

# GDPR, Online Browsing, and Search

Yu Zhao, Pinar Yildirim, Pradeep Chintagunta

The Wharton School, University of Pennsylvania, University of Chicago

April 1, 2022

# General Data Protection Regulation (GDPR)

- Extended consumer data protection & privacy regulations introduced on May 25, 2018 in EU
  - Requiring firms to
    - Obtain consent of consumers to collect and use their data,
    - Encrypt and anonymize consumer data,
    - Provide data breach notifications,
    - Appoint an internal officer to oversee GDPR compliance,
- among others



# Implications of GDPR

- For **consumers**:

- Reduce cost of sharing personal data
- May reduce the hesitancy to visit/transact with lesser-known sites, search goods and services of private nature

# Implications of GDPR

- For **consumers**:
  - Reduce cost of sharing personal data
  - May reduce the hesitancy to visit/transact with lesser-known sites, search goods and services of private nature
- For **firms**:
  - Adds frictions to the information exchange with consumers:
    - costlier to collect and use consumer data, and to target consumers using personal data;
    - harder to send personalized or targeted recommendations
    - may become costlier to advertise for goods and services

# Research Questions

- Broadly: what is the impact of GDPR on consumers and firms?  
Is there an evidence of higher frictions in reaching out to and informing consumers about products and services?
- How did **user behavior online** change after GDPR? Are users more or less active across websites?

# Research Questions

- Broadly: what is the impact of GDPR on consumers and firms?  
Is there an evidence of higher frictions in reaching out to and informing consumers about products and services?
- How did **user behavior online** change after GDPR? Are users more or less active across websites?
- Did GDPR alter consumers' **search** for information and products?
  - **Textual analysis of search queries to measure changes to search episodes**
- Is there **heterogeneity** across firms wrt to the changes experienced following GDPR?

# Empirical Approach

## Data

- Consumer Online Browsing Panel:

# Empirical Approach

## Data

- Consumer Online Browsing Panel:

- Data from a large, multi-national online browsing company, January 31st – September 1st, 2018
- 8 million+ records by 5998 panelists, over 692,600 domains visited

# Empirical Approach

## Data

- Consumer Online Browsing Panel:

- Data from a large, multi-national online browsing company, January 31st – September 1st, 2018
- 8 million+ records by 5998 panelists, over 692,600 domains visited
- Covers EU (UK, Spain) and non-EU regions (United States, Brazil)
- Online browsing and keyword records related to **search & browsing**
  - Desktop alone, mobile app alone, or both
  - Panelist id, time of click, url domain name, duration on a page
  - User-week balanced panel

# Data

## GDPR Policy Date Change

- Data on privacy policy updates of firms:
  - Scraped each domain's privacy policy page
  - If firm is compliant, via phrases like: "GDPR", "general data protection regulation", "data controller", "data protection officer", and "regulation2016/679"
  - Policy change date obtained by extracting dates following these phrases: "*updated at*", "*last modified on*" or "*last updated at/on*"
  - Obtained 35,586 websites' update times

