JIPEL

NYU Journal of Intellectual Property
& Entertainment Law

Statement of Purpose

Consistent with its unique development, the New York University Journal of Intellectual Property & Entertainment Law (JIPEL) is a nonpartisan periodical specializing in the analysis of timely and cutting-edge topics in the world of intellectual property and entertainment law. As NYU's first online-only journal, JIPEL also provides an opportunity for discourse through comments from all of its readers. There are no subscriptions or subscription fees; in keeping with the open-access and free discourse goals of the students responsible for JIPEL's existence, the content is available for free to anyone interested in intellectual property and entertainment law.

The New York University Journal of Intellectual Property & Entertainment Law is published up to three times per year at the New York University School of Law, 139 Mac-Dougal Street, New York, New York, 10012. In keeping with the Journal's open access and free discourse goals, subscriptions are free of charge and can be accessed via www.jipel.law.nyu.edu. Inquiries may be made via telephone (212-998-6101) or e-mail (submissions.jipel@gmail.com).

The Journal invites authors to submit pieces for publication consideration. Footnotes and citations should follow the rules set forth in the latest edition of The Bluebook: A Uniform System of Citation. All pieces submitted become the property of the Journal. We review submissions through Scholastica (scholasticahq.com) and through e-mail (submissions.jipel@gmail.com).

The Journal is also available on WESTLAW, LEXIS-NEXIS and HeinOnline.

CATEGORIZATION OF WORD-MARK DISTINCTIVENESS: POOR REASONING IN UNITED STATES FEDERAL COURTS

THOMAS Y. LU*

United States federal judges have long recognized the importance of word mark distinctiveness in resolving disputes over a word mark's validity. Nonetheless, questions remain as to how rigorously the judges categorize distinctiveness in these cases. To examine this matter, I collected 713 United States federal word mark cases dating from January 1, 2002 to December 31, 2022 and hand-coded related data from external sources such as dictionaries and the United States Patent and Trademark Office (USPTO) trademark search system. I then trained three decision trees with varying subsamples of 19 independent variables in order to interpret the descriptive data above. The results of the decision-tree analyses suggest that federal judges were indeed reasoning poorly when adjudicating word-mark cases. This poor reasoning took four often overlapping forms: (1) a narrow focus on differentiating suggestive distinctiveness from descriptive distinctiveness, (2) an excessive reliance on the linguistic traits of a word mark, such as the dictionary test, as well as on the imagination test and competitor-need test to the exclusion of other critical tests, (3) a fundamental misunderstanding of inherent distinctiveness versus acquired distinctiveness, and (4) inadequate consideration of consumer perception. To counter this poor reasoning, the USPTO should establish practical, clearly defined rules and guidelines with which federal judges can more rigorously assess (1) all categories of distinctiveness, (2) non-linguistic evidence (particularly consumer perception), and (3) inherent distinctiveness.

*Associate Professor in the Department of Business Management, National Sun Yat-sen University; Visiting Fellow, Research Center for Humanities and Social Science, Academia Sinica. (JSD, 2019, from Washington University in St. Louis, School of Law).

I

INTRODUCTION

My central aim in this study is twofold: first, explore the possibility that US federal judges engage in poor reasoning when they are tasked with clarifying word-mark distinctiveness; and second, propose clear, practical solutions where they are

needed. Distinctiveness of a word mark is important in trademark law because this trait helps determine whether a word or other signals, such as color, function as a trademark and therefore merit trademark protection.[1] In 1976, the Second Circuit famously reasoned a standard for determining the level of distinctiveness of a word mark in *Abercrombie & Fitch Co. v. Hunting World, Inc.*[2] In *Abercrombie*, the Second Circuit declared that there are four categories of word marks: (1) fanciful and arbitrary marks (which we group together as one category, but are sometimes discussed separately), (2) suggestive marks, (3) descriptive marks, and (4) generic marks.[3] These four types of marks are listed from most distinctive to least distinctive: the arbitrary mark possesses the highest level of distinctiveness while the generic mark has no distinctiveness and thus merits no protection under trademark law.[4] The opinion further reasoned that arbitrary and suggestive marks possess inherent distinctiveness, which sometimes is presumed.[5]

But what about descriptive marks? Descriptive marks cannot be protected as a trademark unless evidence demonstrates that they possess distinctiveness in the form of secondary meaning.[6] We refer to descriptive marks with secondary meaning as "descriptive-acquired marks," and descriptive marks without secondary meaning, and which are therefore unprotectable, as "purely descriptive marks." For example, the owner may have to present proof of ownership

---

[1] 15 U.S.C. § 1052(f) ("[N]othing in this chapter shall prevent the registration of a mark used by the applicant which has become distinctive of the applicant's goods in commerce."); *see also* Wal-Mart Stores, Inc. v. Samara Bros., 529 U.S. 205, 207 (2000) (considering whether a product's design "is distinctive, and therefore protectible"). Distinctiveness refers to how quickly and clearly the mark identifies the source of the good or service. *See Strong Trademarks*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/trademarks/basics/strong-trademarks [https://perma.cc/Z5AW-YJF7] (last visited Feb. 28, 2025).

[2] Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976).

[3] *Id.*

[4] *Id.* at 9, 14.

[5] *Id.* at 11. *See also* Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 768 (1992) (describing arbitrary and suggestive marks as "inherently distinctive"). The United States Patent and Trademark Office (USPTO) and, more specifically, the Trademark Trial and Appeal Board (TTAB) tend to focus on inherent distinctiveness in determining whether a mark that lacks secondary meaning may be registered on the Principal Register. Federal courts typically consider inherent distinctiveness in the context of infringement litigation, where proof of distinctiveness can have a substantial effect on the outcome of a "likelihood of confusion" analysis. *See* Edward J. Heath & John M. Tanski, *Drawing the Line Between Descriptive and Suggestive Trademarks*, 12 COM. & BUS. LIT. 11, 13 (2010).

[6] *Abercrombie*, 537 F.2d at 9 (citing 15 U.S.C. § 1052(f)); *Two Pesos*, 505 U.S. at 769 ("Marks which are merely descriptive of a product are not inherently distinctive … However, descriptive marks may acquire the distinctiveness which will allow them to be protected under the [Lanham] Act.").

of prior registrations of the said mark.[7] Furthermore, the trademark owner may have to prove that there has been substantially exclusive and continuous use of the mark in commerce for at least five years.[8] Other evidence demonstrating the existence of secondary meaning might pertain to sales, length of time used, unsolicited media coverage, advertising expenditures, the results of consumer surveys, or even the declarations or affidavits of consumers.[9] A finding that third-party use of the mark was widespread during those minimum five years of use would undermine a claim for secondary meaning.[10]

A.   Tests Used to Determine Distinctiveness

Because the arbitrary, suggestive, and descriptive categories each involve unique levels of distinctiveness, it is understandable that federal courts would develop tests to clarify a word mark's distinctiveness. This is particularly true for the differentiation of descriptive-acquired and purely-descriptive marks, since the protectability of the mark turns on its level of distinctiveness.[11] For example, federal courts have developed several tests for differentiating between suggestive and descriptive marks. The most common of these tests are the imagination test, the competitor-need test, and the dictionary test.[12]

The imagination test serves to "measure the relationship between the actual words of a mark and the product to which they are applied."[13] If a mark requires that an observer employ a measurable degree of perception, imagination, or other thought to identify the nature of the product, the mark is considered suggestive.[14] Alternatively, if standing alone, a term conveys information about the

---

[7] *See* TMEP §§ 1212.04–04(e).

[8] 15 U.S.C. §§ 1052(f), 1065, 1127; *Abercrombie*, 537 F.2d at 10; TMEP §§ 1212.05–1212.05(e).

[9] *Abercrombie*, 537 F.2d at 7; TMEP §§ 1212.06–1212.06(e)(iv).

[10] The issue of widespread third-party use applies only to marks categorized as descriptive-acquired or purely descriptive, not to those classified as suggestive or arbitrary, though there might be considerable reason why a third party would use such marks to describe goods or services, particularly given the odd categorization schemes employed by various courts. *See* Joseph Scott Miller, *Abercrombie 2.0—Can We Get There from Here? Thoughts on "Suggestive Fair Use"*, 77 Ohio St. L.J. Furthermore 1, 9–14 (2016).

[11] 15 U.S.C. § 1052(f); *Abercrombie*, 537 F.2d at 10.

[12] *See* Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 792–93 (5th Cir. 1983), *abrogated by* KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111 (2004) (outlining the three tests for identifying and differentiating the distinctiveness of marks).

[13] *Id.* at 792.

[14] *Id.*

characteristics of the goods or service, then it is descriptive.[15] Some federal courts and legal scholars have argued that the imagination test specifically investigates whether or not a mark metaphorically connotes a trait inhering in or associated with the product in question.[16] Other federal courts have even used the imagination test as the "primary criterion" in determining whether a given trademark is suggestive or descriptive.[17]

The competitive-need test measures the extent to which competitors would require a term for the purpose of describing their own product.[18] In other words, is a mark so generally descriptive of a product widely offered by multiple businesses that they would reasonably depend on or at least have natural recourse to the mark when describing their respective products? If the answer is yes, the mark

---

[15] *Id.*

[16] *See, e.g.*, Synergistic Int'l, Inc. v. Windshield Doctor, Inc., No. CV 03-579 FMC (CWx), 2003 U.S. Dist. LEXIS 12660, at *14 (C.D. Cal. Apr. 28, 2003) (finding the mark GLASS DOCTOR for glass installation and repair services to be suggestive given the "creative metaphorical combination of the terms 'Doctor' and 'Glass'"); BigStar Ent., Inc. v. Next Big Star, Inc., 105 F. Supp. 2d 185, 196 (S.D.N.Y. 2000) ("When choosing what to call the article, the creator of the suggestive name meaningfully fixes upon associational terms that will identify the product figuratively and will appeal to the consumer by allusion and metaphor."); Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. REV. 621, 671 (2004) ("Suggestive marks, such as ATLAS for moving services or ROACH MOTEL for insect traps, are textbook metaphors and are described as such by the doctrine."); Laura A. Heymann, *A Name I Call Myself: Creativity and Naming*, 2 U.C. IRVINE L. REV. 585, 603 (2012) ("[T]he inherent strength of a mark (and therefore whether it gets protection ab initio or requires additional evidence) depends on how creative the mark is. The mark might be a commonplace and dull description of the good's qualities or characteristics (and therefore might need to be used by others), or use metaphor to suggest a good's characteristics, or create a new meaning for an existing word."); Laura A. Heymann, *The Grammar of Trademarks*, 14 LEWIS & CLARK L. REV. 1313, 1330–31 (2010) ("[T]he concept of metaphor is fundamental to how most trademarks work. Except for words invented to serve as trademarks—such as 'Kodak' and 'Xerox'—all trademarks, being words in the English language, operate on a level other than a literal one in that they require consumers to use a familiar word or expression in a new and initially unfamiliar context."); Jake Linford, *The False Dichotomy Between Suggestive and Descriptive Marks*, 76 OHIO ST. L.J. 1367, 1372 n.29 (2015) ("Suggestive marks are … metaphorically related to the good or service sold, like using GLEEM to sell toothpaste indirectly invokes the bright, shiny quality one could expect from thoroughly cleaned teeth."); *cf.* Alexandra J. Roberts, *How To Do Things with Word Marks: A Speech Act Theory of Distinctiveness*, 65 ALA. L. REV. 1035, 1048 (2014) (arguing that "fact finders often focus unduly on mark selection, fixing on the employment of double entendre, incongruity, rhyme, metaphor, alliteration, or other rhetorical device as evidence that a mark is distinctive").

[17] Zobmondo Ent., LLC v. Falls Media, LLC, 602 F.3d 1108, 1116 (9th Cir. 2010) (citing Self-Realization Fellowship Church v. Ananda Church of Self-Realization, 59 F.3d 902, 903 (9th Cir. 1995)).

[18] Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 793 (5th Cir. 1983) (citing Union Carbide Corp. v. Ever-Ready, Inc., 531 F.2d 366, 379 (7th Cir. 1976)).

is descriptive.[19] By contrast, a suggestive mark possesses little semantic content that could be deemed necessary for descriptions of a type of product peddled by competitors.

Finally, as its name implies, the dictionary test is rooted in the tendency of some federal courts to start their analysis of a contested mark by consulting a dictionary. After all, dictionaries are the go-to source for people seeking to understand the everyday meaning of words, which is key to determining whether a mark is descriptive.[20] As I will show, the dictionary test and its emphasis on the linguistic traits of disputed word marks have developed into a crutch that federal judges have relied on at the expense of sound reasoning in word-mark cases.

1. *The Imagination Test*

Examples of these three tests can be found in several federal cases. In *Jackpocket v. Lottomatrix*, two companies, which operated online gaming services, were at odds with each other regarding whether Jackpocket's JACKPOCKET trademark was being infringed upon by Lottomatrix Operations, which owned the domain name JACKPOT.COM.[21] In assessing the JACKPOCKET mark, the court specifically explained that "the difference between descriptive and suggestive marks lies in the immediacy of association—how quickly and easily consumers grasp the nature of the product from the information conveyed."[22] Next, the court categorized the JACKPOCKET mark as a suggestive one because "the JACKPOCKET Marks do not immediately describe Plaintiff's products."[23] The judge went on to explain,

> Formed by the juxtaposition of 'jackpot' and 'pocket', 'Jackpocket'
> suggests the nature of Plaintiff's product, the ability to play the lottery
> (and win a jackpot) from one's phone (or pocket). … By virtue of the
> addition of the 'cke' and the connotation of a pocket, it takes some

---

[19] *Id.*

[20] *See, e.g.*, Vision Center v. Opticks, Inc., 596 F.2d 111, 116 (5th Cir. 1980) (using the dictionary definition of the word "center" to support their finding that the term "Vision Center" is descriptive); Am. Heritage Life Ins. Co. v. Heritage Life Ins. Co., 494 F.2d 3, 11 (5th Cir. 1974) (noting that the district court used the dictionary definition of the word "heritage" in support of finding that the term is descriptive).

[21] Jackpocket, Inc. v. Lottomatrix NY LLC, 645 F. Supp. 3d 185, 200–01, 203, 213 (S.D.N.Y. 2022), *aff'd*, No. 23-12-CV, 2024 WL 1152520 (2d Cir. Mar. 18, 2024).

[22] *Id.* at 239.

[23] *Id.* at 240.

"imagination, thought and perception to reach a conclusion as to the nature of" Jackpocket's product.[24]

In his reasoning above, the court seems to have applied the imagination test insofar as he analyzed the "structure" of the JACKPOCKET mark. Nonetheless, if we assume that the imagination test requires federal judges to analyze the "structure" of a mark, future trademark applicants who own some variation of a mark beginning with the letters JACKPO may find it immensely difficult to predict whether their mark is descriptive or suggestive. The imagination test essentially has one basic function: to help trademark applicants and federal judges *roughly differentiate* between descriptive and suggestive marks. The imagination test, however, cannot *precisely differentiate* between any of the five categories of marks. That is, even when armed with the imagination test, trademark applicants and federal judges alike can only take wild "guesses" and can do so only on a case-by-case basis. Thus, I argue that, in all likelihood, the imagination test is of limited use for both trademark applicants and federal judges struggling to determine the distinctiveness of a word mark.

2. The Competitor-Need Test

In *Zobmondo Entertainment v. Falls Media*, the Court applied the competitor-need test, as well as the imagination test, to decide whether the descriptive mark WOULD YOU RATHER … ? had acquired secondary meaning, or if it was purely descriptive.[25] With respect to the competitor-need test, the Ninth Circuit first explained that:

> If competitors have a great need to use a mark, the mark is probably descriptive; on the other hand, if "the suggestion made by the mark is so remote and subtle that it is really not likely to be needed by competitive sellers to describe their goods or services[,] this tends to indicate that the mark is merely suggestive."[26]

Interestingly, the Ninth Circuit "indirectly" cited the district court's opinion related to the case and, partly on the basis of that opinion, concluded that the competitor-

---

[24] *Id.*

[25] Zobmondo Ent., LLC v. Falls Media, LLC, 602 F.3d 1108, 1116–17 (9th Cir. 2010).

[26] *Id.* at 1117 (alteration in original) (quoting Rodeo Collection, Ltd. v. W. Seventh, 812 F.2d 1215, 1218 (9th Cir. 1987)).

need test would not be sufficiently helpful in drawing any final conclusions on the case:

> The district court concluded that the competitors' needs test was "difficult to apply in this case" and declined to consider it because these tests "are merely factors to consider" and other tests favored Zobmondo. Falls Media argues that this was error, and in this case we agree.[27]

In drawing this conclusion about the competitor-need test, the Ninth Circuit seems to have been in agreement with the district court's opinion about the "difficulty" of applying the competitor-need test to analyses of word-mark distinctiveness. Still, the competitor-need test has proven to be an attractive tool used by federal judges in their efforts to analyze distinctiveness.[28]

3.   The Dictionary Test

The dictionary test was applied in *TotalCare Healthcare Services v. TotalMD*.[29] At issue was a conflict between TotalCare Healthcare Services, which had long used the unregistered word mark TOTALCARE, and TotalMD, which had subsequently opened a business called TotalCare Urgent Care.[30] Owing to the companies' identical use of TOTALCARE in different lexical contexts, TotalCare Healthcare Services sought an injunction barring TotalMD and other entities from using the term.[31] Because the issuance of a trademark injunction for a word mark requires that the plaintiff establish a compelling need for special legal protections, TotalCare Healthcare Services needed to prove that the distinctiveness of TOTALCARE was so substantial as to merit such protections.[32] In attempting to do so, the plaintiff contended that the TOTALCARE mark was suggestive, and, somewhat predictably, the defendant argued that the mark was purely descriptive.[33] Considering the two sets of arguments, the court concluded that TOTALCARE

---

[27] *Id.* at 1117.

[28] *See, e.g.*, Firefly Digit. Inc. v. Google Inc., 817 F. Supp. 2d 846, 861–62 (W.D. La. 2011) (finding that the mark WEBSITE GADGET is purely descriptive in part because the district court deemed the component terms virtually indispensable to the vocabulary of the website industry).

[29] TotalCare Healthcare Servs. v. TotalMD, LLC, 643 F. Supp. 3d 636, 643 (N.D. Tex. 2022).

[30] *Id.* at 640–41.

[31] *Id.* at 640.

[32] *Id.* at 641–42.

[33] *Id.* at 642.

was suggestive, not descriptive, and grounded this conclusion in the following justification:

> TOTALCARE does not describe any product, business, industry, or characteristic. Though it may evoke a nebulous quality of service, it is not a word that has a dictionary definition like 'speedy', 'reliable', 'green', or 'menthol'. This mark is different from marks like 'Urgent Care', 'Vision Center,' or 'Bank of Texas' in that what it describes is left up to the imagination and not plain on its face.[34]

In formulating this conclusion, the court seems to have combined the imagination test with the dictionary test, using definitions that would apparently prove whether or not a particular "imagination" would take hold in a person's mind. Specifically, the court compared several dictionary terms analogously with TOTALCARE. The comparison, however, failed to clarify how one can rigorously identify what category a word mark falls under and what level of distinctiveness the word mark possesses. Specific defining characteristics of arbitrary, suggestive, descriptive-acquired, purely descriptive, and generic word marks would remain unknown. Therefore, similar to the cases cited above in reference to the imagination test and the competitor-need test, the case of TOTALCARE teaches us that the dictionary test, though useful, cannot satisfactorily predict categories and levels of distinctiveness in all circumstances.

Yet another case can shed further light on this matter. In *UMG Recordings v. OpenDeal*, the court needed to analyze the distinctiveness of UMG's registered trademark REPUBLIC RECORDS.[35] Ultimately, the judge decided that two-word phrase was an arbitrary mark:

> Here, it is undisputed that UMG owns a valid trademark registration in the 'Republic Records' mark and its stylized flag logo, which cover various music-related goods and services. And, the word 'Republic' has a specific, well-known meaning, but it has no intrinsic relationship to records or music-related goods or services. The 'Republic Records' mark

---

[34] *Id.* at 644.

[35] UMG Recordings, Inc. v. OpenDeal Inc., No. 21 CIV. 9358 (AT), 2022 WL 2441045, at *3 (S.D.N.Y. July 5, 2022).

is, therefore, entitled to protection both as an arbitrary mark, and based on its valid registration.[36]

By exploring the meaning of 'republic', the court seems to have been relying on the dictionary test to analyze the arbitrary distinctiveness of the mark, as well as, though perhaps to a lesser extent, on the imagination test, insofar as the analysis of a possible intrinsic relationship between the word 'republic' and the registered good or service. Nonetheless, the conclusion rested on problematic reasoning and on unconvincing evidence because she overlooked to consider the concept of consumer perception into the analysis of the distinctiveness.

These cases may suggest that federal judges, when analyzing word mark distinctiveness, frequently rely on either one or a combination of the three abovementioned tests—the imagination test, competitor-need test, and dictionary test. However, the apparent prevalence of these three tests in federal courts suggests that judges, rather than establishing clear standards determining categories and levels of distinctiveness, focus chiefly—and perhaps quite arbitrarily—on distinctions between suggestive and descriptive marks. This rests on the premise that when assessing distinctiveness, federal judges may be engaging in poor reasoning—a topic that has seldom been explored in the literature.[37]

To explore the degree to which federal judges have engaged in poor reasoning when adjudicating word-mark disputes, I have divided the remainder of this study into four parts followed by the conclusion. Part II addresses the research on word-mark distinctiveness. Part III addresses the methods, variables, and data for this study's descriptive and decision-tree analyses. Part IV presents the results of the analyses. My focus will be on the three decision trees' various periods and outputs and on the importance of independent variables for accurate categorization. Most importantly, I will discuss how my comparative analysis of the decision trees enabled me to uncover two critical patterns: first, linguistic attributes consistently played key roles in federal judges' categorization of word-mark distinctiveness;

---

[36] *Id.* at *4 (citation omitted).

[37] *See* Linford, *supra* note 16, at 1409 (finding that trademark law exaggeratedly differentiates between suggestive and descriptive marks); Christopher Buccafusco, Jonathan S. Masur & Mark P. McKenna, *Competition and Congestion in Trademark Law*, 102 Tex. L. Rev. 437, 494 (2024) (arguing that, although boundary problems are an inescapable facet of all categorization methods, courts cannot rigorously make the factual distinctions necessary for the legal distinctions in trademark law).

second, only the decision-tree results for suggestive distinctiveness and descriptive-acquired distinctiveness had relatively high correspondence rates with the actual case data. With these results in hand, I discuss in Part V their roots and implications. Also in Part V, I discuss practical, executable mechanisms by which we can substantially diminish federal judges' tendency to resort to poor reasoning in trademark-distinctiveness cases. In Part VI, the conclusion, I summarize the study's findings and discuss its limitations.

II
LITERATURE REVIEW

The research about trademark distinctiveness has been widely discussed from the doctrinal perspective. However, only a few scholars have used empirical methods to identify and flesh out the characteristics of each category of distinctiveness. More importantly, to my knowledge, no study has explored the possibility that federal judges engage in poor reasoning when they adjudicate cases involving the distinctiveness of disputed word marks.

In his early research about distinctiveness, Graeme Dinwoodie discussed how a product's geometrically designed shape can take on the role of trademark status, an issue that came to the fore in the Supreme Court decision in *Two Pesos v. Taco Cabana* in 1992.[38] Furthermore, Dinwoodie suggested that the *Abercrombie* taxonomy may be helpful in determining word mark distinctiveness but not helpful in determining a geometric product's distinctiveness.[39] Thus, Dinwoodie reformulated the concept of distinctiveness by developing a concept referred to as "predictive inquiry": its purpose is to help researchers investigate, among other things, the scope of protections available for a product's trade dress (i.e., the product's appearance).[40] Although Dinwoodie's research in the mid-1990s filled an important research gap regarding the distinctiveness of non-linguistic and non-pictorial marks, his presentation of the *Abercrombie* taxonomy was purely

---

[38] Graeme B. Dinwoodie, *Reconceptualizing the Inherent Distinctiveness of Product Design Trade Dress*, 75 N.C. L. REV. 471, 581 (1997) (exploring the possibility of a separate category of trademark distinctiveness called "service dress" for relatively intangible services as opposed to physical products).

