"Loyalty and Agency in Economic Theory" by Richard R. W. Brooks

Discussion prepared by Daniel L. Chen

June 2018

Research Question

Three conceptions of loyalty

- "Structural" - instrumentalized loyalty (e.g., self-interested reciprocity)
  - "discourage disloyalty" (p. 2)
- "Characterological" - screening for loyal types
  - "encourage loyalty" (p. 2)
- "Behavioral" - rule-based loyalty (e.g., duty)
  - self-abnegation (p. 1), obligations (p. 12), commitment (p. 13)

Application to law & economics scholarship

Commons, Coase, Alchian & Demsetz, Marshak, Williamson, Buchanan, ...

Research Question

Three conceptions of loyalty

- "Structural" - instrumentalized loyalty (e.g., self-interested reciprocity)
  - "discourage disloyalty" (p. 2)
- "Characterological" - screening for loyal types
  - "encourage loyalty" (p. 2)
- "Behavioral" - rule-based loyalty (e.g., duty)
  - self-abnegation (p. 1), obligations (p. 12), commitment (p. 13)

Application to law & economics scholarship

Commons, Coase, Alchian & Demsetz, Marshak, Williamson, Buchanan, ...

Comments

Internal comments

- Clarity of typology (tightness of examples)
- Connection of terminology to economics literature

Theoretical comments

- Behavioral loyalty
  - game & model seems consequentialist, not rule-based
- Thought experiment for existence of rule-based loyalty
  - Shredding criterion for non-consequentialist motives

Textual analysis

- Making doctrinal work rigorous (U Chi L Rev 2017)
- Word embeddings (distinguishing loyalty from obedience)

Comments

Internal comments

- Clarity of typology (tightness of examples)
- Connection of terminology to economics literature

Theoretical comments

- Behavioral loyalty
  - game & model seems consequentialist, not rule-based
- Thought experiment for existence of rule-based loyalty
  - Shredding criterion for non-consequentialist motives

Textual analysis

- Making doctrinal work rigorous (U Chi L Rev 2017)
- Word embeddings (distinguishing loyalty from obedience)

Comments

Internal comments

- Clarity of typology (tightness of examples)
- Connection of terminology to economics literature

Theoretical comments

- Behavioral loyalty
  - game & model seems consequentialist, not rule-based
- Thought experiment for existence of rule-based loyalty
  - Shredding criterion for non-consequentialist motives

Textual analysis

- Making doctrinal work rigorous (U Chi L Rev 2017)
- Word embeddings (distinguishing loyalty from obedience)

Comments

Internal comments

- Clarity of typology (tightness of examples)
- Connection of terminology to economics literature

Theoretical comments

- Behavioral loyalty
  - game & model seems consequentialist, not rule-based
- Thought experiment for existence of rule-based loyalty
  - Shredding criterion for non-consequentialist motives

Textual analysis

- Making doctrinal work rigorous (U Chi L Rev 2017)
- Word embeddings (distinguishing loyalty from obedience)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

Internal comments

How is it different from existing vocabulary

- "Structural" - moral hazard (p. 5)
- "Characterological" - adverse selection, screening (p. 7)
  - "Black" names *limiting outside options* is structural loyalty (p. 9-10)
  - But Fryer-Levitt: "signaling model appears to fall short" (p. 790)
- "Behavioral" - actions independent from considerations of self-interest
  - Hirschman's "reasoned expectation of reform" (is this *instrumentalized* loyalty?)
  - Loyalty game is identical to trust game
  - Berge equilibrium (is there any evidence for this? 186 GS papers.. 58K for social preferences)
  - The utility function seems a special case of altruist motives?

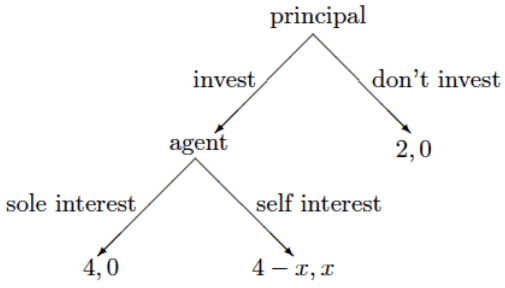
Are the scholars really discussing loyalty

- 1 of the 6 use the term "loyalty" (in institutional economics Section 5.1)

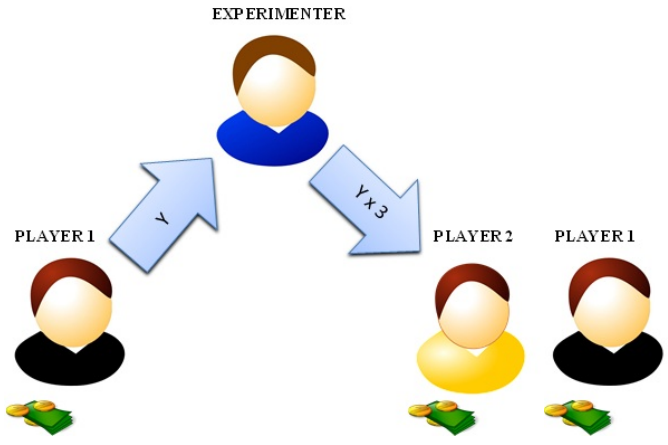
Game and model

Loyalty game is identical to trust game

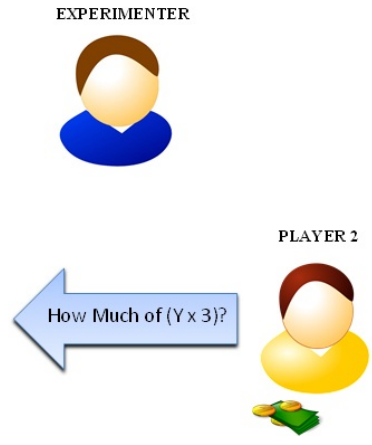
Loyalty Game



Trust Game



Trust Game



Loyalty model is consequentialist

consider a nexus of N contracts between a corporate entity and its various contractual partners, including shareholders, bondholders, various commercial and trade creditors, employees, customers, and so on, indexed by $i = \{1, 2, 3, ..., N\}$. Define the value that the firm derives from each contractual relationship as $v_i(e_i, \cdot)$, where e_i is some measure of non-contractable effort or investment made by the firm's contractual partner, i, that monotonically increases the firm's value of the contract.[48] For instance, the firm

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Why do we need new vocabulary?

Generations of domains of preferences:

- Agent's material consequences (homo oeconomicus) (structural and characaterological loyalty?)
- Agent's and others' material consequences (e.g. Fehr-Schmidt inequity aversion 1999, pure altruism, self-sacrifice, Rabin fairness 1994) (behavioral loyalty?)
- Agent's and others' material consequences, social audience (e.g. Andreoni impure altruism 1989, Sugden reciprocity 1984, McCabe et al signalling intentions 2003 or Benabou-Tirole type 2006, Battigalli et al guilt aversion psychological games 2007)
- Agent's and others' material consequences, social audience, and purely internal consequences (e.g., duty/deontological motivations, Smith's impartial spectator 1761)
- 2 x 4 categorization (agent's v. principal's perspective) (e.g. Ashraf & Bandiera, Acemoglu & Jackson | Wolitzky, etc.)
- Or say what the alternative typology predicts/explains

Typology (Sobel "Interdependent Preferences and Reciprocity" JEL 2005)

- Monetary payoffs of others to enter a decision-maker's utility.
  (e.g., Berge equilibrium, ...)

- Chicago School (preferences over general commodities transformed into consumption goods)
- Identity models (utility function over actions and an identity that incorporates the prescriptions that indicate the identity-appropriate behavior Akerlof-Kranton 2000)

- However
  - Agents choose between quantities (in Chicago models)
    - but do not have preferences over *choices* separate from preferences over *quantities*
  - Agents choose acts (in Identity models)
    - but do not have preferences over *acts* separate from preferences over *consequences of acts*.

Typology (Sobel "Interdependent Preferences and Reciprocity" JEL 2005)

- Monetary payoffs of others to enter a decision-maker's utility. (e.g., Berge equilibrium, ...)
- Chicago School (preferences over general commodities transformed into consumption goods)
- Identity models (utility function over actions and an identity that incorporates the prescriptions that indicate the identity-appropriate behavior Akerlof-Kranton 2000)
- However
  - Agents choose between quantities (in Chicago models)
    - but do not have preferences over *choices* separate from preferences over *quantities*
  - Agents choose acts (in Identity models)
    - but do not have preferences over *acts* separate from preferences over *consequences of acts*.