# Data

## Privacy Policy Updates Data

- Policy update time

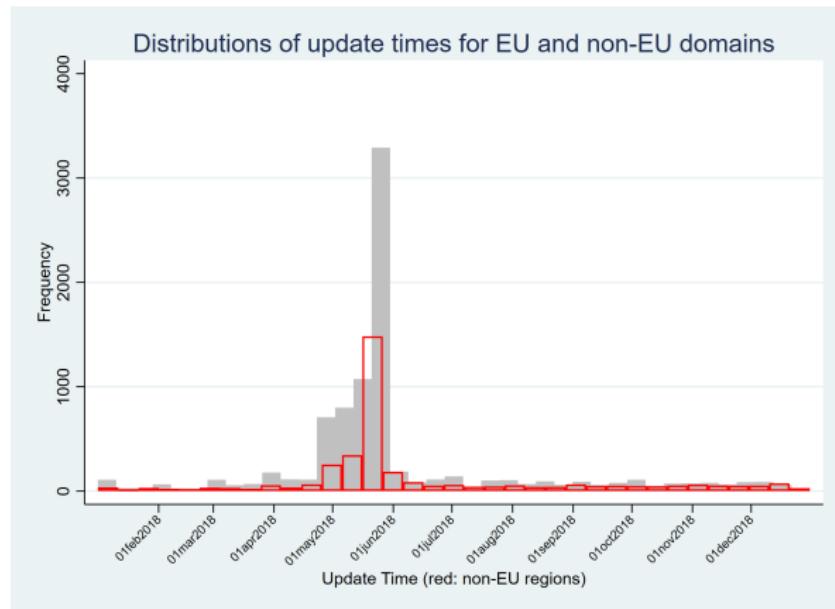


Figure: Distribution of update times, by region

# Changes in Online Consumer Behavior

Compare online behavior before and after GDPR, and across panelists in and outside of EU

$$\log(Y_{it}) = \alpha_1 + \alpha_2 GDPR_t \times \mathbf{1}\{\text{EU}_i = 1\} + \tau_t + \theta_i + \epsilon_{it}$$

- $Y_{it}$ : browsing activity by panelist  $i$  in week  $t$ .
  - Unique domain visits, time spent online, time spent per page
- $GDPR_t$ : GDPR dummy (official date)
- $\text{EU}_i$ : indicates that the panelist is in EU
- Fixed effects – combinations of time, user

# Changes in Online User Behavior

## Desktop

	No. unique domains		Total time (seconds)		Avg. per-page time (seconds)	
	(1)	(2)	(3)	(4)	(5)	(6)
GDPR × EU	0.139*** (0.013)	0.139*** (0.010)	0.370*** (0.031)	0.370*** (0.025)	0.145*** (0.011)	0.145*** (0.009)
EU	-0.300*** (0.010)		-0.926*** (0.022)		-0.327*** (0.008)	
Time period	Yes	Yes	Yes	Yes	Yes	Yes
User		Yes		Yes		Yes
Observations	187,092	187,092	187,092	187,092	187,092	187,092
Adjusted $R^2$	0.04	0.46	0.05	0.36	0.03	0.32

# Changes in Online User Behavior

## Mobile Apps

	Logged no. apps		Logged total time (minutes)		Logged per-app time (minutes)	
	(1)	(2)	(3)	(4)	(5)	(6)
GDPR × EU	0.033*** (0.007)	0.033*** (0.006)	0.104*** (0.015)	0.104*** (0.014)	0.076*** (0.010)	0.076*** (0.009)
EU	-0.265*** (0.005)		-0.411*** (0.011)		-0.124*** (0.007)	
Week	Yes	Yes	Yes	Yes	Yes	Yes
Connection	Yes	Yes	Yes	Yes	Yes	Yes
Device	Yes	Yes	Yes	Yes	Yes	Yes
Operating System	Yes	Yes	Yes	Yes	Yes	Yes
User		Yes		Yes		Yes
Observations	539,604	539,604	539,604	539,604	539,604	539,604
Adjusted $R^2$	0.26	0.45	0.19	0.37	0.11	0.28

# GDPR and Online Consumer Behavior

- After GDPR goes into effect, in the short term, the EU panelists in the data are more active on average:
  - visit 14.9% additional domains
  - browse 0.37% more pages on a domain
  - access 3.4% more apps per week
  - spend 10.9% more total time on the mobile apps

compared to the non-EU consumers

# GDPR and Online Consumer Behavior

- After GDPR goes into effect, in the short term, the EU panelists in the data are more active on average:
  - visit 14.9% additional domains
  - browse 0.37% more pages on a domain
  - access 3.4% more apps per week
  - spend 10.9% more total time on the mobile appscompared to the non-EU consumers
- Increased engagement is consistent with both the enhanced privacy benefits of GDPR *and* the inefficiency firms face to reach out to customers

# GDPR & Online Consumer Search

- Document the implications of privacy protections & informational friction on **consumers' online search**