[39] *Id.* at 475 (suggesting how courts might expand the concept of distinctiveness so that it accounts for spatial products as well as for linguistic or pictorial marks).

[40] *Id.* at 515 (arguing that inherent-distinctiveness analyses are predictive inquiries insofar as they involve speculation about future events).

introductory, so that the issue of word mark distinctiveness remained a markedly confused and confusing issue.

To tackle these ongoing issues besetting word-mark distinctiveness, some scholars have turned to non-legal theory. In the early 2000s, Barton Beebe used the theory of semiotic sensibility to analytically reconceptualize trademark distinctiveness into two forms: source distinctiveness and differential distinctiveness.[41] Source distinctiveness is the extent to which a trademarked symbol is somehow a literal representation of the thing being offered.[42] Differential distinctiveness refers to the differences between a trademarked symbol and other symbols constituting a trademark network.[43] These two forms of distinctiveness (the former referring to the semiotic concept of signification, the latter to the semiotic concept of value) have noticeably distinct functions. In the context of U.S. federal courts, source distinctiveness encourages them to decide whether a particular subject matter merits anti-infringement protection, whereas differential distinctiveness encourages them to investigate the proper scope of anti-infringement protection that should be accorded to a subject matter deserving of protection.[44]

Though passionate about both source distinctiveness and differential distinctiveness, Beebe singled out the latter and links it to a pair of consumer-oriented concepts: consumers' search sophistication (i.e., their ability to distinguish between similar trademarks) and consumers' persuasion sophistication (i.e., their ability to resist commercial inducements).[45] Beebe argued that differential distinctiveness may sometimes form a negative relationship with consumers' search sophistication and a positive relationship with consumers' persuasion

---

[41] *See* Beebe, *supra* note 16, at 625 (noting that there is more confusion than clarity in conventional conceptions of inherent and acquired distinctiveness).

[42] *Id.*

[43] *See id.* ("Corresponding to the semiotic relation of value, differential distinctiveness describes the extent to which a trademark's signifier is distinctive from other signifiers in the trademark system.").

[44] *Id.* at 676 ("While trademark infringement involves the infringement of source distinctiveness, trademark dilution involves the dilution of differential distinctiveness.").

[45] Barton Beebe, *Search and Persuasion in Trademark Law*, 103 Mich. L. Rev. 2020, 2039 (2005) ("Populations with a relatively low degree of search sophistication require the ceding of a relatively broad scope of protection to plaintiff's trademark.").

sophistication, with each of the two relationships ultimately taking the shape of a bell curve.[46]

Jake Linford similarly analyzed distinctiveness from the perspective of a non-legal theory: the theory of semantic shift.[47] Semantic shift, as explained by Linford, is a process whereby a generic term acquires enough source significance to become a trademark.[48] In this process, trademark owners ("speakers") successfully alter the meaning of terms so that the public ("listeners") develop an altered perception of the terms.[49] Linford argued that, to determine whether or not semantic shift has occurred, we must consider two factors: consumer perception and search costs.[50] These two factors, Linford noted, have been neglected by U.S. federal courts tasked with applying to trademark-confusion cases the doctrine of trademark incapacity (i.e., the view that a term, despite having undergone semantic shift, should not qualify as a trademark).[51] Therefore, he claimed that, to counter this neglect, federal courts should adopt and refine the primary-significance test, which is a measure of a once-generic mark's distinctiveness—that is, the extent to which the mark has come to be associated with a product or service.[52]

Linford's subsequent research concerns a specific extreme of distinctiveness: using the theory of linguistic arbitrariness and sound symbolism, he explored fanciful marks: they are marks that have no apparent significance outside their

---

[46] *Id.* at 2049 ("[Trademark] law has operated according to the assumption that, as in search sophistication, the distribution of persuasion sophistication across the general consumer population forms a bell curve.").

[47] *See* Jake Linford, *A Linguistic Justification for Protecting "Generic" Trademarks*, 17 Yale J.L. & Tech 110, 140 (2015).

[48] *Id.* at 112 ("The study of semantic shift in historical and cognitive semantic literatures is the study of how a given word changes over time—first by entering the public lexicon, and then by gaining or losing meanings."). *See also* Stephen Ullmann, Semantics: An Introduction to the Science of Meaning 209–10 (Barnes & Noble, 1979) ("Whenever a new name is required to denote a new object or idea, we can do one of three things: form a new word from existing elements; borrow a term from a foreign language or some other source; lastly, alter the meaning of an old word.").

[49] Linford, *supra* note 47, at 131 ("Semantic shift is motivated by the speaker's need to say new things and communicate more effectively, which encourages the speaker to 'risk' a semantic innovation.").

[50] *Id.* at 144–45 ("Consumers who would not be confused by the competition may pay more for the products they desire because trademark protection can increase costs for competitors, but consumers who have adopted the narrowed meaning will have lower search costs to find the products they desire.").

[51] *Id.* at 170 ("Understanding that the formation of trademark meaning is a form of semantic shift reminds us that sound competition policy cannot neglect the importance of consumer comprehension.").

[52] *Id.* ("The law should instead adopt a primary significance test for determining whether a mark that was once generic has acquired sufficient distinctiveness to merit trademark protection").

function as a trademark (e.g., Exxon, Pepsi).[53] Trademark law treats these marks as inherently and strongly distinctive.[54] Next, Linford introduced two key concepts: linguistic arbitrariness (i.e., the view that no inherent relationship exists between a signifier and the signified)[55] and sound symbolism (i.e., an inherent relationship between the sound of a signifier and the signified).[56] While acknowledging the conventional view that "a fanciful mark will be meaningless until meaning begins to collectively coalesce around the word,"[57] Linford explained that more and more research in linguistics and psychology has detected significant symbolically semantic links between the forms of words (e.g., sounds) and the meanings of the words.[58] Thus, quite simply put, meaning is not always fully independent of word form, and "the sounds of words can convey meaning apart from [the words'] actual definitions."[59] Linford thus concluded that sound symbolism should play a greater

---

[53] *See* Jake Linford, *Are Trademarks Ever Fanciful?*, 105 Geo. L.J. 731, 739–40 (2017). Our paper classifies fanciful marks together with arbitrary rather than its own category.

[54] *Id.* at 742 ("First, the fanciful mark has no inherent lexical meaning when the mark owner first coins it. Because the fanciful mark is an empty vessel, courts see the fanciful mark as automatically source-signifying when used as a mark. Second, that the mark is coined suggests to courts that the mark owner is entitled to the fruits of his or her creativity or at least a presumption that the mark was adopted in good faith. Third, because a fanciful mark has no meaning prior to its conception and use, competitive concerns that animate limits on the protection of descriptive marks or functional trade dress are seen as immaterial or at least less relevant. Fourth, courts treat fanciful marks as inherently distinctive because they are categorically distinguishable from descriptive marks.").

[55] *Id.* at 749.

[56] *Id.* at 750.

[57] *Id.* at 749. *See also* Sam J. Maglio et al., *Vowel Sounds in Words Affect Mental Construal and Shift Preferences for Targets*, 143 J. Experimental Psych. 1082, 1083 (2014) ("Taken together, sound symbolic research to date has documented robust and automatic associations between vowel sounds contained in words and the physical properties of their referents.").

[58] Linford, *supra* note 53, at 749.

[59] *See, e.g.*, Barry Alpher, *Yir-Yoront Ideophones*, *in* Sound Symbolism 161 (Leanne Hinton, Johanna Nichols & John J. Ohala eds., 1995) (reporting evidence of sound symbolism in the Australian language of Yir-Yoront); Brian D. Joseph, *Modern Greek Ts: Beyond Sound Symbolism*, *in* Sound Symbolism 222 (reporting evidence of sound symbolism in modern Greek); Terrence Kaufman, *Symbolism and Change in the Sound System of Huastec*, *in* Sound Symbolism 63 (reporting evidence of sound symbolism in the Mayan language of Huastec); *see also* Russell Ultan, *Size-Sound Symbolism*, 2 Universals of Human Language 525 (Joseph H. Greenberg ed., 1978) (arguing that the majority of the world's languages use sound symbolism); Mark Dingemanse et al., *Arbitrariness, Iconicity, and Systematicity in Language*, 19 Trends in Cognitive Scis. 603, 603 (2015) (reporting on form-to-meaning correspondences across languages); Richard R. Klink, *Creating Brand Names with Meaning: The Use of Sound Symbolism*, 11 Marketing Letters 5, 16–17 (2000) (reporting that the sounds of imaginary brand names influence people's perception of product traits such as size, speed, weight, tactility, and gender); Edward Sapir, *A Study in Phonetic Symbolism*, 12 J.

role than linguistic arbitrariness in guiding federal courts' analysis of arbitrary marks.[60]

Linford explained how the concept of sound symbolism might bolster America's trademark-law regime.[61] First, he argued that, although the *Abercrombie* taxonomy is at times unclear, the cost of abandoning it in favor of sound symbolism would be egregious because the Supreme Court has already fully adopted much of the logic supporting the taxonomy.[62] Furthermore, arbitrary mark analyses that rest solely on sound symbolism might so facilitate the protection of arbitrary marks that competitors would end up facing needlessly high costs stemming from the need to honor these protections.[63] Thus, Linford proposed several ways in which the trademark law regime might harness the concept of sound symbolism without jettisoning the *Abercrombie* taxonomy. For example, federal courts and trademark examiners can examine whether the sounds of an arbitrary mark's syllables, vowels, consonants, and so on suggest product characteristics: the more suggestive the sounds are of the characteristics, the less inherently distinctive the mark would be and thus the less legal protection the mark would be entitled to.[64]

Alexandra J. Roberts adopted speech-act theory to establish tests for trademark distinctiveness, noting that previous research applied the theory to such areas as contract law.[65] After demonstrating that current tests of word-mark distinctiveness are untenably confusing, Roberts integrated speech-act theory into

---

EXPERIMENTAL PSYCH. 225, 228 (1929) (reporting that vowel sounds differ from one another regarding their effect on people's perception of size and that these differences might hold across languages).

[60] Linford, *supra* note 53, at 765.

[61] *Id.* at 748.

[62] *Id.* at 764. The Supreme Court noted that rules requiring "evidence of secondary meaning" can dampen competition, especially for startups and smaller firms. *See* Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 775 (1992).

[63] Linford, *supra* note 53, at 757 ("Firms gain an advantage when the mark connotes product features, because it is easier for consumers to associate the mark with those features.").

[64] *Id.* at 758.

[65] Alexandra J. Roberts, *How To Do Things with Word Marks: A Speech-Act Theory of Distinctiveness*, 65 ALA. L. REV. 1035, 1041 (2014). *See, e.g.*, Peter Meijes Tiersma, *The Language of Offer and Acceptance: Speech Acts and the Question of Intent*, 74 CAL. L. REV. 189, 189–90 (1986); Janet E. Ainsworth, *In a Different Register: The Pragmatics of Powerlessness in Police Interrogation*, 103 YALE L.J. 259, 265 (1993); Martin F. Hansen, *Fact, Opinion, & Consensus: The Verifiability of Allegedly Defamatory Speech*, 62 GEO. WASH. L. REV. 43, 70 (1993); B. Jessie Hill, *Putting Religious Symbolism in Context: A Linguistic Critique of the Endorsement Test*, 104 MICH. L. REV. 491, 511–13 (2005); Jonathan Yovel, *What is Contract Law "About"? Speech Act Theory and a Critique of "Skeletal Promises"*, 94 Nw. L. REV. 937, 938 (2000).

analyses of word-mark distinctiveness and paid special attention to the concept of constative utterance (i.e., statements that are either true or false).[66] She argued that trademark use can be constative in two ways: a source-constative utterance connotes the brand, whereas a goods-constative utterance connotes the product or service, irrespective of the brand.[67] By differentiating between distinctive marks (i.e., source-constative utterances) and merely descriptive marks (i.e., goods-constative utterances), we can differentiate between words that are trademark protected and those that are not.[68] Having persuasively advocated for speech-act theory, Roberts proposed that applying a combination of the fair-use doctrine and constative utterance theory would streamline the questions asked in trademark cases: Can hypothetical competitors rightly use part of a trademarked term to describe their own product?[69]

Theories outside the realm of law have been applied not only to word-mark distinctiveness but also to image distinctiveness. For instance, Dustin Marlan shows that, regarding the task of testing for inherent distinctiveness in logos, product packaging, and other such images, the USPTO and the TTAB often used the *Seabrook* test whereas federal courts used the *Abercrombie* taxonomy.[70] However, these two tests are not problem free: the *Seabrook* test, Marlan argued, focuses solely on thematic variation, which can lead to highly subjective and insufficiently

---

[66] Roberts, *supra* note 65, at 1042. *See also* JOHN L. AUSTIN, HOW TO DO THINGS WITH WORDS 3 (J.O. Urmson ed., 1962); PENELOPE BROWN & STEPHEN C. LEVINSON, POLITENESS: SOME UNIVERSALS IN LANGUAGE USAGE (Cambridge University Press 1987); JONATHAN CULLER, LINGUISTIC THEORY: A VERY SHORT INTRODUCTION 94, 101–02 (Oxford University Press 2000).

[67] Roberts, *supra* note 65, at 1084. *See also* Louis Altman & Malla Pollack, *Callmann on Unfair Competition, Trademarks, & Monopolies* § 18:13 n.14 (4th ed. 2011) (comment by author Altman) ("The fundamental import of the term 'descriptive' … is antithetical to the notion of source-significance.").

[68] Roberts, *supra* note 65, at 1045 (noting that distinctive marks perform an action whereas descriptive marks provide information).

[69] *Id.* at 1082 ("It's crucial that the determination of whether a hypothetical competitor could use a given term descriptively in connection with its own product be based on evidence of whether and how the trademark term is used by the public.").

[70] Dustin Marlan, *Visual Metaphor and Trademark Distinctiveness*, 93 WASH. L. REV. 767, 807–08 (2018). The *Seabrook* test considers four factors: "[1] Whether [the logo or trade dress] was a 'common' basic shape or design, [2] whether it is unique or unusual in a particular field, [3] whether it was a mere refinement of a commonly adopted and well-known form of ornamentation for a particular class of goods viewed by the public as a dress or ornamentation for the goods, or [4] whether it was capable of creating a commercial impression distinct from the accompanying words." Seabrook Foods, Inc. v. Bar-well Foods, Ltd., 568 F.2d 1342, 1344 (C.C.P.A. 1977).

supported conclusions.[71] To support their analyses, judges have been known to cite the Restatement of Trademarks.[72] However, danger lurks in efforts to determine whether a "symbol or design is striking, unusual, or otherwise likely to differentiate the products of a particular producer"[73] because, for instance, a common shape (e.g., the outline of an elephant) might be so unusual in a particular context (e.g., a line of spicy instant noodles) that the shape instantly acquires noteworthy—and perhaps even strong—distinctiveness.[74]

As for the *Abercrombie* taxonomy, there are many questions as to whether federal courts would adopt it and whether it is even adequate for evaluating the inherent distinctiveness of images—as seen with the pronounced lack of clarity in the *Two Pesos* case.[75] Influenced by *Abercrombie*, lower federal courts might ill-advisedly integrate a degree-based hierarchy of strength into their analysis of an image's inherent distinctiveness,[76] leading to the problematic categorization of many logos as arbitrary marks simply because most logos appear on product packaging.[77] Because the *Seabrook* test and the *Abercrombie* taxonomy are not, in themselves, suitable tests for assessing an image's inherent distinctiveness, Marlan

---

[71] *Id.* at 808–09. *See, e.g.*, Amazing Spaces, Inc. v. Metro Mini Storage, 608 F.3d 225, 245–47 (5th Cir. 2010) (finding the mark—a stylized star symbol shaded and set within a circle and used in connection with moving and storage services—to not be inherently distinctive because the symbols attributes did not sufficiently distinguish it from other star-formative logos).

[72] Marlan, *supra* note 70, at 808–09; Restatement (Third) of Unfair Competition § 13 cmt. d (Am. L. Inst. 1995) ("A symbol or graphic design is not inherently distinctive unless the nature of the designation and the manner of its use make it likely that prospective purchasers will perceive the designation as an indication of source. Commonplace symbols and designs are not inherently distinctive since their appearance on numerous products makes it unlikely that consumers will view them as distinctive of the goods or services of a particular seller. Thus, unless the symbol or design is striking, unusual, or otherwise likely to differentiate the products of a particular producer, the designation is not inherently distinctive.").

[73] Marlan, *supra* note 70, at 809; *see also* Lars Smith, *Trade Distinctiveness: Solving Scalia's Tertium Quid Trade Dress Conundrum*, Mich. St. L. Rev. 243, 293 n.300 (2005).

[74] Marlan, *supra* note 70, at 809 ("One issue with deciding whether a 'symbol or design is striking, unusual, or otherwise likely to differentiate the products of a particular producer' is that it is entirely subjective and does not establish anything close to a bright-line rule.").

[75] Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 772 (1992).

[76] *See, e.g.*, Paddington Corp. v. Attiki Imps. & Distribs., Inc., 996 F.2d 577, 583 (2d Cir. 1993) ("Since the choices that a producer has for packaging its products are, as the Fifth Circuit noted, almost unlimited, typically a trade dress will be arbitrary or fanciful and thus inherently distinctive. . . ." (citing Chevron Chem. Co. v. Voluntary Purchasing Grps., Inc., 659 F.2d 695, 697 (5th Cir. 1981), *cert. denied*, 457 U.S. 1126 (1982))).

[77] Marlan, *supra* note 70, at 810 ("When it comes to product packaging especially, the possibilities are virtually limitless and courts are quick to assume anything not resembling the product to be arbitrary.").

turned to the three guiding factors adopted by the metaphor-in-advertising theorist Charles Forceville.[78] Integrated into the imagination test,[79] the three factors can be formulated as questions: (1) Does the image mark clearly represent a person, place, or a thing? (2) Does the image mark contain a visual image that is thematically distinct from any related text or non-visual elements? (3) Can the image mark connote its underlying product or service?[80] If the answer is no to the first and third questions, the image mark is distinctive.

A.   *Concerns of Previous Literature*

In my review of the literature above, I have focused on four distinct lines of inquiry: Beebe's semiotic research on distinctiveness and its link to consumers' search sophistication and persuasion sophistication, Linford's application of semantic-shift theory and sound-symbolism theory to consumers' changing interpretations of marks, Robert's combination of constative-utterance theory and the fair-use doctrine to deepen our grasp of both consumer perception and competitors' right of access to potentially trademarkable words, and Marlan's combination of metaphor-in-advertising theory and the imagination test to make sense of images' inherent distinctiveness. Taken together, these lines of inquiry point to five lingering concerns in the realm of trademark law.

First, to make rigorous determinations about source and differential distinctiveness, judges presiding over federal courts must have a workable understanding of words, meaning, and usage. It is no secret that the inescapable complexities and ambiguities of law, combined with the highly subjective experiences and perspectives of judges, can lead them to misunderstand or misapply these ideas.[81] A consequence of this would be misunderstandings and misapplications of the distinctiveness doctrine.

---

[78] *Id.* at 817; CHARLES FORCEVILLE, METAPHOR IN ADVERTISING 4–6 (1996) ("The first criterion for interpreting something as a visual metaphor is that two 'things' are involved. Thus, two things must be identified: (1) the product or service (i.e., the target) and (2) the 'something else' connoted by the mark that is separate from the product or service (i.e., the source). Second, once it is determined that two 'things' exist, it must be determined which is the target and which is the source.").

[79] Marlan posits that the strengths of the imagination test rest partly on its valuation of the metaphorical nature of marks and thus on its valuation of the symbolism as a crucial aspect of valid trademarks. Marlan, *supra* note 70, at 767, 799–802.

[80] Marlan states that image marks must visually suggest, not describe, the target. *Id.* at 819.

[81] *See* James J. Brudney, *Recalibrating Federal Judicial Independence*, 64 OHIO ST. L.J. 149, 170–73, 177–78 (2003) (pointing out that, in the view of some textualists, courts should harness tools that are resistant

The second concern arises from the above concern: if we cannot assume that federal judges are sufficiently familiar with linguistic concepts, how can we assume that the judges can accurately determine the evidentiary criteria for determining trademark distinctiveness—that plaintiffs and defendants should strive to satisfy—when the judges will be focusing on source and differential distinctiveness? If the evidentiary requirements simply reflect the themes laid out in the *Abercrombie* taxonomy (e.g., advertising expenses, advertising reach, media coverage, consumer surveys), there will be no practical difference between the distinctiveness approach and the *Abercrombie* approach. If, on the other hand, the evidentiary requirements refer to themes outside those stipulated by the *Abercrombie* taxonomy,[82] federal courts may strengthen the rigor with which they decide trademark-distinction cases.[83] The critical catch is this: the principles that we use in assessing source and differential distinctiveness must be clear in their abstractness and must lay out a clear path to identifying the evidentiary criteria that litigants and judges must consider in trademark cases.

A third concern arising from my literature review is that difficulties that persist in determining whether a mark that was once merely descriptive has acquired sufficient distinctiveness. One main reason for the persistence of this concern is rooted in evidentiary challenges: to prove that a mark has acquired distinctiveness, one must prove that consumers regard the mark as essentially a trademark for the applicant's goods.[84] However, there is no settled conclusion as to how much evidence a litigant must present in a federal court in order to prove sufficient consumer recognition. That is, the issue of "sufficiency" remains a stumbling block that has yet to be eliminated.

---

to misapplications by poorly reasoning judges); Amanda Peters, *The Meaning, Measure, and Misuse of Standards of Review*, 13 Lewis & Clark L. Rev. 233, 247–51 (2009).

[82] Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976).

[83] According to Beebe, plaintiffs should establish the differential distinctiveness of their mark by proving that the mark is widely known in the way the plaintiffs want it to be known. This statement suggests that the evidence needed to prove differential distinctiveness is in line with the evidentiary requirements corresponding to the *Abercrombie* taxonomy. *See* Beebe, *supra* note 45, at 2031–33; *cf. Abercrombie* 537 F.2d 4, 9 (2d Cir. 1976).

[84] *See How to Claim Acquired Distinctiveness Under Section 2(f)*, U.S. Pat. & Trademark Off., https://www.uspto.gov/trademarks/laws/how-claim-acquired-distinctiveness-under-section-2f-0 [https://perma.cc/Q5GL-ZNM2] (last visited Mar. 12, 2025).

Similar, if not identical, issues arise with respect to sound symbolism. In harnessing the theory to analyze the connotative relationship between sounds and product features, especially for arbitrary marks, Linford helps us understand why fanciful marks are inherently distinctive.[85] Nonetheless, the strengths of his insights do not amount to a sufficiently thorough set of rules for determining which specific combinations of sounds (be they from syllables, vowels, or consonants) constitute evidence of an arbitrary mark. Thus, neither judges nor trademark stakeholders (e.g., owners, applicants) can identify a current mark as arbitrary, and they certainly cannot know, with certainty, how the alteration of sounds might transform an arbitrary mark into a descriptive mark.

The literature review above highlights a fourth outstanding concern: although, as Roberts has shown, speech-act theory may help link the combined powers of the constative-utterance concept and the fair-use doctrine to trademark distinctiveness,[86] a critical omission remains: the highly problematic nature of speech-act theory. For instance, the seven unresolved issues that John Flowerdew persuasively attributed to speech-act theory promise to constrain, if not derail, the applicability of the constative-utterance concept to trademark distinctiveness.[87] One unresolved issue is the silence in speech-act theory regarding how to calculate not only the precise number of speech acts but more specifically the precise number of speech acts categorizable as constative utterances. Furthermore, the arbitrary categorization of constative utterances as source-constative utterances and goods-constative utterances could easily lead to logical errors. And even if we were to accept this model of categorization as satisfactory, federal judges would still face a host of difficulties in applying the test, particularly if the analysis abandons the *Abercrombie* taxonomy entirely. The confusions that might surface in these contexts are limitless. For instance, how should we define the "hypothetical" competitor? Is a cookie-producing firm a competitor of a cake-producing firm? These and other difficult questions will only swell the workloads of federal judges.

---

[85] *See* Linford, *supra* note 53, at 740.

[86] *See* Roberts, *supra* note 65, at 1042, 1081–82.