Typology (Sobel "Interdependent Preferences and Reciprocity" JEL 2005)

- Monetary payoffs of others to enter a decision-maker's utility. (e.g., Berge equilibrium, ...)
- Chicago School (preferences over general commodities transformed into consumption goods)
- Identity models (utility function over actions and an identity that incorporates the prescriptions that indicate the identity-appropriate behavior Akerlof-Kranton 2000)
- However
  - Agents choose between quantities (in Chicago models)
    - but do not have preferences over *choices* separate from preferences over *quantities*
  - Agents choose acts (in Identity models)
    - but do not have preferences over *acts* separate from preferences over *consequences of acts*.

Typology (Sobel "Interdependent Preferences and Reciprocity" JEL 2005)

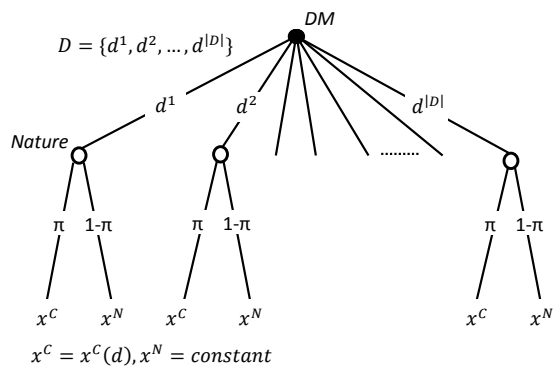
- Monetary payoffs of others to enter a decision-maker's utility. (e.g., Berge equilibrium, ...)
- Chicago School (preferences over general commodities transformed into consumption goods)
- Identity models (utility function over actions and an identity that incorporates the prescriptions that indicate the identity-appropriate behavior Akerlof-Kranton 2000)
- However
  - Agents choose between quantities (in Chicago models)
    - but do not have preferences over *choices* separate from preferences over *quantities*
  - Agents choose acts (in Identity models)
    - but do not have preferences over *acts* separate from preferences over *consequences of acts*.

Duty | Behavioral Loyalty

Hypothetical vs. Categorical Imperative

economic models have thus far focused on the *hypothetical imperative*–preferences over acts because of their consequences–rather than the *categorical imperative*–preferences over acts regardless of their consequences (Kant's axe murderer vignette)

Shredding Criterion for Non-Consequentialist Motivations



Consider preferences over x_1, d (Chen and Schonger, 2015)

- $u(x_1, d) = f(x_1) + b(d)$

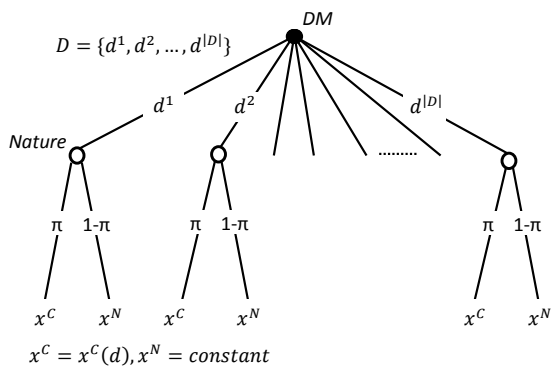
Perceived legitimacy motivates obedience irrespective of sanction likelihood (Tyler 1997)

Duty | Behavioral Loyalty

Hypothetical vs. Categorical Imperative

economic models have thus far focused on the *hypothetical imperative*–preferences over acts because of their consequences–rather than the *categorical imperative*–preferences over acts regardless of their consequences (Kant's axe murderer vignette)

Shredding Criterion for Non-Consequentialist Motivations



Consider preferences over x_1, d (Chen and Schonger, 2015)

- $u(x_1, d) = f(x_1) + b(d)$

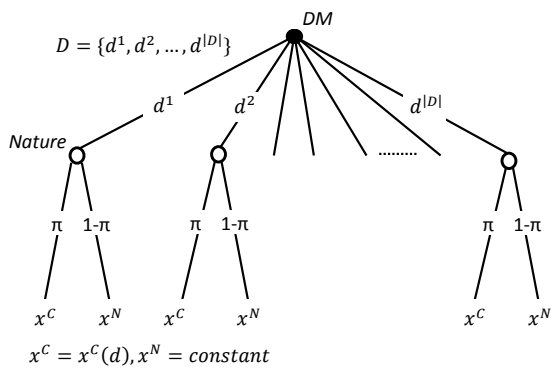
Perceived legitimacy motivates obedience irrespective of sanction likelihood (Tyler 1997)

Duty | Behavioral Loyalty

Hypothetical vs. Categorical Imperative

economic models have thus far focused on the *hypothetical imperative*–preferences over acts because of their consequences–rather than the *categorical imperative*–preferences over acts regardless of their consequences (Kant's axe murderer vignette)

Shredding Criterion for Non-Consequentialist Motivations



Consider preferences over x_1, d (Chen and Schonger, 2015)

- $u(x_1, d) = f(x_1) + b(d)$

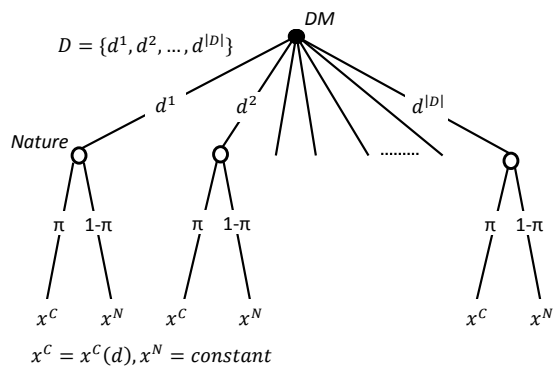
Perceived legitimacy motivates obedience irrespective of sanction likelihood (Tyler 1997)

Duty | Behavioral Loyalty

Hypothetical vs. Categorical Imperative

economic models have thus far focused on the *hypothetical imperative*–preferences over acts because of their consequences–rather than the *categorical imperative*–preferences over acts regardless of their consequences (Kant's axe murderer vignette)

Shredding Criterion for Non-Consequentialist Motivations



Consider preferences over x_1, d (Chen and Schonger, 2015)

- $u(x_1, d) = f(x_1) + b(d)$

Perceived legitimacy motivates obedience irrespective of sanction likelihood (Tyler 1997)

Textual Analysis 1: Commons

Principals, superiors, employers, patrons and the like all, to be sure, expect loyalty. On what basis, according to Commons, will loyalty secured? A traditional rational choice approach would look to incentives (structural loyalty) or to preferences (characterological loyalty), but Commons considered that approach limiting, if not misleading. He instead identified what he thought to be a more promising direction in Wesley Hohfeld's analysis of legal entitlements.[75] Hohfeld's conceptualization of entitlements was, to Commons, nothing short of a general theory of conduct rules,[76] shedding light on "the way in which the common practices of *any* going concern control the individual members of that concern and hold them to the conduct necessary to preserve the existence of the concern."[77] Bentham's actively

"common practices of any concern" = norms?

Textual Analysis 2: Coase

Contractual compliance offers an alternative to this "complex of habits, practices, opinions, promises and customs" operating as "a highly intractable force" (in Commons' words) on persons in association with eachother.[78] These alternatives, however, are not mutually exclusive. Contracts can and often do establish a basis for the behavioral loyalty of agents. Consider, for instance, Coase's foundational 1937 article, "The Nature of the Firm," wherein he conceded "[i]t is true that contracts are not eliminated [by agency] but they are greatly reduced." A "series of contracts" is replaced by one, which is then governed by "the legal relationship normally called that of 'master and servant' or 'employer and employee'." Contract creates the relationship—where "for a certain remuneration (which may be fixed

---

or fluctuating)," an agent "agrees to *obey* the directions of an entrepreneur *within certain limits*"—thereafter it is the conduct rules and norms of masters and servants that control the everyday order and expectations within the relationship.[79] That which is taken for granted in the relationship,

"agrees to obey" follows "for a certain remuneration" = instrumentalized?

Textual Analysis 3: Marschak & Radner

Marschak & Radner (1972) adopt a strong form of behavioral loyalty, at least in the "common interest" multi-party interactions they considered, wherein non-cooperative conduct was assumed away. Williamson (1985), for his own part, assumes an unflinching structural loyalty stance, showing little patience for self-suppressing behavioral loyalty and obedience arguments. While Coase could see the operation some of obedience 'within limits,' Williamson saw it as all-or-nothing. "Obedience is tantamount to

---

non-self-interest seeking,"[92] he writes, suggesting no overlap between self-interest and behavioral loyalty. Recalling Machiavelli's counsel to see "men as they are," he draws on the "primitive response" in all structural loyalty models.[93] Deter the *natural* tendency of men toward disloyalty with "appropriate safeguards." His take on this old stratagem stresses the ad-

"common interest" = common value utility function?

"obedience" = loyalty?

Textual Analysis 4: Coase

Obedience, the correlative to the master's authority, is the essence of what it means to be a loyal servant here. A master-servant relation, to be sure, is not a master-slave one, but that fact does not render the former simply contractual. Slavery and strict contractual compliance do not exhaust the scope of possibility for securing loyalty from servants. Coase appears to identify the loyalty of servants with a broad, though not unlimited, duty of obedience.[82] As such, their actions and choices may follow from behavioral loyalty, separate and apart from incentives provided by the

"obedience" = loyalty?

