# Measuring Regulatory Complexity

Co-Pierre Georg (UCT) and Jean-Edouard Colliard (HEC Paris)

07 January 2024
**AFA**

# Motivation

# Motivation

- Perceived increase in the complexity of financial regulation. For instance:
  - Basel I, 1988: 30 pages
  - Basel II, 2004: 347 pages
  - Basel III, 2011-2014: 616 pages
  - Dodd-Frank Act, 2010: 848 pages
- Calls for simpler regulations, for instance a leverage ratio (Haldane, 2012)
- Conceptualization by BCBS (2013) of a trade-off between:
  - Risk-sensitivity
  - Comparability
  - Simplicity

# Motivation

- Perceived increase in the complexity of financial regulation. For instance:
  - Basel I, 1988: 30 pages
  - Basel II, 2004: 347 pages
  - Basel III, 2011-2014: 616 pages
  - Dodd-Frank Act, 2010: 848 pages
- Calls for simpler regulations, for instance a leverage ratio (Haldane, 2012)
- Conceptualization by BCBS (2013) of a trade-off between:
  - Risk-sensitivity
  - Comparability
  - Simplicity
- **Q**: How can we operationalize this trade-off?

# The complexity of a text is more than just its length

## Our approach...

- We develop a toolkit to quantify the complexity of financial regulation

  ...and test if our measures align with how humans perceive complexity

# Our approach...

- We develop a toolkit to quantify the complexity of financial regulation

  ...and test if our measures align with how humans perceive complexity

- We draw on an analogy between algorithms and regulation:

| **Algorithm** | **Regulation** |
|---|---|
| Take an (arbitrary) input | Take a bank as input |
| Apply a finite sequence of operations | Apply a finite sequence of operations |
| Get resulting output | Get resulting regulatory action |

Not a new analogy: Li et al. (2015), *"Law is Code"*; Katz and Bommarito (2014)

# Our approach...

- We develop a toolkit to quantify the complexity of financial regulation

  ...and test if our measures align with how humans perceive complexity

- We draw on an analogy between algorithms and regulation:

| **Algorithm** | **Regulation** |
|---|---|
| Take an (arbitrary) input | Take a bank as input |
| Apply a finite sequence of operations | Apply a finite sequence of operations |
| Get resulting output | Get resulting regulatory action |

Not a new analogy: Li et al. (2015), *"Law is Code"*; Katz and Bommarito (2014)

- Adapt the well-developed literature on algorithmic complexity

## . . .and what it delivers

- We theoretically distinguish three dimensions of complexity
- We propose measures based on a classification of words according to their "algorithmic" function:
  - $\Rightarrow$ Embeds three measures from the existing (computer science) literature
  - $\Rightarrow$ Yields three **new** measures
- We provide code and dictionary to extend to extend our analysis to other texts
  e.g. complexity of tax rules (Colliard, Eden, and Georg (2021)) and cybersecurity regulation (Georg and Partida (2023))
- We run experiments to validate our (and other) complexity measures

# Related literature - Complexity and economics

- Sophisticated agents strategically exploit complexity (e.g., Carlin 2009)
- Competition between regulators can lead to too much complexity/poor regulatory quality (Asriyan, Foarta, and Vanasco, 2018; Oehmke and Zawadowski, 2018)
- Complexity creates asymmetric information (Arora, Barak, Brunnermeier, and Ge (2009); Celerier and Vallee (2017))
- Risk of capture by sophistication (Hakenes and Schnabel, 2013)
- Opacity to outsiders gives discretion to supervisors (Rochet, 2010)

# Related literature - Complexity and economics

- Sophisticated agents strategically exploit complexity (e.g., Carlin 2009)
- Competition between regulators can lead to too much complexity/poor regulatory quality (Asriyan, Foarta, and Vanasco, 2018; Oehmke and Zawadowski, 2018)
- Complexity creates asymmetric information (Arora, Barak, Brunnermeier, and Ge (2009); Celerier and Vallee (2017))
- Risk of capture by sophistication (Hakenes and Schnabel, 2013)
- Opacity to outsiders gives discretion to supervisors (Rochet, 2010)

$\Rightarrow$ Further theoretical work on this issue hindered by **lack of measures**