[87] John Flowerdew, *Problems of Speech Act Theory from an Applied Perspective*, 40 Language Learning 79, 79 (1990). The seven problems concern: (1) the number of speech acts, (2) the nature of indirect speech acts and the concept of literal force, (3) the size of speech-act realization forms, (4) the contrast between specific and diffuse acts, (5) discrete categories versus scale of meaning, (6) the relationships between locution, illocution, and interaction, and (7) the relationships between wholes and parts in discourse.

Thus, although Roberts has proposed a simple test for the analysis of word-mark distinctiveness, the test lacks the robust persuasiveness that federal judges would expect of such a tool. Before it can be deemed suitable for the court system, the test must address, with sufficient clarity, the specific characteristics attributable to arbitrary, suggestive, descriptive-acquired, purely descriptive, and generic terms.

The fifth and final concern stemming from my literature review pertains to Marlan's three-factor proposal for determining the inherent distinctiveness of images.[88] Though the proposal seemed to fill the gap that *Two Pesos* had failed to bridge, a lingering dilemma is the proposal's inferior status under the umbrella of the imagination test. As I mentioned earlier with regard to the imagination test, any attempt to use the three-factor tool in determining the inherent distinctiveness of visuals might, in the realm of trademark law, create more problems than it resolves. Consider the following scenario: judges and others might be comfortable tackling the first factor (i.e., the clear-representation question) but might then be stymied by the second and third factors (i.e., the "visual vs. non-visual" question and the connotation question) because the considerable degree of subjectivity that these factors permit might encourage federal judges to revert to the *Seabrook* test or the *Abercrombie* taxonomy, which offer comforting legal precedents on which to base a decision. As a result, the judicial system's handling of trademark cases might split into even more divisions if we were to adopt an unamended three-factor approach to determining the inherent distinctiveness of visuals.

Different from Beebe, Linford, Roberts, and Marlan, some scholars have sought to uncover the roots of distinctiveness by means of historical analysis. For instance, consider the genericide doctrine: the phenomenon wherein a once-protectable mark is no longer able to function as a trademark because it became the generic term for an entire category of products instead signifying the specific brand or source of the product.[89] For example, for several decades, the brand name Kleenex has been becoming a generic term for the product category, tissues.[90] Desai and Rierson analyzed the roots of the genericism doctrine back to language

---

[88] Marlan, *supra* note 70, at 819–21.

[89] Xiyin Tang, *Against Fair Use: The Case for a Genericness Defense in Expressive Trademark Uses*, 101 Iowa L. Rev. 2021, 2024 (2016).

[90] Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 Cardozo L. Rev. 1789, 1846–47 (2007). *See also* Megan Garber, *'Kleenex is a Registered Trademark' (and Other Desperate Appeals)*, The Atl. (Sept. 25, 2014), https://www.theatlantic.com/business/

used in the Trade-Mark Act of 1905, through which Congress sought to codify, at the federal level, previous common-law remedies.[91] Influenced by the act's definition of the genericism doctrine, federal courts hearing a trademark case would examine whether the mark of primary significance referred to a product category or to a particular product.[92] In examining these matters, the courts would controversially rely on dictionaries for definitions or on newspapers for how trademarks being used.[93] Desai and Rierson object to these lines of examination on two grounds. First, marks have hybrid functionalities, and source-identifiers are just one. Thus, a narrow focus on *just* the public context or *just* the noncommercial context, without adequate attention paid to the commercial context, is a decidedly fragmentary approach to determining a mark's genericness.[94] Second, Desai and Rierson argue that if federal courts are still regarded as focusing exclusively on noncommercial contexts, word-mark holders seeking to prove fair use will quite reasonably focus on presenting dictionary- and media-based evidence, not product- or service-based evidence.[95]

Historical analyses can bring to light the loopholes on which federal judges have relied while struggling to apply the genericism doctrine. Legal scholars should examine the possible roles played by similar loopholes in arbitrary, suggestive, and descriptive marks. To this end, empirical research on trademark distinctiveness is needed. Beebe has performed scholarship in this direction, as has Thomas R. Lee and his colleagues.[96] Beebe's study focused on the circuit courts' use of differing multifactor tests for determining the likelihood of confusion in trademark

---

archive/2014/09/kleenex-is-a-registered-trademark-and-other-appeals-to-journalists/380733/ [https://perma.cc/4KQQ-XEHC].

[91] Desai & Rierson, *supra* note 90, at 1812. *See* Canal Co. v. Clark, 80 U.S. 311, 323 (1871) (holding that "a generic name, or a name merely descriptive of an article of trade, of its qualities, ingredients, or characteristics, [cannot] be employed as a trade-mark and the exclusive use of it be entitled to legal protection"); Lawrence Mfg. Co. v. Tennessee Mfg. Co., 138 U.S. 537, 547 (1891) (same).

[92] Desai & Rierson, *supra* note 90, at 1855.

[93] *Id.* at 2054.

[94] *Id.* at 1833 ("At best, [the commercial context] demonstrates that the word or term is or may be functioning as a hybrid trademark, while shedding little light on which understanding of the term constitutes its 'primary significance' to the consumer in a commercial context.").

[95] *Id.* at 1855. *See also* Ralph H. Folsom & Larry L. Teply, *Trademarked Generic Words*, 70 Trademark Rep. 206, 236 (1980).

[96] *See* Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 Cal. L. Rev. 1581, 1584 (2006); *see also* Thomas R. Lee, Eric D. DeRosia & Glenn L. Christensen, *An Empirical and Consumer Psychology Analysis of Trademark Distinctiveness*, 41 Ariz. St. L.J. 1033, 1038 (2009).

litigation.[97] Beebe collected and analyzed all (331) reported federal district court opinions from trademark infringement cases involving a multifactor tests from 2000 to 2004.[98] Beebe's analysis revealed that although federal courts always acknowledged the non-dispositive nature of the multifactor test and the importance of considering all factors, in actuality, federal judges tended to consider only a few decisive factors.[99] That is, despite the injunction against ignoring factors, these judges were tempted to decide likelihood-of-confusion cases in a more "efficient" way.[100] Regarding the specific core factors of trademark strength and inherent distinctiveness, Beebe found that 44% of the 331 opinions lacked any rigorous assessment of the given mark's potentially inherent distinctiveness.[101] Moreover, only 58% of the 331 opinions used the *Abercrombie* taxonomy, and, of these, 29 simply cited a prior *Abercrombie* case rather than categorize the mark's distinctiveness according to the taxonomy.[102] Finally, Beebe uncovered in the opinions a series of contradictions between the analyses of acquired distinctiveness and the analyses of inherent distinctiveness: federal courts would simultaneously declare a mark to be inherently weak yet commercially strong.[103] Thus, Beebe argued that inherent distinctiveness has broken down because it has been trumped by acquired distinctiveness.[104]

Beebe's empirical research on multifactor tests for likelihood of confusion in trademark litigation seems to have accidentally unearthed a curious loophole that has enabled—and perhaps even encouraged—federal judges to avoid conducting rigorous analyses of trademark strength and distinctiveness. Nonetheless, the explanation that federal judges are simply attempting to decide likelihood of confusion cases in a more efficient way does not address why the judges would ignore a clearly stipulated rule governing how one should determine

---

[97] Beebe, *supra* note 96, at 1581, 1584.

[98] *Id.* at 1584.

[99] *Id.* at 1619 (explaining that "a plaintiff will not bring an action for trademark infringement unless the facts of its case are such that it will win at least a few of the multifactor test factors").

[100] *See id.* at 1614; *see, e.g.*, Jens Förster, E. Tory Higgins & Amy Taylor Bianco, *Speed/Accuracy Decisions in Task Performance: Built-In Tradeoff or Separate Strategic Concerns*, 90 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 148, 149 (2003) (discussing speed-vs.-accuracy decisions from the perspective of regulatory focus theory).

[101] Beebe, *supra* note 96, at 1635.

[102] *Id.*

[103] *Id.* at 1636.

[104] *Id.*

distinctiveness under the *Abercrombie* taxonomy. Furthermore, a key issue for trademark applicants is the challenge of designing a word mark that a judge would regard as a strong mark in the likelihood of confusion analysis, regardless of consideration of other factors in the multifactor test. Thus, both federal judges and trademark applicants would greatly benefit from clear guidelines governing how one should calculate trademark distinctiveness under the *Abercrombie* taxonomy.

A pertinent empirical study about distinctiveness comes to us courtesy of a 2009 paper by Thomas R. Lee and his colleagues.[105] They adopted a consumer psychology model built on the theory of perceptual schema and used it to test the hypotheses contained within the *Abercrombie* taxonomy.[106] Specifically, the study consisted of three constituent empirical studies focusing on consumers' perception of a word mark's distinctiveness. The first constituent study, completed by 210 participants, involved an online questionnaire adapted from the TEFLON test.[107] First, participants were shown a product package featuring a mark consisting of both a picture and words. The participants were then asked whether the mark on the package "is a brand name," "is not a brand name," or "I don't know, or I have no opinion."[108] Participants who stated that the mark "is a brand name" were coded as having identified the mark as source indicating.[109] In the first constituent study, Lee and his colleagues found that, in typical trademark use involving product packaging, descriptive marks could be as highly source-indicating as suggestive

---

[105] Lee, DeRosia & Christensen, *supra* note 96, at 1035–36 (addressing the *Abercrombie* assumption that the power to indicate a source belongs, in descending order, to fanciful marks, arbitrary marks, suggestive marks, descriptive marks, and generic marks, and that there are no source-indicating differences among types of descriptive marks).

[106] *Id.* at 1033. "Perceptual schemas" are mental frameworks built through past perceptual experiences that guide current perception. *Id.* at 1074. Lee's paper was interested specifically in "brand perceptual schemas"—or consumer perception of visual cues in the marketplace for the goal of identifying a product's source. *Id.* at 1075.

[107] *Id.* at 1086. The *Teflon* test, generally accepted for evaluating secondary meaning, was first formulated to evaluate the distinctiveness of the TEFLON brand in E. I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc., 393 F. Supp. 502, 526 (E.D.N.Y. 1975). *See, e.g.*, Schwan's IP, LLC v. Kraft Pizza Co., 379 F. Supp. 2d 1016, 1024 (D. Minn. 2005); March Madness Athletic Ass'n, LLC v. Netfire, Inc., 310 F. Supp. 2d 786, 809 (N.D. Tex. 2003). However, Lee and his colleagues stated that "the TEFLON test cannot serve as a straightforward measure of source indication because (1) that test assumes that if a word is not a brand name, then it must be a generic term; and (2) it presents participants with bare words rather than presenting trademarks in a realistic commercial context." Lee, DeRosia & Christensen, *supra* note 96, at 1086.

[108] Lee, DeRosia & Christensen, *supra* note 96, at 1088.

[109] *Id.*

marks.[110] This finding is inconsistent with our conventional understanding of the *Abercrombie* taxonomy, according to which a descriptive mark is less source-indicating than a suggestive mark.

The above inconsistency prompted Lee and his colleagues to conduct a second constituent study, focusing on the hypothesis that descriptive marks are less source-indicating than suggestive marks.[111] This study revolved around pita chip snacks, vitamin food supplements, laundry stain removers, and packaged cookies. Each product was presented in typical product packaging.[112] The procedure used was identical to the method for the previous study (participants responding to an online survey).[113] The results of the second constituent study were the same as those of the first: descriptive marks and suggestive marks exhibited similar source indication.[114]

Lastly, because Lee and his colleagues had paired a picture with words in the first two constituent studies, the question naturally arose as to whether the non-linguistic parts affected the results concerning source indication.[115] Thus, the researchers set out to conduct a third constituent study, this time testing whether the picture or any other non-linguistic elements played a key role in source indication.[116] The procedure for this study was the same as the one established in the first study, with the exception that only 120 participants completed the survey.[117] The results indicated that the non-linguistic elements other than the words significantly persuaded consumers to perceive the given descriptive mark as source-indicating.[118] This might help explain why in the previous studies descriptive marks were exhibiting similar source-indication levels as suggestive marks.

The research conducted by Lee and his colleagues inspired me to pursue an alternative approach to conducting empirical research on word-

---

[110] *Id.* at 1092.
[111] *Id.* at 1094.
[112] *Id.* at 1092.
[113] *Id.*
[114] *Id.* at 1095.
[115] *Id.* at 1096.
[116] *Id.*
[117] *Id.*
[118] *Id.* at 1098.

mark distinctiveness. However, there are potential concerns with Lee's research. First, the theory of perceptual schema has been the target of considerable criticism concerning the theory's appreciable ambiguity, vagueness, and weak applicability.[119] According to Thorndyke and Yekovich, the theory is "so vaguely specified that it is able to explain *post hoc* virtually any set of available data."[120] That is, the theory itself has no specified process constraints.[121] Thus, researchers might judiciously regard, with skepticism, the theory's ability to yield consistent results.

Additionally, even if we can overcome the process-constraint limitations currently plaguing the theory of perceptual schema, flaws in the cited research persist. For example, regarding the second constituent study, its pita chip snacks, vitamin food supplements, laundry stain removers, and packaged cookies cover only two trademark categories: Class 20 (furniture products) and Class 29 (meat and processed-food products).[122] Whether packaged products in other trademark categories would have yielded identical or at least similar results is an issue worthy of investigation.

A third concern is the practicality or usefulness of the results stemming from Lee and his colleagues' research. The two most striking findings were that (1) suggestive word marks and descriptive word marks might have identical source-indicating effects, and (2) the non-linguistic characteristics in a mark might alter consumers' perception of the mark's distinctiveness. These two interesting findings, though they might assist federal courts in navigating the *Abercrombie* taxonomy, provide no clear guidance for the analysis of distinctiveness as a whole. Thus, while it is constructive for federal courts to consider the effects of non-linguistic characteristics when analyzing word-mark distinctiveness, the judges must still take into account precedents when trying to "analogously" ascertain, for

---

[119] *See* Katherine L. Plant & Neville A. Stanton, *The Explanatory Power of Schema Theory: Theoretical Foundations and Future Applications in Ergonomics*, 56 ERGONOMICS 1, 4–5 (2012); Milton Lodge, Kathleen M. McGraw, Pamela Johnston Conover, Stanley Feldman & Arthur H. Miller, *Where Is the Schema? Critiques*, 85 AM. POL. SCI. REV., 1357, 1357 (1991); Charles H. Shea & Gabriele Wulf, *Schema Theory: A Critical Appraisal and Reevaluation*, 37 J. MOTOR BEHAV. 85, 96 (2005).

[120] Perry W. Thorndyke & Frank R. Yekovich, *A Critique of Schema-based Theories of Human Story Memory*, 9 POETICS 23, 40 (1980).

[121] *Id.* at 41 (noting that schema theory suffers from poor predictive powers and from such excessively vague specifications that it yields only results consistent with the theory).

[122] Lee, DeRosia & Christensen, *supra* note 96, at 1081–82, 1092.

instance, whether a previous ruling categorizing the 'COCA' mark as a suggestive mark should encourage a judge presiding over a current case to categorize the similarly spelled 'CACA' mark as suggestive.

Table 1, which summarizes my literature review findings, shows that most of the studies do not touch on the critical issue of whether or not—and if so, to what extent—federal judges rationally ignore the role of vagueness in the categorization of trademark distinctiveness. Beebe's research indicates the presence of such ignorance, but his study, in addition to having a small sample, focuses on the test for likelihood of confusion.[123] More generally, the literature has provided no clear guidance by which stakeholders, whether it be federal judges or trademark applicants or owners, can rigorously categorize word-mark distinctiveness in the context of the *Abercrombie* taxonomy. Overall, Table 1's summary of the literature demonstrates not only the *status quo* with regard to trademark distinctiveness but also the value of exploring the possible existence of rational ignorance among federal judges who focus excessively on differences between suggestive and descriptive distinctiveness. As I shall demonstrate in the present study, rational ignorance is a problem in trademark litigation, and the reasons for resolving the problem will become evident.

---

[123] Beebe, *supra* note 96, at 1581. *See generally* Olga Ampuero & Natalia Vila, *Consumer Perceptions of Product Packaging*, 23 J. Consumer Mktg. 100 (2006) (discussing the impact of package positioning on consumer perception).

Main Research Topics	Literature	Research Focuses	Found Evidence of or Solutions to Rational Ignorance
Product design and trade-dress distinctiveness	Graeme B. Dinwoodie, *Reconceptualizing the Inherent Distinctiveness of Product Design Trade Dress*, 75 N.C. L. Rev. 471 (1997)	Trade-dress distinctiveness	No
An investigation, based on non-legal theory, into trademark law and distinctiveness	Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. Rev. 621 (2004) Barton Beebe, *Search and Persuasion in Trademark Law*, 103 Mich. L. Rev. 2020 (2005) Jake Linford, *A Linguistic Justification for Protecting "Generic" Trademarks*, 17 Yale J.L. & Tech Jake Linford, *Are Trademarks Ever Fanciful*, 105 Geo. L.J. 731 (2017) Alexandra J. Roberts, *How To Do Things with Word Marks: A Speech-Act Theory of Distinctiveness*, 65 Ala. L. Rev. 1035 (2014) Dustin Marlan, *Visual Metaphor and Trademark Distinctiveness*, 93 Wash. L. Rev. 767 (2018)	Inherent and acquired distinctiveness, and specific categorizations of distinctiveness (e.g., generic, fanciful)	No
Historical analysis of the roots of distinctiveness	Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 Cardozo L. Rev. 1789 (2007)	The doctrine of genericism	No
Empirical analysis of confusion and distinctiveness	Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 Calif. L. Rev. 1581 (2006) Thomas R. Lee, Eric D. DeRosia & Glenn L. Christensen, *An Empirical and Consumer Psychology Analysis of Trademark Distinctiveness*, 41 Ariz. St. L.J. 1033 (2009)	Tests for confusion and consumers' perception of linguistic and non-linguistic elements of marks	Beebe found that (1) only 58% of 331 opinions used the *Abercrombie* taxonomy and (2) 29 of the 58%, rather than categorize the marks' distinctiveness, only cited previous *Abercrombie* cases.

Table 1: Summary of the literature review

A central takeaway here is that little or none of the literature has either analyzed each category of word mark distinctiveness or laid out a plan for resolving the lack of clarity in the categorization of word mark distinctiveness. Unfortunately, the three tests—the imagination test, competitive-need test, and dictionary test—are incapable of effectively rigorously assessing all forms of word mark distinctiveness. To make matters worse, the three tests quite possibly encourage federal judges to focus on differences between suggestive and descriptive marks at the expense of clarifying the importance of other types of distinctiveness, and the Supreme Court failed to provide much-needed guidance when it had the opportunity to do so in *Booking.com*.[124]

Therefore, in the present study, I explore (1) the degree to which federal judges excessively focus on differences between suggestive and descriptive marks, (2) the linguistic patterns of this excessive focus, (3) the distractive influence that the focus has on the neglected task of categorizing all types of distinctiveness, and (4) practical, comprehensive solutions to this problem of poor judicial reasoning.

III
Methods, Variables, and Data

Because my two central aims in this study are to identify patterns of poor reasoning exhibited by federal judges in word mark dispute cases and to provide workable solutions to the problem, it is necessary that I first observe to what extent word mark precedents across federal jurisdictions and under the umbrella of the *Abercrombie* taxonomy are related to this poor reasoning. In order to analyze the caselaw, I adopted methods of capable of dealing with categorical, rather than numerical, data. In the following paragraphs, I introduce these methods, address the variables to be analyzed, discuss the sources and my collection of the data, and conclude with my approach to the hand-coding of values.

A.   Methodology

In this study, I rely extensively on the decision-tree method, which is a machine-learning tool for data categorization. Below, I explain why I chose this

---

[124] In *Booking.com*, the Supreme Court addressed whether "Booking.com" was capable of being source indicating or if it was a generic term. *See* U.S. Pat. & Trademark Off. v. Booking.com B. V., 591 U.S. 549, 555 (2020).

tool to identifying patterns of rational ignorance among federal judges presiding over trademark cases. No method is perfect, however, and thus, I also address the limitations of decision trees and explain how I dealt with those limitations. Finally, I introduce the dependent and independent variables of this paper and explain why I chose them for my decision-tree analysis.

(1) An introduction to decision trees and their suitability for the present study

A decision tree is "a non-parametric supervised learning algorithm" and can be used for categorical output variables (classification trees) and continuous output variables (regression trees).[125] As the name implies, the tool has a hierarchical, tree-like structure, which consists of a root node, branches, internal nodes, and leaf nodes.
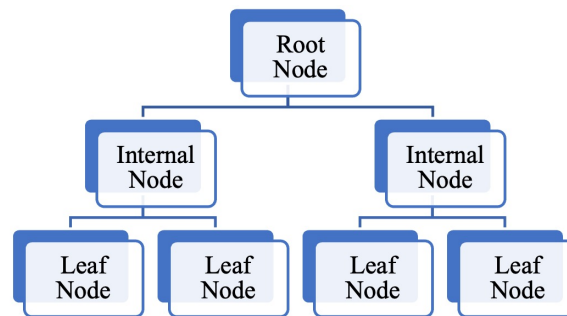
In the diagram above, the decision tree begins with the initial decision, known as the root node. It is distinctive in that it has only outgoing branches, not incoming branches. Its outgoing branches lead to the internal nodes, which are also referred to as decision nodes. Internal nodes involve evaluations of features and feed into leaf nodes, which are terminal and represent all final possible outcomes.[126] Based on a divide-and-conquer strategy, decision trees perform greedy searches (i.e., searches in pursuit of the best outcome at a given moment). Once identified, the best outcomes form optimal split points, and this splitting continues downward along the tree until the dataset on which the tree is based has been exhaustively classified.[127]

---

[125] *What is a Decision Tree?*, IBM, https://www.ibm.com/think/topics/decision-trees [https://perma.cc/YN3B-46K2] (last visited Mar. 3, 2025).

[126] Yan-yan Song & Ying Lu, *Decision Tree Methods: Applications for Classification and Prediction*, 27 SHANGHAI ARCHIVES PSYCH. 130, 131 (2015).

[127] *What is a Decision Tree?*, IBM, https://www.ibm.com/think/topics/decision-trees [https://perma.cc/YN3B-46K2] (last visited Mar. 3, 2025).

There are three main reasons why I have relied on the decision-tree method: (1) it categorizes data with a focus on objectively assessed features, (2) it handles categorical, as well as numerical data despite incomplete values, and (3) it supports non-linearity, which is a chief characteristic of my data. I discuss these three reasons in detail below.

First, a categorization of data that permits objective assessments of important features is key in the present study, where my focus is on how federal judges, when making decisions about word-mark distinctiveness, may either ignore or at least fail to clarify certain types of distinctiveness under the *Abercrombie* taxonomy. To explore how federal judges possibly fail to categorize forms of distinctiveness and how these judges instead excessively emphasize differences between suggestive and descriptive marks in relation to the imagination test, competitive-need test, and dictionary test, I obtained historical data in the form of case precedents and then analyzed the dataset. Part of this analysis rested on a decision-tree algorithm: its categorization output was central to my analysis of the historical data.[128] With decision trees, we can assess federal judges' reliance on the three aforementioned tests by establishing not only variables relating to those tests but also variables that may not be relevant to the tests. If a decision tree treats the former variables as key categorizable features, I can reasonably infer that federal judges would be inclined to use those tests to analyze word-mark distinctiveness. If, by contrast, the decision tree does not treat the variables as key categorizable features, I can just as reasonably infer that federal judges care more about other variables than about those involved in distinctiveness tests. Moreover, the algorithmic results pertaining to feature-based categorizations of distinctiveness can indicate two important points: (1) whether the features might enable federal judges to clarify categories of distinctiveness, and (2) whether federal judges, by focusing on suggestive and descriptive distinctiveness, have historically ignored the task of clarifying categories of distinctiveness, particularly if the hypothetical features in question were all related to the three tests above. In short, decision trees are well suited for the focus of the present study.

---

[128] *See* Jonathan P. Kastellec, *The Statistical Analysis of Judicial Decisions and Legal Rules with Classification Trees*, 7 J. EMPIRICAL LEGAL STUD. 202, 206–07 (2010).