Textual Analysis 5: Buchanan

From this simple fact, Buchanan concludes that "the attempted separation between economics and morals was, at best, an illusion that simply cannot be sustained."[107] To demonstrate the inescapable link 'between economics and morals,' he initially appears to advance a behavioral loyalty argument, very much in line with the writings of Amartya Sen on commitment. "For many, perhaps most, of those who claim membership in socially organized communities," Buchanan writes, "a descriptive model of behav-

---

ior would require recognition of the presence of endogenous constraints on choice options."[108] Questions unaddressed or unanswerable, empirically or theoretically by the standard models, revealed the limits of the conventional rational self-interested approach. "Why do many persons seem deliberately to act contrary to the dictates of their own preference orderings? How can those alternatives be rejected that seem, objectively, to promise higher utility yields than those alternatives eligible for choice?"[109]

"deliberately to act contrary to preference orderings" = heuristics? club good?

Textual Analysis 6: Alchian & Demsetz

An effective structural loyalty tool is therefore available, argue Alchian & Demsetz, "[i]f one could enhance a common interest in nonshirking in the *guise* of a team loyalty or team spirit, the team would be more effi-

"loyalty"

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

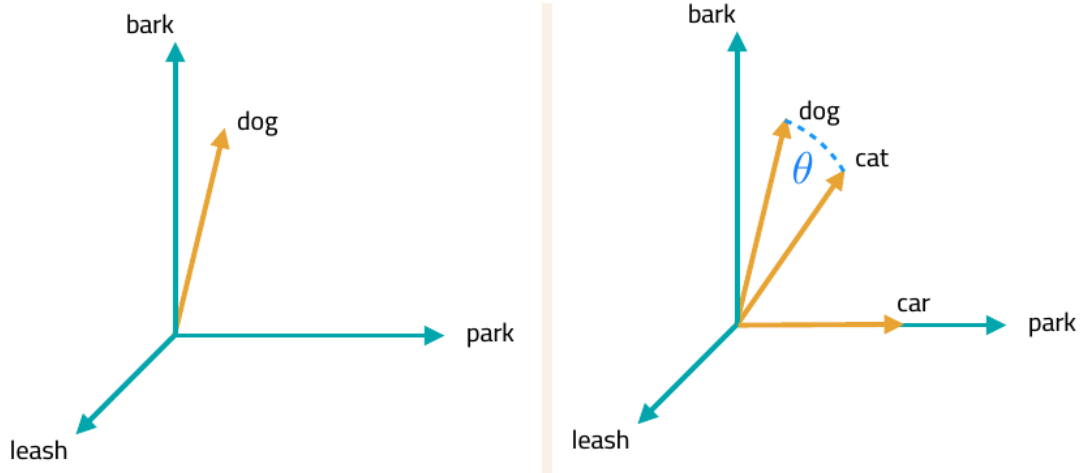
1. Finding the degree of similarity between two words.
```
model.similarity('woman','man')
0.73723527
```
2. Finding odd one out.
```
model.doesnt_match('breakfast cereal dinner
lunch';.split())
'cereal'
```
3. Amazing things like woman+king-man =queen
```
model.most_similar(positive=
['woman','king'],negative=['man'],topn=1)
queen: 0.508
```
4. Probability of a text under the model
```
model.score(['The fox jumped over the lazy
dog'.split()])
0.21
```

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Relatedness between words

How does it work? Predict given a word using surrounding words

Words as Vectors



- Use cosine similarity as a measure of relatedness:

$$\cos\theta = \frac{v_1 \cdot v_2}{||v_1||||v_2||}$$

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

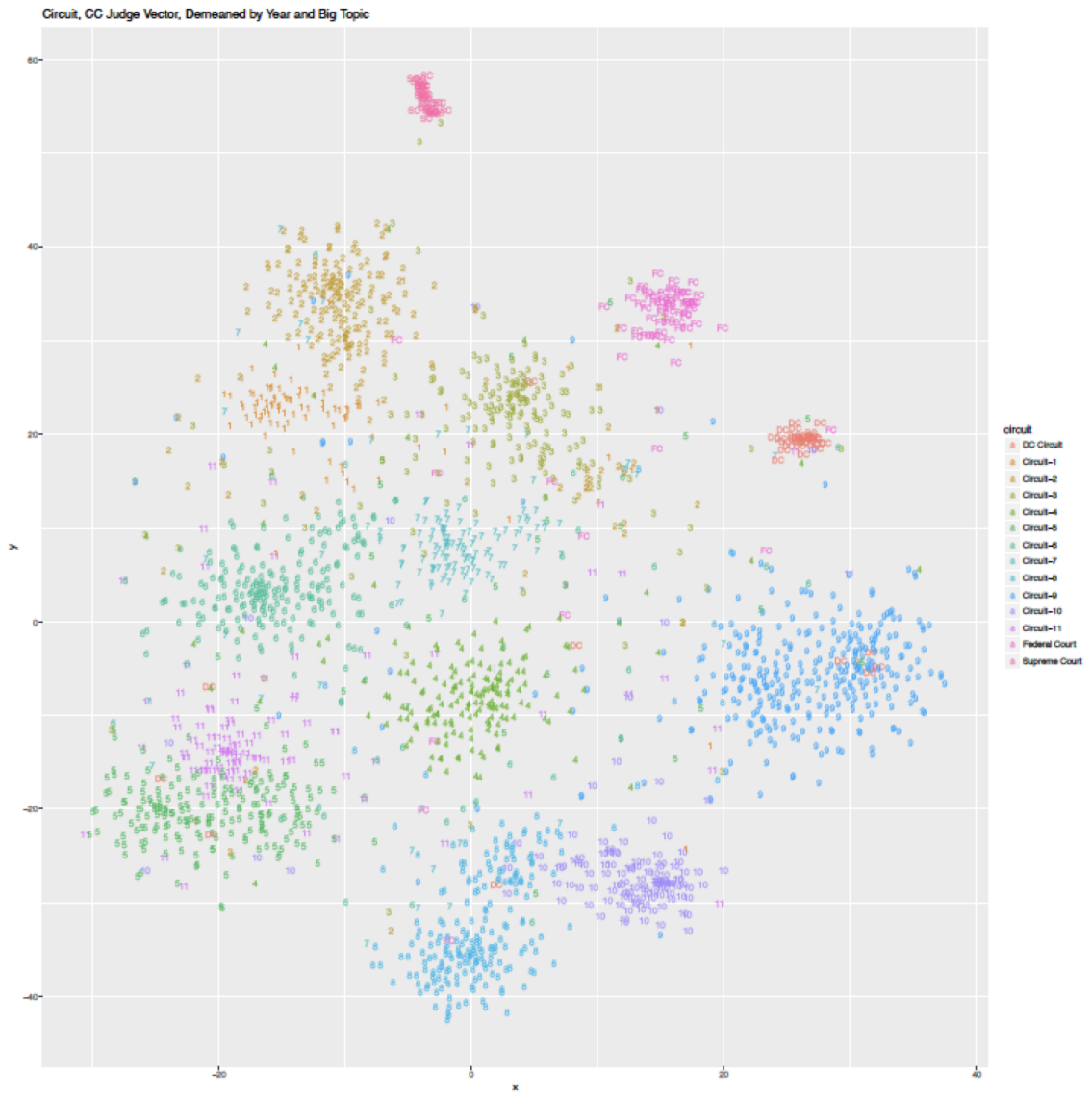
- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

