0001 for all MWR values.

This table excludes highly similar terms: ‘plaintiff’ is included, for example, but ‘the plaintiff’ and ‘plaintiff has’ are excluded.

MWR, Mann-Whitney Rho.

Research has indicated that to the extent the jury focuses on issues including the safety of cigarettes, corporate misconduct, sending a message to the tobacco companies, the plaintiff’s chances are enhanced. Conversely, if the jury focuses on the individual plaintiff, his choices, his actions, his environment and history, a defense verdict is more likely. In essence, the well-prepared plaintiff tries us; we try the plaintiff.40

In virtually every trial since the 1980s, the goal of both sides has been to emphasise the agency of the opposing side, while diminishing or hiding their own agency (see table 3, where terms are ranked by frequency of use). Algorithmic analysis shows that this strategic divergence has left its imprint on the pronouns used in courtroom rhetoric. MWR, which rates terms highly if they consistently appear more often in plaintiffs’ or defense documents, identifies he and she, as well as Mr, Mrs and Ms, as among the most distinctive terms for the defense. Such terms almost always refer to the smoker or his or her family. By contrast, we found that they (usually referring to the industry, see again table 1) is the second most distinctive term for the plaintiffs by MWR.

Pronoun politics: putting the smoker on trial (while making the industry invisible)

In 1983 or 1984, lawyers from Shook, Hardy & Bacon were running mock trials to develop strategies for upcoming litigation, notably Cipollone.40 They concluded that to counter plaintiffs’ attacks, the best strategy would be to focus on the individual smoker:

Research has indicated that to the extent the jury focuses on issues including the safety of cigarettes, corporate misconduct, sending a message to the tobacco companies, the plaintiff’s chances are enhanced. Conversely, if the jury focuses on the individual plaintiff, his choices, his actions, his environment and history, a defense verdict is more likely. In essence, the well-prepared plaintiff tries us; we try the plaintiff.40

In virtually every trial since the 1980s, the goal of both sides has been to emphasise the agency of the opposing side, while diminishing or hiding their own agency (see table 3, where terms are ranked by frequency of use). Algorithmic analysis shows that this strategic divergence has left its imprint on the pronouns used in courtroom rhetoric. MWR, which rates terms highly if they consistently appear more often in plaintiffs’ or defense documents, identifies he and she, as well as Mr, Mrs and Ms, as among the most distinctive terms for the defense. Such terms almost always refer to the smoker or his or her family. By contrast, we found that they (usually referring to the industry, see again table 1) is the second most distinctive term for the plaintiffs by MWR.

Cipollone.40 They concluded that to counter plaintiffs’ attacks, the best strategy would be to focus on the individual smoker:

Research has indicated that to the extent the jury focuses on issues including the safety of cigarettes, corporate misconduct, sending a message to the tobacco companies, the plaintiff’s chances are enhanced. Conversely, if the jury focuses on the individual plaintiff, his choices, his actions, his environment and history, a defense verdict is more likely. In essence, the well-prepared plaintiff tries us; we try the plaintiff.40

In virtually every trial since the 1980s, the goal of both sides has been to emphasise the agency of the opposing side, while diminishing or hiding their own agency (see table 3, where terms are ranked by frequency of use). Algorithmic analysis shows that this strategic divergence has left its imprint on the pronouns used in courtroom rhetoric. MWR, which rates terms highly if they consistently appear more often in plaintiffs’ or defense documents, identifies he and she, as well as Mr, Mrs and Ms, as among the most distinctive terms for the defense. Such terms almost always refer to the smoker or his or her family. By contrast, we found that they (usually referring to the industry, see again table 1) is the second most distinctive term for the plaintiffs by MWR.