Second, decision trees are attractive for their ability to read both categorical and numerical data, even when some values are incomplete or missing.[129] This functionality can help me to analyze the federal-court data herein because the federal cases that I am considering involve content that is mostly categorical. For instance, federal cases involve a dependent variable—the distinctiveness of a given mark—that constitutes a piece of categorical data. Of course, not all variables are dependent: international trademark classes (abbreviated as ICs) are an independent variable that I examine here. If word marks are registered in the USPTO's Trademark Search System (TSS),[130] the TSS records registration includes the word mark's ICs. This independent variable is also a type of categorical data. I was well aware that, because some owners of a trademark may not register it with the TSS, missing values would surface during my collection of data. The fact that the decision-tree algorithm could help me overcome this obstacle was the second main reason why I chose this tool for the present study.

Third, decision trees can support non-linearity,[131] a trait that is all-important for the present study insofar as non-linearity is one of the main characteristics of my data. Consider, for instance, the fact that the dependent variable of this paper is word mark distinctiveness while one of the independent variables is "first-year use": it constitutes a type of categorical data. Changes in "first-year use" do not form a linear relationship with categorizations of word-mark distinctiveness. This point was fundamental in my decision to pass over multiple linear regression in favor of the decision-tree method.

---

[129] Decision trees use something called surrogate splits to overcome the problem of missing values: "These surrogate splits act as backup choices when the primary attribute for a split has missing values. The algorithm identifies the next best attribute that can provide a similar separation as the primary attribute." Aishwarya Kurre, *How Decision Trees Handle Missing Values: A Comprehensive Guide*, PICKL.AI (Aug. 16, 2023), https://www.pickl.ai/blog/how-decision-trees-handle-missing-values-a-comprehensive-guide/ [https://perma.cc/CAN5-E3KB].

[130] The TSS replaced the Trademark Electronic Search System (TESS) on November 30, 2023. Because I collected the bulk of the present study's data prior to this date, I used the TESS in most situations. The TSS with respect to trademark information is almost identical to the TESS. For more details about the system substitution, *see Trademark Search System Updates*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/trademarks/search/trademark-search-system-updates [https://perma.cc/22YL-H7YP] (last accessed Mar. 21, 2025).

[131] Kastellac, *supra* note 128, at 209.

(2) Methodological processes

Having reviewed decision trees, we can now consider the methodological processes by which I identified the potential patterns of poor reasoning exhibited by federal judges in work-mark dispute cases. First, I based this study's variables on the literature about word mark distinctiveness and on the three abovementioned tests: the imagination test, competitor-need test, and dictionary test. My decision to base the variables on the three tests was rooted in one of my central research objectives: to determine whether or not federal judges truly sought to clarify the boundaries separating and defining all types of distinctiveness. By examining the judges' application of the tests, I would be able to achieve this objective: judges who apply *only one or some combination of these three tests* likely strive to differentiate between suggestive and descriptive distinctiveness; judges who apply *alternative tests in addition to one or some combination of these three tests* likely analyze all spectrums of distinctiveness, not just suggestive and descriptive distinctiveness.

The second step in the methodological process was to collect federal trademark cases. To this end, I consulted the Lexis Nexis database for the period extending from January 1, 2002 to December 31, 2022 and filtered out the decisions in which the presiding federal judges made no reference to types of word mark distinctiveness. I then hand-coded the data relating to the variables established in the first step. Specifically, I searched for or independently calculated the values for data obtained from various authoritative sources, including most notably the abovementioned federal trademark cases, the TSS, and the Corpus of Contemporary American English (COCA). Part of the task I faced was to code these values as either categorical or numerical data for the training of the decision trees. Finally, upon completion of the hand-coding, I commenced the analysis phase of this study. The results of the analysis would shed light on any patterns of poor reasoning exhibited by federal judges in trademark litigation.

(3) Variables for training the decision tree

The dependent variable in the present study is federal judges' categorization of marks according to their distinctiveness. The five possible categories are arbitrary marks (I treat fanciful and arbitrary marks as a single category although *Abercrombie* treats them as separate), suggestive marks, descriptive-acquired marks (i.e., marks possessing acquired distinctiveness, also known as secondary meaning), purely descriptive marks (i.e., marks possessing no acquired

distinctiveness), and generic marks.[132] I established the independent variables on the basis of the literature review and the three tests. Most of these variables come from sources that reflect three types of information: (1) information directly related to federal cases involving disputed word marks, (2) information related to the linguistic characteristics of the disputed word marks, and (3) mostly TSS-based information related primarily to product-and-service categories for the disputed word marks. In total, I selected nineteen (19) independent variables for training.

Four independent variables stem from the federal cases: (1) decision year, (2) jurisdiction, (3) judge gender, and (4) judge tenure (in years). Regarding the gender variable, because some of the courts that I studied were presided over by three judges (e.g., circuit courts), not one judge (e.g., district courts), I used the majority, or dominant, gender for the multi-judge courts. Independent Variable 3 (gender) and Independent Variable 4 (tenure) serve to investigate whether gender differences and work experience affect judges' categorization of distinctiveness.[133]

Eight independent variables concern the linguistic characteristics of the litigated word marks in the federal cases: (5) word-formation category,[134] (6) dictionary status,[135] (7) plosive status (i.e., does the word mark start with a plosive, which is to say, a B, C, D, G, K, P, or T sound),[136] (8) word count, (9) syllable count, (10) vowel count, (11) consonant count, and (12) COCA frequency (i.e., the frequency with which a given word mark appeared in the COCA database). Two points should be made here. First, regarding Independent Variable 7 (plosive status), I decided to consider the opening plosive characteristics of the litigated word marks because research has shown that recollection and

---

[132] Abercrombie & Fitch Co. v. Hunting World, Inc. 537 F.2d 4, 9 (2d Cir. 1976). *Abercrombie* does not refer to descriptive-acquired marks and purely descriptive marks as separate categories, and instead refers to them together as "descriptive marks" and notes that these marks are only capable of functioning as a trademark if they acquire distinctiveness via secondary meaning. *See id.* at 10. I treat them as separate categories for the purpose of the analysis.

[133] For discussion of whether gender differences and work experience affect judges' categorization of distinctiveness, see María L. Sanz de Acedo Lizárraga, María T. Sanz de Acedo Baquedano & María Cardelle-Elawar, *Factors That Affect Decision Making: Gender and Age Differences*, 7 Int'l J. Psych. & Psych. Therapy 381 (2007).

[134] William O'Grady & John Archibald, Contemporary Linguistic Analysis: An Introduction (Pearson Ed. Can., 8th ed. 2015). Types of word formation are inflection, derivation, cliticization, suppletion, compounding, conversion, blending, clipping, and acronyms and initialisms.

[135] The dictionary status refers to whether the word mark can be found in the dictionary.

[136] Bruce G. Vanden Bergh et al., *Sound Advice on Brand Names*, 61 Journalism Q., 835, 835 (1984).

recognition of words tend to be stronger when words begin with plosives than when words begin with other sounds.[137] It is possible, therefore, that the initial sound of an uttered word might assist in the categorization of distinctiveness. Second, concerning Independent Variable 12 (COCA frequency): the COCA database contains more than one billion words, including 20 million words for each year in the period extending from 1990 through 2019.[138] Because the database estimates the frequency of word usage in several categories (e.g., conversational contexts, works of fiction, magazines, newspapers, academic contexts, web texts, TV and film), I decided to incorporate frequency into the study in order to determine whether this independent variable might affect judges' categorization of the distinctiveness of litigated word marks. Not incidentally, the COCA database was used by Beebe and Fromer in their research on word mark depletion and congestion.[139]

The remaining seven independent variables concern TSS-based information: (13) International Class (IC) count, (14) U.S. trademark class count (word-mark owners can choose multiple classes to register in the TSS), (15) word-mark product or service,[140] (16) first-year use, (17) duration, (18) third-party registration count, and (19) categorization of third-party registration count (few $\leq$ 10, medium = 11–60, large = 61–100, super large $\geq$ 100). The estimations of Independent Variable 18 (third-party registration count) are quite complicated, so I will discuss the matter in greater depth in the section on coding processes. At this point, let me simply note that I established Independent Variable 19 (categorization of third-party registration count). To prevent the decision tree from being dominated

---

[137] *Id.* at 839.

[138] In COCA's official website, it explains that "the corpus contains more than one billion words of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, TV and movies subtitles, blogs, and other web pages." CORPUS OF CONTEMP. AM. ENG., https://www.english-corpora.org/coca/ [https://perma.cc/3DE9-WREA].

[139] Barton Beebe & Jeanne C. Fromer, *Are We Running Out of Trademarks? An Empirical Study of Trademark Depletion and Congestion*, 131 HARV. L. REV. 945, 975–76 (2018); *see also Word Frequency Data: Based on 450 million Word COCA Corpus*, WORD FREQUENCY DATA, https://www.wordfrequency.info/100k.asp [https://perma.cc/Q636-XNAV] (last visited Mar. 21, 2025); *see also* Mark Davies, *The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English*, 25 LITERARY & LINGUISTIC COMPUTING 447, 453 (2010).

[140] According to the USPTO's website about the International Trademark Classes, Class 1 through Class 34 are related to goods. *See* U.S. Pat. & Trademark Off., *Goods and Services*, https://www.uspto.gov/trademarks/basics/goods-and-services [https://perma.cc/V8GK-YWB5] (last visited Mar. 21, 2025).

by an excessive number of large and super-large third-party registrations, I defined Independent Variable 19 (categorization of third-party registration count) as a polytomous variable derived from Independent Variable 18. Table 2 summarizes all twenty variables (the one dependent variable and the nineteen independent variables).

Dependent Variable	Word-mark distinctiveness decisions of federal judges
Independent Variables	• **Information about the federal cases** 1. decision year 2. jurisdiction 3. judge gender 4. judge tenure • **Linguistic characteristics of word marks** 5. word-formation category 6. dictionary status 7. plosive status 8. word count 9. syllable count 10. vowel count 11. consonant count 12. COCA frequency • **Information recorded in the TSS system** 13. IC count 14. US class count 15. product or service 16. first-year use 17. duration 18. third-party registration count 19. categorization of third-party registration count

Table 2: Summary of dependent and independent variables

Some important variables about word-mark distinctiveness were not suitable for this study because their measured values were inaccessible (e.g., word-mark marketing expenses). The absence of these data sets from the decision-tree training constitutes a notable limitation of the present study.

B.   Data Collection, Coding Processes, and Filtration

The third-party sources of data were critical for my analysis of the independent variables. Here, I will explain the collection and hand-coding steps for these independent variables. I will also explain this study's data-filtration steps, which helped shape the final dataset.

1.   Dependent Variable

I hand-coded the dependent-variable data. Ideally, the decisions of the federal judges would clearly identify any categories of distinctiveness assigned to a given litigated word mark. I coded the five categories of distinctiveness thus: arbitrary = A, suggestive = S, descriptive-acquired = DA, purely descriptive = D, and generic = G.

Beebe's research suggests that federal judges would not necessarily address word mark distinctiveness in assessing likelihood-of-confusion in trademark litigation.[141] This possibility points to two problems that might complicate efforts to study the opinions of judges: judges might issue opinions that offer neither clear reasoning nor clear consequences regarding the distinctiveness of the litigated word mark. To train the decision tree in the present study, I needed to filter out cases tainted by the first problem (no clear reasoning in the ruling) because they in no way facilitate my effort to determine how judges categorized word mark distinctiveness.

As for the second problem (no clear consequences of the ruling), judges assessing word mark distinctiveness sometimes waffled between a "suggestive and descriptive" label or between an "arbitrary and suggestive" label. To deal with this lack of decisiveness with respect to "suggestive and descriptive" equivocation, I coded the court's decision as 'DA' because the descriptive-acquired category is weaker than the suggestive category (i.e., suggestive distinctiveness is always stronger than descriptive distinctiveness, be it acquired or not).[142] As for why I did not choose the ultra-conservative path and code the "suggestive and descriptive" equivocation as simply 'D' (purely descriptive), the simple answer is that, in

---

[141] Beebe, *supra* note 96, at 1635 (stating that "courts failed to specify whether or not the mark at issue was inherently distinctive in 40% of the 192 preliminary injunction and bench trial opinions sampled and in 50% of the 139 summary judgment opinions sampled, for an overall failure rate of 44% in the 331 opinions examined").

[142] *See* Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 10–11 (2d Cir. 1976).

most cases involving word-mark disputes, federal judges who use 'suggestive' and 'descriptive' interchangeably are treating the 'descriptive' category as stronger than 'purely descriptive'. I applied the same conservative reasoning to "arbitrary and suggestive" equivocation: I conservatively coded it 'S' so as to avoid an overestimation of the distinctiveness.

For an illustration of my coding process, consider how I handled *International IP Holdings, LLC and Innovation Ventures, LLC v. Green Planet, Inc.*[143] Presiding over the case was Judge Cleland, who offered his two cents on the disputed word-mark 5-HOUR ENERGY:

> It is clear the mark is not fanciful or arbitrary because its name describes or at least suggests what the product is supposed to do—provide "energy" for "five hours." In fact, when Plaintiffs first applied to register the 5-hour ENERGY trademark, the Patent and Trademark Office denied the application on the grounds that the name was descriptive. The court agrees that due to the suggestive or descriptive nature of the 5-hour ENERGY mark, the Friendly test indicates that the mark is inherently weak.[144]

Judge Cleland's reasoning, which clearly waffles between a suggestive categorization and a descriptive categorization (e.g., "the suggestive or descriptive nature"), led me to hand-code the judge's categorization of the 5-HOUR ENERGY word mark as a "descriptive-acquired" mark. Hence, I entered the code 'DA' into the dataset.

2. *Independent Variables*

2.1. Information About the Federal Cases

To find federal cases to analyze for this study, I used the LexisNexis database and searched for the term 'strength of the mark.'[145] Next, I designated the practice

---

[143] *See* Int'l IP Holdings, LLC v. Green Planet, Inc., No. 13-13988, 2016 WL 1242275, at *2–12 (E.D. Mich. Mar. 30, 2016), *opinion withdrawn and vacated*, No. 213CV13988RHCRSW, 2017 WL 1538621 (E.D. Mich. Mar. 9, 2017).

[144] *Id.* at *6.

[145] I decided to use the term 'strength of the mark' because, in his research, Beebe found that some judges erroneously omitted strength of the mark analyses from their likelihood of confusion analyses, see Beebe, *supra* note 100, at 1633–34, and that some judges, rather than categorize a mark's distinctiveness,

area as "trademark law" and selected the time period extending from January 1, 2002 to December 31, 2022. I chose not to start the period with the year 1977, the year following the *Abercrombie* decision, because from 1977 to 2001, the Supreme Court issued several rulings that greatly affected the landscape of the trademark-distinctiveness regime. In 1992, in *Two Pesos, Inc. v. Taco Cabana, Inc.*, Justice White argued that secondary meaning should not be analyzed in a trade dress case because secondary meaning incentivizes competitors of the originator of a trade dress to "appropriate the originator's trade dress in other markets prior to the establishment of the secondary meaning and to deter the originator from expanding into and competing in these areas."[146] In 1995, in *Qualitex v. Jacobson Products*, the Court again addressed secondary meaning.[147] In *Qualitex*, Justice Breyer reasoned that colors cannot be inherently distinctive.[148] However, colors could constitute descriptive trademarks because they could take on secondary meaning over time in the course of use in the marketplace.[149] Finally, in 2000, in *Wal-Mart Stores, Inc. v. Samara Brothers, Inc.*, Justice Scalia made a similar declaration, holding that, although color itself is not inherently distinctive, it could be inherently distinctive if the color is part of a product's packaging whose main function is to identify the product's source.[150] By contrast, if the color and the words are part of a product design, they are not inherently distinctive because consumers "are aware of the reality that the feature is intended not to identify the source."[151]

---

would simply cite an *Abercrombie* case. Beebe, *supra* note 96, at 1635. These findings suggest to me that analyses of trademark confusion cases should not ignore the strength-of-mark factor and that analyses of mark distinctiveness should take into consideration the *Abercrombie* taxonomy.

[146] Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 777 (1992). We should note that the *Two Pesos* case dealt with the function of secondary meaning, especially its effect on competitors. *Id.* at 765. An analysis of a descriptive word mark under the *Abercrombie* taxonomy, which requires that trademark owners prove the existence of secondary meaning in their mark, might do well to consider Justice White's opinion when the analysis turns to the effects that secondary meaning can have on competitors.

[147] *See* Qualitex Co. v. Jacobson Prods. Co., Inc., 514 U.S. 159, 164 (1995).

[148] *Id.* at 172.

[149] *Id.*

[150] Wal-Mart Store, Inc. v. Samara Brothers, Inc., 529 U.S. 205, 211 (2000).

[151] *Id.* at 208. As in *Two Pesos*, these issues surrounding secondary meaning and color might have a great bearing on how we analyze secondary meaning in relation to word-mark distinctiveness. In the *Wal-Mart Stores* case, the Court's analysis of inherent distinctiveness led them to separate the concept of product packaging from the concept of product design—an analytical step that might hold promise for analyses of word-mark distinctiveness. After all, word marks can be part of product packaging or product design—a distinction that, though nuanced, can result in varying levels of distinctiveness. *See id.* at 208.

The above Supreme Court cases reveal how the years following 1977 brought with them major changes to the American judiciary's conception and treatment of trademark distinctiveness. To avoid a situation in which those changes hopelessly complicate my analysis of word-mark distinctiveness, I very deliberately made sure that the present study's data would not derive from the period covering those cases.

In terms of jurisdiction, U.S. trademark applicants usually have two choices to register their trademarks. The first choice, as outlined in the Lanham Act, is to register a trademark as a federal trademark.[152] Another choice is to register a trademark as a state trademark in individual state trademark offices.[153] Unfortunately for researchers like me, it is extremely difficult to collect state-registered trademarks from across all the state governments because there is no central database containing this information. Thus, although some trademark cases can be found on, for example, LexisNexis, I decided to side-step this complicating issue entirely by collecting only federal cases.

In total, I collected 1,212 cases. These cases have two main characteristics. First, the information on the cases is a mix of textual data and numerical data. To determine whether or not federal judges exhibited patterns of careless reasoning with respect to categorizations of distinctiveness after the *Wal-Mart Stores* case, I realized that I would need more information than would simply appear in a conventional filing of a federal court case. Thus, I sought out a broader array of sources for data related to word-mark distinctiveness. Moreover, as Beebe discovered in his empirical study on trademark-confusion cases, some federal judges, rather than analyze the strength-of-mark factor and the distinctiveness-of-mark factor, would simply cite *Abercrombie* in the context of the given case.[154] Therefore, I was well aware that I would have to filter out such cases from the 1,212 I had initially collected.

Regarding the two independent variables of "dominant judge gender" and "average judge tenure," I had to hand-code this information by performing Google searches. Fortunately, the career information about every federal judge can be

---

[152] 15 U.S.C. § 1051 (2012).

[153] *See State Trademark Information Links*, U.S. Pat. & Trademark Off., https://www.uspto.gov/trademarks/basics/state-trademark-information-links [https://perma.cc/7KDH-CSV9] (last visited Feb. 23, 2025).

[154] Beebe, *supra* note 96, at 1635.

accessed at Ballotpedia, a widely consulted digital encyclopedia of American politics.[155] For cases decided by one judge, I quite simply coded gender as 'M' for male judges and 'F' for female judges. Likewise in these cases, I calculated these judges' tenure simply by calculating the number of years that would have passed between a given judge's confirmation by Congress and the year of a given case's decision.

Consider, for instance, *Phat Fashions v. Phat Game Athletic Apparel, Inc.*[156] The case was decided by Judge Lawrence K. Karlton in the United States District Court for the Eastern District of California in 2002.[157] Judge Karlton had been confirmed to serve as a federal judge in 1979,[158] so I calculated the presiding years by subtracting 1979 from 2002 and arrived at the desired answer: 23 years. The gender of Judge Karlton was male,[159] so I coded it 'M'. United States District Court for the Eastern District of California is a part of the Ninth Circuit, so I coded the jurisdiction '9'.

Matters grew a little more complicated for cases decided by three federal judges. I would code the gender of the judges 'M' ('F') if at least two of the judges were male (female). To calculate the tenure of the three judges, I calculated the average tenure of all three individuals. For instance, in *Entrepreneur Media v. Smith*,[160] three circuit court judges decided the case: Judge Betty B. Fletcher,[161] Judge Thomas G. Nelson,[162] and Judge Marsha S. Berzon.[163] Two of the three were female, so I coded their gender 'F'. Average tenure was the sum of the three judges'

---

[155] *See* Ballotpedia, https://ballotpedia.org/Main_Page [https://perma.cc/HU43-ZJBK] (last visited Feb. 24, 2025).

[156] *See generally* Phat Fashions, L.L.C. v. Phat Game Athletic Apparel, Inc., No. 01C1771, 2002 U.S. Dist. LEXIS 15734 (E.D. Cal. Mar. 20, 2002). This case was included in the present study's dataset.

[157] *Id.* at *1.

[158] *Judge Lawrence K. Karlton*, Ballotpedia, https://ballotpedia.org/Lawrence_Karlton [https://perma.cc/L8PP-ZTFW] (last visited Feb. 24, 2025).

[159] *Id.*

[160] Entrepreneur Media v. Smith, 279 F.3d 1135 (9th Cir. 2002).

[161] *Judge Betty B. Fletcher*, Ballotpedia, https://ballotpedia.org/Betty_Binns_Fletcher [https://perma.cc/T4H8-4NW9] (last visited Feb. 24, 2025).

[162] *Judge Thomas G. Nelson*, Ballotpedia, https://ballotpedia.org/Thomas_G._Nelson_(Federal_judge) [https://perma.cc/2LXA-Z9LL] (last visited Feb. 24, 2025).

[163] *Judge Marsha S. Berzon*, Ballotpedia, https://ballotpedia.org/Marsha_Berzon [https://perma.cc/7R3B-G58H] (last visited Feb. 24, 2025).

tenure (calculated according to the simple arithmetic formula above) divided by the total number of judges (three): in this case, average tenure was 13 years.[164]

2.2. Information About the Linguistic Characteristics of Words

To code the linguistic characteristics of the disputed word marks in federal cases, I followed three steps: (1) investigate whether the alleged word mark consists of a single word or multiple words (because word-mark owners can register either a single word or multiple words as a trademark, I had to treat them differently when coding); (2) for disputed single-word marks, investigate whether the word mark can be found in dictionaries—if yes, code the word mark as a dictionary word, but if no, observe which types of word formation (e.g., acronyms, blending) most accurately reflect the word mark;[165] (3) if no word formation satisfactorily reflects the word mark, code it as a coined word. As for word marks consisting of multiple words, perform step one and, for the second step, observe whether one of the multiple words in a single word mark might fall under a particular word-formation category: if yes, code the word mark "compound + type of formation"; if no, code the word mark only as "compound." The following flow chart summarizes the coding processes for word formation:

---

[164] Until 2002, Judge Betty B. Fletcher's presiding years were 23 years; Judge Thomas G. Nelson's presiding years were 12 and Judge Marsha S. Berzon were 2 years. Therefore, (23 + 12 + 2)/ 3 = 12.3 years. For ease of coding, I rounded numbers to the nearest integer (e.g., 12).