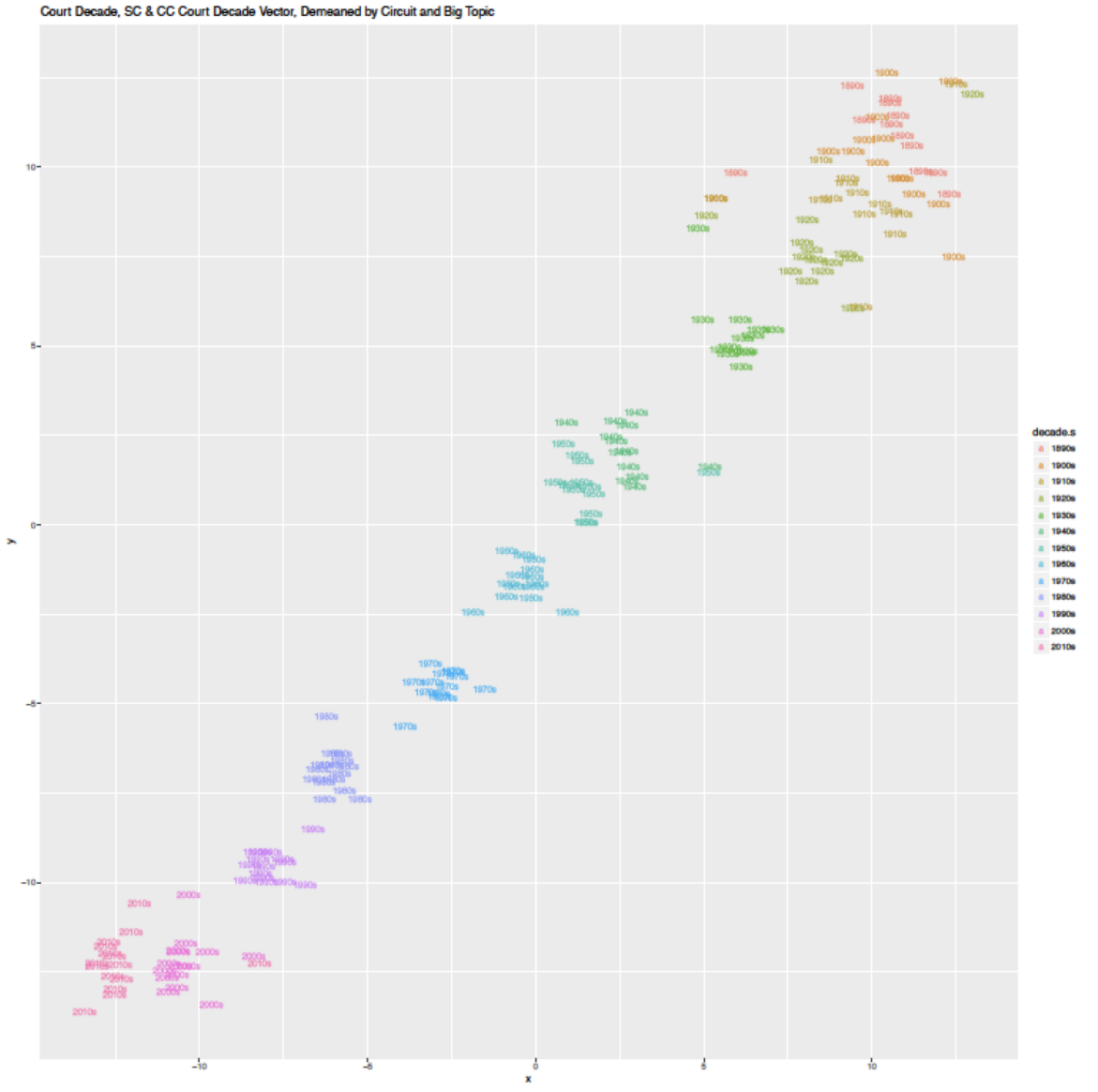
Visual Structure of Case Vectors by Circuit

