

Economics of Readability: Privacy Policies and the GDPR

Jacopo Gambato¹ Bernhard Ganglmair¹ Julia Krämer²

¹ZEW, University of Mannheim, and MaCCI

²Erasmus University Rotterdam

German Law and Economics Association
Nancy, France
July 7, 2022



UNIVERSITY
OF MANNHEIM



Are privacy policies used as non-price competition strategy?

Are privacy policies used as non-price competition strategy?

What are the economic incentives for drafting “proper” privacy policies?

Are privacy policies used as non-price competition strategy?

What are the economic incentives for drafting “proper” privacy policies?

What is the effect of the GDPR on readability and scope of policies? Why?

Who Really Reads Privacy Policies?

- Strong public response to “data leaks” and other scandals (Facebook and Cambridge Analytica)
- People seem to care about privacy (Spiekermann et al., 2001) ...
- ... yet we see overwhelming evidence that nobody reads privacy policies (Obar et al., 2016) or the fine print more generally (Bakos et al., 2014)

Dear user!

We use Google Admob to show ads. Ads support our work, and enable further development of this app. In line with the new European Data Protection Regulation (GDPR), we need your consent to serve ads tailored for you.

Can your data be used to show ads tailored for you?

YES, CONTINUE TO SHOW RELEVANT ADS

NO, SHOW ADS THAT ARE IRRELEVANT

REMOVE ADS

You can change this setting anytime in the "About" window.

[Learn how your data is used](#)

“If internet users were to read every privacy policy they agree to, it would take them on average 201 hours per year to finish.” (McDonald and Cranor, 2008)

Because It's Just Gibberish? The Legislator's Response

1. Disclose what data is collected (Art. 13–14 GDPR)
 - Must inform data subjects about which data is collected (and on what legal basis)
2. Comply with new transparency standard (Art. 12(1) GDPR)
 - information to be provided “*in a concise, transparent, intelligible and easily accessible form, using clear and plain language*”
3. Avoid the jargon, make it clear (Art. 29 Working Party)
 - “information should not contain overly legalistic language/terminology” (Transparency Guidelines)
 - **a lot of legal uncertainty, no real guidance by CJEU**

What Did the Regulators Accomplish?

- Mannheim Privacy Policy Panel (MaPPP)
 - novel data set with privacy policies for German firms (2014–2021)
- Methods from natural language processing
 - **informational content** and the **readability** of these policies
- Firm and industry-level information (size, business model, market concentration)
 - incentives and underlying mechanisms

... to examine the firm-level response to the GDPR in 2018.

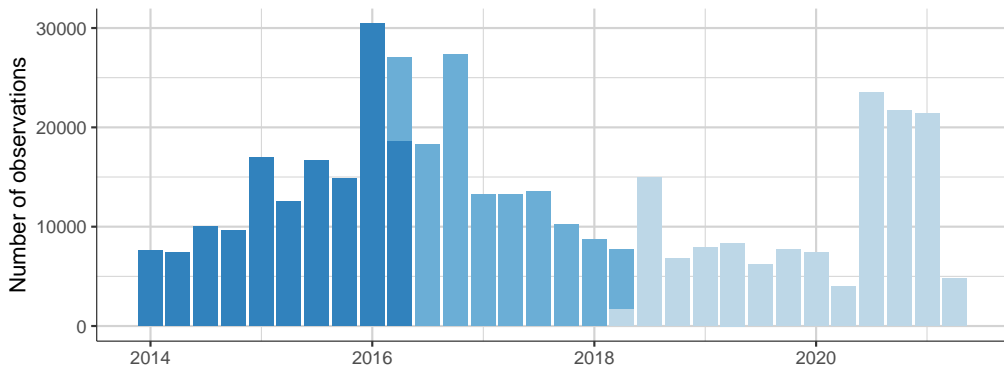
Main Findings

- Policies change a lot, both in scope and in readability
- We observe a convergence in readability:
 - Firms with difficult-to-read policies drafted easier policies
 - Firms with easy-to-read policies drafted more difficult policies
- Readability with B2C-focused firms has not deteriorated more than with others

Data

Mannheim Privacy Policy Panel (MaPPP)

- Mannheim Privacy Policy Panel
 - Consists of 1.7m privacy policies by more than 500,000 German firms between 2012 and 2021.
 - Data source: Internet Archive via the Wayback Machine
- Trimming top/bottom 2% (by word count)
- Subsample of firms with at least one observation in each of three time periods:
 - 2014–2016 (before GDPR announcement)
 - 2016–2018 (between announcement and implementation)
 - 2018–2021 (after implementation)
- Sample: **400,178 privacy policies** posted by **40,350 firms**



Shades of blue highlight three phases (before publication; after adoption; between); 2016:Q2 and 2018:Q2 have observations in both phases.