[165] For details about word-formation categories (namely, inflection, derivation, cliticization, suppletion, compounding, conversion, blending, clipping, and acronyms), see WILLIAM O'GRADY & JOHN ARCHIBALD, CONTEMPORARY LINGUISTIC ANALYSIS: AN INTRODUCTION (Pearson Ed. Can., 8th ed. 2015).
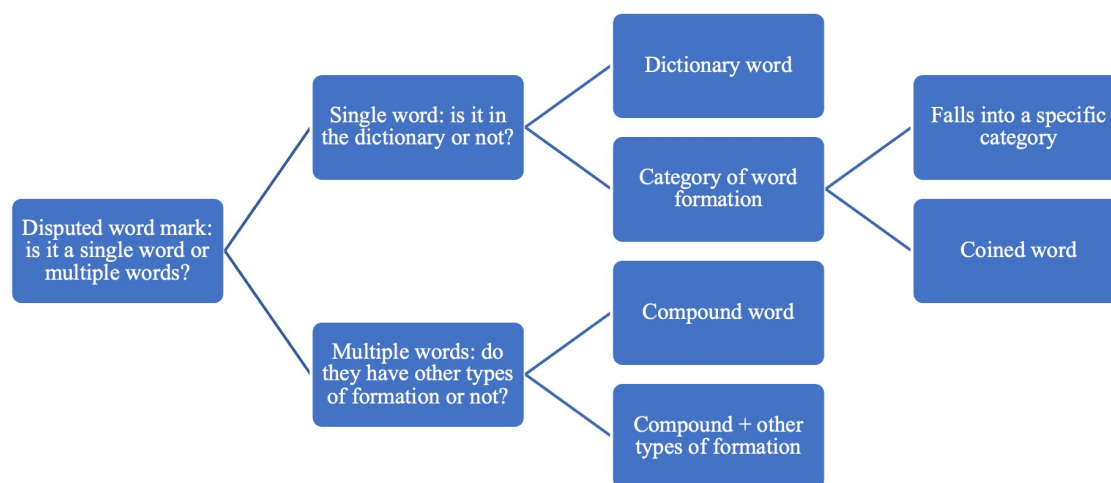
I was compelled to add two additional categories of word formation—dictionary word and coined word—because of the popularity of the dictionary test among federal judges. The addition of these two categories enabled me to observe whether or not a word mark's status as a dictionary word would have a bearing on judges' categorization of the word mark's distinctiveness. To determine a word mark's dictionary status, I consulted three distinct online dictionaries: Merriam-Webster, The Dictionary of American Family Names, and A Dictionary of Geography.[166] If just one of the three dictionaries featured the word mark, I coded it, with respect to formation, as a dictionary word. Second, because a disputed word mark might fall into more than one category, I would account for all the categories during my coding of the mark. For instance, in *New York City Triathlon, LLC v. NYC Triathlon Club, Inc.*,[167] the disputed word mark was NYC TRIATHLON. The formation of this word mark happened to fall into two categories: acronyms and compound words. To thoroughly understand which type of word formation would be important for categorization, I coded

---

[166] *See* MERRIAM-WEBSTER, https://www.merriam-webster.com/ [https://perma.cc/5RSB-YGA4] (last visited Feb. 27, 2025); *Dictionary of American Family Names*, OXFORD REFERENCE, https://www.oxfordreference.com/display/10.1093/acref/9780195081374.001.0001/acref-9780195081374 [https://perma.cc/W92P-SCKD] (last visited Mar. 21, 2025); *A Dictionary of Geography*, OXFORD REFERENCE, https://www.oxfordreference.com/display/10.1093/acref/9780199680856.001.0001/acref-9780199680856 [https://perma.cc/7J3G-RBVA] (last visited Mar. 21, 2025).

[167] N.Y.C. Triathlon, LLC v. NYC Triathlon Club, Inc., 704 F. Supp. 2d 305, 311 (S.D.N.Y. 2010).

the mark "acronyms+compounds" in the dataset. A point worth noting is that TRIATHLON has an entry in any standard English-language dictionary. Thus, the code "acronyms+compounds" fails to account for the dictionary status of NYC TRIATHLON. Thus, in the given case, I assigned the code 'Y' to NYC TRIATHLON.

Finally, in terms of a disputed word mark's word count, syllable count, vowel count, consonant count, and plosive status, I hand-coded all this information on the basis of personal observation. An interesting point to address is that some disputed word marks might include punctuation (e.g., an exclamation mark) or other symbols. For the purposes of the present study, I did not code for these symbols, even though they should not be regarded as irrelevant to the topic of word-mark distinction. For instance, in *Women, Action & the Media Corp. v. Women in the Arts & Media Coalition, Inc.*,[168] the disputed word mark was 'WAM!' Excluding the exclamation mark, 'WAM' is a single non-plosive (NP) word with one syllable, one vowel, and two consonants. I coded the mark '1' for word count, syllable count, and vowel count and '2' for consonant count.

2.3. TSS Information and Estimating Both Duration and Third-party Use

The TSS system has abundant trademark information about disputed word marks. For coding purposes, I would first and foremost locate the "earliest-use" information about a word mark litigated in a federal case. Three steps guided me in this process. In the first step, I sought to identify the "true" owner of the disputed word mark. Logically, a plaintiff may sue a defendant on the grounds of likelihood of confusion if the plaintiff owned the registered word mark prior to the defendant's alleged use of the mark; in turn, the plaintiff may be countersued by the defendant for trademark infringement because the defendant had registered similar marks prior to the plaintiff's use of the given word mark, thus presenting a situation in which the plaintiff's mark lacked distinctiveness.[169] In this scenario,

---

[168] *See* Women, Action & the Media Corp. v. Women in the Arts & Media Coal., Inc., No. CIV.A. 13-10089-RWZ, 2013 WL 3728414, at *1 (D. Mass. July 12, 2013).

[169] *See, e.g.*, Eurotech Inc. v. Cosmos Eur. Travels, 213 F. Supp. 2d 612, 622 (E.D. Va. 2002). At the center of the case was a dispute over the ownership of the domain name COSMOS.COM. The plaintiffs, including the current owner of the disputed domain name, sought a court declaration confirming their propriety rights with respect to the use and ownership of the domain name. The defendant—the owner of the registered trademark COSMOS—filed counterclaims against the plaintiffs for trademark infringement and unfair competition in violation of the Lanham Act.

the "true" owner of the disputed word mark could be the plaintiff or defendant. Such complexity in federal cases required that I hand-code relevant data.

The second step in locating the "earliest-use" information would be to perform a keyword search of the TSS, with the keyword being the name of the mark's true owner. These searches were quite time-consuming because the true owner of a mark might have registered it several times in more than one year. To complicate matters even further, a previously registered word mark might have subsequently had its registration cancelled or invalidated. Thus, in perusing the TSS database, I had to keep an eye open not just for currently registered marks but for all possible marks, including live ones and dead ones. Only in this way was I able to obtain accurate information about the earliest use of disputed word marks.

For the third and final step, once I identified a disputed mark's true owner and obtained the "earliest-use" information, I coded the information as it pertained to international trademark classes (ICs), US trademark classes, product-and-service classes, and first-year use. One detail that merits our attention with regard to first-year use is that, in some instances, the TSS may register a disputed word mark yet not record the year of the mark's first use. To deal with this matter, I would replace the missing "first year of commercial word-mark use" information with the "filing year" information, which thus served as a proxy for the missing information. If the TSS database contained information about a word-mark owner's priority year (i.e., period of priority), I would use this information as a proxy for the "filing year" information because the priority year discloses a more accurate timing of the use for a word mark. Finally, having obtained the necessary information pertaining to the first year of commercial use and the federal-case year, I was in a position to estimate the duration of the disputed word mark.

To better understand the coding processes discussed above, consider again the example of WAM!. When studying the WAM! legal case, I found that the true owner of the disputed mark was Women Action & the Media, the plaintiff in the case.[170] A search of the TSS revealed that WAM! was registered for International Class 35 and U.S. Classes 100, 101, and 102.[171] Thus, I coded the classes categorically and recorded the number of classes corresponding to WAM!

---

[170] *See Women, Action & the Media Corp.*, 2013 WL 3728414, at *10.

[171] WAM!, Registration No. 4,275,416.

(1 for international classes and 3 for US classes). Moreover, because International Class 035 falls under the service category,[172] I coded the class 'S' in my dataset. Next, given that the first year of commercial use was June 1, 2004, I used the code '2004' to estimate the duration of the disputed mark.[173] Finally, given that the WAM! case was decided in 2013, I coded the duration of the disputed word mark as '9' (2013 minus 2004).[174] The following flow chart summarizes the coding process that I followed when estimating length of use (i.e., 'duration') on the basis of TSS-registered information.

Find the true owner → Find the earliest-use information → Code the information and estimate the duration
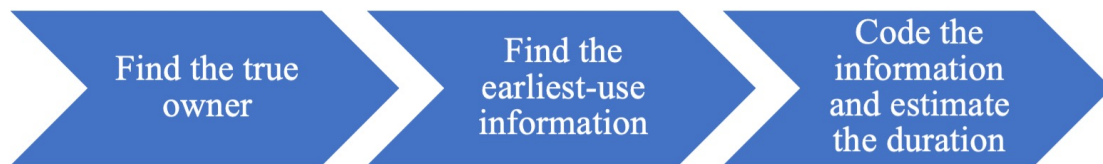
One nuance of the coding process for estimations of length is linked to the missing registration values in the TSS. Some federal cases that I collected for this study had opinions about the distinctiveness of the disputed word mark, yet, in these cases, the true owner had failed to register the mark prior to the opinion in the TSS. Thus, for these cases, I would encounter missing values for four key variables: IC count, US class count, first-year use, and duration would be the missing values. To deal with this situation, I capitalized on an advantage of decision-tree algorithms—their ability to deal with missing values through a deft

---

[172] The categorization of goods and services marks can be found in USPTO's website. *See Goods and Services*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/trademarks/basics/goods-and-services [https://perma.cc/5A2T-82W6] (last visited Mar. 30, 2025).

[173] *See* WAM!, Registration No. 4,275,416.

[174] *See generally Women, Action & the Media Corp.*, 2013 WL 3728414.

use of surrogate splits.[175] Thus, for disputed word marks not registered in the TSS, I coded the missing values as 'N/A' (i.e., not available).

The most difficult part of these various coding processes was the task of coding for third-party registration of word marks that were similar to a disputed word mark litigated in a federal court. I could not find the exact and correct number of third-party registrations from the TSS. The main reason for this limitation is that the TSS does not allow users to select a specific year for word-mark searches. Further complicating this matter is the fact that the TSS keeps updating information for each disputed word mark. Thus, it could easily come to pass that a valid and registered word mark today becomes a cancelled or abandoned word mark tomorrow, and *vice versa*. As a result, it is impossible for the present study and for similarly structured studies to obtain correct and stable numbers related to the third-party registration of word marks that are similar to disputed word marks. For my part, I was able only to make "rough" estimations about these third-party registrations, and I did so by assuming that marginal daily changes in "live" use and "dead" use for word marks were small—that is, relatively stable. I based this assumption on previous findings that these daily changes tend to be minor.[176]

Using the above assumptions, I followed three steps to collect information about the third-party registration of word marks. In the first step, I would search the TSS by selecting the "owner" search category and entering the name of the true owner into the search field. The term 'true owner' refers to any entity, usually a company, that was directly or indirectly related to one of the studied federal cases and that was determined, by a federal judge, to be the rightful ('true') owner of a previously disputed word mark. In response to my search-engine query, the TSS would present information about historical trademark data corresponding to the true owner (e.g., sometimes the current true owner was not the original true owner, owing perhaps to a bankruptcy, a merger, and so on). Using this information, I could

---

[175] Nicholas J. Tierney et al., *Using Decision Trees To Understand Structure in Missing Data*, 5(6) BMJ OPEN 1, 3–4 (2015) (explaining how to address missing values for variables that are required for a split by using surrogate splits, which rest on alternative variables whose splitting property is similar to that of the missing-value variables).

[176] The assumption that the marginal daily changes of word-mark registration is minor could be inferred from Beebe's research about the word-mark depletion. *See* Beebe & Fromer, *supra* note 139, at 978 (explaining the assessment of word mark depletion by addressing the difficulty that the depletion does not necessarily entail a decline in the number of potential marks that remain available for registration because an entity may register a mark that has already been claimed by another).

count the number of disputed word marks that belonged to the true owner. In the second step, I would search the TSS again, this time by selecting the "wordmark" search category and entering the given true owner's disputed word mark itself into the search field. In response to this second query, the TSS would present all the word marks—whether live or dead, and whether belonging to the "true owner" or a "third-party"—that were identical to or contained the searched-for disputed word mark. Because I assumed that the daily marginal changes in live and dead uses were relatively stable (see above), I omitted the "dead word mark" count from the total number of the live word marks. In the third step, I would estimate the number of third-party registered word marks that contained the disputed word mark. Because the TSS furnished me with (1) the precise number of disputed word marks belonging to a given true owner (a number that was not always 'one') and (2) the precise number of live word marks containing the disputed word mark but belonging to third-parties or to the given true owner, I could estimate the number of purely third-party registrations by subtracting **the total number of search results involving those disputed word marks belonging to the given true owner** from **the total number of search results involving the disputed word mark generally.** The following flow chart visually summarizes the above coding steps for the estimation of the third-party registrations of each disputed word mark.

Count the number of disputed word marks belonging to the true owner → Count the total number of related live word marks → Estimate the number of third-party uses
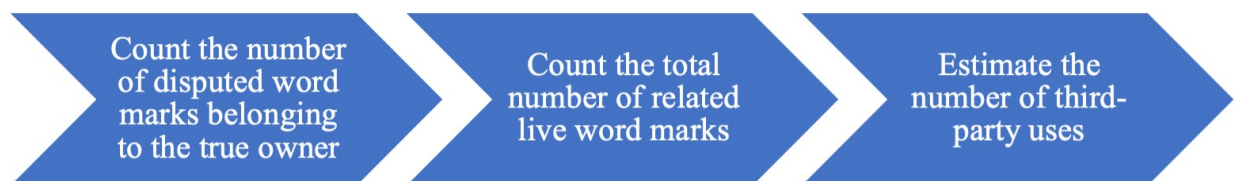
3. Description of the Data and Preliminary Observations

Using the various data-collection and hand-coding processes discussed above, I ended up with 713 valid federal court cases with which to train the decision trees in this study. Before the training could commence, I needed to acquire a birds-eye

view of both the dependent-variable descriptive data and the independent-variable descriptive data.

3.1.   Descriptive Data for the Dependent Variable

Table 3 presents the results pertaining to federal judges' interpretation of disputed word-marks' distinctiveness. I analyzed 713 valid federal cases concerning disputed word marks: in 279, the marks were found to be suggestive (S); in 141, the marks were found to be descriptive with acquired distinctiveness (DA); in 140, the marks were found to be purely descriptive (D); in 135, the marks were found to be arbitrary (A); and in 18, the marks were found to be generic (G). As a percentage, the most common type of distinctiveness in court judgements was suggestive distinctiveness (39%), whereas the least common type of distinctiveness was, quite predictably, generic distinctiveness (2.5%). The predictability of the latter result rests on the simple fact that word-mark owners suing another owner over its word mark would clearly avoid characterizing the disputed word mark as generic. After all, a generic word mark—that is, a word marks that is least likely to be distinctive—has little to no chance of being successfully registered in the TSS under the scrutiny of the USPTO.

Distinctiveness (Validity)		Numbers of Counts	Percentage	Valid Percentage	Accumulated Percentage
Valid	A	135	18.9	18.9	18.9
	D	140	19.6	19.6	38.6
	DA	141	19.8	19.8	58.3
	G	18	2.5	2.5	60.9
	S	279	39.1	39.1	100.0
	Total	713	100.0	100.0	

Table 3: Descriptive data for the dependent variable

3.2.   Descriptive Data for the Independent Variables

In this study's dataset involving independent variables, some values are continuous while others are categorical. Table 4 presents the minimum and maximum values, the averages, and the standard deviations for the variables. Tables 5 through 13 present the categorial data.

3.2.1 Continuous independent variables

Table 4 below presents eleven variables that are coded as having continuous values. Information about the eleven variables came from the following sources: average judge tenure came from various sources covering federal cases; linguistic information pertaining to words, syllables, vowels, consonants, and frequency came from COCA; numerical data pertaining to ICs and US classes, word-mark duration, and third-party registrations of disputed word marks came from the TSS.

As for judge tenure, the statistics reveal that, at the time the judges rendered their decision in a word-mark case, they had accumulated close to thirteen (13) years of experience, with a standard deviation of almost nine (9) years. In other words, most federal judges with roughly 13 years of experience will have had at least one opportunity to preside over a case concerning word-mark distinctiveness.

As for the linguistic elements studied herein (words, syllables, vowels, consonants and frequencies from the COCA database), the statistics reveal several interesting points. First, the average disputed word mark in this study's sample consists of about 2 words, 3 syllables, 3 vowels, and 6 consonants. From these results, we can infer that the true owners of the disputed word marks in our sample were inclined to use short words, perhaps because short words are generally more memorable than long words. Second, the COCA frequencies for the disputed word marks varied substantially because the standard deviation was so high (1,886,913). Combined with the previously discussed findings, the high standard deviation for the COCA frequencies indicates that although most disputed word marks had similar characteristics (they were short and memorable), these similarities in no way translated into similar frequencies of mass-media use.

Finally, as for the information obtained from TSS regarding the 713 federal cases, we acquired 644 valid pieces of TSS-registration data, as 69 disputed word marks were not registered. The valid TSS registrations had an average of one (1) IC and three (3) US classes. Word-mark duration in the sample was, on average, about nineteen (19) years, with a standard deviation of twenty-one (21) years. From these results, we can infer that quite a few of the disputed word marks have an incontestable degree—or at least a high degree—of distinctiveness.[177] This inference is consistent with the statistical results of my dependent-variable analysis,

---

[177] For more details about the Declaration of Incontestability of a Mark, see 15 U.S.C. § 1065 (2015).

which show that very few disputed word marks that judges found to be distinctive were categorized by those judges as generic. As for the number of third-party registrations of word marks that are identical or similar to a disputed word mark, my statistical analysis reveals a large standard deviation (681). In other words, there were huge differences in the numbers of third-party TSS registrations of word marks (on the low end, there were 0 third-party registrations, and on the high end, 13,500 third-party registrations). This spread is similar to the one characterizing the COCA frequencies. The high standard deviations and high maximum values prove that many word-mark owners have found it almost impossible to protect their word marks from use by third parties.

	N	Minimum value	Maximum value	Average	Standard deviation
Judge Tenure	713	1.0	44.0	13.125	9.0372
Word Count	713	0	8	1.81	1.060
Syllable Count	713	0	16	3.33	2.152
Vowel Count	713	0	19	3.87	2.569
Consonant Count	713	0	30	6.64	3.912
COCA Frequency	713	0	50360812	76926.92	1886913.089
IC Count	644	1	11	1.39	1.029
US Class Count	644	1	21	3.97	3.136
Duration	644	0	185	18.92	20.854
Third-Party Registration Count	713	0	13500	126.28	681.261
Valid N (listwise)	644				

Table 4: Independent variables with continuous values

3.2.2 Categorical independent variables

Some of this study's independent variables took the form of categorical data pertaining to three basic areas: the federal cases themselves, linguistics, and TSS-based information. For the cases, I identified the year a decision was rendered in a case, the jurisdiction in which the case was held, and the gender of most of the federal judges presiding over the cases. For linguistic topics, I identified the word-formation categories of each disputed word mark, whether or not the word mark began with a plosive, whether or not some or all of the words in a word mark could be found in dictionaries, and the first year of commercial use for the word mark. Below, I discuss each of these topics in greater detail.

3.2.3 Categorical independent variables for federal cases

As shown in Chart 1 and Table 5 below, the year with the greatest number of federal trademark-distinction cases (54) was 2021 and the year with the lowest number (21) was 2009. From 2002 to 2022, the average annual number of court decisions was 33. Most (52%) of the court decisions in these distinctiveness cases were issued by the U.S. Courts of Appeals for the Second Circuit and for the Ninth Circuit. The remaining 48% of the decisions were issued in other circuits. Because the whole sample was skewed toward the Second and the Ninth Circuits, the decision tree that I trained with this sample might reflect the subjective tendencies of judges from these two circuits with respect to their categorization of word-mark distinctiveness.
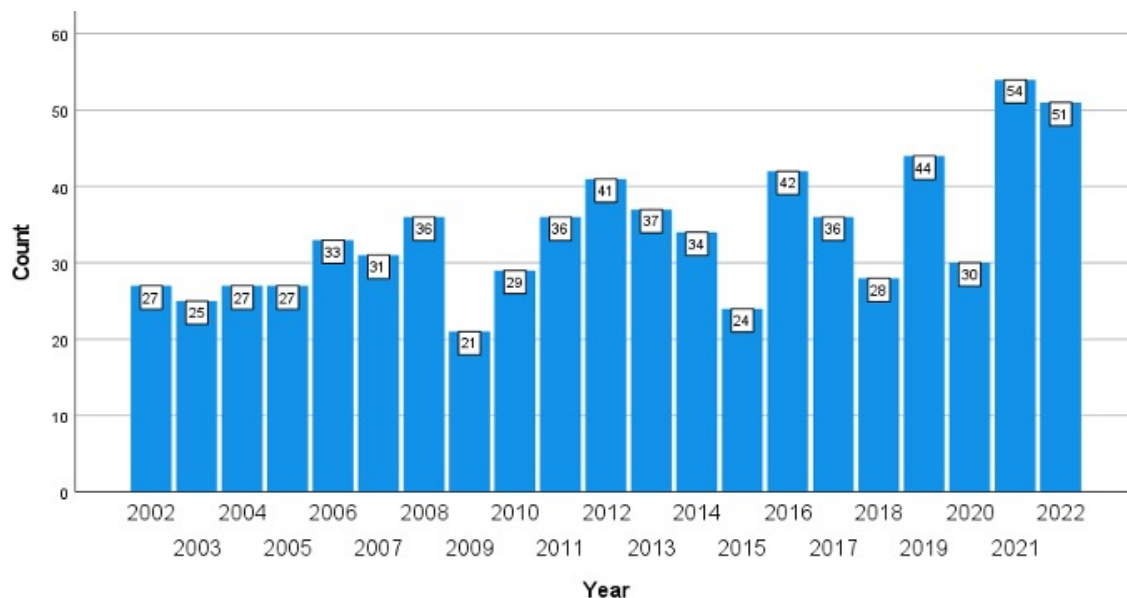
Chart 1: Word-mark distinction cases from January 1, 2002 to December 31, 2022

Circuit Controls				
	Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid 0	4	0.6	0.6	0.6
1	17	2.4	2.4	2.9
2	122	17.1	17.1	20.1
3	45	6.3	6.3	26.4
4	44	6.2	6.2	32.5
5	34	4.8	4.8	37.3
6	60	8.4	8.4	45.7
7	28	3.9	3.9	49.6
8	20	2.8	2.8	52.5
9	249	34.9	34.9	87.4
10	15	2.1	2.1	89.5
11	75	10.5	10.5	100.0
Total	713	100.0	100.0	

Table 5: Distribution of jurisdictions (Note: '0' refers to the United States Court of Appeals for the Federal Circuit)

Regarding gender, about 70% of the judges (judge trios) were male (dominant male). Although the limited data prevented me from ascertaining with any certainty whether or not gender was significantly associated with the court decisions regarding word-mark distinctiveness decisions, the topic of whether or not—and if so, in what ways and to what extent—gender shaped and continues to shape distinctiveness rulings should be of interest to legal scholars.

		Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid	F	210	29.5	29.5	29.5
	M	503	70.5	70.5	100.0
	Total	713	100.0	100.0	

Table 6: Gender (dominant gender) of federal judges (federal-judge trios)

3.2.4. Categorical independent variables for linguistics

Table 7 presents key results from my analysis of categorical independent variables for linguistics. As we can see, 74.9% of the disputed word marks in this study's sample feature dictionary words; put another way, only 25.1% of the disputed word marks were purely coined terms. Because, as I noted earlier, it is reasonable to assume that federal judges heavily rely on the dictionary test to

analyze word-mark distinctiveness, an intriguing path of analysis is to investigate whether there is a significantly positive relationship between a disputed word mark's dictionary roots and federal judges' willingness to categorize the word mark as distinctive.

		Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid	N	179	25.1	25.1	25.1
	Y	534	74.9	74.9	100.0
	Total	713	100.0	100.0	

Table 7: The dictionary status of disputed word marks

Table 8 reveals that 14 categories of word formation emerged from the study's sample. However, the disputed word marks were not evenly distributed across these categories. In descending order, the top five categories of word formation for the disputed word marks were compound words (400), dictionary words (138), coined words (62), blend words (42), and acronyms (33). The top two categories—compound words and dictionary words—accounted for a whopping three-fourths of the data in the sample, a fact that might have substantially skewed the decision-tree training process. Compound words were, by far, the major formation because each word mark offered more than one opportunity for a variation. For instance, one disputed word mark in the sample was 'THERMA-SCAN'. This word mark consists of a coined word, 'THERMA' (which, despite its being coined, is not particularly unique), and a very common dictionary word, 'SCAN'. Thus, one could reasonably expect that many variations of this compound word are possible. Moreover, the 'THERMA-SCAN' example and Table 8 suggest that many compound-word marks consist of at least one dictionary word: 19.4% of the disputed word marks fall under the dictionary-word word-formation category, but this percentage grows to 75.5% if we combine the dictionary category with the compound-word category (56.1%). A topic worthy of investigation is whether or not federal judges tend to hold that word-formation categories, which are part and parcel of the dictionary test, determine the specific type of distinctiveness that corresponds to a disputed word mark.

		Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid	Acronyms	33	4.6	4.6	4.6
	Acronyms+Compounds	1	0.1	0.1	4.8
	Blend Words	42	5.9	5.9	10.7
	Clipped Words	7	1.0	1.0	11.6
	Clipped Words+Compounds	21	2.9	2.9	14.6
	Clitics	2	0.3	0.3	14.9
	Clitics+Compounds	2	0.3	0.3	15.1
	Coined Words	62	8.7	8.7	23.8
	Compounds	400	56.1	56.1	79.9
	Converted Words	1	0.1	0.1	80.1
	Derived Words	2	0.3	0.3	80.4
	Dictionary Words	138	19.4	19.4	99.7
	Inflected Words	1	0.1	0.1	99.9
	Inflected+Compounds	1	0.1	0.1	100.0
	Total	713	100.0	100.0	

Table 8: Word-formation categories for disputed word marks

Finally, I investigated how many of the disputed word marks in the sample began with a plosive. As noted earlier, the literature strongly suggests that, for consumers, plosive words are much more memorable than non-plosive words, a finding that could have a significant bearing on the distinctiveness level of a word mark.[178] Table 9 presents my findings regarding plosives: 66.5% of the disputed word marks did not begin with a plosive. Thus, when viewed from the opposite angle, the findings suggest that only 33.5% of these word marks possessed this sound-based mechanism capable of enhancing a word mark's ability to be memorable. A topic meriting further inquiry is whether federal judges might, as consumers do, pay attention to the sounds of uttered words. If judges take sound into consideration, plosives and similar mechanisms might influence the judges' assessment of word-mark distinctiveness.

		Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid	NP	474	66.5	66.5	66.5
	P	239	33.5	33.5	100.0
	Total	713	100.0	100.0	

Table 9: Plosive word marks in the sample

---

[178] *See* Vanden Bergh et al., *supra* note 136, at 837.

3.2.5 Categorical independent variables for TSS data

Table 10 sheds light on the classes of TSS-registered disputed word marks: 47% were registered as products, 36.6% were registered as services, and 6.7% were registered as both products and services. Of all the disputed word marks, 9.7% had no registration status in the TSS.

		Numbers of counts	Percentage	Valid percentage	Accumulated percentage
Valid	P	335	47.0	47.0	47.0
	P&S	48	6.7	6.7	53.7
	S	261	36.6	36.6	90.3
	Total	644	90.3	90.3	
Missing Value	N/A	69	9.7	9.7	
Total		713	100.0	100.0	

Table 10: The IC status of disputed word marks

Finally, I treated the size classification of third-party registrations in the TSS (i.e., "categorization of third-party registration count") as a polytomous variable, which I established on the basis of the number of third-party registrations (i.e., "third-party registration count"). Chart 2 breaks down the statistical distribution of the polytomous variable across the four categories (i.e., few, medium, large, super large): 495 disputed word marks (69.4%) were classified as few, 98 (13.7%) as super large, 94 (13.2%) as medium, and 26 (3.7%) as large. These statistical results for the categorization of third-party registration count point to an intriguing question: why is it that almost seventy percent of the disputed word marks in this study's sample correspond to only a few third-party registrations in the TSS even though federal judges varied significantly in their categorization of the disputed word marks' distinctiveness?
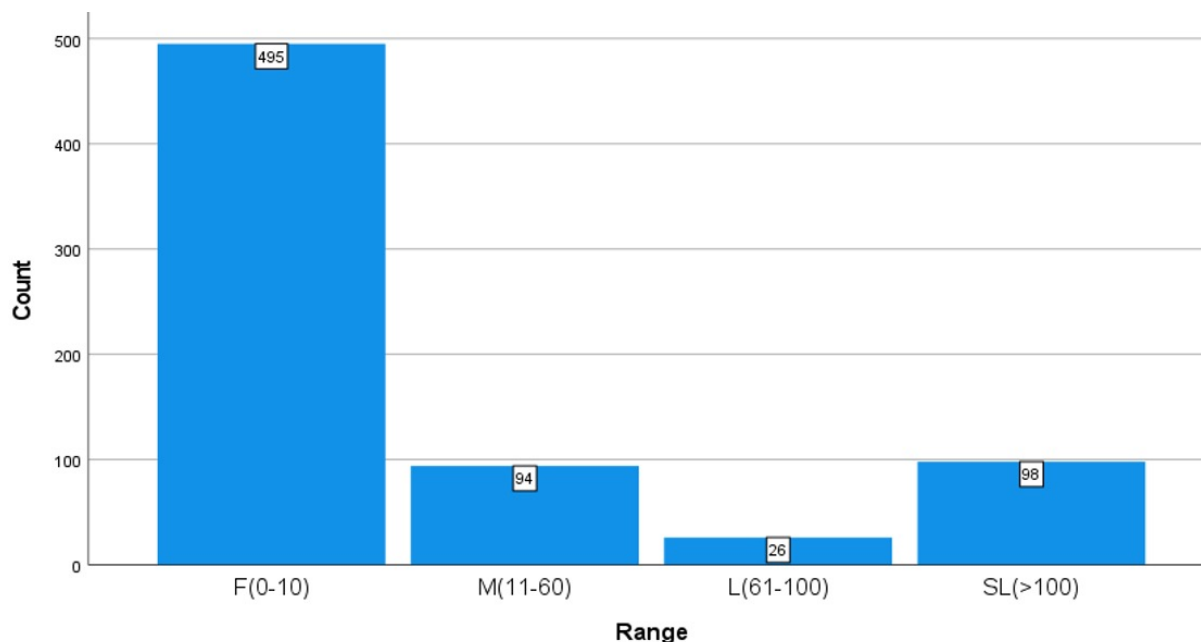
Chart 2: The categorization of third-party registration count for disputed word marks

3.2.6 Summary of observations about both dependent and independent variables

Overall, the descriptive statistics concerning the dependent and independent variables reveal some noteworthy patterns. First, all of the variables exhibit an uneven distribution of data in most instances. As for the dependent variable (i.e., judges' categorization of word-mark distinctiveness), only 2.5% of the disputed word marks were judged to be generic whereas 39.1% were categorized as suggestive marks. As for all of the independent variables (e.g., duration, third-party registration count, COCA frequency), statistical analyses of the data reveal the existence of high standard deviations (20 for duration, 681 for third-party registration count, and 1,886,913 for COCA frequency). Moreover, other independent variables (e.g., jurisdiction, judge gender, word-formation category, plosive status, first-year use) were significantly skewed in the direction of one or a few specific categories. A second noteworthy pattern is that the words in the disputed word marks were quite similar to one another linguistically. The evidence for this finding stems from the comparatively small standard deviations for the continuous values corresponding to word count (1.06), syllable count (2.152), vowel count (2.569), and consonant count (3.912). Finally, the results for the TSS data reveal that most of the disputed word marks, despite having diversely

categorized word-mark distinctiveness in judicial rulings, had small numbers of third-party registrations (fewer than 10).

IV
DECISION-TREE ANALYSIS

Having described the present study's dataset in Part III, I now turn my attention to analyzing the three decision trees that I trained with the data above. As noted, a central objective in this study is to determine whether or not—and if so, in what ways—federal judges neglected certain types of distinctiveness in favor of three privileged tests (i.e., the imagination test, the competitive-need test, and particularly the dictionary test). The algorithmic powers of decision trees assisted me in uncovering any such patterns. Thus, I set out to compare three time periods with one another, and for this task, I employed three decision trees: Decision Tree 1 (January 1, 2002–December 31, 2022), Decision Tree 2 (January 1, 2002–December 31, 2010), and Decision Tree 3 (January 1, 2011–December 31, 2022). Using various groupings of independent variables, these decision trees shed light on the logic underlying judges' categorization of word-mark distinctiveness. We should keep in mind a few points: first, it is not necessarily the case that the more important a feature is, the higher its node will be on a decision tree; second, differences in categorization criteria can affect decision-tree results.[179] For these two reasons, one can ascertain neither the importance of a feature nor the performance of a decision tree simply by observing the tree. To gain insights into these matters, one must have in hand two important outputs: the charted importance of independent variables and the charted results of trees' categorization of disputed word marks.

The charted importance of independent variables reveals both the amount of weight and the order of importance assignable to independent variables chosen by the algorithm. This information, for the present study, is key to understanding judges' categorization of word-mark distinctiveness. To ascertain the importance of a variable, one can measure the extent to which the removal of a variable triggers

---

[179] *See* Bahzad Taha Jijo & Adnan Mohsin Abdulazeez, *Classification Based on Decision Tree Algorithm for Machine Learning*, 2 J. APPLIED SCI. & TECH. TRENDS 20, 21 (2021) (noting various types of decision-tree algorithms, including the Iterative Dichotomies 3, or ID3, tree and the Classification and Regression Tree, or CART).

a decrease in a tree's ability to mirror the descriptive data drawn from the actual court decisions. Dan Steinberg explains that the importance of a variable

> is based on the sum of the improvements in all nodes in which the variable appears as a splitter (weighted by the fraction of the training data in each node split). Surrogates are also included in the importance calculations, which means that even a variable that never splits a node may be assigned a large importance score.[180]
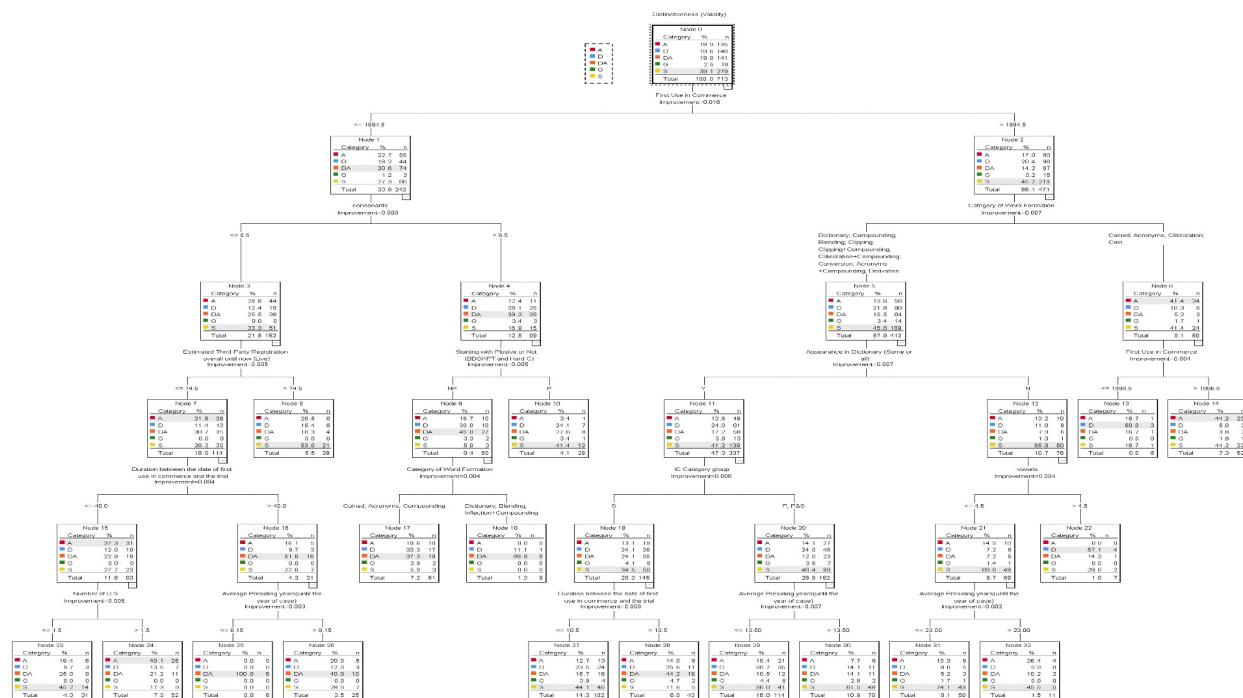
The above explanation helps clarify why one must tease out the differences between a decision tree's independent "splitter" variables and the charted importance of independent variables.

By comparing the results of the decision trees' categorization of word-mark distinctiveness with the judges' corresponding decisions, I focused on the rate at which the decision-tree results mirrored the actual decisions (i.e., the correspondence rate). Once in possession of this information, I could better grasp the extent to which federal judges, in discernable patterns, (1) may have failed to clarify the standards for all types of distinctiveness and (2) may have excessively focused on differences between suggestive and descriptive distinctiveness.

A.   Observations of Decision Tree 1

As noted above, Decision Tree 1 was trained for the overarching period extending from January 1, 2002 to December 31, 2022. Consisting of 5 layers with 32 nodes, the tree yielded several important findings. First of all, it chose 11 of the 19 independent variables for the task of categorizing word-mark distinctiveness. An independent variable—judge tenure—appeared three times in the tree between the fourth and fifth layers. Several other independent variables appeared two times in the tree: word-mark duration appeared between the third and fourth layers and between the fourth and fifth layers, first-year use appeared between the starting point and the first layer and between the second and third layers, and word-formation category appeared between the first and second layers and between the third and fourth layers.

---

[180] Dan Steinberg, *CART: Classification and Regression Trees*, in THE TOP ALGORITHMS IN DATA MINING 179, 190 (2009).

Decision Tree 1: Data trained from January 1, 2002 to December 31, 2022

Second, according to Table 11, Decision Tree 1 did not assign equal importance to all the selected independent variables. In descending order, the most important independent variables, according to Decision Tree 1, are word-mark duration, first-year use, and word-formation category. Interestingly, by comparing the list of independent variables selected by Decision Tree 1 with Table 11, which ranks their importance, we can see that two independent variables appear neither in Decision Tree 1 nor in its importance chart—judge gender and jurisdiction. Their absence suggests that they did not play a key role in judges' categorization of word-mark distinctiveness. Of course, caution should be taken in drawing any firm conclusions, as other factors merit consideration (e.g., the original data were concentrated in the second and ninth circuits).

Furthermore, some independent variables that appeared in Table 11 do not appear in the decision tree. These variables include syllable count, COCA frequency, word count, IC count, and third-party registration count. The absence of these five variables from Decision Tree 1 might entail that the tree delegated their capabilities to "surrogate" independent variables. For instance, the syllable count of a disputed word mark might be identical to the vowel count of the mark, so that the vowel count in Decision Tree 1 functions partly as a substitute for

syllable count. Similarly, IC count might be identical to US class count, since they both serve as expressions of classes of registered word marks in the TSS. The same explanation might apply to COCA frequencies, whose function might be satisfactorily covered by third-party registration count, since both of the variables similarly concern general word usage.

Independent Variable	Importance	Normalized Importance
Duration	0.033	100.0%
First-Year Use	0.029	88.1%
Word-Formation Category	0.023	69.4%
Vowel Count	0.022	65.7%
Consonant Count	0.021	63.2%
Judge Tenure	0.021	62.0%
Categorization of Third-Party Registration Count	0.019	55.8%
COCA Frequency	0.018	53.8%
Syllable Count	0.017	52.4%
US Class Count	0.016	47.1%
Word Count	0.011	33.4%
Dictionary Status	0.010	31.3%
Product or Service	0.008	25.1%
IC Count	0.007	21.7%
Categorization of Third-Party Registration Count	0.006	18.9%
Plosive Status	0.005	15.1%

Table 11: The importance of independent variables based on Decision Tree 1

Finally, it is important to see how Decision Tree 1 categorizes distinctiveness in comparison with how the judges categorized distinctiveness. Table 12 presents two sets of data: federal judges' categorization of distinctiveness as observed and described by me (i.e., "observed categorization" from the descriptive data) and Decision Tree 1's categorization of distinctiveness (i.e., "interpreted categorization"). In presenting these comparative results, Table 12 reveals, in percentage form, the degree to which Decision Tree 1's categorizations mirror the judge's categorizations (i.e., the correspondence rate). First, consider the 135 cases where judges ruled that disputed word marks possessed arbitrary distinctiveness. Decision Tree 1 made only 48 such categorizations, for a correspondence rate of

35.6%. Of the remaining 87 categorizations, 65 involved suggestive distinctiveness, 21 involved descriptive-acquired distinctiveness, and 1 involved purely descriptive distinctiveness. Taken together, these results indicate that Decision Tree 1 did not differentiate arbitrary marks from other marks—especially from suggestive marks—as often as judges did.

Now let us consider the 279 cases where judges attributed suggestive distinctiveness to a disputed word mark. Decision Tree 1 made this same attribution in 229 of the 279 judicial decisions, for a correspondence rate of 82.1%. Of the 50 non-corresponding categorizations by the tree, 32 involved arbitrary distinctiveness, 15 involved descriptive-acquired distinctiveness, and 3 involved descriptive distinctiveness. This second set of Decision Tree 1 results indicates not only that the tree effectively mirrored the judges' categorization of suggestive marks but also that the judges themselves did a good job of accurately identifying the distinctiveness of suggestive word marks.

As for descriptive-acquired distinctiveness, judges ruled that this categorization applied to disputed word marks in 141 federal cases. Decision Tree 1 did so in only 62 of these 141 cases, for a correspondence rate of 44%. Of the non-corresponding predictions, 64 involved suggestive distinctiveness, 13 involved arbitrary distinctiveness, and 2 involved descriptive distinctiveness. Interestingly, these results are similar to Decision Tree 1's categorizations for arbitrary distinctiveness, suggesting that this tree sometimes had difficulty identifying the difference especially between descriptive-acquired distinctiveness and suggestive distinctiveness.

Regarding the fourth category of distinctiveness (i.e., descriptive distinctiveness), let us recall that in 140 federal cases, judges ruled that a disputed word mark possessed this form of distinctiveness. As for Decision Tree 1, it made a corresponding categorization in a mere 7 of these 140 decisions, for a success rate of only 5%. Of the 133 non-corresponding categorizations, 91 involved suggestive distinctiveness, 32 involved descriptive-acquired distinctiveness, and the remaining 10 involved arbitrary distinctiveness. These results obviously indicate that Decision Tree 1 has significant algorithmic difficulties in differentiating descriptive marks from other types of marks. A hypothesis we might reasonably infer from this high degree of non-correspondence is that federal judges may perceive many parallels between descriptive marks and suggestive marks.

Finally, as for generic distinctiveness, Decision Tree 1 mirrored not even one of the 18 judicial generic-distinctiveness categorizations. Instead, 13 of the 18 non-corresponding categorizations involved suggestive distinctiveness, 4 involved descriptive-acquired distinctiveness, and 1 involved arbitrary distinctiveness. These results, constituting a correspondence rate of 0%, are not difficult to make sense of, as this study's sample had only 18 judicial rulings to work with in this category. With such small numbers for the training process, decision trees can easily miscategorize. Moreover, as I emphasized earlier, it is rare to see a judge grant distinctiveness to a word mark on the basis of generic traits, as there is a general assumption that a purely generic word mark cannot possess trademark status in nature.

Observed	Interpreted					
	A	D	DA	G	S	Percent Correct
A	48	1	21	0	65	35.6%
D	10	7	32	0	91	5.0%
DA	13	2	62	0	64	44.0%
G	1	0	4	0	13	0.0%
S	32	3	15	0	229	82.1%
Overall Percentage	14.6%	1.8%	18.8%	0.0%	64.8%	48.5%

Growing Method: CRT
Dependent Variable: Distinctiveness (Validity)

Table 12: The categorization results for Decision Tree 1

B.   Observation of Decision Tree 2

As noted above, Decision Tree 2 was trained for the initial period extending from January 1, 2002 to December 31, 2010. One topic of interest in the present study is the possible role that the *Abercrombie* taxonomy played in federal trademark-confusion decisions across various recent historical periods. Decision Tree 2, which covers the first ten years of the overarching period under investigation, involves 256 federal cases. Decision Tree 2 consists of 5 layers with 14 notes. Of the 19 independent variables, the tree chose only 5 during the training process: word-mark duration, consonant count, third-party registration count, categorization of third-party registration count, and word-formation category. Two of these independent variables appear twice in the Decision Tree 2: word-mark duration appears between the starting point and the first layer and between the second and third layers, and third-party registration count appears between the second and third layers and between the third and fourth layers.
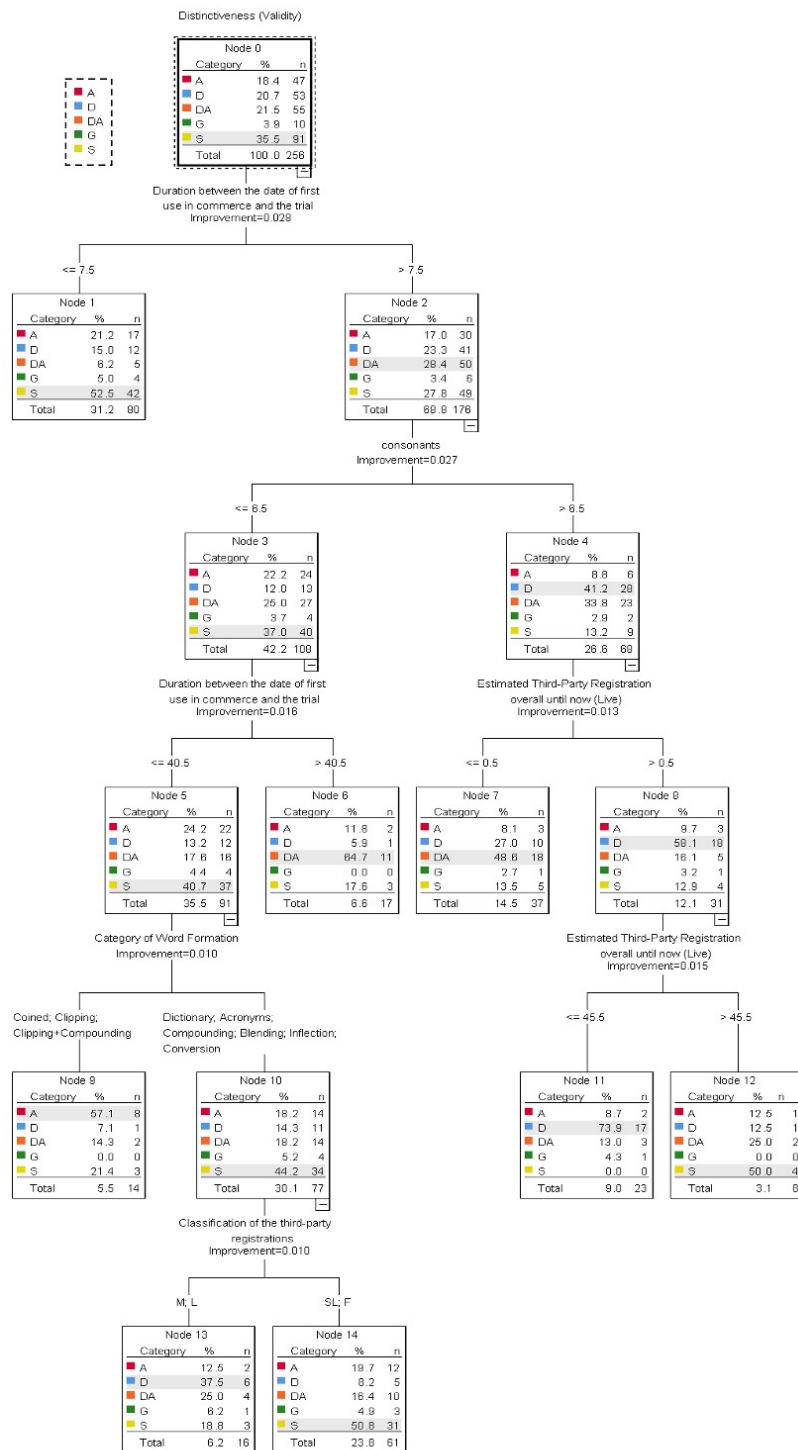
Decision Tree 2: Data trained from January 1, 2002 to December 31, 2010

As with Decision Tree 1, the chart of importance for Decision Tree 2 reveals the contributions that each selected independent variable made to the tree's categorizations. By comparing this chart of importance with the variables in the decision tree, we can shed light on how the decision tree might have delegated the functions of a rejected independent variable to a selected independent variable. Thus, it is that some independent variables appear in the chart of importance but not in the decision tree.

