Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

Appendix

A. Data Collection

A.1 Fee Exemptions from Federal District Courts

In Table A.1, I provide a list of the district courts which granted fee exemptions to me. This list provided the basis for the sampling of one district court from each geographic circuit.

Table A.1: Fee Exemptions Obtained from Federal District Courts

1st Circuit	*2nd Circuit*	*3rd Circuit*	*4th Circuit*	*5th Circuit*	*6th Circuit*
D. Me.	D. Conn.	D.N.J.	D. Md.	W.D. La.	E.D. Ky.
D. Mass.	N.D.N.Y.	E.D. Pa.	D.N.C.	S.D. Miss.	W.D. Ky.
D.N.H.	W.D.N.Y.	M.D. Pa.	D.S.C.	E.D. Tex.	E.D. Mich.
D.R.I.	D. Vt.	W.D. Pa.	W.D. Va.	N.D. Tex.	W.D. Mich.
			S.D. W.Va.	W.D. Tex.	N.D. Ohio
					W.D. Tenn.
7th Circuit	*8th Circuit*	*9th Circuit*	*10th Circuit*	*11th Circuit*	*D.C. Circuit*
C.D. Ill.	E.D. Ark.	E.D. Cal.	D. Col.	M.D. Ala.	D.D.C.
S.D. Ill.	N.D. Iowa	N.D. Cal.	D. Kan.	S.D. Ala.	
N.D. Ind.	S.D. Iowa	S.D. Cal.	D.N.M.	M.D. Fla.	
S.D. Ind.	E.D. Mo.	D. Mont.	N.D. Okla.	N.D. Fla.	
W.D. Wis.	D. Neb.	D. Or.	W.D. Okla.	N.D. Ga.	
	D.S.D.	E.D. Wash.	D. Utah		
		W.D. Wash.			

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

A.2 Fee Exemptions from Federal District Courts

In Table A.2, I provide a list of the randomly sampled district courts which were included in the dataset, along with the number of observations for each sampled court.

Table A.2: Sampled Federal District Court

Circuit	District Court	Observations in Dataset
1st	D.R.I.	116
2nd	W.D.N.Y.	326
3rd	D.N.J.	453
4th	S.D. W.Va.	210
5th	W.D. Tex.	370
6th	W.D. Ky.	199
7th	S.D. Ill.	373
8th	D.S.D.	159
9th	E.D. Cal.	273
10th	D. Col.	458
11th	S.D. Ala.	244

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

A.3 Nature of Suit Codes

I identified potentially eligible cases from each sampled district using the nature of suit codes found in Table A.3.

Table A.3: Nature of Suit Codes

Cause of Action	Observations in Dataset
Civil Rights:	1,617
440 Other Civil Rights	797
441 Voting	6
442 Employment	776
443 Housing/Accommodations	37
444 Welfare	1
448 Education	0
Torts:	1,564
310 Airplane	31
315 Airplane Product Liability	8
320 Assault	18
330 Federal Employers Liability	53
340 Marine	90
345 Marine Product Liability	0
350 Personal Injury: Motor Vehicle	214
355 Motor Vehicle Product Liability	36
360 Other	306
362 Personal Injury: Medical Malpractice	129
365 Product Liability	677
367 Health Care/Pharmaceutical Personal Injury/Product Liability	0
368 Asbestos	2

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

A.4 Sampling

Using the nature of suit codes found in Appendix A.3, I identified eligible cases from each district in the sample for a total of 15,329 cases. Based on resource constraints, my goal was to sample approximately 7,665 with the understanding that a significant portion of these complaints would be ineligible because the original complaint had not been filed in federal court (a determination that could not be made before the sampling). I stratified the sampling by issue area, time period, and court to ensure sufficient coverage across these dimensions (Jawale 2012). I divided the target number of observations by the number of district courts. I then automatically included all cases in districts with fewer observations than this number. I continued doing this until there were no courts with fewer observations than their potential share. Among the remaining courts, I divided the remaining observations equally and sampled based on the share that remained. Thus, the districts were sampled with uneven probability. Finally, I identified and oversampled cases in which a motion to dismiss pursuant to Rule 12(b)(6) had been filed to allow for future research regarding outcomes. In order to account for this non-random sampling, I used appropriate probability weights in all analyses.

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

B. Coding Specificity

B.1 Specificity Scale

The research team carried out the coding using the following scale:

Table B.1: Specificity Scale

1. **Broadest possible pleading** - simply names the causes of action the plaintiff is claiming without discussing elements or facts.
2. **Very broad pleading** - names the causes of action the plaintiff is claiming and the elements of the claims but no specific facts.
3. **Broad** - provides basic factual information regarding the parties and nature of the claims.
4. **Somewhat broad** - provides a few specific factual details regarding the causes of actions.
5. **Middle, broad** - provides some specific factual allegations regarding the causes of actions generally.
6. **Middle, specific** - provides some specific factual allegations regarding some elements of the causes of action.
7. **Somewhat specific** - provides some specific factual allegations regarding most elements of the causes of action.
8. **Specific** - provides detailed specific factual allegations regarding most elements of the causes of action.
9. **Very specific** - provides detailed factual allegations regarding most elements of the causes of action.
10. **Most specific possible** - provides detailed factual allegations covering all elements of the causes of action.