Table 2

<table>
<thead>
<tr>
<th>Plaintiffs’ attorneys will rely rarely say:</th>
<th>Defence</th>
<th>Plaintiffs</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>had the ability to quit</td>
<td>241</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>(nothing the defendants) said or did</td>
<td>220</td>
<td>4</td>
<td>0.02</td>
</tr>
<tr>
<td>plaintiff has to prove</td>
<td>175</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>(if you, the jury) get this far (in the verdict form)</td>
<td>176</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>smoking decisions</td>
<td>137</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>her burden</td>
<td>132</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>warnings from</td>
<td>108</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>knew that smoking was dangerous</td>
<td>95</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>decision to quit</td>
<td>89</td>
<td>1</td>
<td>0.01</td>
</tr>
</tbody>
</table>

P values < 0.0001 for all FS.

This table contains expressions appearing at least 50 times more often in plaintiffs’ than in defence closings or vice versa, sorted by term frequencies. If scores for very similar expressions appeared in close proximity we selected the more expressive ones (eg, ‘had the ability to quit’ instead of ‘had the ability to’). Names of attorneys are excluded (‘Mr Bigger’, for example).

A score of 0.99 means that 99% of all instances of a given term will be in plaintiffs’ documents, with corpora normalised to have the same number of words.
The main actor in the industry’s narrative is the smoker: what he or she heard, saw or did and what his or her family members remember. This basic story has remained the same ever since the industry developed its ‘common knowledge’ defence, which holds that smokers have only themselves to blame for any harms they may have suffered.4 5 25 “If Mr. Barbose (the smoker) had to assign responsibility, would he? Nobody made Mr. Barbose's decisions for him.”41 The industry’s lawyers try to highlight what the smoker in question knew and how he or she failed to act: “He knew that smoking was dangerous…. He chose to smoke and he never tried to quit and the reason he never tried to quit was through no fault of RJR.”42 But also: “she quit when she had a heart attack, she never smoked again, never smoked again.”43 Industry attorneys want juries to believe that smoking is a free choice: when a smoker wants to quit, they can.

Tobacco’s lawyers also work hard to ignore or even to erase the very existence of the industry by using carefully crafted terms. Consider the difference between the terms smoker (FS 0.44) and customer (FS 0.94). Customer and its plural are taboo terms for the industry—plaintiffs use them almost 20 times more often than the defense—because such terms link the smoker to the industry. This same pattern holds true for killed (by someone or something, FS 0.85) versus passed away (FS 0.17), or product (of the industry, FS 0.79) versus cigarettes (FS 0.53). Defence attorneys use terms that draw attention away from the fact that the cigarettes smoked by the plaintiff were manufactured by large corporations bonded in conspiracy.25 27

Attorneys working for the plaintiffs, of course, use different rhetoric. Plaintiffs centre their case around what the industry did and what they need to be punished for: “Are we going to stand for liars and companies that treat people like that just because they’re a corporation in America, just in the name of profits? That’s going to be for you to decide.”44

### Putting the ‘common’ in ‘common knowledge’—by blaming family members

Cigarette makers claim that smokers have known about the dangers inherent in smoking since at least 1966, when caution labels were first placed on packs.4 5 25 To buttress this argument, industry lawyers have often hired historians to testify that everyone knew about such harms, thanks to a purported ‘deluge’ of publicity.4 24 27 These historians never look at cigarette ads or the industry’s own documents, focusing instead on warnings in articles that give this impression of an ‘avalanche’ of warnings, with the industry itself essentially impotent and invisible. As Louis Kyriakoudes summarises: “From the testimony of industry historians, one would never understand how it came to be that anyone ever smoked.”45

In recent Engle progeny trials, instead of historians, the industry often uses the plaintiff and his or her family to put on its common knowledge defence (table 4). Husbands and wives, brothers and sisters, mothers and fathers are all brought in to testify that the smoker must have known that smoking was dangerous: “Ladies and Gentlemen, I submit to you that the evidence demonstrates that Ellen Tate knew that smoking was dangerous to her. And she knew it when her sister Marcia knew it. She knew it when her husband Mr. Fazio knew it. She knew it when her friend Mona knew it. She knew it when the warnings went on the packages, and she saw them every day.”46 Family members are used to suggest that a smoker must have been aware of the harms caused by smoking; he or she made an informed decision when starting and continuing to smoke.