# Related literature - Existing measures

- Literature needs measures that can be computed at scale (on a cross-section of regulatory texts)
- Main available measure is about the quantity of regulation (e.g., Al-Ubaydli and McLaughlin (2017))
- Alternative: Measures of text readability, but these are not complexity measures
- Many other measures of complexity not usable at scale, rather case studies

# Related literature - Existing measures

- Literature needs measures that can be computed at scale (on a cross-section of regulatory texts)
- Main available measure is about the quantity of regulation (e.g., Al-Ubaydli and McLaughlin (2017))
- Alternative: Measures of text readability, but these are not complexity measures
- Many other measures of complexity not usable at scale, rather case studies

$\Rightarrow$ Our framework can be used at scale

# Framework

# Some definitions

- Different people have different things in mind when using the word "complexity"
- Complexity is clearly multidimensional

  $\Rightarrow$ Our measures aim at capturing a subset of dimensions only
- Some preliminary definitions useful to clarify:
  - What we are aiming at
  - Mapping between algorithms and regulations
- Also useful for the policy debate: What type of complexity is problematic?

# Regulation as an algorithm

**Definition**
A *regulatory problem*: Ideal mapping $f$ from an entity $\epsilon$ to some action $a$.

**Definition**
A *regulatory text*: A list $\tilde{f}$ of words representing the mapping $f$.

**Definition**
A *supervision mechanism*: Observed mapping $\hat{f}$ from any entity $\epsilon$ to some action $a$.

**Definition**
*Perfect supervision*: Mapping $\tilde{f}^*$ that exactly corresponds to $\tilde{f}$.

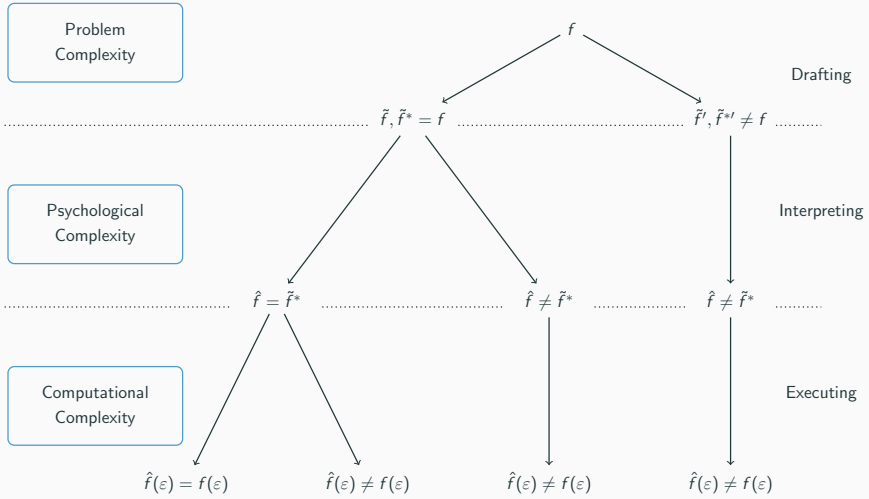
# Complexity

*Regulation* is the combination of a regulatory problem, a regulatory text, and a supervision mechanism

A regulation has a higher level of *regulatory complexity* if either:

1. All else equal, it is more likely that $\hat{f}(\epsilon) \neq f(\epsilon)$;
2. All else equal, ensuring that $\hat{f}(\epsilon) = f(\epsilon)$ is more costly (supervision hours, data, etc.)

# Three dimensions of complexity



# Measuring the complexity of algorithms

# Algorithms as operands and operators: The Halstead measures

- Pioneering work of Halstead (1977)
- Define an algorithm as a list of operands and operators:
  - Operands: variables, constants...
  - Operators: $+$, $-$, $=$, *if*, *end*, etc.
- We go further and distinguish:
  - Regulatory operators
  - Logical operators
  - Mathematical operators

# Example: Compute the absolute value

Algorithm to compute the absolute value of a number:

$$\text{if } x \geq 0, y \equiv x$$
$$\text{if } x < 0, y \equiv -x$$

- $\eta_{OD} = 3$ unique *operands* $(x, y, 0)$, $N_{OD} = 8$ total operands
- $\eta_R = 1$ unique *regulatory* operator $(\equiv)$, $N_R = 2$

$$\Rightarrow \text{Quantity: } quantity = N_R$$

- $\eta_L = 1$ unique *logical* operator $(if)$, $N_L = 2$
- $\eta_M = 3$ unique *mathematical* operators $(\geq, <, -)$, $N_M = 3$

$$\Rightarrow \text{Length: } length = N_{OD} + N_R + N_L + N_M$$

# A new measure: Potential volume

- What would be the shortest program in the best possible programming language?