Table 13 below shows that the three most important independent variables contributing to Decision Tree 2's categorization of distinctiveness are, in descending order, word-mark duration, third-party registration count, and first-year use. Next, a comparison between Decision Tree 2 and the chart of importance reveals that three independent variables appear in neither the tree nor the chart: jurisdiction, decision year, and dictionary status.

Eleven independent variables appear in the chart of importance but not in Decision Tree 2: judge tenure, judge gender, word count, vowel count, syllable count, COCA frequency, IC count, US class count, word-mark product or service, plosive status, and first-year use. In other words, Decision Tree 2 chose only 5 of the 16 independent variables in the chart of importance, indicating that the 5 chosen variables could serve as surrogates for most of the independent variables regarding the task of categorizing word-mark distinctiveness.

Independent Variable	Importance	Normalized Importance
Duration	0.060	100.0%
Third-Party Registration Count	0.045	74.3%
First-Year Use	0.044	73.3%
Word-Formation Category	0.039	64.7%
Consonant Count	0.037	61.8%
Word Count	0.029	48.8%
Vowel Count	0.026	42.9%
Categorization of Third-Party Registration Count	0.022	36.9%
Product or Service	0.019	31.1%
US Class Count	0.013	21.8%
Number of IC	0.012	19.2%
Syllable Count	0.011	18.9%
Judge Tenure	0.009	15.5%
COCA Frequency	0.009	15.1%
Plosive Status	0.004	6.4%
Judge Gender	0.004	6.2%
Growing Method: CRT		
Dependent Variable: Distinctiveness (Validity)		

Table 13: The importance of independent variables based on Decision Tree 2

Regarding Decision Tree 2's categorization of the disputed word marks, several points merit our attention and are summarized in Table 14 below. First, during this initial period, federal judges ruled that 47 disputed word marks possessed arbitrary distinctiveness. Decision Tree 2 placed only 8 of these 47 word marks in the category of arbitrary distinctiveness, for a correspondence rate of only 17%. By contrast, the tree placed 30 of these word marks in the category of suggestive distinctiveness. These results indicate that Decision Tree 2 can discern almost no difference between the arbitrary marks and suggestive marks identified by the federal judges. This finding suggests that most court-identified arbitrary marks in the sample between 2002 and 2010 may have had characteristics similar to those of suggestive marks. Second, as for the 91 disputed word marks deemed by courts to be in possession of suggestive distinctiveness, Decision Tree 2 categorized 77 of them as suggestive marks, for an impressive correspondence rate of 84.6%. This impressive statistic indicates that the independent variables in Decision Tree 2 were able to yield suggestive-mark categorizations highly similar to those made by judges during this period.

Third up are the 55 disputed word marks that, in the eyes of federal courts during this initial period, possessed descriptive acquired distinctiveness. Decision Tree 2 agreed in 29 of these 55 cases, for a correspondence rate of 52.7%. Notably, the tree assigned suggestive distinctiveness to almost one-third (17) of the 55 word marks. These results indicate that the independent variables in Decision Tree 2 are sometimes capable of yielding categorizations identical to those made by courts with respect to descriptive acquired distinctiveness, but that these same variables can lead the tree to conclude that disputed word marks are suggestive. In other words, perhaps DA marks and suggestive marks share similar characteristics. Similar results characterize Decision Tree 2's handling of the courts' 53 descriptive distinctiveness word marks from this period. The tree mirrored the courts in 23 of the 53 cases, for a correspondence rate of 43.4%. A similar number (18 of 53) were categorized by the tree as having suggestive distinctiveness. These results can be interpreted much as the descriptive-acquired results were.

Finally, regarding the 10 word marks that courts designated as generic, Decision Tree 2, like Decision Tree 1, categorized none of them as generic. One reason for this outcome might be the smallness of the sample, and of course, another reason might be the nature of generic distinctiveness: it seldom serves as a basis for distinctiveness. Thus, in 7 of the 10 cases, Decision Tree 2 found evidence of suggestive distinctiveness.

|		Interpreted				
Observed	A	D	DA	G	S	Percent Correct
A	8	4	5	0	30	17.0%
D	1	23	11	0	18	43.4%
DA	2	7	29	0	17	52.7%
G	0	2	1	0	7	0.0%
S	3	3	8	0	77	84.6%
Overall Percentage	5.5%	15.2%	21.1%	0.0%	58.2%	53.5%
Growing Method: CRT Dependent Variable: Distinctiveness (Validity)						

Table 14: Categorization results for Decision Tree 2

C. Observation of Decision Tree 3

Decision Tree 3 underwent training for a twelve-year period (January 1, 2011 to December 31, 2022) encompassing 457 federal decisions. The decision tree, which has 5 layers with 28 notes, chose 9 of the 19 independent variables: judge tenure, word-formation category, word count, syllable count, consonant count, COCA frequency, IC count, first-year use, and categorization of third-party registration count. Some of these 9 independent variables appear three times in

Decision Tree 3 (e.g., judge tenure appears between the first and second, the second and third, and the fourth and fifth layers). Other variables appear twice (e.g., word-formation category appears between the starting point and the first layer and between the fourth and fifth layers, IC count and COCA frequency appear twice between the third and fourth layers.)
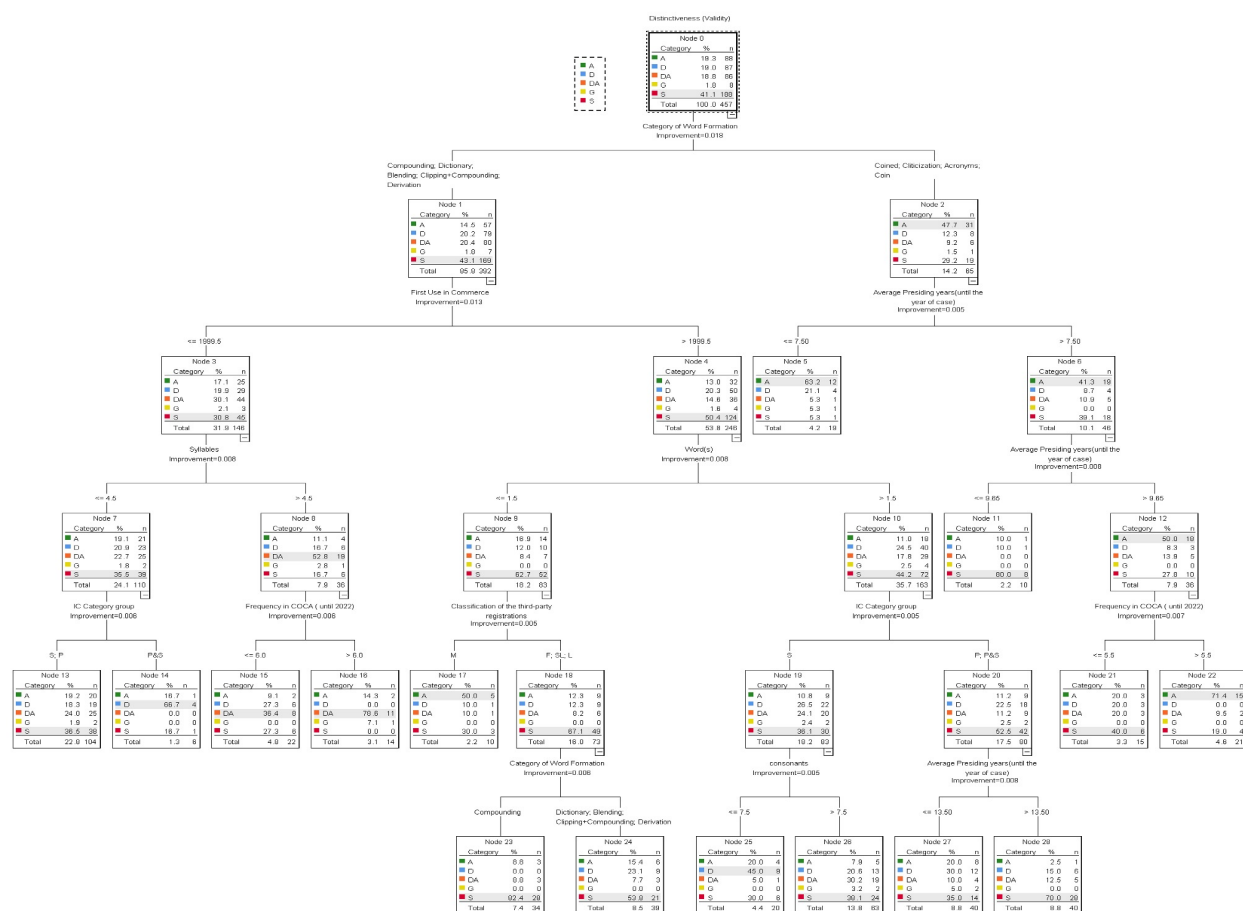
Decision Tree 3: Data trained from January 1, 2011 to December 31, 2022

Table 15 presents the chart of importance for Decision Tree 3. As we can see, two variables (i.e., decision year and jurisdiction) appear in neither Decision Tree 3 nor Table 15, and six other variables (i.e., judge gender, vowel count, plosive status, dictionary status, third-party registration count, and word-mark duration) appear only in the chart of importance but not in the decision tree. The latter result indicates that Decision Tree 3's nine independent variables can serve as surrogates for the six aforementioned independent variables.

Independent Variable	Importance	Normalized Importance
Consonant Count	0.034	100.0%
Syllable Count	0.033	98.2%
Vowel Count	0.033	97.5%
COCA Frequency	0.030	89.5%
Word-Formation Category	0.028	82.8%
First-Year Use	0.028	82.2%
Judge Tenure	0.026	76.6%
Word Count	0.024	71.5%
US Class Count	0.023	68.8%
Duration	0.023	67.0%
Third-Party Registration Count	0.020	60.3%
Categorization of Third-Party Registration Count	0.017	49.6%
Product or Service	0.015	45.3%
IC Count	0.012	34.6%
Dictionary Status	0.008	25.0%
Judge Gender	0.008	22.4%
Plosive Status	0.003	7.8%

Growing Method: CRT
Dependent Variable: Distinctiveness (Validity)

Table 15: The importance of independent variables based on Decision Tree 3

Turning our attention to the categorization results for Decision Tree 3, as summarized in Table 16, we can see that, during this period, there were 88 federal cases in which disputed word marks were placed under the category of arbitrary distinctiveness. Decision Tree 3 concurred with the courts in 32 of these cases, for a correspondence rate of 36.4%. In an impressive 47 of these 88 rulings, however, the tree assigned suggestive distinctiveness to the disputed word marks. These results indicate that arbitrary distinctiveness mimics suggestive distinctiveness within the framework of Decision Tree 3's independent variables.

Similar to the previous trees, Decision Tree 3 mirrored the federal courts regarding categorizations of suggestive distinctiveness. Specifically, of the 188 suggestive-distinctiveness rulings for this period, 167 were identically categorized by Decision Tree 3, resulting in a correspondence rate of 88.8%. In contract, Decision Tree 3 mirrored the federal courts in only 19 of the 86 descriptive-acquired distinctiveness decisions, for a correspondence rate of only 22.1%. As for the remaining 67 court cases, the tree settled on suggestive distinctiveness 62 times. These results indicate that Decision Tree 3, with its unique independent-variable

profile, has a strong tendency to treat as suggestive those word marks previously categorized by courts as descriptive-acquired. These results also indicate that suggestive marks and descriptive-acquired marks are substantially similar to each other. More interestingly, the results for Decision Tree 3's handling of descriptive distinctiveness are almost identical to the results for Decision Tree 3's handling of descriptive-acquired distinctiveness: of the 87 disputed word marks that federal courts during this period placed under the category of descriptive-acquired distinctiveness, only 13 were similarly categorized by Decision Tree 3, for a correspondence rate of 14.9%. This means that the tree selected other categories for 74 of the 87. As it turns out, 63 of these 74 "other categories" selections rested on suggestive distinctiveness. These results are almost the same as the ones associated with Decision Tree 3's handling of descriptive-acquired distinctiveness.

Finally, and again in line with the previous trees, Decision Tree 3 had to deal with a very small number of generic-distinction court rulings. Of the 8 generic-distinction categorizations made by judges between 2011 and 2022, not 1 was mirrored by Decision Tree 3. The reasons cited with respect to the first two trees apply to the third tree.

Observed	Interpreted					
	A	D	DA	G	S	Percent Correct
A	32	5	4	0	47	36.4%
D	5	13	6	0	63	14.9%
DA	4	1	19	0	62	22.1%
G	1	0	1	0	6	0.0%
S	8	7	6	0	167	88.8%
Overall Percentage	10.9%	5.7%	7.9%	0.0%	75.5%	50.5%

Growing Method: CRT
Dependent Variable: Distinctiveness (Validity)

Table 16: Categorization results for Decision Tree 3

D. Comparative Analysis and Key Findings

1. A Comparison of the Three Decision Trees

Table 17, below, presents a side-by-side comparison of all the independent variables and their ordering in Decision Tree 1, Decision Tree 2, and Decision Tree 3.

	Decision Tree 1 (Jan. 1, 2002–Dec. 31, 2022)	Decision Tree 2 (Jan. 1, 2002–Dec. 31, 2010)	Decision Tree 3 (Jan. 1, 2011–Dec. 31, 2022)
Layer 1	first-year use	duration	word-formation category
Layer 2	consonant count, word-formation category	consonant count	first-year use, judge tenure
Layer 3	third-party registration count, plosive status, dictionary status, first-year use	duration, third-party registration count	syllable count, word count, judge tenure
Layer 4	duration, word-formation category, IC count	word-formation category, third-party registration count	IC count (two times), COCA frequency (two times), categorization of third-party registration count
Layer 5	IC count, judge tenure (three times), duration	categorization of third-party registration count	word-formation category, consonant count, judge tenure

Table 17: Comparison of all independent variables and their ordering in the three decision trees

From Table 17 above, several key observations can be made. First, when comparing Layer 1 and Layer 2 in all three decision trees, we can see that first-year use (Layer 1 of Tree 1, Layer 2 of Tree 3) and word-formation category (Layer 1 of Tree 3 and Layer 2 of Tree 1) were commonly chosen by the trees to occupy first and second orders for the categorization of word-mark distinctiveness. Duration and consonant count, which appear in Layer 1 of Tree 2 and Layer 2 of Tree 1 and Tree 2, were also common choices of the decision trees. **These findings are significant because they infer that those linguistic characteristics may be the first factor for federal judges to categorize the word-mark distinctiveness.**

Second, when comparing the three trees with one another regarding Layer 2 and Layer 3, we can make the following observation: to categorize distinctiveness, the three decision trees chose third-party registration count (Layer 3 of Tree 1 and Tree 2) and several linguistic variables (plosive status, dictionary status, syllable count, and word count in Layer 3 of Tree 1 and Tree 3). The independent variables chosen by the three decision trees for Layer 1 and Layer 2 still play key categorization roles in Layer 3. Interestingly, judge tenure appears again in Layer 3 of Tree 2. **The significance of these findings aligned with the previous paragraph that linguistic characteristics still played a key role for federal judges to categorize the distinctiveness.**

Third, a comparison of all three trees regarding Layer 3 and Layer 4 reveals the following important points: Tree 1 and Tree 3 chose IC count and similar word-mark registration variables for Layer 4's categorization of distinctiveness. Moreover, for Layer 4, Tree 2 chose third-party registration count and Tree 3 chose categorization of third-party registration count, but neither of these independent variables was chosen by Tree 1. The variables (i.e., attributes) chosen by all three trees for Layer 1 and Layer 2 still play key roles in Layer 4's categorization function. A particularly interesting finding is that COCA frequency appears in Layer 4 of Tree 3. **The significance of these findings is that third-party uses (i.e., word-mark registration and COCA frequency) may not be first factor for federal judges to categorize the distinctiveness.**

Finally, a comparison of all three trees regarding Layer 4 and Layer 5 reveals an interesting fact: to categorize distinctiveness in Layer 5, IC count is the only variable that are not chosen repetitively in other Layers. Other variables (i.e., attributes) in Layer 5 of all three trees such as judge tenure, duration, word-formation category, consonant count, and categorization of third-party registration count have been chosen to categorize distinctiveness in Layers 1, 2, and 4. **This finding infers that both linguistic characteristics and third-party uses may be factors for federal judges to categorize the distinctiveness in the long run.**

2.   Comparison of the Three Charts of Importance and Findings

As I noted earlier, it is not necessarily the case that the more important a decision tree's feature is, the higher its node will be. Thus, a comparative analysis of the three charts of independent-variable importance is advisable because, in this way, we can better understand the order and the weight of the importance of the three samples. The order of importance reveals, in descending order, the information gain that each independent variable in a decision tree is capable of. Likewise, the weight of importance refers to each independent variables' contribution to the output of a decision tree.

Regarding order of importance, Tree 1's five most important independent variables can serve as a benchmark from which we can determine that duration was the most important independent variable in Tree 1 and Tree 2, but ranked tenth in Tree 3. First-year use was the second most important independent variable in Tree 1 and the third most important in Tree 2, but ranked sixth in Tree 3. Word-formation category was the third most important independent variable in Tree 1, the

fourth in Tree 2, and the fifth in Tree 3. Vowel count was the fourth most important independent variable in Tree 1, the seventh in Tree 2, and the third in Tree 3. Finally, consonant count was the fifth most important independent variable in both Tree 1 and Tree 2, but ranked first in Tree 3. The findings above point to an interesting pattern: **duration and first-year use were the most important independent variables in Tree 1 and Tree 2. However, in Tree 3 (i.e., the subsample for the 2011–2022 period), independent linguistic variables were more important than both duration and first-year use.**

I found similarities among the three charts regarding their respective weight-of-importance measures: the weight of importance for all the independent variables spans a range between 0.003 (the lowest weight) and 0.06 (the highest weight). In particular, the weight of importance attached to duration, which is the most important independent variable in Tree 1 and Tree 2, is 0.033 in Tree 1 and 0.06 in Tree 2. In Tree 3, the most important independent variable is consonant count, whose weight of importance measures 0.034. By contrast, plosive status, which is the least important independent variable in Tree 1 and Tree 3, has weights of importance measuring, respectively, 0.005 and 0.003. In Tree 2, the least important independent variable is judge gender, weighing in at 0.004. **These individual weights reflect a pattern in which independent variables possessing a relatively high weight of importance were insufficient, in this study's three decision trees, for the task of categorizing word-mark distinctiveness. The individual weights reflect another pattern, as well: some of the independent variables possessing a relatively low weight of importance seem to have been irrelevant to the categorization of word-mark distinctiveness.**

3.   *Comparison of the Categorization Results for All Three Decision Trees and Findings*

When comparing the three decision trees regarding their respective categorization results, we can glean important information about the trees' correspondence rates. First, **suggestive distinctiveness has the highest rates of correspondence across all three decision trees** (the rates were over 80% in each tree), while **generic distinctiveness has the lowest rates of correspondence across all three trees** (0%). Descriptive-acquired distinctiveness achieved the second highest rates of correspondence in Tree 1 (44%) and Tree 2 (52.7%), but ranked third in Tree 3 (22.1%). **Interestingly, all three trees tended to categorize as suggestive distinctiveness the disputed word marks that federal courts had**

placed under the category of descriptive-acquired distinctiveness. Arbitrary distinctiveness has the third highest correspondence rate in Tree 1 (35.6%) and Tree 3 (36.4%) but ranks fourth in Tree 2 (17.0%). **In line with the previously cited types of distinctiveness, the three trees tended to categorize as suggestive distinctiveness the disputed word marks that federal courts had previously placed under the category of arbitrary distinctiveness.** Finally, descriptive distinctiveness ranks fourth for its correspondence rate in Tree 1 (5%) and Tree 3 (14.9%), but ranks second in Tree 2 (43.4%). **This pattern is similar to the previously cited patterns for correspondence rates: the three trees tended to place word marks in the category of suggestive distinctiveness.**

V
DISCUSSION AND SOLUTIONS

The results of this study's decision-tree algorithms, when compared with the results of this study's case analyses, lead to two conclusions: (1) Between 2002 and 2022, federal judges relied heavily on the linguistic features of disputed word marks when categorizing the distinctiveness of the marks. (2) The chief consequence of this reliance was that judges tended to miscategorize marks, either as suggestive or as descriptive. More specifically, judges' excessive reliance on the dictionary test reflected an unwillingness or an inability to make full, rigorous use of the *Abercrombie* taxonomy.[181] In what follows, I discuss the roots of this poor judicial reasoning and explain why we must not turn a blind eye to this problem.

1. The dictionary test and the roots of federal judges' poor reasoning in word-mark disputes

In the present study, patterns reflecting the importance of the trees' independent variables reveal that linguistic variables were more influential in the decision-making of federal judges. In short, my analysis of the decision-tree nodes and the patterns of importance related to independent variables has led me to infer that when determining word park distinctiveness, federal judges made immoderate use of the dictionary test.

Though of practical importance, if relied on excessively, the dictionary test can induce judges to focus at length on differentiating between suggestive

---

[181] *See* Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976).

and descriptive distinctiveness. Two decision-tree patterns support this tentative conclusion. First, a brief look at the categorization performances of all three decision trees shows that the highest and the second-highest correspondence rates between the trees' categorizations and judges' categorizations involved suggestive distinctiveness and descriptive-acquired distinctiveness. By contrast, there were lower rates of correspondence with respect to the arbitrary and descriptive types of distinctiveness. These patterns indicate that federal judges focused almost exclusively on differences between suggestive distinctiveness and descriptive-acquired distinctiveness. Second, the low correspondence rates characterizing the arbitrary and descriptive types of distinctiveness indicate that federal judges, perhaps because of their over-reliance on the dictionary test, had been experiencing difficulties when attempting to differentiate both arbitrary and descriptive distinctiveness from suggestive distinctiveness. With the dictionary test readily at hand and no other tests available for the holistic analysis of distinctiveness, the aforementioned correspondence-rate patterns indicate something further: that federal judges seem to have been unwilling or unable to clarify and harness each category of distinctiveness in the *Abercrombie* taxonomy.

2. The dictionary test and misconceptions of inherent distinctiveness

Given that federal judges' unwillingness or inability to thoughtfully harness the *Abercrombie* taxonomy may stem from their overreliance on the dictionary test, we can now consider a critical question: why did federal judges rely on the dictionary test yet fail to clarify each and every type of distinctiveness? To answer this question, we should observe how the federal judges interpreted the concept of inherent distinctiveness. Most federal judges seem to have relied on the dictionary test to grasp the concept of inherent distinctiveness. For instance, in *Virgin Enterprises v. Nawab*, the Second Circuit interpreted the inherent distinctiveness, saying

> Considering first inherent distinctiveness, the law accords broad, muscular protection to marks that are arbitrary or fanciful in relation to the products on which they are used, and lesser protection, or no protection at all, to marks consisting of words that identify or describe the goods or their attributes.[182]

---

[182] *See* Virgin Enters. Ltd. v. Nawab, 335 F.3d 141, 147 (2d Cir. 2003). This case was included in the present study's dataset.

The dataset in this study abounds with similar examples from other federal jurisdictions.[183] Yet even merely from the *Virgin Enterprises* case, we can infer that the dictionary test was serving as the sole judicial test for determining the arbitrariness of a word mark. Nonetheless, it is poor practice for federal judges to directly apply the dictionary test to analyses of inherent distinctiveness because a critical lens through which the concept of distinctiveness must be analyzed is the consumer: word meaning and word formation, by themselves, can in no way answer the question of whether or not a word mark is distinctive.[184] One reason why federal judges would nevertheless rely exclusively or at least excessively on the dictionary test might stem from the judges' misconception of inherent distinctiveness, which could, in turn, lead the judges to neglect the complex spectrum of distinctiveness under *Abercrombie*.