Cigarette industry lawyers learnt how important family members can be to their cause in the 1980s, in the course of running mock trials in preparation for the Cipollone case. To the surprise of these lawyers, family members turned out to be the industry’s best assets: mock jurors “rated Rose (Cipollone) and her family as plaintiffs worst witness (by far) and defendants best (by far).” (p128)48 In fact, the family’s evidence “was so powerful that nothing much could be added.” To harness the power of this testimony, it was essential to establish that the smoker had been told repeatedly, even as a child, about the dangers of smoking, and that in later life he or she had been warned by his or her spouse and children. It was through these witnesses that the industry could establish ‘awareness’. This same strategy continues today.

Care and concern by family members are weaponised by industry lawyers in a number of different ways. Warnings from relatives are used to portray the smoker as a rational actor: “His wife, his stepdaughter, his own doctors told him many times over the years, ‘Smoking is dangerous. You should quit.’ ... But William Starbuck enjoyed smoking cigarettes and had no real interest in quitting.”49 Family care and concern can also be used to paint the smoker as reckless: “He is responsible for
the decision to continue smoking and to not even try to quit in the face of begging and encouragement from his children and from his wife.47 Parents who were smokers can also be blamed for passing on the habit to their children: “Mrs. Cohen stole her first cigarette from her father. Mrs. Cohen started smoking because her friends smoked. And that is the greatest predictive factor ... if you have got a parent who smokes, you have got friends who smoke, chances are you will become a smoker, not (from) seeing cartoons on TV with ads...”

Talking about ‘free choice’ while avoiding the term ‘free choice’

One of our more surprising discoveries is that free choice is a highly distinctive term for plaintiffs, appearing 450 times in plaintiffs’ closings but only 32 times in defence closings (table 5). This is surprising, because the industry for many years has claimed that smoking is a free choice.3 7 20 In court, however, attorneys for the plaintiffs have appropriated the term. They often cite a 1980 Tobacco Institute document, which concludes that “We can’t defend continued smoking as ‘free choice’ if the smoker in question must have made an informed decision: ‘The evidence is that Mr. O’Hara knew the risks, saw the warnings on the packs, made the decision to smoke.’”31 People smoke, according to this argument, because they enjoy it: “She was warned about the dangers and addictive nature of cigarettes were common knowledge.”30

Cigarette industry lawyers try to avoid this conundrum by accepting that while smoking may be addictive in principle, it never is for any actual smoker confronting them in court: “Nicotine is addictive. Cigarettes are addictive.... What you have to decide is whether or not Mrs. Lloyd was addicted.”31 To discredit the plaintiffs, the industry deploys a Catch-22: If you managed to quit, you were never addicted. And if you were not able to quit, you were not sufficiently motivated. Big Tobacco lawyers will question whether any smoker was ever really addicted: “Mr. Barbos was not addicted. He was in control over his smoking decisions on the packs, made the decision to smoke.”31 People smoke, according to this argument, because they enjoy it: “She was warned about the dangers and addictive nature of cigarettes were common knowledge.”30

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Exculpating the industry by weaponising the smoker’s friends and family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defence attorneys will often say:</td>
<td></td>
</tr>
<tr>
<td>Defence</td>
<td>Plaintiff</td>
</tr>
<tr>
<td>relatives*</td>
<td>5342</td>
</tr>
<tr>
<td>father/mother</td>
<td>1670</td>
</tr>
<tr>
<td>husband/wife</td>
<td>1614</td>
</tr>
<tr>
<td>brother/sister</td>
<td>705</td>
</tr>
<tr>
<td>friends</td>
<td>755</td>
</tr>
<tr>
<td>common knowledge</td>
<td>72</td>
</tr>
</tbody>
</table>

P values <0.0001 for all FS and MWR score unless otherwise noted.