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

Examples of complaints coded in the 10-point scale include:

Table B.2: Specificity Scale Example Complaints

Specificity Scale	**Case Citation**
2	*Miller v. Warrior & Gulf Navigation Co.*, No. 1:07-cv-00113 (S.D. Ala. 2007)
3	*Forkell v. Wright Tree Service, Inc.*, No. 3:09-cv-00486 (W.D. Ky. 2010)
4	*Slawson v. Tri-State Concrete Pumping Co., Inc.*, No. 3:07-cv-00485 (S.D. W.Va. 2007)
5	*Archard v. Potter*, No. 1:08-cv-00746 (S.D. Ala. 2009)
6	*Nguyen v. Boeing Company, Inc.*, No. 1:09-cv-02208 (D. Col. 2009)
7	*Baca v. Calderon*, No. 2:07-cv-01216 (E.D. Cal. 2007)
8	*Martin v. Johnson & Johnson*, No. 2:06-cv-05719 (D.N.J. 2006)
9	*Tarr v. September Schools*, No. 1:07-cv-01684 (D. Col. 2007)
10	*Edwards v. Gulf Stream Coach*, No. 1:09-cv-00474 (S.D. Ala. 2009)

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

B.2 Conversion to a Binary Measure of Specificity

The data was initially collected with the high level of demarcation, found in Table B.1, to allow for the possible expansion of the data set with the finer grain measure. However, estimating the classifier with 362 observations for an ordinal scale of 1-10 creates the risk of data sparsity and model dependence. Furthermore, small differences in specificity are more difficult to detect that the differences between generally specific and general pleadings. A comparison of the recall of the measures and ordinal and binary level indicates these concerns are warranted and that the binary measure is preferable in the interest of inference: the recall using the ordinal scale is 0.46, while the binary scale increases recall to 0.83, over double that of the ordinal scale.

Furthermore, the decision to use the binary measure does not drive the main results of the model. In fact, they are stronger using the ordinal scale for both the classification and analysis. The results of analyses on the classified observations can be found in Table B.3:

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

Table B.3: Ordinal Logistic Regression - Specificity of Classified Observations

Variable	**Coefficient**	**Std. Error**	**p-value**
Pre-*Twombly*, Torts	-0.62*	0.22	0.01
Post-*Twombly*, Torts	-0.88*	0.24	<0.01
Pre-*Iqbal*, Torts	-0.64*	0.16	<0.01
Pre-*Twombly*, Civil Rights	0.57*	0.26	0.03
Post-*Twombly*, Civil Rights	0.83*	0.27	<0.01
Pre-*Iqbal*, Civil Rights	0.55*	0.21	<0.01
Post-*Iqbal*, Civil Rights	-0.45*	0.16	<0.01
District Median Ideology	0.39	0.45	0.38
Alabama Southern	-0.34	0.23	0.13
California Eastern	0.29	0.18	0.10
Colorado	0.02	0.20	0.91
Illinois Southern	1.00*	0.52	0.05
Kentucky Western	-0.55	0.49	0.26
New York Western	0.41	0.29	0.15
Rhode Island	1.31*	0.47	0.01
South Dakota	-0.69	0.50	0.17
Texas Western	-0.18	0.21	0.39
West Virginia Southern	-1.14*	0.30	<0.01
Cutpoint 4	-3.11	0.36	
Cutpoint 5	-1.11	0.35	
Cutpoint 6	-0.52	0.34	
Cutpoint 7	0.02	0.35	
Cutpoint 8	0.84	0.35	
Cutpoint 9	3.06	0.38	
Cutpoint 10	5.16	0.46	

n=2,629
*Indicates significant at the .05, two-tailed level

The main difference is that ideology is no longer significant, though the estimate is still in the same direction.

C. Classifier

C.1. Classifier Features

The elements of the classifier are:

Word Count

Length generally indicates specificity (Huber and Shipan 2002; Louis and Nenkova 2011). The potential relationship between the length of complaints and the specificity demanded by *Twombly* and *Iqbal* has been identified by judges and attorneys. For example, Judge Posner noted: "Since a plaintiff must now show plausibility, complaints are likely to be longer – and legitimately so – than before *Twombly* and *Iqbal*" (*Kadamovas v. Stevens*). A complaint littered with vague claims, however, may be extended due to the number of causes of action (Boyd et al. 2013). Therefore, I included a measure for the length of the complaint and the length divided by the number of claims. The word count was obtained via the Linguistic Inquiry and Word Count (LIWC), a commercially available text analysis program. Research assistants hand-coded the number of claims.