Firm and Industry-Level Information

- Mannheim Enterprise Panel (MUP):
 - Firm size information
 - Industry classifiers
- German Monopolies Commission (Heidorn and Weche, 2021):
 - Industry level concentration measures (HHI, CR-6)
- Community Innovation Survey (2017):
 - Business model information (B2B vs. B2C)
Question 1.8: “Who are the main customers of your enterprise?”
 - Classification from Srinivasan et al. (2011) for robustness

How do we measure regulatory effects?

Measuring Volume and Legal Basis

- Informational volume (Art. 13–14 GDPR)
 - Number of sentences and paragraphs
 - Number of *unique* words
- Legal basis (Art. 13(1)(c) GDPR)
 - Count of *articles* cited in the policies

	Mean	Std.	Min	Max
Unique words	395.99	263.85	22	2039
Number of paragraphs	37.69	53.04	1	815
Number of sentences	77.28	72.07	2	556
Number of article citations	3.41	7.84	0	93

Measuring Readability

- What makes texts easier to read?
 - word length (in syllables)
 - average sentence length (in words)
 - share of big words (words with at least 5 syllables)
- Components used in *readability indices*, with different weights
- We construct three widely used indices/scores:
 - *Flesch-Kincaid Grade Level*: standard measure of readability (= number of years of formal education needed)
 - *Neue Wiener Sachtextformel*: readability index (= number of years) specifically for German texts
 - *Flesch Reading Ease Index* (adapted for German texts; inverted)
- *Higher values mean more difficult texts*

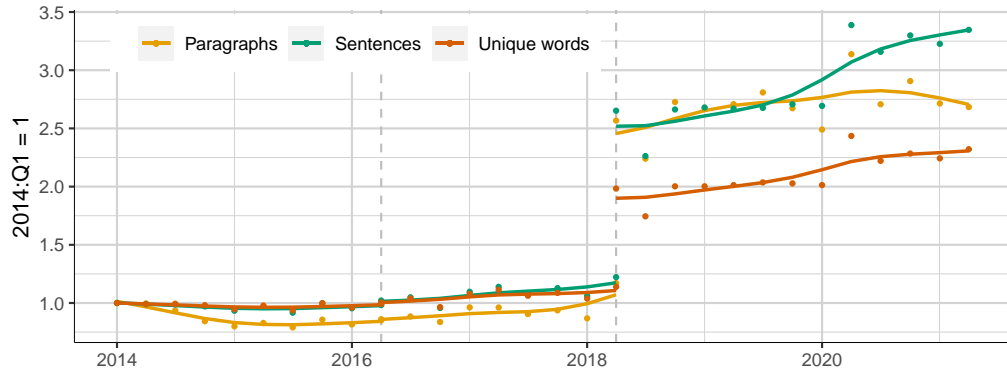
	Mean	Std.	Min	Max
Word length (in syllables)	2.16	0.07	1.4	3
Sentence length (in words)	17.9	3.16	4.2	213.8
Share of big words (5+ syllables)	0.2	0.04	0	0.7
Flesch-Kincaid grade level	16.84	1.56	4.7	95
Neue Wiener Sachtextformel	12.93	0.88	3.9	53.6
inverted (German) Flesch Reading-Ease score	50.25	5.43	0	272