*Relatives include the terms wife, husband, mother, father, grandmother, grandfather, sister/s, brother/s, daughter/s, son/s, granddaughters, grandsons, uncle/s and aunt/s.

†P value <0.001.

FS, frequency score; MWR, Mann-Whitney Rho.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Talking about ‘free choice’ while avoiding the term ‘free choice’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defence attorneys will often say:</td>
<td></td>
</tr>
<tr>
<td>Defence</td>
<td>Plaintiff</td>
</tr>
<tr>
<td>quit smoking</td>
<td>3410</td>
</tr>
<tr>
<td>motivated</td>
<td>487</td>
</tr>
<tr>
<td>enjoyed smoking</td>
<td>307</td>
</tr>
<tr>
<td>had the ability to quit</td>
<td>241</td>
</tr>
<tr>
<td>did not want to quit</td>
<td>215</td>
</tr>
</tbody>
</table>

Plaintiffs’ attorneys will often say:

| free choice | 32 | 450 | 0.94 | 0.86 | “Once addicted, it is no longer an unconstrained free choice for the smoker because of the physiological and psychological need for cigarettes containing nicotine.”77 |
Ignore the truth, focus only on the facts

Another surprising rhetorical divergence is truth versus facts. The plaintiffs focus on documenting the truth about the industry and its misconduct, including its decades of casting doubt on research linking cigarettes and disease. The defendants do not so much deny this history as simply ignore it, while insisting that the jury focus on the facts of the specific case, which shifts all agency away from the industry. This strategy, too, can be found in the industry’s 1980s litigation training manuals: “Our great strength is the particular plaintiff—her specific disease and her personal option to quit. Our potential weakness is the ‘universal’ level: general causation, failure to warn the public, and alleged deception: advertising, industry research, lobbying.”

Focusing on individual facts isolates a particular trial from its larger context and allows the industry to appeal calm and rational, presenting ‘just the facts’ while claiming that plaintiffs want to arouse emotions. Defence attorneys sometimes even suggest that plaintiffs are using scare tactics (table 6): “I submit you heard those things (from the plaintiffs) to make you angry, to get you mad, to distract you from the facts of Mr. Ahrens’ case.... Dr. Proctor talked to you for 4 days in the hopes of getting you so mad that you would decide the case based on emotion... If you keep your eye on the ball, and focus on Mr. Ahrens and the facts of his life, Mrs. Ahrens loses.”

By contrast, truth—about the tobacco industry and its conduct—is a crucial weapon in the plaintiffs’ rhetorical armamentarium, along with doubt and conspiracy. Their argument is that for decades, the tobacco industry has been engaged in a conspiracy to create doubt and hide the truth: “So when (the smoker) started way back in that other time and place, they were years into conspiracy, hiding addiction, hiding the truth (while offering) false, fraudulent, safe-seeming fixes.”

The plaintiffs’ lawyers claim they do not need to show how the industry’s disinformation campaign affected their client in particular: “You do not have to put anything particular directly in the man’s hands, because there was nothing personal about the public doubt campaign, there was nothing personal about the Traveling Truth Squad, the College of Tobacco Knowledge, Anne Browder, and all these different folks that went around the country saying that it is an open controversy... meant for the population in whole, of which Mr. Banks was a part.”

This radical divergence between general ‘truths’ and specific ‘facts’ shows that it is wrong to think of a denialist enterprise like Big Tobacco having some kind of antipathy towards facts. The broader truth is that compartmentalised facts, suitably framed, can be harnessed to serve a vital purpose in the industry’s efforts to thwart justice. Cigarette makers are very good at creating macromyths out of microfacts, which also helps explain why they have such an aversion to any mention of the ‘truth’, both as a concept and a linguistic expression.