# A new measure: Potential volume

- What would be the shortest program in the best possible programming language?
  $\Rightarrow$ Answer: simple call for a built-in function in that language

**Example**: compute the absolute value of a number

$$y \equiv abs(x)$$

- 2 operators: $=, abs$
- 2 operands: $x, y$
- Potential volume: $potential = 2 + \eta_{OD} = 4$

# Level

- The level of a program is:

$$
\begin{aligned}
L &= \frac{potential}{length} = \frac{2 + \eta_{OD}}{N_{OD} + N_R + N_L + N_M} \\
&= \frac{2 + \eta_{OD}}{\eta_{OD} + \eta_R + \eta_L + \eta_M} \times \frac{\eta_{OD} + \eta_R + \eta_L + \eta_M}{N_{OD} + N_R + N_L + N_M}
\end{aligned}
$$

- The level thus depends on the number of intermediary operations and the number of repetitions
- In the context of regulation:
  - High $L$ corresponds to efficient but specialized language
  - Measure can be part of a trade-off between transparency and length

# Summary of existing and proposed measures

| Name | Source | Formula | Complexity Dimension |
|---|---|---|---|
| Length | e.g., Haldane and Madouros (2012) | $N$ | Psychological |
| Cyclomatic complexity | McCabe (1976) | $N_L$ | Psychological |
| Quantity of regulations | Al-Ubaydli and McLaughlin (2017) | $N_R$ | Problem |
| Potential volume | This paper and Halstead (1977) | $2+\eta_{OD}$ | Problem |
| Operator diversity | This paper | $\eta_{OR}$ | Psychological |
| Level | This paper and Halstead (1977) | $\frac{2+\eta_{OD}}{N}$ | Psychological |
| Average word length | e.g., Amadxarif *et al.* (2019) | - | Psychological |
| Lexical diversity | Maas (1972) | - | Psychological |
| Readability metric | (Kincaid *et al.*, 1975) | - | Psychological |
| Shannon's entropy | e.g., Katz and Bommarito (2014) | - | Psychological |
| PageRank | Amadxarif *et al.* (2019) | - | Psychological |
| Network Centralities (Eigenvector, Betweenness, Closeness) | Boulet *et al.* (2011) | - | Psychological |
| Regulation Index | Simkovic and Zhang (2020) | - | Computational |
| RegIn | Kalmenovitz (2021) | - | Computational |

# Proof of Concept: Basel I

# Application to a text

- Can we apply these concepts to actual regulations?
- Testing ground: Basel I capital ratios; The are already very “algorithmic”

# Application to a text

- Can we apply these concepts to actual regulations?
- Testing ground: Basel I capital ratios; The are already very "algorithmic"
- Two approaches:
  - "Translate" the regulation into an actual algorithm
  - Adapt the measure to regulatory texts directly
- We compare both approaches

# Basel I - Annex 2 - Algorithm vs. Text

| Basel I Algorithm | Basel I Text |
|---|---|
| `IF (`<br>`  ASSET_CLASS == "cash"`<br>`) THEN:`<br>`  risk_weight = 0.0;` | Cash shall have a 0% risk weight |
| `IF (`<br>`  ASSET_CLASS == "claims" AND`<br>`  (ISSUER == "`<br>`central_governments" OR ISSUER == "`<br>`central_banks") AND`<br>`  DENOMINATION == "national" AND`<br>`  FUNDING_CURRENCY == "national"`<br>`) THEN:`<br>`  risk_weight = 0.0;` | Claims on central governments and central banks denominated in national currency and funded in that currency shall have a 0% risk weight |
| `IF (`<br>`  ASSET_CLASS == "claims" AND`<br>`  (ISSUER == "`<br>`central_governments" OR ISSUER == "`<br>`central_banks") AND`<br>`  ISSUER_COUNTRY == "oecd"`<br>`) THEN:`<br>`  risk_weight = 0.0;` | Other claims on OECD central governments and central banks shall have a 0% risk weight |