3. The harms posed by federal judges' poor reasoning in word-mark cases

Some skeptics might argue that the task in any rigorous analysis of word mark distinctiveness is to differentiate between suggestive and descriptive because the latter requires proof of secondary meaning whereas the former does not.[185] This line of reasoning would seem to suggest that federal judges need not clarify the lines separating arbitrary, suggestive, descriptive, and generic distinctiveness from one another. This skepticism suffers from two inescapable fallacies: the "lesser importance of arbitrariness" fallacy and the "lesser importance of genericness" fallacy. First, the evidence in the present study suggests that federal judges hold the view—perhaps unthinkingly, perhaps not—that judges don't think the marks themselves are less important, but that clearly delineating them from others might be less important to strength of the mark analysis. A huge problem arising from

---

[183] *See, e.g.*, Kellogg Co. v. Toucan Golf, Inc., 337 F.3d 616, 626 (6th Cir. 2003) (showing that the court determined both the word mark TOUCAN SAM and its logo to be fanciful, which is to say, arbitrary); Aceto Agr. Chems. Corp. v. Bayer Aktiengesellschaft, No. 10 CIV. 1770 AJN, 2012 WL 3095060, at *5 (S.D.N.Y. July 30, 2012), *aff'd*, 531 F. App'x 103 (2d Cir. 2013) (defining a fanciful mark as made-up, a descriptive mark as expressive of the traits or functions of a product or service, and a suggestive mark as expressive in a way that depends on people's interpretive perceptions); Stark v. Diageo Chateau & Estate Wines Co., 907 F. Supp. 2d 1042, 1060 (N.D. Cal. 2012) (noting that arbitrary marks, though perhaps common, are not descriptive of a good or service, whereas fanciful—arbitrary—marks are unusual insofar as they are either made up or no longer commonly used).

[184] A word mark, if it is distinctive, must enable consumers either to identify the source of a good or to know that the good comes from a unique source. *See* J. Thomas McCarthy, McCarthy On Trademarks and Unfair Competition § 3:9 (4th ed. 2007).

[185] *See* Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 12 (2d Cir. 1976).

this view is that, if courts ignore the line separating arbitrary distinctiveness from suggestive distinctiveness, trademark owners will be hard pressed to predict with any degree of accuracy whether a court might deem their word marks to be weak, even though the trademark owners might have no need to prove secondary meaning. For example, in a recent case pitting Teetex against Zeetex, led the presiding judge to analyze the strength of the plaintiff's word mark TEETEX.[186] The judge addressed the matter as follows:

> Teetex is at best a suggestive mark. The suffix 'tex' suggests textiles, but the name does require some imagination to associate the mark with the product. Although stronger than a descriptive or generic mark, suggestive marks are still "presumptively weak."[187]

The judge categorized the suggestive word mark as a "weak" mark, even though the word mark owner was under no obligation to prove secondary meaning.[188] Moreover, because the judge compared the strength of suggestive marks with the strength of generic and descriptive marks and declared that they are all weak marks, we can infer that fanciful marks and arbitrary marks are "strong" marks. Therefore, if the specific characteristics of the fanciful or arbitrary marks and suggestive marks are not well known, there is a significant risk that judges will miscategorize the marks.

The second fallacy rests on the unstated view—again, perhaps held unthinkingly—that generic distinctiveness is not as important as suggestive or descriptive distinctiveness. However, in cases where the line between a descriptive mark and a generic mark is unclear, this fallacy could prove to have serious consequences because genericization of a word mark could trigger a loss of trademark protection.[189] The issue of trademark genericness has caught the attention of the Supreme Court, which has provided some guidance for lower courts. However, the guidance provides little help in the way of boundary

---

[186] Teetex LLC v. Zeetex, LLC, No. 20-CV-07092-JSW, 2022 WL 1203097, at *3–4 (N.D. Cal. April 22, 2022). This case was included in the present study's dataset.

[187] *Id.* at 4.

[188] *See id.* at *3–4.

[189] *In re* Merrill Lynch, Pierce, Fenner, & Smith, Inc., 828 F.2d 1567, 1569 (Fed. Cir. 1987) ("Generic terms [are] by definition incapable of indicating source … and can never attain trademark status.").

delineation for descriptive marks and generic marks. In *Booking.com*, the Supreme Court explained three characteristics of a generic term:

> First, a "generic" term names a "class" of goods or services, rather than any particular feature or exemplification of the class. Second, for a compound term, the distinctiveness inquiry trains on the term's meaning as a whole, not its parts in isolation. Third, the relevant meaning of a term is its meaning to consumers.[190]

While the Supreme Court in *Booking.com* attempted to elaborate on what makes a mark generic, lower court applications of this guidance make plain that a serious lack of clarity persists. For example, in *Snyder's Lance, Inc. v. Frito-Lay North America, Inc.*, a district court addressed whether PRETZEL CRISPS was generic:

> Unlike booking.com (the combined mark identifies a specific company at that internet address) and American Airlines (consumers understand that there are numerous separately named airlines in the United States and don't refer to them collectively as "American Airlines"), there is no additional meaning that results from the combination of the generic terms that make up PRETZEL CRISPS in the minds of consumers. "Pretzel" "crisps" are pretzels in the shape or form of a cracker and "pretzel crisps", viewed together, would be perceived as the same thing. In sum, the Court finds that the combined term PRETZEL CRISPS adds no additional meaning to consumers that suggests the mark is not primarily a generic name.[191]

The reasoning above suggests that the district court followed a two-step process: first, he compared two marks—the Booking.com and American Airline marks—with PRETZEL CRISPS; then, he analyzed whether "pretzel crisps" had additional meaning when the mark's constituent parts ("pretzel" and "crisps") were combined. This logic, at its core, was still a form of reasoning by analogy—taking previously disputed marks and determining whether their characteristics were at all similar to those of "PRETZEL CRISPS". This type of analysis, however, does not elaborate what characteristics are specific to generic marks, what characteristics

---

[190] U.S. Pat. & Trademark Off. v. Booking.com B.V., 591 U.S. 549, 556 (2020).

[191] Snyder's Lance, Inc. v. Frito-Lay N. Am., Inc., 542 F. Supp. 3d 371, 384 (W.D.N.C. 2021) (citations omitted).

are specific to descriptive marks, and what characteristics are shared by both marks. Therefore, although the Supreme Court had sought to provide more than a modicum of guidance for analysis of generic marks, lower courts have still found it necessary to rely on analogical reasoning to decide whether or not a given mark is a generic mark. As for word mark applicants and owners, they are still struggling to "guess" whether their word marks are analogous to previous marks and are still preparing to bear the risks that accompany generic word marks.

The fact that federal courts have long prioritized the categorization of suggestive and descriptive marks to the exclusion of arbitrary and generic marks, leading to the existence of these two fallacies, should not suggest that federal trademark litigation rarely deals with questions of arbitrary or generic distinctiveness. In fact, the opposite is true. Of the present study's 713 cases covering the period from January 2002 to December 2022, 153 cases centered on disputes regarding the arbitrary or generic status of plaintiffs' or defendants' marks—a figure that amounts to an impressive 21 percent of the sample. My point here is not that judges ignore or never deal with arbitrary and generic marks, but rather that judges consistently rely on three tests (i.e., the imagination test, the competitor-need test, and—perhaps most conspicuously—the dictionary test) that are ill suited for the proper analysis of arbitrary and generic distinctiveness. Thus, judges at the federal level can benefit greatly from a better understanding of arbitrary marks and generic marks—an area of inquiry that has been neglected in favor of a dangerously narrow focus on suggestive and descriptive marks.

The above fallacies that help explain the focus on suggestive versus descriptive analyses also help explain why federal judges have consistently engaged in poor reasoning when hearing cases related to word mark disputes. To address this problem, federal judges hearing these types of cases should consider all categories of distinctiveness and should thus harness the full powers of the *Abercrombie* taxonomy in trademark law. A number of judges have cited and thoughtfully used *Abercrombie* to categorize the distinctiveness of various word marks. However, as I have postulated, the existence of three tests—the imagination test, competitor-need test, and dictionary test—may convince federal judges that there is no need to grapple with the vague lines that fuzzily delineate the various categories of distinctiveness. The *Abercrombie* taxonomy is difficult to understand and apply, so if judges have a superficially compelling—yet ultimately fallacious—reason to sidestep the taxonomy, they may very well do so. My position

is that the *Abercrombie* taxonomy, though complex, should be and is a coherent and comprehensible set of principles that, if studied by federal judges, can be understood and applied in ways that will greatly diminish the poor reasoning that has long plagued rulings in trademark litigation. *Abercrombie* is nevertheless insufficient: the judiciary is in need of an even fuller set of tools for categorizing distinctiveness.

4. Rethinking the analytical approaches to distinctiveness and solutions to poor judicial reasoning

Because *Abercrombie* rigorously defined the concept of distinctiveness and, with equal rigor, laid out a taxonomy of distinctiveness categories, the case has been the subject of many studies from diverse perspectives, as I discussed in the literature review. The findings of these studies indicate that, although the *Abercrombie* taxonomy is useful, it falls short of the spectrum of tools that federal judges need for a comprehensive analysis of distinctiveness. One area in which *Abercrombie* is particularly deficient is that of consumer perception. Judges should seek empirical, concrete data on consumer perception rather than rely on purely abstract legal theories and on easily citable precedents. Nonetheless, there is great hesitancy regarding judges' application of surveys and other studies of perception to analyses of distinctiveness because lack of familiarity with these fields of knowledge may lead to the judges' incorrect interpretation of the results.[192] However, the reality of this challenge does not justify judges' current overreliance on the dictionary test and judges' misconceptions about the supposed inherent nature of distinctiveness. Until the U.S. judiciary properly Until judges are incentivized to take a more rigorous approach to trademark analyses, they will continue to engage in poor legal reasoning in trademark litigation.

How can we successfully address this poor judicial reasoning? To answer this question, we must understand why judges engage in the poor reasoning to begin with. One explanation might be found in the concept of rational ignorance.[193]

---

[192] *See* David L. Faigman, *Judges as "Amateur Scientists"*, 86 B.U. L. Rev. 1207, 1209 (2006) (arguing that judges who lack a fundamental understanding of science cannot render reliable judgements in cases requiring scientific knowledge).

[193] The phrase 'rational ignorance' appears mainly in discussions related to political economics and public-choice theory. *See, e.g.*, Jonathan R. Macey, *Cynicism and Trust in Politics and Constitutional Theory*, 87 Cornell L. Rev. 280, 306 (2002) ("Rational ignorance and other collective action problems make it difficult for even well-educated citizens to effectively monitor the performance of government."); John O. McGinnis,

People who engage in rational ignorance refuse to acquire knowledge when the perceived cost of acquiring the knowledge seems to exceed the expected potential benefit that the knowledge would provide.[194] When applying this concept to federal cases involving word mark distinctiveness, we can see that judges might embrace rational ignorance because they perceive the cost of establishing, say, a new rule to clarify all types of distinctiveness as much higher than the benefit to be derived from the new rule. Put more specifically, **federal judges may rely excessively on the dictionary test, which itself excessively privileges the concept of inherent distinctiveness, because the benefits of this reliance are perceived to be much higher than the costs of establishing a new rule even if the new rule would improve judicial reasoning.**

Empirical evidence supports this explanation, as federal judges have long cited *Abercrombie* to justify their categorizations of word marks' distinctiveness, yet the test most frequently used is often only the dictionary test, which narrowly differentiates between suggestive and descriptive distinctiveness in trademark likelihood-of-confusion cases. Clear rules for determining other types of distinctiveness remain neglected.

The precise effects attributable to the vagueness or uncertainty of legal rules remain a matter of considerable debate.[195] However, the current reliance on unclear rules governing word-mark distinctiveness rests on two empirically discernable

---

Reviving Tocqueville's America: The Rehnquist Court's Jurisprudence of Social Discovery, 90 Cal. L. Rev. 485, 503 n.81 (2002) ("'Rational ignorance' describes the systematic tendency of diffuse citizens to pay little attention to political information."). The theory of rational ignorance has also been adopted in U.S. patent law. In this respect, Professor Lemley explains that the basic idea of rational ignorance is that any person will spend only a certain amount of time or money to obtain information. If obtaining that information costs more than the information is worth, the person will (or should) rationally choose to remain ignorant of it. *See* Mark Lemley, *Rational Ignorance at the Patent Office*, 95 Nw. L. Rev. 1, 3 n.6 (2001).

[194] Post-war discussions about rational ignorance in the context of cost seem to have originated with the political economist Anthony Downs. *See* Anthony Downs, *An Economic Theory of Democracy*, J. Pol. Econ. 135, 139 (1957); *see also* George J. Stigler, *The Economics of Information*, 69 J. Pol. Econ. 211, 213 (1961). For applications of the concept of rational ignorance in law, *see, e.g.*, Melvin Aron Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 Stan. L. Rev. 211, 241 (1995).

[195] *See* Anthony D'Amato, *Legal Uncertainty*, 71 Cal. L. Rev. 1, 3 (1983); Richard R. Brooks & Warren F. Schwartz, *Legal Uncertainty, Economic Efficiency, and the Preliminary Injunction Doctrine*, 58 Stan. L. Rev. 381, 382 (2005); Giuseppe Dari-Mattiacci & Bruno Deffains, *Uncertainty of Law and the Legal Process*, 163 J. Institutional & Theoretical Econ. 1, 4 (2007); Matthias Lang, *Legal Uncertainty: A Selective Deterrent* 1 (Preprints of the Max Planck Inst. for Rsch. on Collective Goods, Working Paper No. 2014/17), https://www.econstor.eu/handle/10419/106905 [https://perma.cc/95YF-G3XD]; Jiwon Lee,

fallacies, whether stated or not. Therefore, drawing on the descriptive case analysis and the decision-tree analysis above, I propose two methods by which we can diminish the problem of poor reasoning in federal trademark-confusion cases.

Method 1: The USPTO can decrease the cost of establishing a new rule by comprehensively and clearly articulating the main factors that contribute to distinctiveness.

Federal judges usually introduce the concept of word mark distinctiveness and cite the *Abercrombie* taxonomy without clarifying all types of distinctiveness. The doctrine of *stare decisis* can shed light on this situation, stating that judges cannot easily establish a new rule, especially if the rule will require that they substantively alter their existing approach to handling cases.[196] Therefore, one way to improve judges' knowledge of distinctiveness is to improve, rather than replace, the existing rules; that is, Courts should not dissolve the *Abercrombie* taxonomy but establish alternatives to it.

One source of alternatives is the U.S. Patent and Trademark Office's Trademark Manual of Examination Procedure (TMEP).[197] The latest version of the TMEP, published in May 2024, elaborates five factors for determining the inherent distinctiveness of "repeating-pattern" marks: does the repeated use of a mark (1) constitute a common or widely used pattern, (2) create a distinct commercial impression, (3) comprise elements of a distinct nature, (4) reflect industry practices, and (5) refer to a type of product or service.[198] Though useful, these factors have two drawbacks. First, they are specifically used for determining the inherent distinctiveness of repeating-pattern marks, not word marks. Second, even if these factors could be used for determining the inherent distinctiveness of word marks, not one of the factors focuses on actual consumer perception, as would be gleaned

---

David Schoenherr & Jan Starmans, *The Economics of Legal Uncertainty* (Eur. Corp. Governance Institute, Working Paper No. 669/2022, 2024).

[196] Amy Coney Barrett, *Stare Decisis and Due Process*, 74 U. Colo. L. Rev. 1011, 1015 (2003) (suggesting that courts of appeals feel the restrictions imposed by horizontal stare decisis more strongly than do district courts or the Supreme Court).

[197] The TMEP is published to provide trademark examining attorneys, trademark applicants, attorneys, and other trademark stakeholders with a reference work on the practices and procedures relative to prosecution of applications to register marks in the USPTO. *See Trademark Manual of Examining Procedure – Files and Archives*, https://www.uspto.gov/trademarks/guides-and-manuals/tmep-archives [https://perma.cc/8M9W-8ARC] (last visited Mar. 21, 2025).

[198] TMEP §§ 1212.19(e)(i)(A)–(E).

from surveys, declarations, affidavits, and the like. Thus, to lower the cost of establishing a new rule for federal judges, the USPTO could first separate inherent-distinctiveness factors from acquired-distinctiveness factors. This step would go far in reducing judges' overreliance on imagination and dictionary tests, both of which emphasize inherent distinctiveness.

Knowledge of consumer perceptions can help clarify the strength of a word mark.[199] With this concept in mind, the USPTO could calculate the different degrees to which consumer knowledge has a bearing on, say, arbitrary distinctiveness versus suggestive distinctiveness, and this knowledge can be obtained from consumer data (e.g., survey data) in the TMEP.[200] In the previous scenario, arbitrary distinctiveness requires a greater presence of consumer recognition from a specific source than does suggestive distinctiveness. Once the USPTO clearly articulates the requirements and guidelines for identifying the presence (or absence) of inherent distinctiveness in the TEMP, not only federal Judges but also trademark applicants and owners will finally have clear, workable criteria for determining which types of evidence point to the existence of inherent distinctiveness. With these improvements in place, more importantly, federal judges would be far less likely to misconstrue and mishandle the concept of inherent distinctiveness, thus greatly reducing the problem of judicial overreliance on a limited spectrum of the available tests. The end result would be better reasoning in trademark-confusion cases.

The second method that can reduce the problem of poor judicial reasoning in federal trademark-confusion cases is essentially geared toward lowering judges' ill-advised prioritization of linguistics-related evidence.

> **Method 2: The USPTO can decrease the benefits of relying on the dictionary test by lowering the incentives that judges currently have to prioritize linguistics-related evidence over other types of evidence**

This method can best be implemented by the USTPO in conjunction with the TMEP. As I have demonstrated throughout this study, federal judges have

---

[199] Phillip Johnson, *Enhanced Distinctiveness and Why "Strong Marks" Are Causing Us All Confusion*, 55 INT'L REV. INTELL. PROP. & COMPETITION L. 185, 186 (2023) (arguing that consumers would not be easy to be confused by the stronger mark, proven by the psychological and marketing evidence).

[200] Jake Linford, *Democratizing Access to Survey Evidence of Distinctiveness*, *in* RESEARCH HANDBOOK ON TRADEMARK LAW REFORM 225, 226 (Dinwoodie & Janis, eds., 2021).

relied on the dictionary test when analyzing the categories of distinctiveness, a reliance that, being excessive, leads to and stems from incorrect perceptions of inherent distinctiveness. An important consequence of this overreliance is that judges underestimate the importance of consumer perception when analyzing the extent of a disputed word mark's inherent distinctiveness.

Any solution to this problem must contend with a highly predictable obstacle: federal judges will not easily change their tried-and-true habits for determining distinctiveness. The doctrine of *stare decisis* makes this point clear. Therefore, the TMEP can also specify that, for determining all types of distinctiveness, consumer perceptions (as gleaned from surveys, declarations, affidavits, and the like) are superior to linguistic evidence. I propose that this specification, if made clearly and without equivocation, will greatly incentivize federal judges to lessen their reliance on the dictionary test. Let us consider such a specification in greater detail: To incentivize judges in this direction, the TMEP can provide comprehensible (i.e., clear and practical) guidance for calculating the weight of evidence required for analyses of *inherent distinctiveness* and the corresponding weight of evidence required for analyses of *acquired distinctiveness*. Because inherent distinctiveness, which requires a word mark to identify the source of product or service when consumers see the word mark at the first time, is less easily established than acquired distinctiveness, the weight that judges assign to consumer perception should be greater—perhaps much greater—than the weight that judges assign to linguistic evidence. The guidance for relative weight could be couched in quantitative terms: for instance, seventy percent for consumer-perception evidence, and the remaining thirty percent for linguistic evidence.

It is reasonable to expect that, once the TMEP clearly and rigorously establishes the superiority of consumer-perception evidence, federal judges will gradually or perhaps even quickly decrease their reliance on the dictionary test when analyzing categories of word-mark distinctiveness. It is thus also reasonable to expect that, in turn, there will be a diminution of poor reasoning in federal cases concerning trademark-confusion disputes.

VI
Conclusion and Limitations of Research

In this study, I have performed a descriptive analysis and a decision-tree analysis of federal trademark litigation covering a roughly twenty-year period

extending from 2002 through 2022. The results of these analyses reveal that federal judges have consistently engaged in poor reasoning when dealing with questions of word mark distinctiveness. Specifically, the judges excessively focus on differentiating between suggestive and descriptive distinctiveness, most likely because the judges have a misplaced preference for inherent distinctiveness as opposed to acquired distinctiveness and for linguistic evidence (e.g., the dictionary test) as opposed to consumer-perception evidence. This poor reasoning, regardless of whether it is a consequence of rational ignorance or simple ignorance, is a problem that demands our attention and that merits practical, implementable solutions. To this end, I have proposed that the USPTO (1) should summarize the main factors of inherent distinctiveness and acquired distinctiveness in the TMEP and (2) should offer judges a set of USTPO guidelines that ends the judiciary's long-standing prioritization of linguistic evidence in a way that elevates the importance of consumer-perception evidence.

As with all studies, the current one has its fair share of limitations, many of which can be addressed in more future research. First, I wanted to integrate into this study's analyses the marketing-expense data for word-mark owners. Unfortunately, this category of data is very difficult to collect. Though regularly used by federal judges in trademark cases, much of the relevant data are kept secret from the public. The lack of marketing-expense data in the present study thus constitutes a major research limitation insofar as my descriptive and decision-tree analyses had to do without satisfactory inputs of data for this topic. Second, the attributes of the competitive-need test are hard to measure. Because researcher-conducted surveys are necessary to determine whether a competitor would likely use the words in a disputed word mark, I was able to conduct only rough measures of third-party registrations for each disputed word mark. My aim, through the decision-tree analyses, was to get a sense of whether third-party registrations had played a key role in the distinctiveness decisions of federal judges. My rough measures, though better than nothing, may have biased the results of the decision-tree analyses, making it that much more difficult to speculate about both the degree to which federal judges rely on the competitive-need test and the causes of their poor reasoning. Until such time as the USPTO's TSS grants interested parties dynamic access to comprehensive, correct third-party registration numbers for each disputed word mark, this research limitation will persist unabated.

Third, although the decision tree is a powerful tool for dealing with non-linear data, such as the data pertaining to the federal cases and related variables addressed in the present study, the decision tree is by no means perfect. In particular, the issue of whether all data points are classified as homogeneous is dependent largely on the complexity of the decision tree in question. According to the article "What Is a Decision Tree?" on the IBM website, "Smaller trees are more easily able to attain pure leaf nodes. … However, as a tree grows, it becomes increasingly difficult to maintain this purity, and it usually results in too little data falling within a given subtree"—a problem that sometimes causes overfitting.[201] These problems should not lead one to conclude that smaller is always better: if too simple, a decision tree can easily overlook important patterns in the data—a problem that results in the opposite of overfitting: underfitting.

How can researchers avoid the problem of overfitting? Early stopping and pruning might help. Early stopping during the training can prevent a decision tree from taking in—and learning from—too much noisy data. However, knowing when to stop is tricky, as too early a pause in the training will yield inaccurate results. As for pruning, it essentially entails a reduction in the size of a decision tree: the parts that are pruned off are presumably parts that contribute little or not at all to the tree's classificatory powers. Just as there are ways to avoid overfitting, there are ways to avoid underfitting: among the proposed approaches are increased levels of dataset features, decreased levels of noisy data, and longer periods of decision-tree training.[202]

The purpose of reducing incidences of overfitting and underfitting is to strengthen the interpretive or predictive powers of decision trees. In the present study, I used decision trees for a purely interpretive, not predictive, purpose. I wanted to better understand the possible presence of poor reasoning in federal judges' handling of trademark-confusion cases. It is almost certainly the case

---

[201] *See What is a Decision Tree?*, IBM, https://www.ibm.com/topics/decision-trees [https://perma.cc/47DX-99E4] (last visited Feb. 24, 2025). Overfitting is an excessive adherence to training data, resulting in a model that cannot adequately generalize. The reasons for overfitting include insufficient training data size, excessive irrelevant data ("noise"), excessively lengthy training on a subset of the data, and excessive model complexity, which prompts the model to train on the noisy data. *See also* Byron Boots, *Decision Trees: Overfitting*, https://courses.cs.washington.edu/courses/cse446/20wi/Lecture4/04a_Overfitting.pdf [https://perma.cc/39Y2-EU7B] (last visited Mar. 21, 2025).

[202] Mark Last, Oded Maimon & Einat Minkov, *Improving Stability of Decision Trees*, 16 INT'L J. PATTERN RECOGNITION & A.I. 145, 148 (2002).

that early stopping and pruning for all three decision trees in the present study would have unacceptably distorted the results. Thus, I left the decision trees intact. Though less than ideal, this course of action was, as far as I can tell, the best one available.

<div align="center">Acknowledgements</div>