HEC PARIS

The Impact of Privacy Regulation on the Online Advertising Market

Prof. Dr. Klaus M. Miller
HEC Paris

Hi! Paris Summer School
July 8, 2022

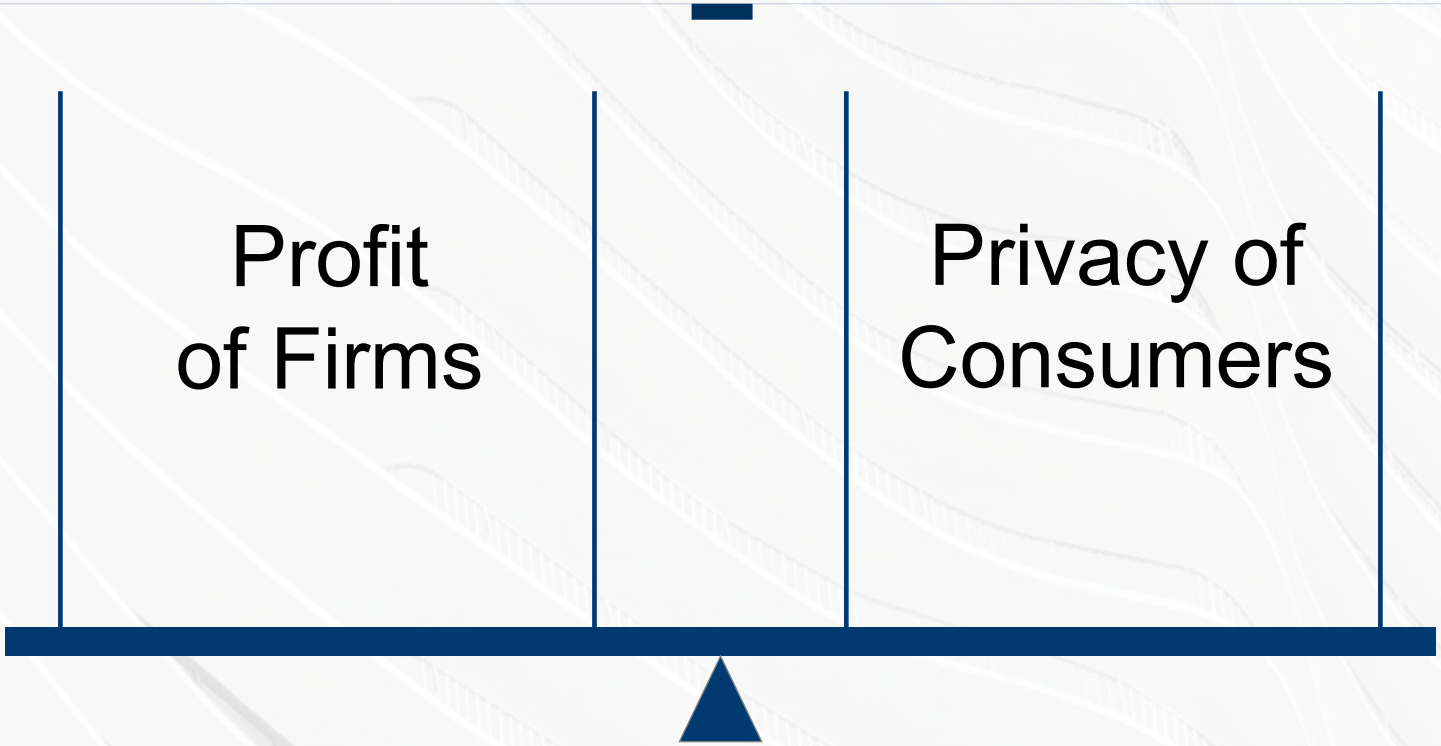
Supported by the European Research Council (ERC)