# Algorithmic vs. text-based measures

**Table 1:** Correlation coefficients between algorithmic and text-based measures.

| | Pearson | Spearman |
|---|---|---|
| *length* | 0.76 | 0.84 |
| *cyclomatic* | 0.41 | 0.64 |
| *quantity* | 1 | 1 |
| *potential* | 0.82 | 0.8 |
| *diversity* | 0.4 | 0.48 |
| *level* | 0.39 | 0.43 |

- *quantity* and *potential* measure problem complexity
- Hence, they should not depend on the representation chosen
- For these two at least one can work directly on a text

# Algorithmic vs. text-based measures

**Table 2:** Top 5 words in each category in Basel I (text version).

| Operands | | Operators: Regulatory | | Operators: Logical | | Operators: Mathematical | |
|---|---|---|---|---|---|---|---|
| risk weight | 19 | have | 19 | and | 12 | up to | 2 |
| claims | 15 | shall | 19 | other | 6 | above | 1 |
| banks | 10 | | | or | 5 | all | 1 |
| OECD | 10 | | | outside | 4 | over | 1 |
| central | 9 | | | excluding | 2 | | |

# Comparing text-based measures

**Table 3:** Pairwise correlations between complexity measures, across the 19 items of Basel I (text version).

| | *length* | *cyclomatic* | *potential* | *diversity* | *level* |
|---|---|---|---|---|---|
| *length* | 1 | 0.65 | 0.88 | 0.63 | -0.62 |
| *cyclomatic* | 0.65 | 1 | 0.48 | 0.89 | -0.69 |
| *potential* | 0.88 | 0.48 1 | 0.44 | -0.24 | |
| *diversity* | 0.63 | 0.89 | 0.44 | 1 | -0.72 |
| *level* | -0.62 | -0.69 | -0.24 | -0.72 | 1 |

- Measures positively correlated, except level.
- Yet, length seems insufficient to capture all dimensions.

# Experimental Validation

# Motivation

- How do we know complexity measures capture complexity?
- We borrow another idea from computer science - coding experiments:
  - Programmers write codes to solve a problem.
  - One computes different measures of the complexity of their codes.
  - Good measure of complexity should predict time taken, occurrence of bugs, etc.

# What we do

- We built a random “Basel-I like” regulation generator.

## What we do

- We built a random “Basel-I like” regulation generator.
  (Yes, really)

## What we do

- We built a random "Basel-I like" regulation generator.
  (Yes, really)
- We selected 38 regulations that were not too unreasonable.
- We then asked a cohort of HEC Paris MSc International Finance students to compute capital requirements:
  - 10 iterations with different randomly selected regulations.
  - Same artificial bank balance sheet.
- We measure the answer given and the time taken.

## What we do

- We built a random "Basel-I like" regulation generator.
  (Yes, really)
- We selected 38 regulations that were not too unreasonable.
- We then asked a cohort of HEC Paris MSc International Finance students to compute capital requirements:
  - 10 iterations with different randomly selected regulations.
  - Same artificial bank balance sheet.
- We measure the answer given and the time taken.

$\Rightarrow$ `https://regulatorycomplexity.org/`

# Validating the measures

- We have 118 students indexed by $i$, answering 9 questions indexed by $t$.
- We have different measures of complexity $\mu$ based on regulation $R_{i,t}$.
- Logit regression:

$$\Pr(correct_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t)$$

- On the subset of correct answers we run OLS:

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t$$

# Explaining the likelihood of providing a correct answer

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *length* | -0.037*** | -0.055*** | 0.005 | -0.030*** | -0.065*** | -0.031*** |
| | (-7.95) | (-5.69) | (0.62) | (-2.75) | (-7.85) | (-5.16) |
| *cyclomatic* | | 0.068** | | | | |
| | | (2.13) | | | | |
| *quantity* | | | -0.554*** | | | |
| | | | (-5.92) | | | |
| *potential* | | | | -0.017 | | |
| | | | | (-0.70) | | |
| *diversity* | | | | | 0.438*** | |
| | | | | | (4.16) | |
| *level* | | | | | | 1.275 |
| | | | | | | (1.55) |
| Pseudo-$R^2$ | 0.243 | 0.248 | 0.277 | 0.244 | 0.260 | 0.246 |

(i) Cyclomatic and length measure psych. complexity; (ii) Adding one rule (to 4.79 average) decreases prob. of correct answer from 68.6% to 53.8%; (iii) Level has different impact on more sophisticated participants, hence overall ambiguous;