Polarity

Louis and Nenkova (2011) find that non-neutral words, such as negative and positive words, are associated with general statements. This is not surprising as such language is generally associated with subjective rather than objective description. To capture polarity, I used the LIWC measures for the percent of words related to negative and positive emotion (Pennebaker et al. 2007).

Word Specificity

To capture the specificity of the individual words, I utilized five measures:

WordNet Depth

The first measure is based on the publicly available content-analytic algorithm WordNet. The WordNet lexical database provides groups of words (synsets) that share a common meaning (i.e., synonyms) (Miller 1995). These synsets are linked to other synsets through semantic relationships. All synsets are ultimately linked to the root synset, which is the word with the broadest meaning within the group. Thus, the distance of a word from the root synset provides a measure of specificity: the closer a word is to the root synset, the broader the meaning (*see* Louis and Nenkova 2011). Using the algorithm via Python, I measured the longest distances between each recognized noun in the complaint to a root synset. Then I calculated the mean distances for each factual allegation. This measure indicates the average specificity of the words used in the complaint.

Inverse Document Frequency

Additionally, the classifier includes measures based on inverse document frequencies ("IDF") to capture how many complaints the word appears in, out of all relevant complaints (Louis and Nenkova 2011, citing Joho and Sanderson 2007). I measured the IDF of the words in complaints by each cause of action using the NLTK Toolkit in Python. The classifier includes the IDF for the 50th percentile and 95th percentile of the distributions. There is insufficient variation in the 5^{th} percentile to warrant its inclusion. This allowed me to capture how rare the median words are in addition to the rarest set of words. The more unusual the words, the more likely the document is specific.

Specific Types of Words

I also include measures for the percent of numbers described in text and number of proper nouns, as both indicate detailed descriptions. The numbers measure was calculated via LIWC. Proper nouns were identified via Python as words tagged as nouns that are capitalized and not at the beginning of a sentence.

<u>Causation</u>

Establishing how a defendant caused harm to the plaintiff is a defining feature of civil claims in the American legal system (Hart and Honoré 1959). Thus, the percentage of causation words found in factual allegations is an important marker of a specific claim. The measure is based on the LIWC dictionary for causation language. To the extent that litigants are describing causal mechanisms, they should be more likely to be specific in their allegations.

Certainty

Additionally, the new pleading standard was crafted to deal with speculative claims. Speculative claims tend to be housed in more tentative language to avoid potential sanctions. Thus, the percentage of language reflecting certainty is potentially a useful measure of the type of specificity the Supreme Court sought. This measure was calculated via LIWC.

Number of Claims

One way of conceptualizing complaint specificity is the number of causes of action (Boyd et al. 2013). When a plaintiff is pleading broadly, he will often use a shotgun effect alleging many claims that do little more than restate the elements. More specific complaints should contain fewer causes of action. Therefore, I include the count of claims in the classifier. Where the drafting attorney included headers identifying claims, the count was based on those demarcations. Where the attorney did not identify separate claims, a member of the research team read the text and counted the distinct claims based on common elements of a cause of action.

Many components of my classifier are the same as Louis and Nenkova (2011), but there are some differences. First, my unit of analysis is the factual allegations section of the complaint rather than a sentence in a news summary. I also consider two additional types of language that I believe are relevant in pleadings: causation and certainty language. A count for the number of claims is also included (Boyd et al. 2013). Additionally, I use the target words for negative and positive emotions in the Linguistic Inquiry and Word Count (LIWC), a commercially available text analysis program, to help identify polar words. Louis and Nenkova (2011) also use measures based on the General Inquirer (Stone, Dunphy, and Smith 1966) and MPQA Subjectivity Lexicon (Wilson, Wiebe, and Hoffmann 2009). I also do not include all measures of NE+CS. Specifically, dollar signs and plural nouns are not obviously related to specificity in complaints. Finally, they use other syntax features that apply to news summaries that I do not believe carry over to legal writing and, thus, do not use. Finally, Louis and Nenkova (2011) include language models based on unigrams, bigrams, and trigrams, which I do not, due to the unavailability of a corpus of

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

relevant text upon which to build the dictionary.

Supplemental Material (not copyedited or formatted) for: Morgan L.W. Hazelton. 2021.
"Judicial Impact and Factual Allegations: How the Supreme Court Changed Civil Procedure through the Plausibility Standard."
Journal of Law and Courts 9(1). DOI: https://doi.org/10.1086/712653.