2022

2022

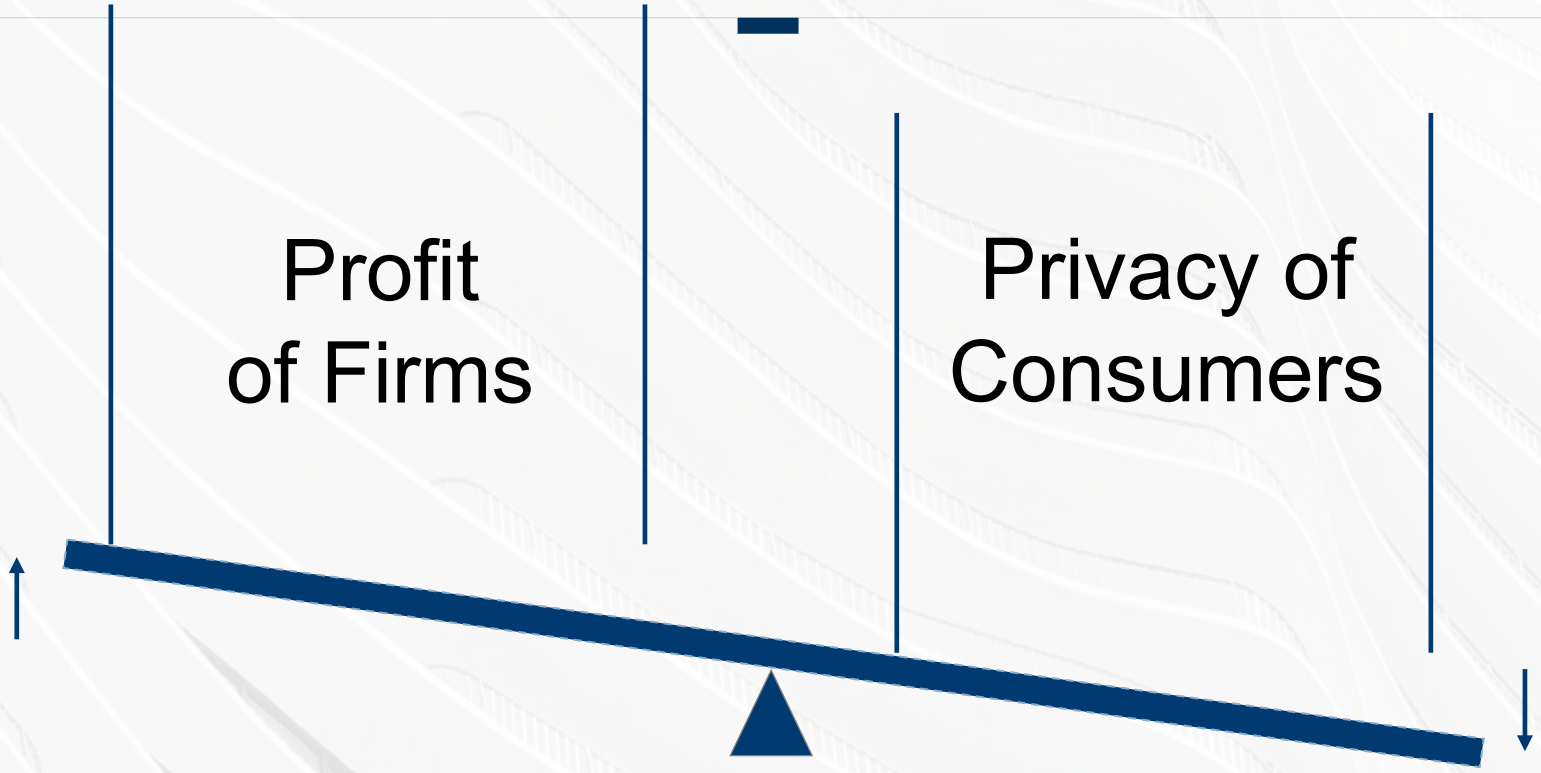
Fundamental Trade-Off Between Firm Profit and Consumer Privacy