LIMITATIONS

Our results reflect the demands of Engle progeny cases: Engle plaintiffs are required to show that addiction was a ‘legal cause’ of the smoker’s illness or death, for example, which requires showing that he or she was addicted. This means that disputes over addiction and related concepts will feature prominently in all such trials, even if they are of lesser import in other tobacco trials. Our algorithms are best at capturing patterns that remain consistent from trial to trial. For example, the industry will consistently blame non-tobacco causes for a smoker’s disease, but those causes will vary from trial to trial. If a smoker died from oesophageal cancer, defence attorneys will postulate non-tobacco causes like alcohol or the human papillomavirus: “The plaintiff must prove in this case that smoking was the cause of oesophageal cancer, I suggest you that it was alcohol. That’s the evidence.” Our algorithms do not capture these disputations of specific causation because they are usually tailored to the individual case and do not produce consistent patterns we could identify.

DISCUSSION

The closing arguments analysed in this study help us better understand the world that cigarette industry lawyers want jurors to imagine. In this alternate reality, all agency lies with the individual consumer and none with the producer. And every smoker has always been fully informed, weighing costs and benefits before taking up or continuing to smoke. In this alternate universe conjured by the defence, the specific facts of an individual case weigh heavier than larger truths about the industry’s deception or...
the pharmacological grip of nicotine. And inconspicuous black-and-white warnings on the sides of packs deliver a bigger punch than colourful and ubiquitous ads for cigarettes.

Put in these blunt terms, the alternate reality created by the industry is grotesque and myopic. Industry lawyers have known since the 1980s that they need to blame without appearing to blame, to say without saying. ‘Free choice’ is a good example. Cigarette makers claim that smokers weigh the risks and benefits of smoking and then make an informed decision to smoke. Advancing this argument, however, requires carefully avoiding the term ‘free choice’ itself—since cigarette makers have always had choices that are clearly freer than those of an addicted smoker. Instead, the industry’s lawyers focus on what he or she (the smoker) knew, wanted, did or did not do; they renarrate the smoker’s life—and the industry’s lawyers focus on what he or she (the smoker) knew, wanted, did or did not do; they renarrate the smoker’s life—and only the smoker’s life—as filled with conscious decisions, creating the appearance of a life replete with choices.

Tobacco control scholars are familiar with myriad forms of deceptive language used by cigarette makers: code words like ‘Zephyr’ for cancer or ‘Borstar’ for benzpyrene or ‘Compound W’ for nicotine; we know about the industry’s many euphemisms: ‘smoking and health’ for cigarettes and death, ‘young adult’ for teens, or ‘tar and nicotine’ for cancer and addiction. Computational methods open up new ways to explore the industry’s strategic rhetoric, helping us understand how even seemingly trivial lexical items like pronouns or references to family members have become legal weapons. The methods developed here could be fruitfully applied to other courtroom documents, including expert witness testimony. By comparing the language of opposing experts, we may be able to reverse engineer whatever lawyerly coaching may have occurred. And by comparing rhetorical patterns against trial outcomes, it might even be possible to discover winning strategies previously undiscovered.

Acknowledgements We would like to thank Crystal Lee for helpful comments, Felicia Schuessler for help with data entry and cleaning, Rachel Taketa for helping us identify closing arguments in newly added documents and Ruth Malone for her editorial advice.

Contributors Both authors contributed by conceptualising the project and writing and revising the report. Stephan Risi wrote the code and accompanying website for the project.

Funding This work was supported by the State of California’s Tobacco-Related Disease Research Program (TRDRP) high impact pilot award “Fighting Big Tobacco with Big Data,” award number 25IP-0017.

Competing interests RNP has served as an expert witness for plaintiffs in cigarette litigation.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

9 Malone RE. Word wars and tobacco control: saying what needs saying that we don’t yet know how to say, or saying it better. Tobacco Control 2010;19:261–2.