Figure 1: Centered by Topic-Year, Averaged by Judge, Labeled by Court



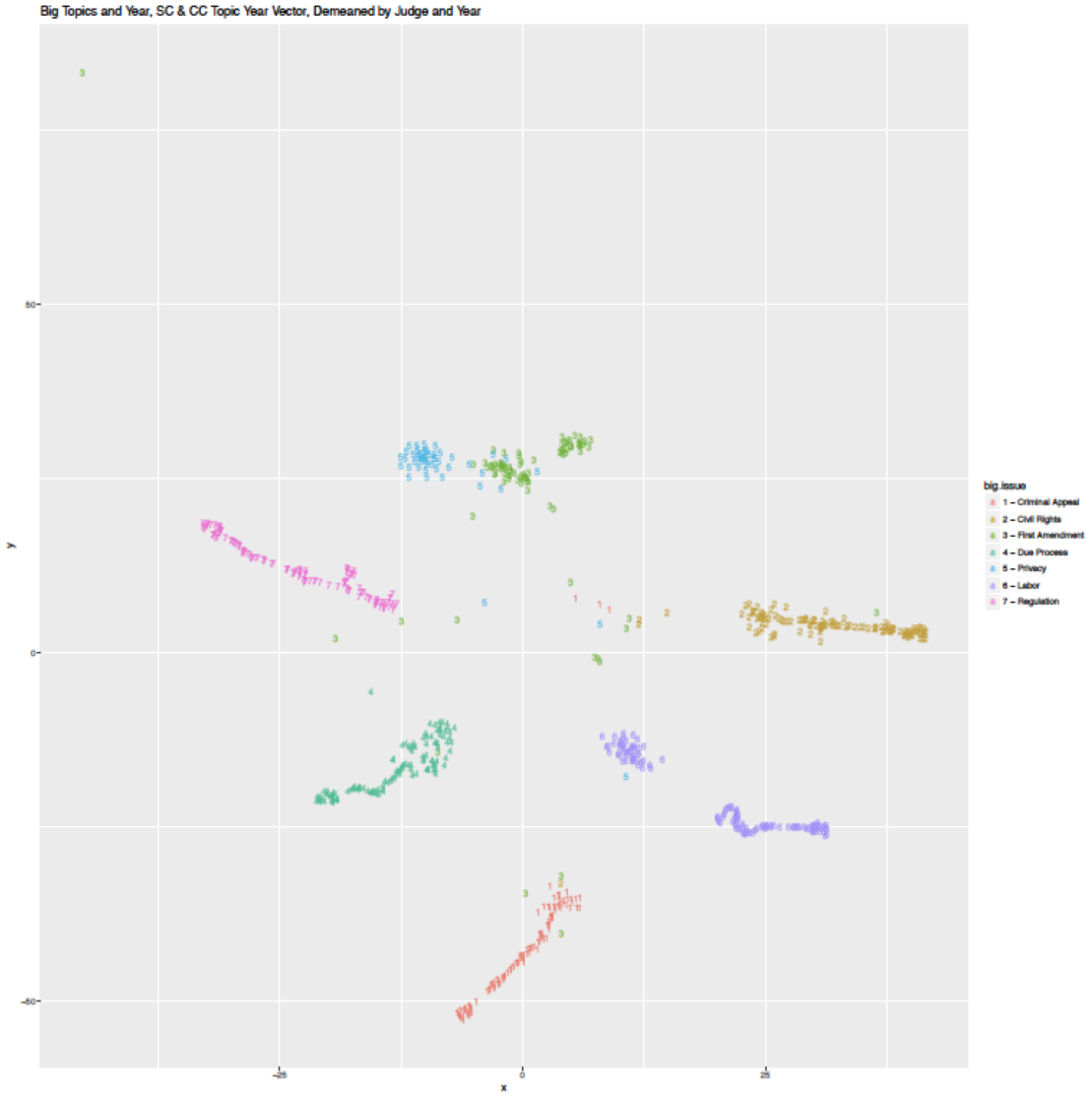
Case Vectors by Decade

Figure 2: Centered by Court-Topic, Averaged by Court-Year, Labeled by Decade



Case Vectors by Topic

Figure 3: Centered by Judge-Year, Averaged by Topic-Year, Labeled by Topic

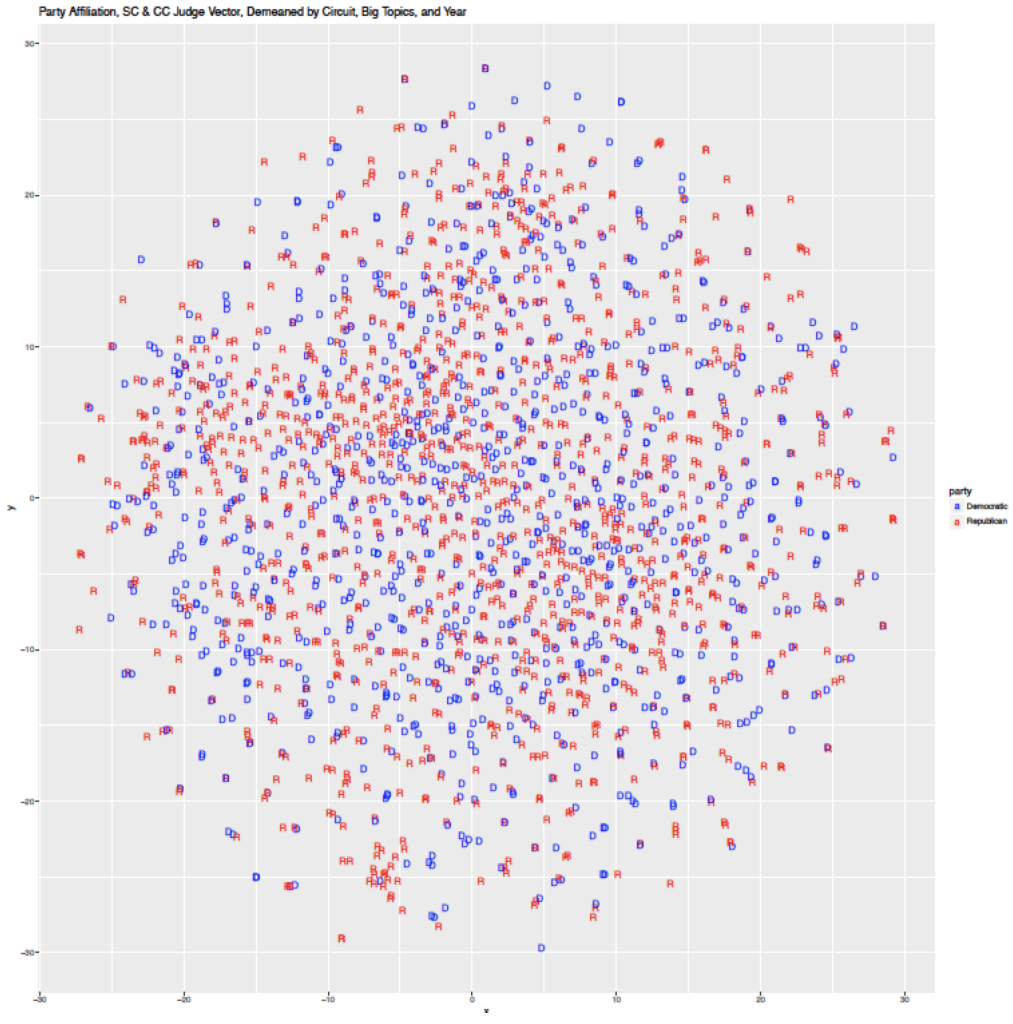


Case Vectors by Birth Cohort

Figure 5: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Judge Birth Cohort

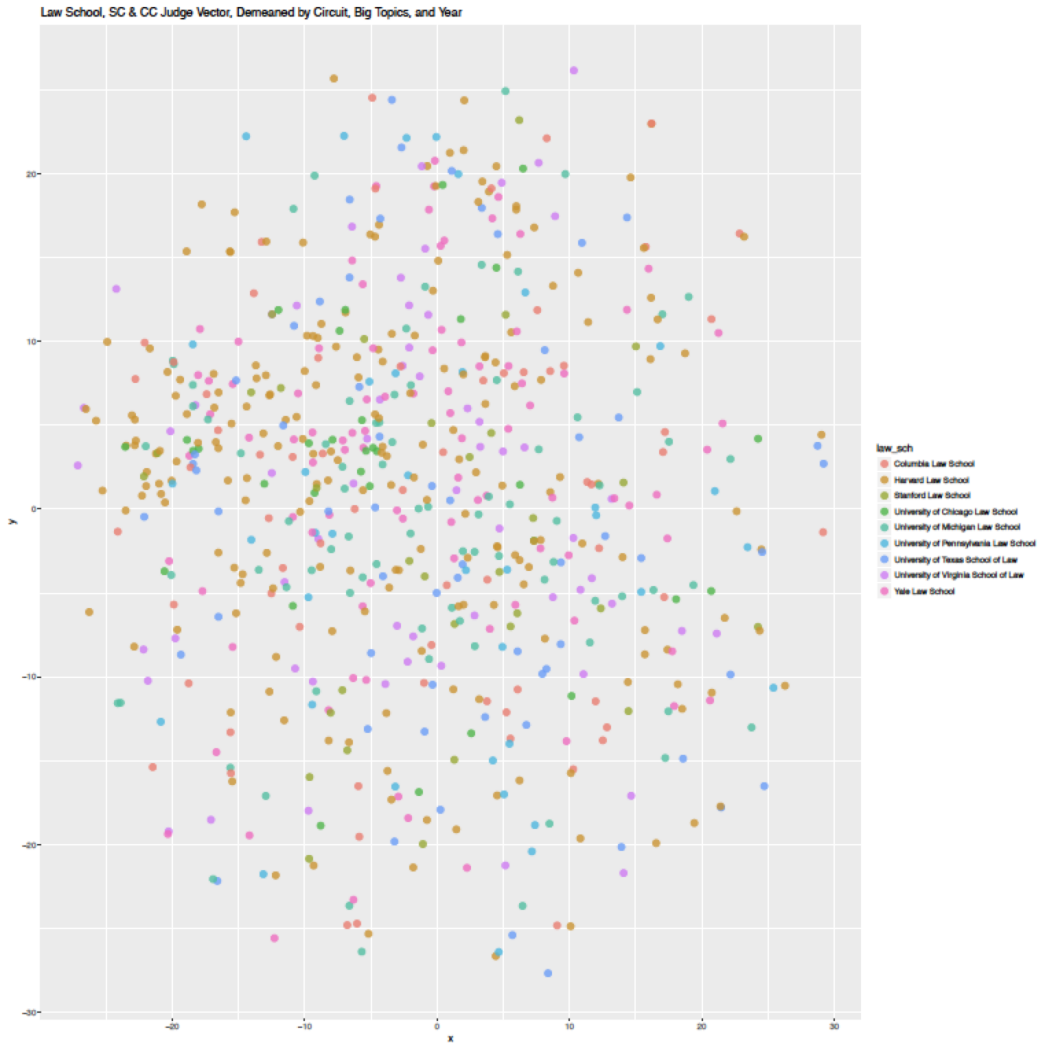
Case Vectors by Party

Figure 4: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Political Party



Case Vectors by Law School

Figure 6: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Law School Attended



Relatedness between judges

Circuit Judge Name	Similarity	Rank	Circuit Judge Name	Similarity	Rank
POSNER, RICHARD A.	1.000	1	TONE, PHILIP W.	0.459	16
EASTERBROOK, FRANK H.	0.663	2	SIBLEY, SAMUEL	0.459	17
SUTTON, JEFFREY S.	0.620	3	SCALIA, ANTONIN	0.456	18
NOONAN, JOHN T.	0.596	4	COLLOTON, STEVEN M.	0.445	19
NELSON, DAVID A.	0.592	5	DUNIWAY, BENJAMIN	0.438	20
CARNES, EDWARD E.	0.567	6	GIBBONS, JOHN J.	0.422	21
FRIENDLY, HENRY	0.566	7	BOGGS, DANNY J.	0.420	22
KOZINSKI, ALEX	0.563	8	*BREYER, STEPHEN G.*	0.414	23
GORSUCH, NEIL M.	0.559	9	GOODRICH, HERBERT	0.412	24
CHAMBERS, RICHARD H.	0.546	10	LOKEN, JAMES B.	0.410	25
FERNANDEZ, FERDINAND F.	0.503	11	WEIS, JOSEPH F.	0.408	26
EDMONDSON, JAMES L.	0.501	12	SCALIA, ANTONIN (SCOTUS)	0.406	27
KLEINFELD, ANDREW J.	0.491	13	BOUDIN, MICHAEL	0.403	28
WILLIAMS, STEPHEN F.	0.481	14	RANDOLPH, A. RAYMOND	0.397	29
KETHLEDGE, RAYMOND M.	0.459	15	MCCONNELL, MICHAEL W.	0.390	30

Document vectors demeaned by court, year, and topic, then aggregated by judge.

Contrast with N-Gram Approach

Law-Econ Style = Cosine distance to JSTOR JEL K

Rank	Judge	Law-Econ Style
1	CARDAMONE, RI.	2.85
2	DUNCAN, ALLYSON	2.69
3	MILLER, JUSTIN	2.57
4	SMITH, EDWARD	2.55
5	**GARLAND, MERRICK**	**2.33**
6	**WHITE, BYRON**	**2.25**
7	GARTH, LEONARD I	2.21
8	WOODROUGH, J.	2.13
9	O'SULLIVAN, CLIFF	2.00
10	ROBB, ROGER	1.78
11	PREGERSON, HARRY	1.77
12	STALEY, AUSTIN	1.64
13	HENDERSON, A.	1.50
14	MOTZ, DIANA	1.45
15	BIGGS, JOHN JR.	1.44
16	CLARK, CHARLES	1.44
17	REED, STANLEY	1.42
18	JACKSON, HOWELL	1.41
19	SIMONS, CHARLES	1.40
20	MILLER, SHACKEL.	1.38
21	WOODBURY, PETER	1.38
22	JONES, JOHN	1.27
23	HICKS, XENOPHON	1.25
24	SUHRHEINRICH, R.	1.24
25	**POSNER, RICHARD**	**1.23**
	GORSUCH, NEIL M.	**-0.84**
	SOTOMAYOR, SONIA	**-1.02**
	SCALIA, ANTONIN	**-1.28**