Trade-Off between Profit and Privacy



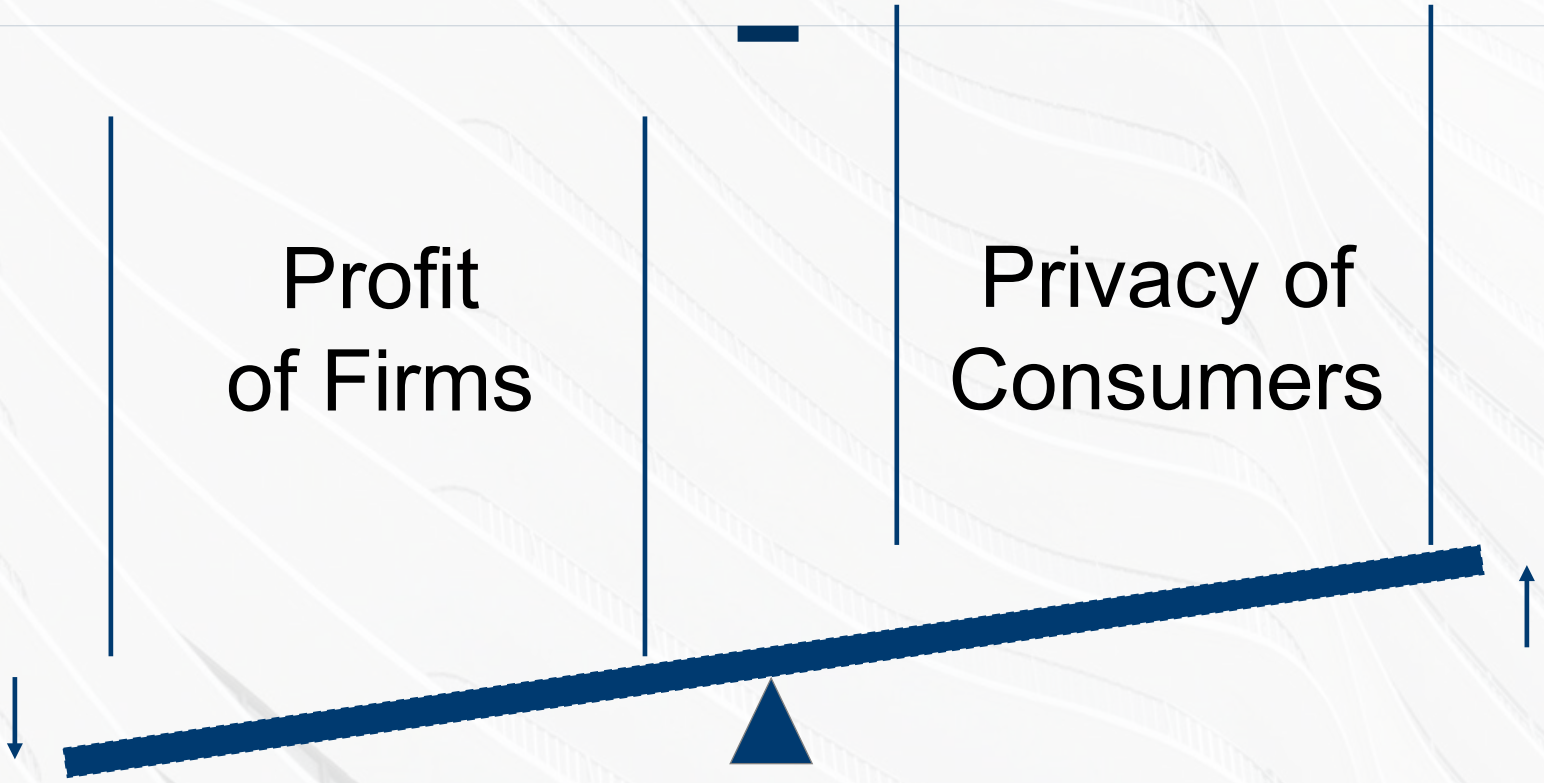
Appropriate Use of Tracking Technologies

Heavy Use of Tracking Technologies



Heavy Use of Tracking Technologies

Light Use of Tracking Technologies



Light Use of Tracking Technologies

Aim and Structure of Tutorial

Our Plan for Today

Timing	Content
08:30 – 10:00	**Session I: Introduction & Overview**
10:00 – 10:30	Break
10:30 – 12:00	**Session II: Empirical Results**
12:00	End of Tutorial

*Note: This schedule is a rough guideline and may be subject to change based on class interaction and discussion.

Planned Contents

- **Session I: Introduction and Overview**
  - Online Advertising Market
    - Actors
    - Generation of Ad Prices
    - Tracking, Profiling, Targeting
  - Importance of GDPR for Online Advertising Market

- **Session II: Empirical Results**
  - Impact of the GDPR on User Tracking
  - Quality of User Tracking Data
  - Impact of the GDPR on the Value of User Tracking

- **Summary and Outlook**

Your Lecturer For Today

Your Lecturer: Klaus M. Miller

Current Position

- Since Fall 2021: Assistant Professor of Quantitative Marketing, HEC Paris, France
  Hi! Paris Chair for the Study of Artificial Intelligence in Business & Society