	Obs.	Word length	Sentence length	Big words	Flesch- Kincaid	Neue WS	German Flesch
Privacy policies (MaPPP)	400178	2.16 (0.07)	17.9 (3.16)	0.2 (0.04)	16.84 (1.56)	12.93 (0.88)	50.25 (5.43)
Simple-language news (nachrichtenleicht.de)	293	1.75 (0.12)	10.89 (1.92)	0.03 (0.03)	9.28 (1.67)	7.93 (1.6)	33.12 (7.54)
Speeches and statements: Angela Merkel	1128	1.83 (0.07)	18.16 (2.3)	0.3 (0.03)	13.08 (1.12)	9.69 (0.88)	45.16 (4.47)
Decisions by German Consti- tutional Court (BVerfG)	9358	1.96 (0.09)	16.35 (2.91)	0.49 (0.04)	13.86 (1.79)	10.52 (1.26)	50.73 (6.75)
Wikipedia (German)	10000	1.9 (0.2)	20.63 (14.48)	0.12 (0.04)	14.82 (6.02)	11.09 (3.22)	51.52 (18.23)
Wikipedia (English)	10000	1.71 (0.16)	19.78 (6.57)	0.05 (0.03)	12.27 (3.21)	9.1 (2.28)	39.67 (11.58)
GDPR/DS-GVO (Wikipedia)	1	2.1	18.66	0.12	16.46	12.46	61.46
GDPR/DS-GVO (official)	1	2.24	40.39	0.18	26.54	17.53	91.17

Results:

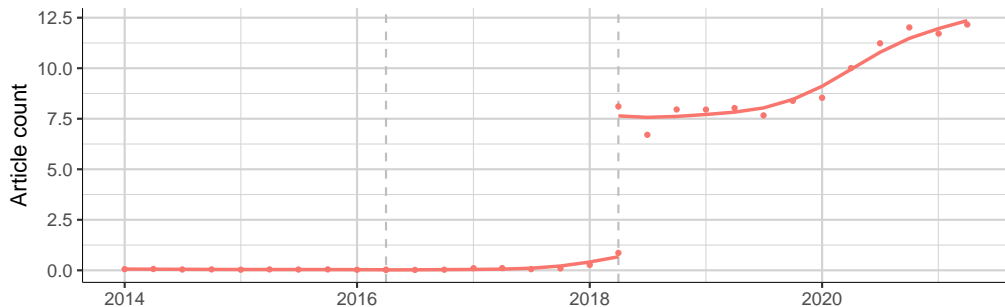
1. Effects of GDPR adoption
2. Convergence of (firm-level) readability
3. The role of a firm's business model

[1] Policies Do Change: Longer, More Information



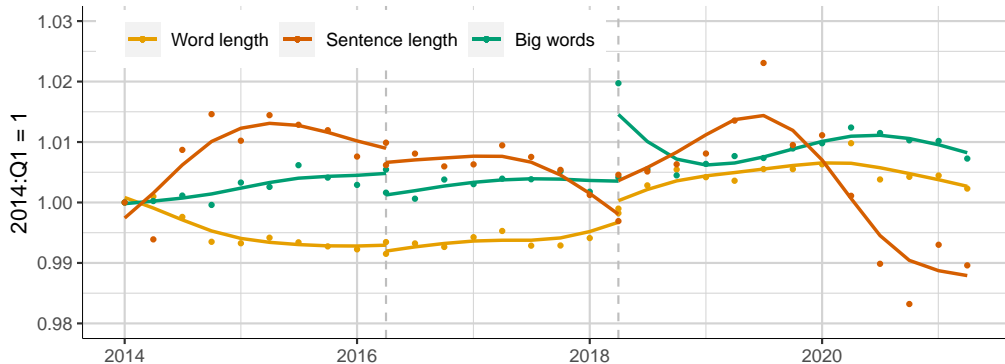
Length, informational volume (Art. 13–14 GDPR); quarterly averages (2014:Q1 = 1)

[1] Not Just Simple Adjustments



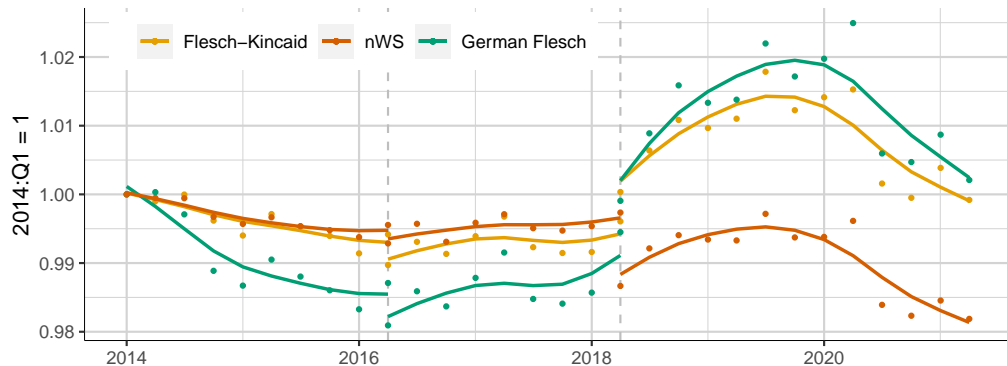
Count of article citations (Art. 13(1)(c) GDPR); quarterly averages