C.2. Estimation of the Classifier

The results of the model that underlies the classifier are in Table C.1:

Table C.1: Logistic Classifier – Allegation Specificity

Variable	Coefficient	Std. Error	p-value
Word Count	0.002*	0.001	0.017
Word Count by Claim	0.001	0.001	0.135
Positive Emotion	0.210	0.168	0.211
Negative Emotion	0.043	0.097	0.658
WordNet Depth	1.065*	0.466	0.022
Inverse Document Frequency 50%	2.354	1.945	0.226
Inverse Document Frequency 95%	-0.774*	0.357	0.030
Numerals	0.136	0.130	0.295
Proper Nouns	-0.002	0.003	0.556
Causation	0.171	0.165	0.300
Certainty	-0.011	0.296	0.969
Claims	-0.052	0.126	0.679
Intercept	-10.611*	3.683	0.004

n=362
*Indicates significant at the .05, two-tailed level

Ultimately, the role of the classifier is to create accurate classifications, as opposed to confirming or denying hypotheses regarding the use of language. Regardless, the estimates are interesting. The roles of Word Count and WordNet Depth are significant and operated as anticipated. Inverse Document Frequency for the 95% of the complaint operated in the opposite way than expected: the rarer these words are, the less likely the complaint is to be specific holding all other variables constant. This unexpected estimate is likely the results of the inclusion of the Inverse Document Frequency for the 50% of the complaint. Holding the rarity of the words in the middle of the distributions constant, those complaints with exceedingly rare words near the tail tend to be less specific.

D. Alternative Specifications: Data and Models

D.1 Alternative Specifications: Hand-Coded, Binary Dependent Variable

In Table D.1, are the results for the main model run on the hand-coded observations with the binary measure of specificity as the dependent variable. While the results of this analysis are largely consistent with the main model in the paper, there are some differences including the fact that the sign for district median ideology is in the opposite direction (negative). These differences are likely due to the relative paucity of data for the hand-coded observations versus classified observations: the hand-coded data is only around one-fifth of the observations compared with the classified data. Based on the number of parameters that are being estimated, the smaller dataset is more likely to produce results based on data sparsity and model dependence.

Table D.1: Logit - Specificity of Hand-Coded Observations

Variable	**Coefficient**	**Std. Error**	**p-value**
Pre-*Twombly*, Torts	-1.14*	0.50	0.02
Post-*Twombly*, Torts	-0.73	0.51	0.15
Pre-*Iqbal*, Torts	-0.89*	0.40	0.03
Pre-*Twombly*, Civil Rights	1.59*	0.67	0.02
Post-*Twombly*, Civil Rights	0.52	0.65	0.43
Pre-*Iqbal*, Civil Rights	1.16*	0.57	0.04
Post-*Iqbal*, Civil Rights	-0.64	0.40	0.11
District Median Ideology	-5.59*	2.27	0.01
Alabama Southern	-1.40	0.72	0.05
California Eastern	0.42	0.49	0.40
Colorado	-1.50	0.79	0.06
Illinois Southern	-5.32*	2.55	0.04
Kentucky Western	-6.31*	2.29	0.01
New York Western	-2.57*	1.28	0.05
Rhode Island	-3.69	2.11	0.08
South Dakota	-4.62	2.42	0.06
Texas Western	-2.23*	0.77	<0.01
West Virginia Southern	-0.88	0.85	0.30
Intercept	4.70*	1.75	0.01

n=552

*Indicates significant at the .05, two-tailed level

D.2 Alternative Specifications: Classified, Binary Dependent Variable

In Table D.2, are the results of the main model using a binary measure of specificity derived from the fitted values using .5 at the cut point. These results are consistent with those reported in the Table 2 using the fitted values.

Table D.2: Logit - Specificity of Classified Observations

Variable	Coefficient	Std. Error	p-value
Pre-*Twombly*, Torts	-0.42*	0.21	0.04
Post-*Twombly*, Torts	-0.53*	0.21	0.01
Pre-*Iqbal*, Torts	-0.51*	0.18	0.01
Pre-*Twombly*, Civil Rights	0.36	0.29	0.2
Post-*Twombly*, Civil Rights	0.54	0.28	0.06
Pre-*Iqbal*, Civil Rights	0.46	0.26	0.07
Post-*Iqbal*, Civil Rights	-0.20	0.18	0.27
District Median Ideology	1.57*	0.68	0.02
Alabama Southern	0.01	0.26	0.96
California Eastern	0.86*	0.24	<0.01
Colorado	0.62*	0.27	0.02
Illinois Southern	2.27*	0.77	<0.01
Kentucky Western	0.90	0.7	0.2
New York Western	1.43*	0.41	<0.01
Rhode Island	2.64*	0.68	<0.01
South Dakota	0.67	0.74	0.37
Texas Western	0.21	0.26	0.43
West Virginia Southern	-0.71	0.42	0.09
Intercept	-0.70	0.5	0.16

n=2,629
*Indicates significant at the .05, two-tailed level