# GDPR & Online Consumer Search

- Document the implications of privacy protections & informational friction on **consumers' online search**
- Identify the **topics/products** users are searching for, compare effort consumers put into search:

# GDPR & Online Consumer Search

- Document the implications of privacy protections & informational friction on **consumers' online search**
- Identify the **topics/products** users are searching for, compare effort consumers put into search:

$$\log(\text{Search effort}_{ikt}) = \gamma_0 + \gamma_1 \text{UK}_i \times \text{GDPR}_t + \eta_i + \theta_t + \nu_k + \varepsilon_{it}$$

- $\text{Search\_effort}_{ikt}$ : various measures of search effort of panelist  $i$  related to topic or product category  $k$  in week  $t$ .
- Control for individual, week, and topic fixed effects

# Identify Consumer Search Episodes

- Search incidences belong to the same search episode if they fall under the same semantic topic (cluster)
  - Keyword data: single word, multi-word phrase, short sentence
    - Examples: "car", "supermarket", "smart watch with gps maps", "best twenty albums of all time", "truck"
- Represent words meanings with a skip-gram model

# Identify Consumer Search Episodes

- Search incidences belong to the same search episode if they fall under the same semantic topic (cluster)
  - Keyword data: single word, multi-word phrase, short sentence
    - Examples: “car”, “supermarket”, “smart watch with gps maps”, “best twenty albums of all time”, “truck”
- Represent words meanings with a skip-gram model
  - Cosine similarity
  - Similar words to “car”:
    - (US) “vehicle”, “truck”, “lease”
    - (UK) “garage”, “property”, “vehicle”

# Identify Consumer Search Episodes

- Search incidences belong to the same search episode if they fall under the same semantic topic (cluster)
  - Keyword data: single word, multi-word phrase, short sentence
    - Examples: “car”, “supermarket”, “smart watch with gps maps”, “best twenty albums of all time”, “truck”
- Represent words meanings with a skip-gram model
  - Cosine similarity
  - Similar words to “car”:
    - (US) “vehicle”, “truck”, “lease”
    - (UK) “garage”, “property”, “vehicle”
- Cluster the queries: UK: 50 clusters, US: 65 clusters (topics)

# Identifying Consumer Search Episodes

Table: Top phrases (bi-grams) in US and UK clusters

US			UK		
cluster label*	sample keywords	cluster size**	cluster label	sample keywords	cluster size**
1	yahoo lottery, illinoi lottery	42,701	1	auction site, vehicle tax	1,752
2	galaxi j3, ipod touch	22,200	2	travel destinations	74,098
3	indiana realtors, counti property	18,537	3	2018 world cup, wimbledon 2018	13,952
4	shorthair burmes, munchkin cat	48,059	4	updat windows, file explor	28,192
5	news lates, tenni news	3,775	5	cash back, price guarantee	20,407
:	:	:	:		
40	nyse, nasdq	45,100	30	harri potter, children author	45,809
41	enterpris car, inn suit	60,790	31	watch seri, free watch	2,345
42	account generator, number generator	1,737	32	song lyrics, bob dylan	9,082
43	crochet pattern, dress women	41,510	33	news latest, today brexit	3,265
44	tv direct, tv channel	4,301	34	clearance final, gift vouchers	18,387
45	login att, online login	11,522	35	towel rail, hang basket	48,440

\*: Order of the cluster is not meaningful

\*\*: Each item in a cluster is the union of queries submitted by a panelist in an hour

# GDPR & Online Search

## Search for Information

	Cosine similarity (1)	log(search effort) (2)
GDPR × UK	0.004*** (0.001)	0.0473*** (0.000684)
Constant	0.459*** (0.000)	0.202*** (0.000202)
Week FE	Yes	Yes
Panelist FE	Yes	Yes
Day of week	Yes	(not applicable)
Hour of day	Yes	(not applicable)
Topic FE		Yes
No. obs	4,590,420	8,203,932
Adjusted $R^2$	0.14	0.19