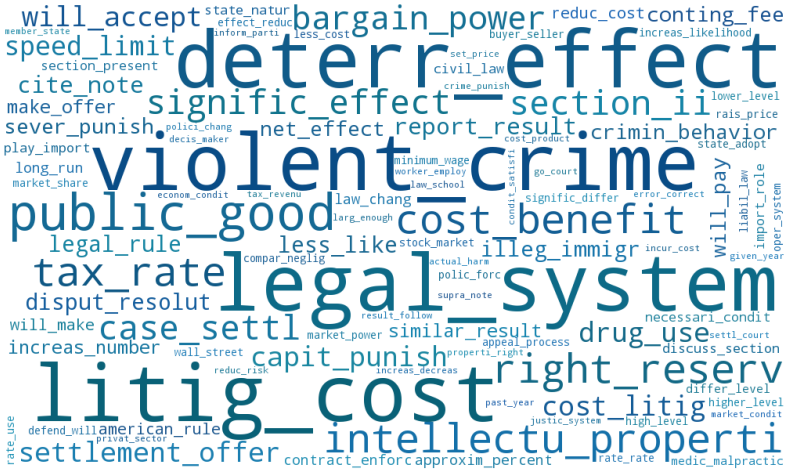
Contrast with N-Gram Approach

Law-Econ Style = Cosine distance to JSTOR JEL K

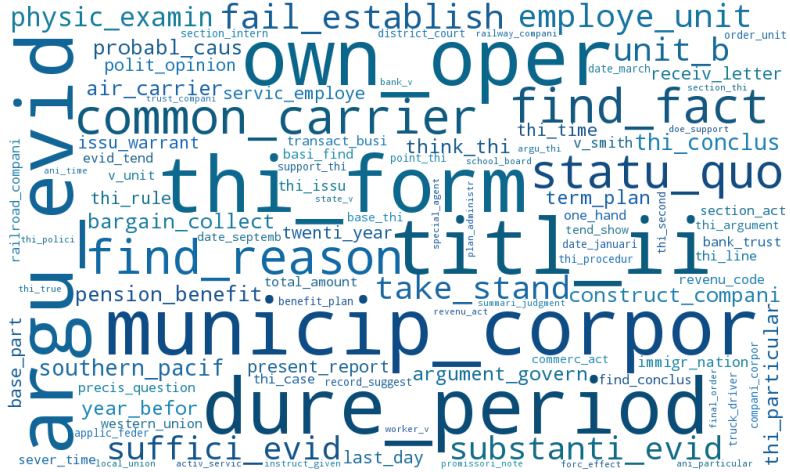
Rank	Judge	Law-Econ Style
1	CARDAMONE, RI.	2.85
2	DUNCAN, ALLYSON	2.69
3	MILLER, JUSTIN	2.57
4	SMITH, EDWARD	2.55
5	**GARLAND, MERRICK**	**2.33**
6	**WHITE, BYRON**	**2.25**
7	GARTH, LEONARD I	2.21
8	WOODROUGH, J.	2.13
9	O'SULLIVAN, CLIFF	2.00
10	ROBB, ROGER	1.78
11	PREGERSON, HARRY	1.77
12	STALEY, AUSTIN	1.64
13	HENDERSON, A.	1.50
14	MOTZ, DIANA	1.45
15	BIGGS, JOHN JR.	1.44
16	CLARK, CHARLES	1.44
17	REED, STANLEY	1.42
18	JACKSON, HOWELL	1.41
19	SIMONS, CHARLES	1.40
20	MILLER, SHACKEL.	1.38
21	WOODBURY, PETER	1.38
22	JONES, JOHN	1.27
23	HICKS, XENOPHON	1.25
24	SUHRHEINRICH, R.	1.24
25	**POSNER, RICHARD**	**1.23**
	GORSUCH, NEIL M.	**-0.84**
	SOTOMAYOR, SONIA	**-1.02**
	SCALIA, ANTONIN	**-1.28**

Law-and-Economics Language (N-gram)

- All JSTOR economics articles (1960-) JEL K (1990-) JLE (1960-)
  - Highest and lowest frequencies for two-grams in $\geq$ 1000 cases:



Most similar to Law-Econ Corpus



Least similar to Law-Econ Corpus

- Law-Econ: deterrent effect, cost-benefit, public goods, bargaining power, litigation costs
  - violent crime, criminal behavior, capital punishment, illegal immigration
- Non-LE: find reason, find fact, fail establish, substantive / sufficient / argue evidence
  - evidence and other constitutional theories of interpretation seem less salient

("Ideas Have Consequences: The Impact of Law and Economics on American Justice", Ash, Chen, Naidu)

Law-and-Economics Vectors

- externalit*, transaction_costs, efficien*, deterr*, cost_benefit, capital, game_theo, chicago_school, marketplace, law1economic, law2economic identified by Ellickson (2000)

- One of the sentences that is closest to “economics” in is: “The discussion then turned to economics.”

Law-and-Economics Vectors

- externalit*, transaction_costs, efficien*, deterr*, cost_benefit, capital, game_theo, chicago_school, marketplace, law1economic, law2economic identified by Ellickson (2000)

- One of the sentences that is closest to “economics” in is: “The discussion then turned to economics.”

Loyalty and Obedience Vectors

Loyalty

Obedience

- "Loyalty" and "obedience" don't seem very related
- Loyalty associated with certain native american tribes

Loyalty and Obedience Vectors

Loyalty

Obedience

- "Loyalty" and "obedience" don't seem very related
- Loyalty associated with certain native american tribes

Loyalty and Obedience Vectors

Loyalty

Obedience

- "Loyalty" and "obedience" don't seem very related
- Loyalty associated with certain native american tribes

Implicit (or Explicit) Attitudes

- Google translate
  - “he/she is a doctor”(turkish) -> “he is a doctor” (english)
  - “he/she is a nurse”(turkish) -> “she is a nurse” (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.
- We ask whether gender and racial bias varies across judges.

Implicit (or Explicit) Attitudes

- Google translate
  - "he/she is a doctor"(turkish) -> "he is a doctor" (english)
  - "he/she is a nurse"(turkish) -> "she is a nurse" (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.

- We ask whether gender and racial bias varies across judges.

Implicit (or Explicit) Attitudes

- Google translate
  - "he/she is a doctor"(turkish) -> "he is a doctor" (english)
  - "he/she is a nurse"(turkish) -> "she is a nurse" (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.

- We ask whether gender and racial bias varies across judges.

Word Embedding Association Test (Science 2017)

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure
Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter
Implicit Racism Target Words	
european, white, caucasian	black, african, negro

- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} - \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} - \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Word Embedding Association Test (Science 2017)

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure
Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter
Implicit Racism Target Words	
european, white, caucasian	black, african, negro

- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} - \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} - \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Word Embedding Association Test (Science 2017)

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure
Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter
Implicit Racism Target Words	
european, white, caucasian	black, african, negro

- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

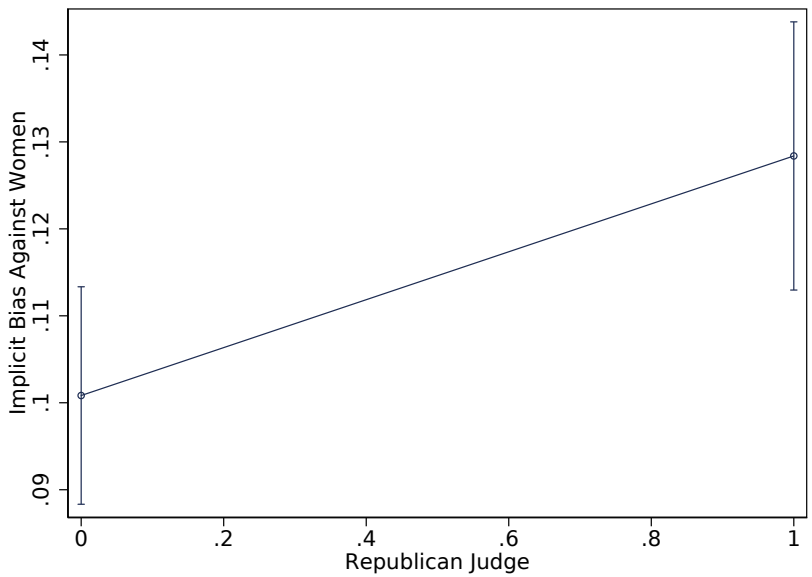
$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} - \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