Visiting Scholar

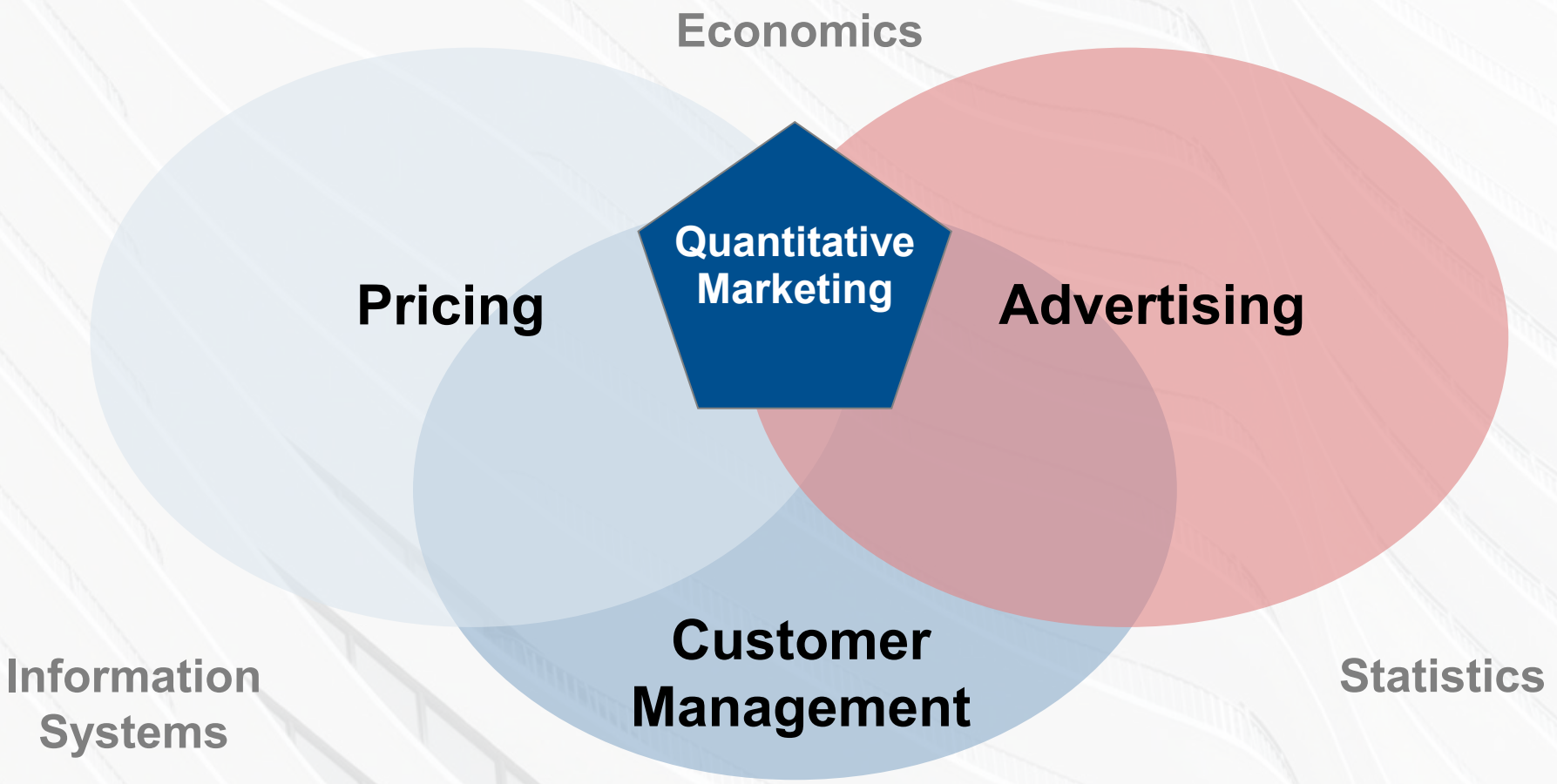
- 2019: Graduate School of Business, Stanford University, USA
- 2009: Graduate School of Business, Stanford University, USA
- 2008: Wharton School, University of Pennsylvania, USA

Prior Positions

- 2015 – 2021: Assistant Professor of Quantitative Marketing, Goethe University, Frankfurt, Germany
- 2011 – 2015: Head of Business Intelligence & Analytics, Ringier AG, Zurich, Switzerland
- 2013 – 2014: Lecturer in Quantitative Marketing, University of Zurich, Switzerland

- 2008 – 2012: Post Doctoral Research Scholar, University of Bern, Switzerland
- 2005 – 2008: Research and Teaching Associate, University of Bern, Switzerland

Scientific Competence in Quantitative Marketing & Analytics



Top Publications in International Scientific Journals

Shunyao Yan, Klaus Miller, Bernd Skiera (2022):
How Does the Adoption of Ad Blockers Impact News Consumption?,
Journal of Marketing Research.