- Semantic similarity of consecutive search terms for the panelists from UK is higher
- Number of queries related to the same topic is also higher

# GDPR & Online Search

## Product Search

- Parse URLs and look for mentions of product category names
- Group related product search weekly

Figure: Illustration: Parse URLs to Identify Product Browsing Records

User ID	URL	Time	duration	category
0007921a2577d346	www.sears.com/jewelry-pendants-necklaces/b-10201927Stone%20Type=Amethyst...	1/29/2018 9:10	16	Clothing and accessories
0007921a2577d346	www.sears.com/jewelry-pendants-necklaces/b-10201927Stone%20Type=Amethyst...&subCatView=true&unitNo=XXXXXXXX	1/29/2018 9:11	11	Clothing and accessories
0007921a2577d346	www.sears.com/jewelry-pendants-necklaces/b-10201927Price=025&Stone...&filterList=XXXXXX XXXX&subCatView=true&searsTab=true	1/29/2018 9:11	9	Clothing an accessories
0007921a2577d346	www.sears.com/deals/whats-cool-tools.html#/grid	1/29/2018 9:15	68	
0007921a2577d346	www.sears.com/deals/whats-cool-tools.html#/grid?oldBy=sears	1/29/2018 9:15	11	
0007921a2577d346	www.sears.com/crsp/mx/checkout#/checkout/	1/29/2018 9:19	36	

A checkout



User ID	Check out	Time	Total_time	Num_pages	Num_domains	category
0007921a2577d346	www.sears.com/crsp/mx/checkout#/checkout/	1/29/2018 9:19	16+11+9=36	3	1	Clothing and accessories

# GDPR & Online Search

## Product Search

	(1) No. pages	(2) Total time (seconds)	(3) No. domains
GDPR × UK	0.101*** (0.003)	0.262*** (0.007)	0.060*** (0.002)
Constant	0.501*** (0.001)	1.297*** (0.003)	0.307*** (0.001)
Week FE	Yes	Yes	Yes
Product category FE	Yes	Yes	Yes
Panelist FE	Yes	Yes	Yes
Observations	1,264,140	1,264,140	1,264,140
Adjusted $R^2$	0.21	0.21	0.26

- Users in EU region input higher number of queries visit higher number of product pages, spend more time to search for a product under one category.

# GDPR & Online Search

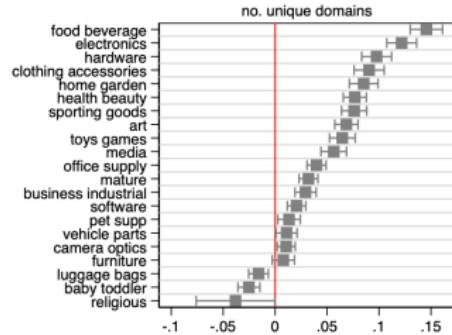
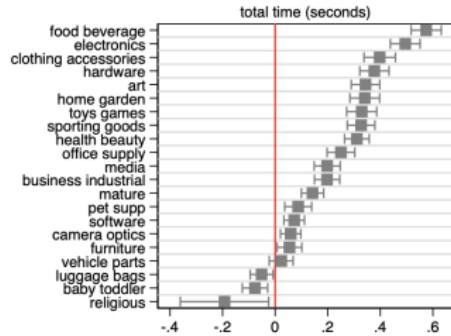
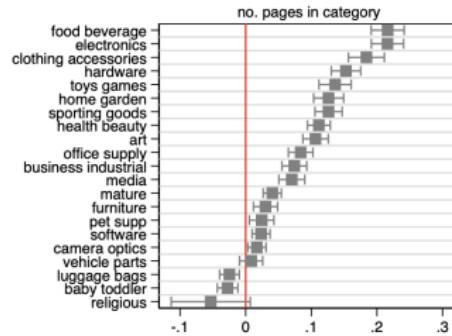
## Product Search

After GDPR, compared to the panelists in the US, panelists in the US:

- input higher number of queries (4.8%)
- visit higher number of product pages (10%)
- visit a larger set of domains (6%)
- spend more time to search for a product under one category (26%)

# Change in search effort by product category

Product Category



Estimates and 95% CI of GDPR X UK



# GDPR & Online Search

## Product Search with Successful Checkouts

GDPR & Search, Conditional on Successful Checkout			
	log(no. pages) (1)	log(total time (seconds)) (2)	log(no. domains) (3)
GDPR × UK	-0.099*** (0.034)	-0.183*** (0.051)	-0.034* (0.018)
GDPR	-0.046 (0.102)	-0.151 (0.153)	0.004 (0.053)
UK	-0.044* (0.025)	-0.123*** (0.037)	0.039*** (0.013)
Product category FE	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes
No. obs	12,284	12,284	12,284
Adjusted $R^2$	0.10	0.10	0.036

- UK panelists view 1.9% fewer product pages, spend 0.21% less time, and visit 1.34% fewer domains in the 48-hours before a successful transaction

# Consumer Traffic & Domain Size

- Provisions that protect smaller domains
- Compare the outcomes for smaller vs. larger firms

# Consumer Traffic & Domain Size

- Provisions that protect smaller domains
- Compare the outcomes for smaller vs. larger firms

$$\log(\text{traffic}_{jt}) = \gamma_0 + \gamma_1 \text{GDPR}_{jt} + \gamma_2 \text{GDPR}_{jt} \times \text{EU-penet}_j + \theta_j + \tau_t + \epsilon_{jt}$$

# Consumer Traffic & Domain Size

- Provisions that protect smaller domains

- Compare the outcomes for smaller vs. larger firms

$$\log(\text{traffic}_{jt}) = \gamma_0 + \gamma_1 \text{GDPR}_{jt} + \gamma_2 \text{EU-penet}_j + \theta_j + \tau_t + \epsilon_{jt}$$

- $\text{traffic}_{jt}$ : unique users that site  $j$  receives in week  $t$

- $\text{EU-penet}_j$ : proxy for the relative exposure of site  $j$  to GDPR regulations

- Share of EU panelists among all panelists visiting  $j$ , in the first quarter of 2018

# Consumer Traffic & Domain Size

GDPR &amp; Consumer Traffic, by Domain Size

	(1) Pre-GDPR traffic below 90th percentile (3)	(2) Pre-GDPR traffic above 90th percentile
GDPR × EU-penet	-0.043*** (0.001)	0.077*** (0.010)
GDPR	0.016** (0.005)	-0.023 (0.015)
Constant	0.148*** (0.003)	2.413*** (0.008)
Domain FE	Yes	Yes
Week FE	Yes	Yes
Observations	1,085,400	67,356
Adjusted $R^2$	0.50	0.86
Mean of DV	0.278	32.121

- Domains visited by number of users see more traffic, whereas the effects for smaller domains may be more negative

# Conclusion

- Preliminary, short term evidence consistent with higher informational frictions following GDPR
  - Consumers spend more time online, visit more sites
- Spend more time searching for information and products, keeping search topic/product category fixed
  - However successful checkouts consist of shorter search
- Increased traffic → increased concentration: does not seem to benefit the small domains

Thank you!

Questions/comments:  
[pyild@wharton.upenn.edu](mailto:pyild@wharton.upenn.edu)

# Appendix

Table: Browsing pattern, long and short term effects

Lengths, post-period	2 weeks	4 weeks	8 weeks	16 weeks
after GDPR	-.05571*** (.007305)	-.06819*** (.006189)	-.06819*** (.006278)	-.06819*** (.00618)
after GDPR $\times$ EU	.05221*** (.01024)	.03391*** (.008718)	.03391*** (.008843)	.03391*** (.008705)
Constant	3.903*** (.001841)	3.905*** (.001891)	3.896*** (.002127)	3.88*** (.002549)
Observations	92650	102308	121314	158600
Adjusted $R^2$	.71	.71	.72	.71

Pre-period lengths are fixed across columns

Standard errors in parentheses