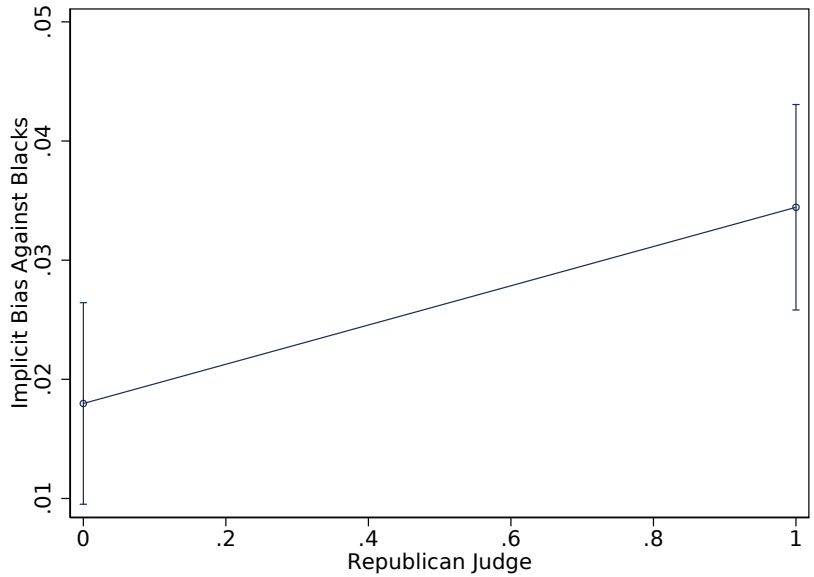
$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} - \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

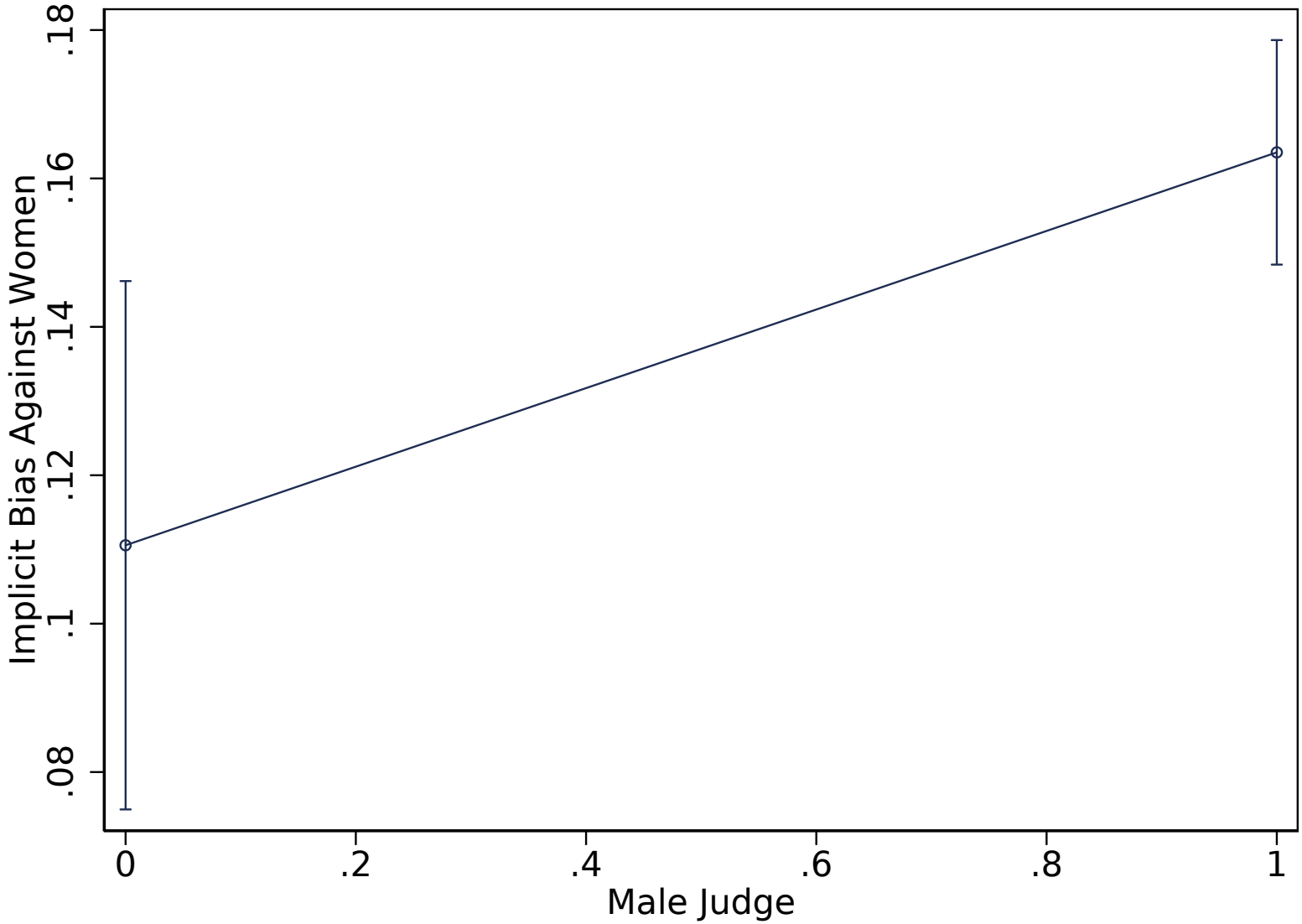
("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Republican judges have higher gender bias and race bias

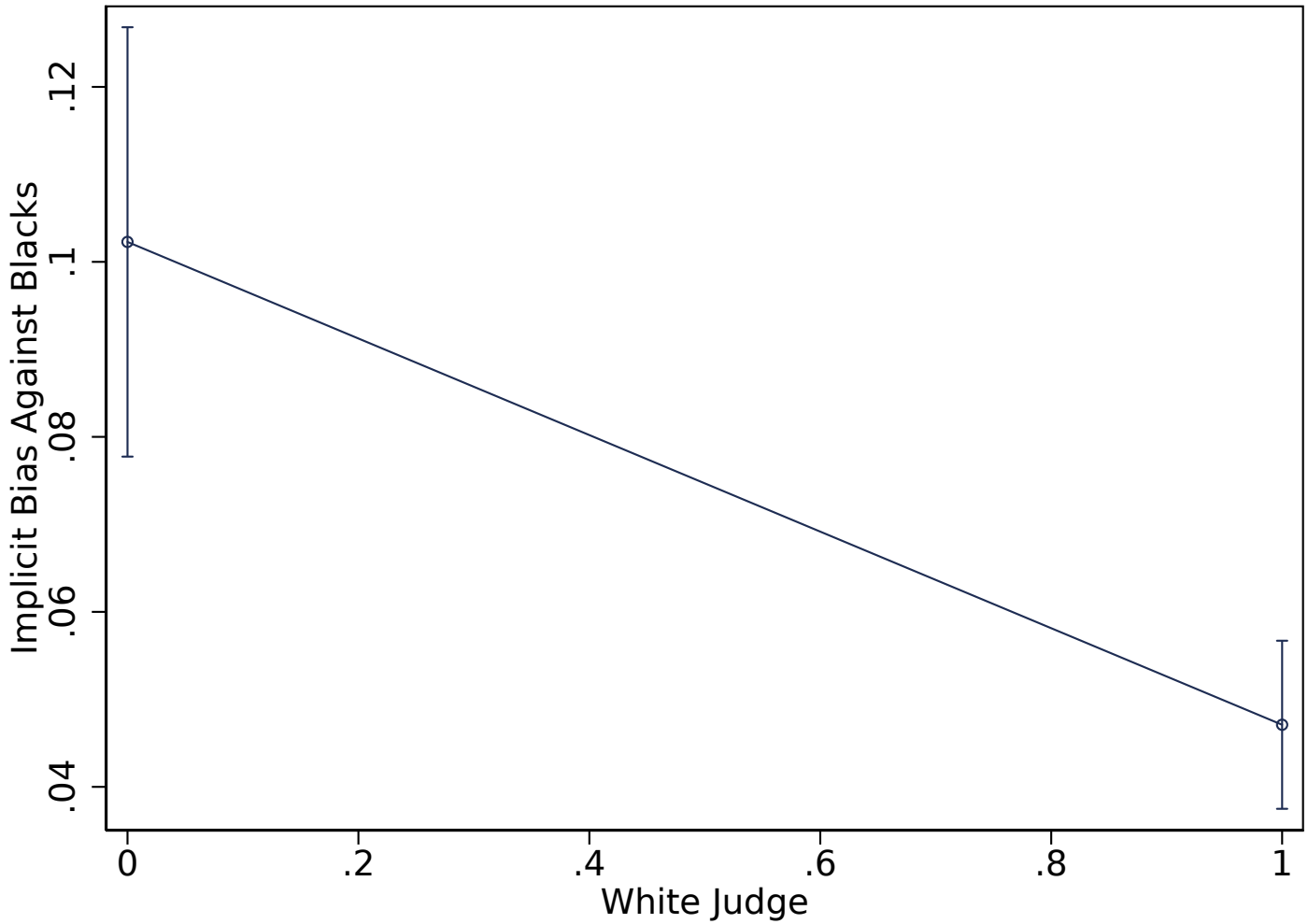




Male judges have higher gender bias than female judges



White judges have *lower* race bias than black judges



Both the words and the IAT work at an unconscious level, in contrast to the decisions which are more conscious.