[1] Some Deteriorating of Readability Factors



Readability factors; quarterly averages (dots) and fitted (spline) (2014:Q1 = 1)

[1] Partial Disagreement of Readability Indices



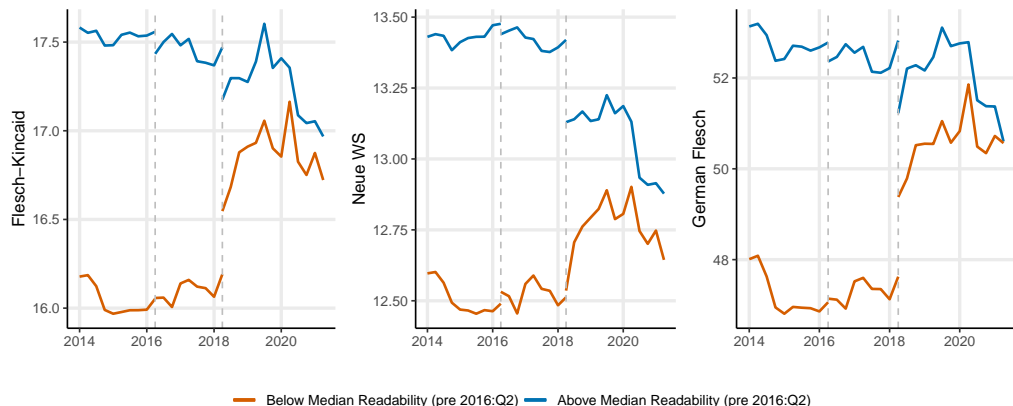
Readability indices; quarterly averages (dots) and fitted (spline) (2014:Q1 = 1)

[1] Multivariate Regression Results

	Word length		Sentence length		Big words	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-GDPR (=1)	0.0157*** (0.0006)	0.0127*** (0.0007)	0.0259 (0.0305)	0.1427*** (0.0194)	0.0005 (0.0004)	0.0002 (0.0004)
Text Controls	No	Yes	No	Yes	No	Yes
R ²	0.625	0.654	0.552	0.932	0.311	0.311
	log Flesch-Kincaid		log Neue WS		log German Flesch	
	(7)	(8)	(9)	(10)	(11)	(12)
Post-GDPR (=1)	0.0086*** (0.0008)	0.0096*** (0.0005)	-0.0066*** (0.0006)	-0.0018*** (0.0005)	0.0153*** (0.0010)	0.0146*** (0.0007)
Text Controls	No	Yes	No	Yes	No	Yes
R ²	0.563	0.886	0.601	0.815	0.581	0.796

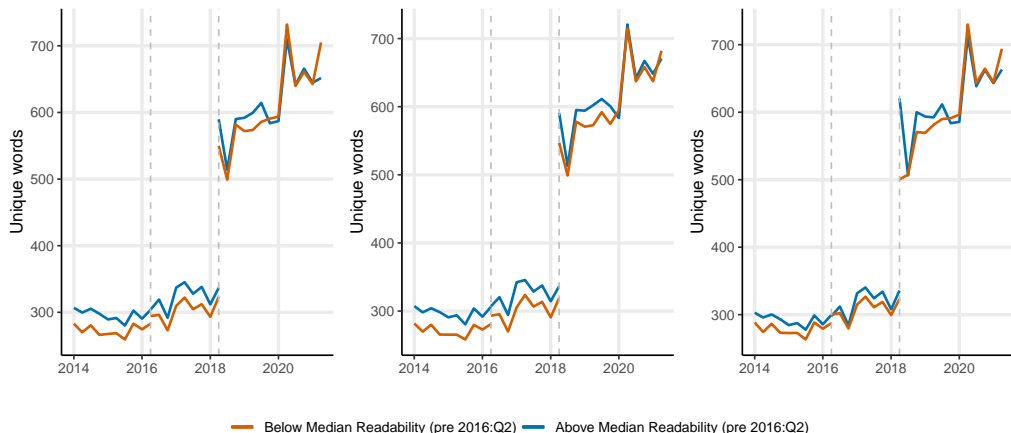
All models include firm and year FE. Text controls: length and volume