# Explaining the time take to provide a correct answer

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *length* | 3.388*** | 5.371*** | 2.234*** | 0.556 | 2.946*** | 4.183*** |
| | (14.39) | (10.46) | (4.39) | (0.88) | (6.81) | (12.74) |
| *cyclomatic* | | -8.084*** | | | | |
| | | (-4.33) | | | | |
| *quantity* | | | 13.072** | | | |
| | | | (2.56) | | | |
| *potential* | | | | 6.807*** | | |
| | | | | (4.79) | | |
| *diversity* | | | | | 6.899 | |
| | | | | | (1.22) | |
| *level* | | | | | | 165.462*** |
| | | | | | | (3.44) |
| $R_a^2$ | 0.445 | 0.461 | 0.450 | 0.465 | 0.445 | 0.455 |

(i) Adding one unique operand (to average 16.66) increases time from average 132s to 145s; (ii) Results largely robust to alternative specifications (filtering, winsorizing);

# Large-scale Application - The Dodd-Frank Act

# Application

- Our approach is useful only if it can be applied to:
  - Variety of existing regulations (empirical research on regulatory complexity)
  - Newly proposed regulatory texts (normative use)

- In both cases one needs:
  - Automated computation of measures based on text analysis
  - A large dictionary of regulatory operands/operators

# Dodd-Frank Act

- We consider the different titles of the 2010 Dodd-Frank Act.
- Very broad regulatory text, covering all areas of finance
- Large vocabulary of terms to classify
- Many policy discussions on the complexity of this Act