Trump nominees have high race and gender bias

President Donald J. Trump's Supreme Court List

Amy Coney Barrett of Indiana, U.S. Court of Appeals for the Seventh Circuit

Keith Blackwell of Georgia, Supreme Court of Georgia

Charles Canady of Florida, Supreme Court of Florida

Steven Colloton of Iowa, U.S. Court of Appeals for the Eighth Circuit

Allison Eid of Colorado, U.S. Court of Appeals for the Tenth Circuit

Britt Grant of Georgia, Supreme Court of Georgia

Raymond Gruender of Missouri, U.S. Court of Appeals for the Eighth Circuit

Thomas Hardiman of Pennsylvania, U.S. Court of Appeals for the Third Circuit

Brett Kavanaugh of Maryland, U.S. Court of Appeals for the District of Columbia Circuit

Raymond Kethledge of Michigan, U.S. Court of Appeals for the Sixth Circuit

Joan Larsen of Michigan, U.S. Court of Appeals for the Sixth Circuit

Mike Lee of Utah, United States Senator

Thomas Lee of Utah, Supreme Court of Utah

Edward Mansfield of Iowa, Supreme Court of Iowa

Federico Moreno of Florida, U.S. District Court for the Southern District of Florida

Kevin Newsom of Alabama, U.S. Court of Appeals for the Eleventh Circuit

William Pryor of Alabama, U.S. Court of Appeals for the Eleventh Circuit

Margaret Ryan of Virginia, U.S. Court of Appeals for the Armed Forces

David Stras of Minnesota, U.S. Court of Appeals for the Eighth Circuit

Diane Sykes of Wisconsin, U.S. Court of Appeals for the Seventh Circuit

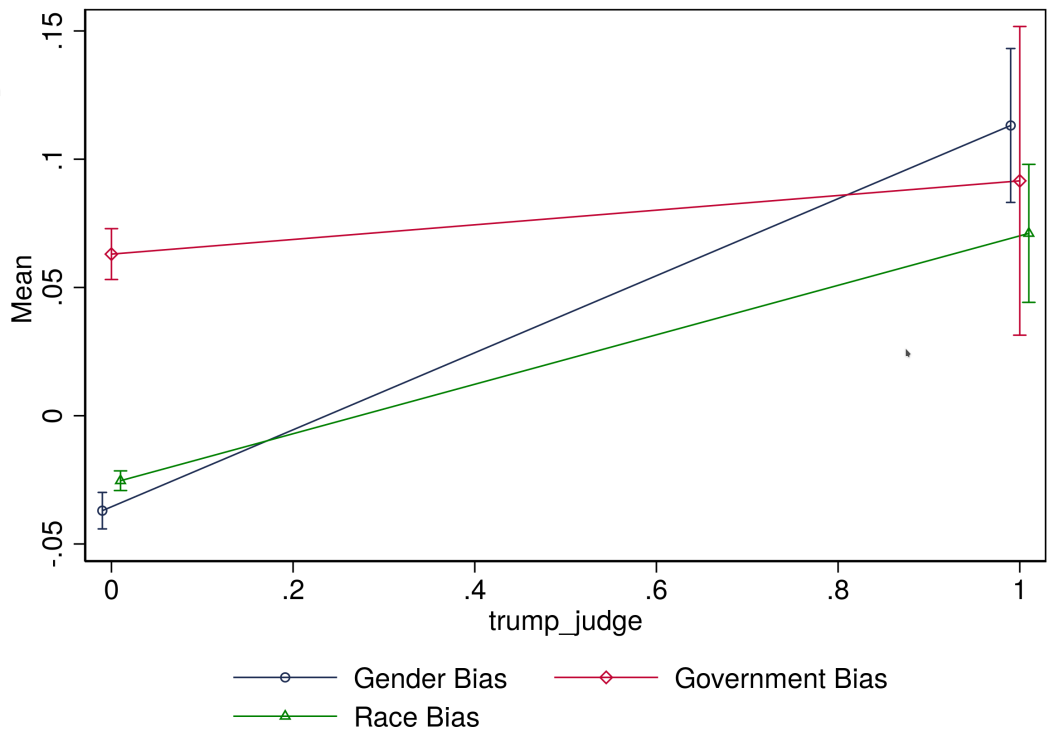
Amul Thapar of Kentucky, U.S. Court of Appeals for the Sixth Circuit

Timothy Tymkovich of Colorado, U.S. Court of Appeals for the Tenth Circuit

Robert Young of Michigan, Supreme Court of Michigan (Ret.)

Don Willett of Texas, Supreme Court of Texas

Patrick Wyrick of Oklahoma, Supreme Court of Oklahoma



Making Doctrinal Work Rigorous

Word Embeddings

- to identify the original meaning from the text
  - loyalty or ?
  - loyalty $\approx$ obedience?
- see if the vectors cluster along three distinct categories
  - (structural, characterological, behavioral) being distinct?
  - or whether there are omitted categories?
- see if typology is novel *or* just redefinition of close concepts
- see how judges use the concepts
  - or show why we care about what law & econ scholars write?

Making Doctrinal Work Rigorous

Word Embeddings

- to identify the original meaning from the text
  - loyalty or ?
  - loyalty $\approx$ obedience?
- see if the vectors cluster along three distinct categories
  - (structural, characterological, behavioral) being distinct?
  - or whether there are omitted categories?
- see if typology is novel *or* just redefinition of close concepts
- see how judges use the concepts
  - or show why we care about what law & econ scholars write?

Making Doctrinal Work Rigorous

Word Embeddings

- to identify the original meaning from the text
  - loyalty or ?
  - loyalty $\approx$ obedience?
- see if the vectors cluster along three distinct categories
  - (structural, characterological, behavioral) being distinct?
  - or whether there are omitted categories?
- see if typology is novel *or* just redefinition of close concepts
- see how judges use the concepts
  - or show why we care about what law & econ scholars write?

Making Doctrinal Work Rigorous

Word Embeddings

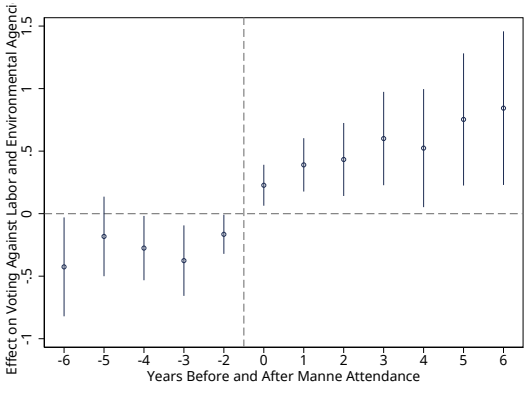
- to identify the original meaning from the text
  - loyalty or ?
  - loyalty $\approx$ obedience?
- see if the vectors cluster along three distinct categories
  - (structural, characterological, behavioral) being distinct?
  - or whether there are omitted categories?
- see if typology is novel *or* just redefinition of close concepts
- see how judges use the concepts
  - or show why we care about what law & econ scholars write?

Making Doctrinal Work Rigorous

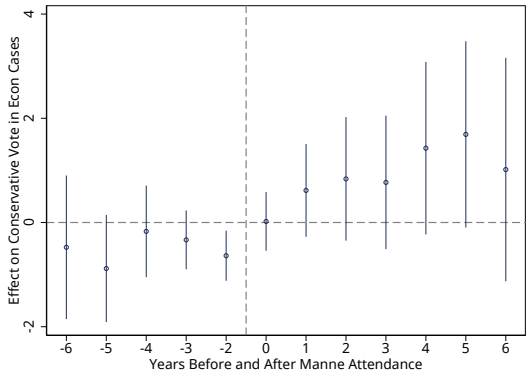
Word Embeddings

- to identify the original meaning from the text
  - loyalty or ?
  - loyalty $\approx$ obedience?
- see if the vectors cluster along three distinct categories
  - (structural, characterological, behavioral) being distinct?
  - or whether there are omitted categories?
- see if typology is novel *or* just redefinition of close concepts
- see how judges use the concepts
  - or show why we care about what law & econ scholars write?

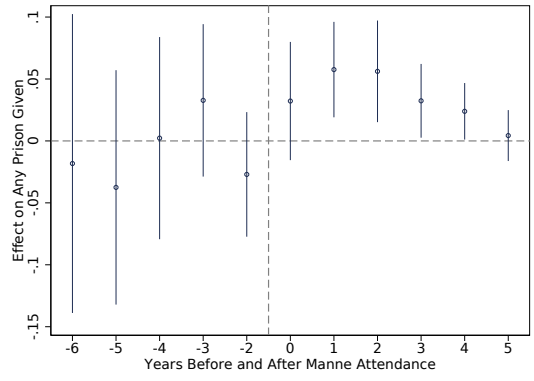
Impact of Law and Economics Training



Rejecting Labor/Environmental Regulations



Conservative Votes



Sentencing Decisions