[2] Convergence: Hard-to-Read Policies Get Easier



Readability (quarterly averages) for below (red) and above (blue) median prior to 2016:Q2

[2] But: Readability and Information are *Not* the Same Thing



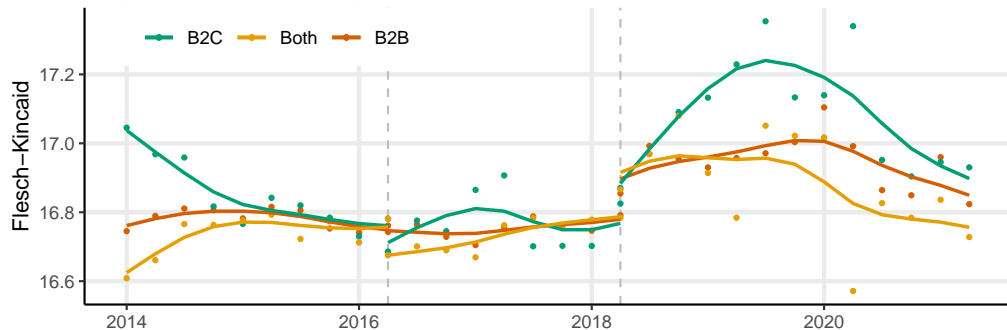
Unique words (quarterly averages) for below/above median of readability prior to 2016:Q2

[2] Multivariate Regression Results (Split Samples)

	<i>log</i> Flesch-Kincaid		<i>log</i> Unique words	
	Readable	Not	Readable	Not
Post-GDPR (=1)	0.0319*** (0.0012)	-0.0135*** (0.0011)	0.5356*** (0.0074)	0.4925*** (0.0073)
R ²	0.524	0.506	0.756	0.752
	<i>log</i> Neue WS		<i>log</i> Unique words	
	Readable	Not	Readable	Not
Post-GDPR (=1)	0.0112*** (0.0009)	-0.0229*** (0.0008)	0.5530*** (0.0076)	0.4775*** (0.0071)
R ²	0.500	0.547	0.753	0.755
	<i>log</i> German Flesch		<i>log</i> Unique words	
	Readable	Not	Readable	Not
Post-GDPR (=1)	0.0475*** (0.0014)	-0.0134*** (0.0012)	0.5377*** (0.0076)	0.4922*** (0.0072)
R ²	0.556	0.495	0.747	0.760

All models include firm and year FE. GDPR effect (=1) is a % change.

[3] Does Readability for Consumers Deteriorate?



Quarterly averages (fitted spline) for Flesch-Kincaid grade level.

[3] Evidence Multivariate Regressions Says No!

	<i>log</i> Flesch-Kincaid		
	B2C	B2C and B2B	B2B
Post-GDPR	0.0074*** (0.0013)	0.0140*** (0.0024)	0.0085*** (0.0015)
R ²	0.564	0.564	0.562
	<i>log</i> Neue WS		
	B2C	B2C and B2B	B2B
Post-GDPR	-0.0074*** (0.0010)	-0.0036** (0.0018)	-0.0071*** (0.0011)
R ²	0.603	0.597	0.601
	<i>log</i> German Flesch		
	B2C	B2C and B2B	B2B
Post-GDPR	0.0142*** (0.0015)	0.0210*** (0.0027)	0.0155*** (0.0017)
R ²	0.580	0.578	0.582

All models include firm and year FE. GDPR effect (=1) is a % change.

Summary

Track Record of the GDPR is Mixed

- The average privacy policy has gotten more difficult to read after the adoption of the GDPR.
- But we observe an improvement of policies by firms that used to be among the hardest to read.
- We do not find a particular role for firms in primarily B2C focused markets
- Takeaway for regulators:
 - Choice of the “right” readability index is crucial as indices often disagree

Next: a Better Look at a Complex Issue


- *So far:* firm FE as control to single out general effect of GDPR
- *Next:* more in-depth look at heterogeneity: effect of firm size and market concentration
- Interaction of characteristics: effect of GDPR on different firms in different kinds of markets
- ... and open to further ideas and suggestions

Thank you!

Find the paper soon

- through the internet search engine of your choice
- on our websites
- at <https://ssrn.com>

b.ganglmair@gmail.com |  @ganglmair

jacopo.gambato@zew.de |  @JGambato_econ

j.k.kramer@law.eur.nl |  @Julia_Kraemer