Not secure | regulatorycomplexity.pythonanywhere.com/words/index?
cgeorg
On
Title 0
Title 1
Title 2
Title 3
Title 4
Title 5
Title 6
Title 7
Title 8
Title 9
Title 10
Title 11
Title 12
Title 13
Title 14
Title 15
Title 16
class = "ex1">TITLE 0.
class = "ex4">SEC. 1. Short title; table of contents.
class = "ex5">(a) Short title. This Act may be cited as the "Dodd-Frank Wall Street Reform and Consumer Protection Act".
class = "ex5">(b) Table of Contents. The table of contents for this Act is as follows:
class = "ex4">SEC. 2. Definitions.
class = "ex5">As used in this Act, the following definitions shall apply, except where the context otherwise requires or as otherwise specifically provided in this Act:
class = "ex6">(1) Affiliate. The term "affiliate" means any affiliate (as determined in accordance with section 3 of the Federal Deposit Insurance Act (12 U.S.C. 1813)).
class = "ex6">(2) Appropriate Federal banking agency. Unless the context otherwise requires, the term "appropriate Federal banking agency" has the same meaning as in section 3(q) of the Federal Deposit Insurance Act (12 U.S.C. 1813(q)), as amended by title III.
class = "ex6">(3) Board of governors. The term "Board of Governors" means the Board of Governors of the Federal Reserve System.
class = "ex6">(4) Bureau. The term "Bureau" means the Bureau of Consumer Financial Protection established by title X.
class = "ex6">(5) Commission. The term "Commission" means the Securities and Exchange Commission, except that for purposes of title VII, Commodity Futures Trading Commission.
class = "ex6">(6) Commodity futures terms. The terms "futures commission merchant", "swap", "swap dealer", "swap execution facility", "derivatives clearing organization", "board of trade", "commodity trading advisor", "commodity pool", and "commodity pool operator" have the same meanings as given the terms in section 1a of the Commodity Exchange Act (7 U.S.C. 1 et seq.).
class = "ex6">(7) Corporation. The term "Corporation" means the Federal Deposit Insurance Corporation.
class = "ex6">(8) Council. The term "Council" means the Financial Stability Oversight Council established under title I.
class = "ex6">(9) Credit union. The term "credit union" means any federal credit union, State credit union, or State-chartered credit union, as those terms are defined in section 101 of the Federal Credit Union Act (12 U.S.C. 1752).
class = "ex6">(10) Federal banking agency. The term
class = "ex7">(A) "Federal banking agency" means, individually, the Board of Governors, the Office of the Comptroller of the Currency, and the Corporation; and
class = "ex7">(B) "Federal banking agencies" means any or all agencies defined in subparagraph (A), collectively.
class = "ex6">(11) Functionally regulated subsidiary. The term "functionally regulated subsidiary" has the same meaning as in section 5(c)(5) of the Bank Holding Company Act of 1956 (12 U.S.C. 1844(c)(5)).
class = "ex6">(12) Primary financial regulatory agency. The term "primary financial regulatory agency" means
class = "ex7">(A) the appropriate Federal banking agency, with respect to institutions described in section 3(q) of the Federal Deposit Insurance Act, except to the extent that an institution or the activities of an institution are otherwise described in subparagraph (B), (C), (D), or (E);
class = "ex7">(B) the Securities and Exchange Commission, with respect to
class = "ex8">(i) any broker or dealer that is registered with the Commission under the Securities Exchange Act of 1934, with respect to the activities of the broker or dealer that require the broker or dealer to be registered under that Act;
class = "ex8">(ii) any investment company that is registered with the Commission under the Investment Company Act of 1940, with respect to the activities of an investment company that require the investment company to be registered under that Act;
class = "ex8">(iii) any investment adviser that is registered with the Commission under the Investment Advisers Act of 1940, with respect to the investment advisory activities of such company and activities that are incidental to such advisory activities;
class = "ex8">(iv) any clearing agency registered with the Commission under the Securities Exchange Act of 1934, with respect to the activities of the clearing agency that require the agency to be registered under such Act;
class = "ex8">(v) any nationally recognized statistical rating organization registered with the Commission under the Securities Exchange Act of 1934;
class = "ex8">(vi) any transfer agent registered with the Commission under the Securities Exchange Act of 1934;
class = "ex8">(vii) any exchange registered as a national securities exchange with the Commission under the Securities Exchange Act of 1934;
class = "ex8">(viii) any national securities association registered with the Commission under the Securities Exchange Act of 1934;
class = "ex8">(ix) any securities information processor registered with the Commission under the Securities Exchange Act of 1934;
class = "ex8">(x) the Municipal Securities Rulemaking Board established under the Securities Exchange Act of 1934;
class = "ex8">(xi) the Public Company Accounting Oversight Board established under the Sarbanes-Oxley Act of 2002 (15 U.S.C. 7211 et seq.);
class = "ex8">(xii) the Securities Investor Protection Corporation established under the Securities Investor Protection Act of 1970 (15 U.S.C. 78aaa et seq.); and
class = "ex8">(xiii) the security-based swap execution facility, security-based swap data repository, security-based swap dealer or major security-based swap participant registered with the Commission under the Securities Exchange Act of 1934, with respect to a security-based swap activities if the person is required to be registered under such Act;
class = "ex7">(C) the Commodity Futures Trading Commission, with respect to
class = "ex8">(i) any futures commission merchant registered with the Commodity Futures Trading Commission under the Commodity Exchange Act (7 U.S.C. 1 et seq.), with respect to the activities of the futures commission merchant that require the futures commission merchant to be registered under that Act;
class = "ex8">(ii) any commodity pool operator registered with the Commodity Futures Trading Commission under the Commodity Exchange Act (7 U.S.C. 1 et seq.), with respect to the activities of the commodity pool operator that require the commodity pool operator to be registered under that Act, or a commodity pool, as defined in that Act;
class = "ex8">(iii) any commodity trading advisor or introducing broker registered with the Commodity Futures Trading Commission under the Commodity Exchange Act (7 U.S.C. 1 et seq.), with respect to the activities of the commodity trading
INSTRUCTIONS
WORDS CLASSIFIED
LOG OUT
DOWNLOAD HTML
Word Count
Total: 1754
Classified: 0
Remaining: 1754
Words Classified:

# Output

- We obtain and make available a dictionary of:
  - 5,872 operands
  - 429 operators (230 logical, 161 regulatory, 38 mathematical)
  - 2,799 other words (incl. 2,450 legal references, 222 function words, 127 ambiguous words)
- We also make the dashboard available, if classifying extra words is necessary

$\Rightarrow$ `https://github.com/cogeorg/RegulatoryComplexity_Public`

# Conclusion

- New framework to conceptualize and measure regulatory complexity

- Our work opens up several interesting avenues of research:
- Making a regulation easier to understand does not necessarily make it less costly to process and vice versa
  - Empirical studies on the determinants and impact of regulatory complexity
  - Applied tools to help policymakers manage complexity