Reto Hofstetter, Klaus Miller, Harley Krohmer, John Zhang (2021):
A De-Biased Direct Question Approach to Measuring Consumers' Willingness to Pay,
International Journal of Research in Marketing.

Klaus Miller, Reto Hofstetter, Harley Krohmer, John Zhang (2011):
How Should Consumers' Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches,
Journal of Marketing Research.

Reto Hofstetter, Klaus Miller, Harley Krohmer, John Zhang (2013):
How Do Consumer Characteristics Affect the Bias in Measuring Willingness to Pay for Innovative Products,
Journal of Product Innovation Management.

Selected Research Project Partners

Contact Information

Prof. Dr. Klaus M. Miller

HEC Paris

1 Rue de la Libération

78350 Jouy-en-Josas

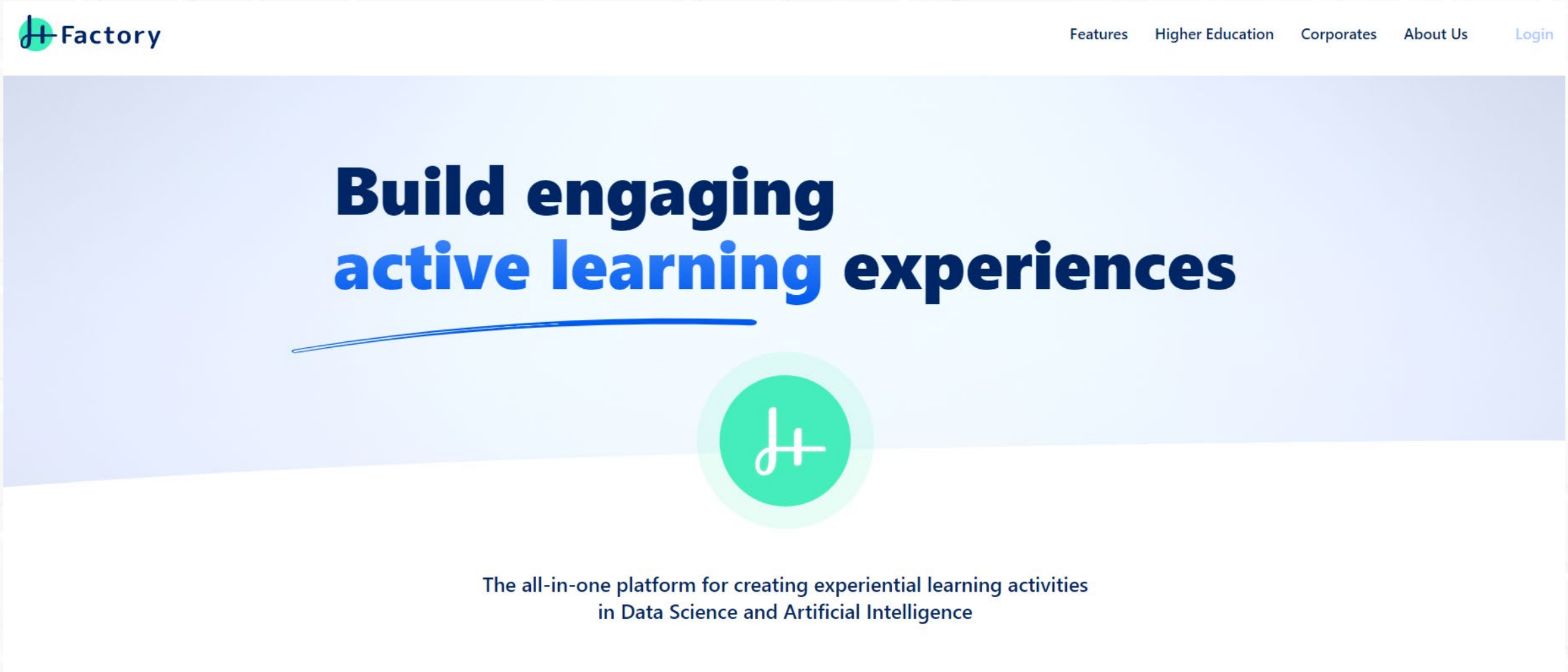
Email: millerk@hec.fr

Web: klausmiller.com

@klausmiller

Additional Materials

Course Material



Our Book

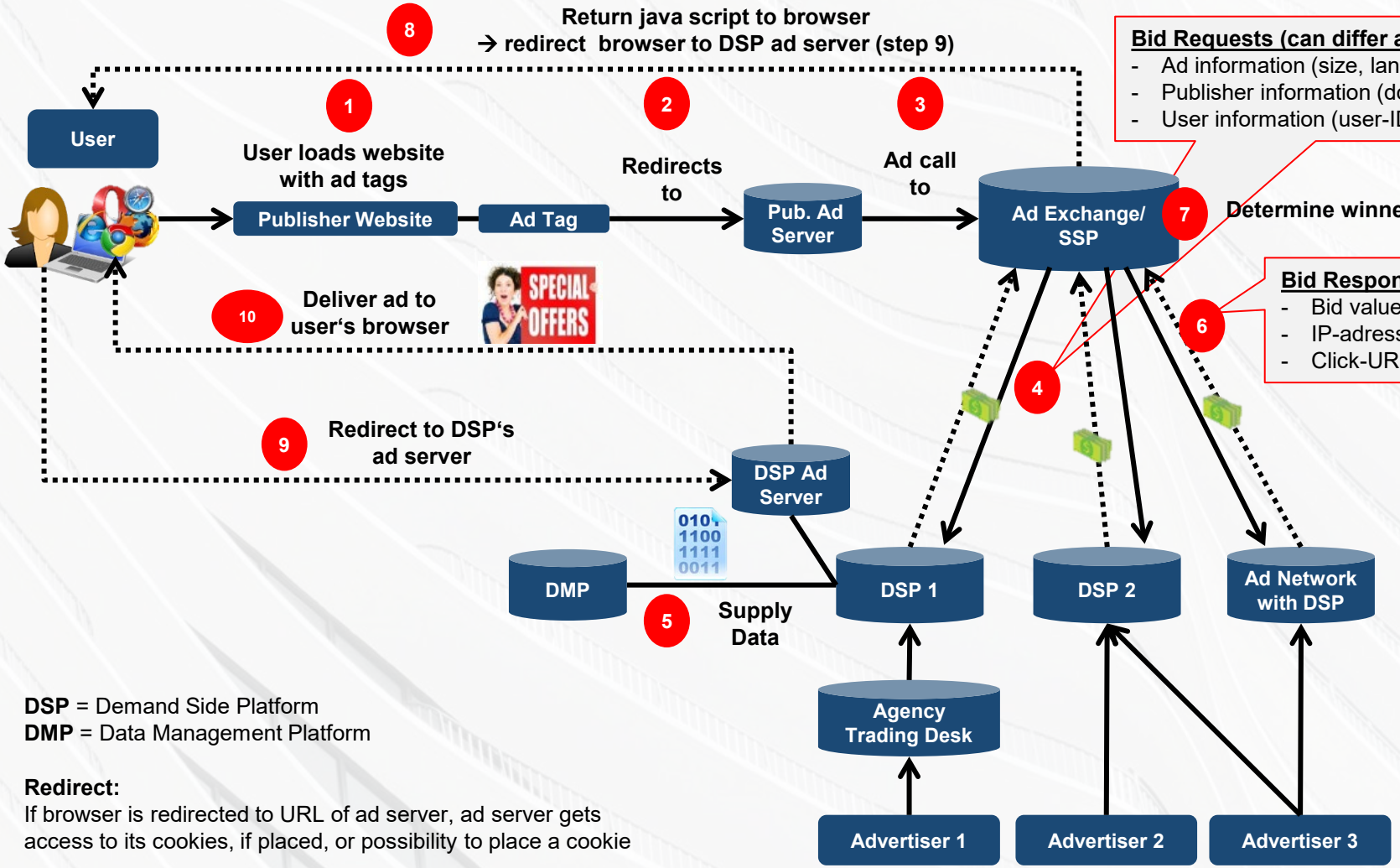
- Published: January, 2022
- gdpr-impact.com
- Also available
  - Interactive PDF
  - Kindle eBook
  - Paperback

Session I:

Introduction and Overview of Online Advertising Industry

At Least 3 Parties in the Online Advertising Market

Process to Sell Online Display Ads



Bid Requests (can differ across bidders):
- Ad information (size, language)
- Publisher information (domain)
- User information (user-ID)

Bid Responses
- Bid value
- IP-adress of ad server
- Click-URL for ad

DSP = Demand Side Platform
DMP = Data Management Platform

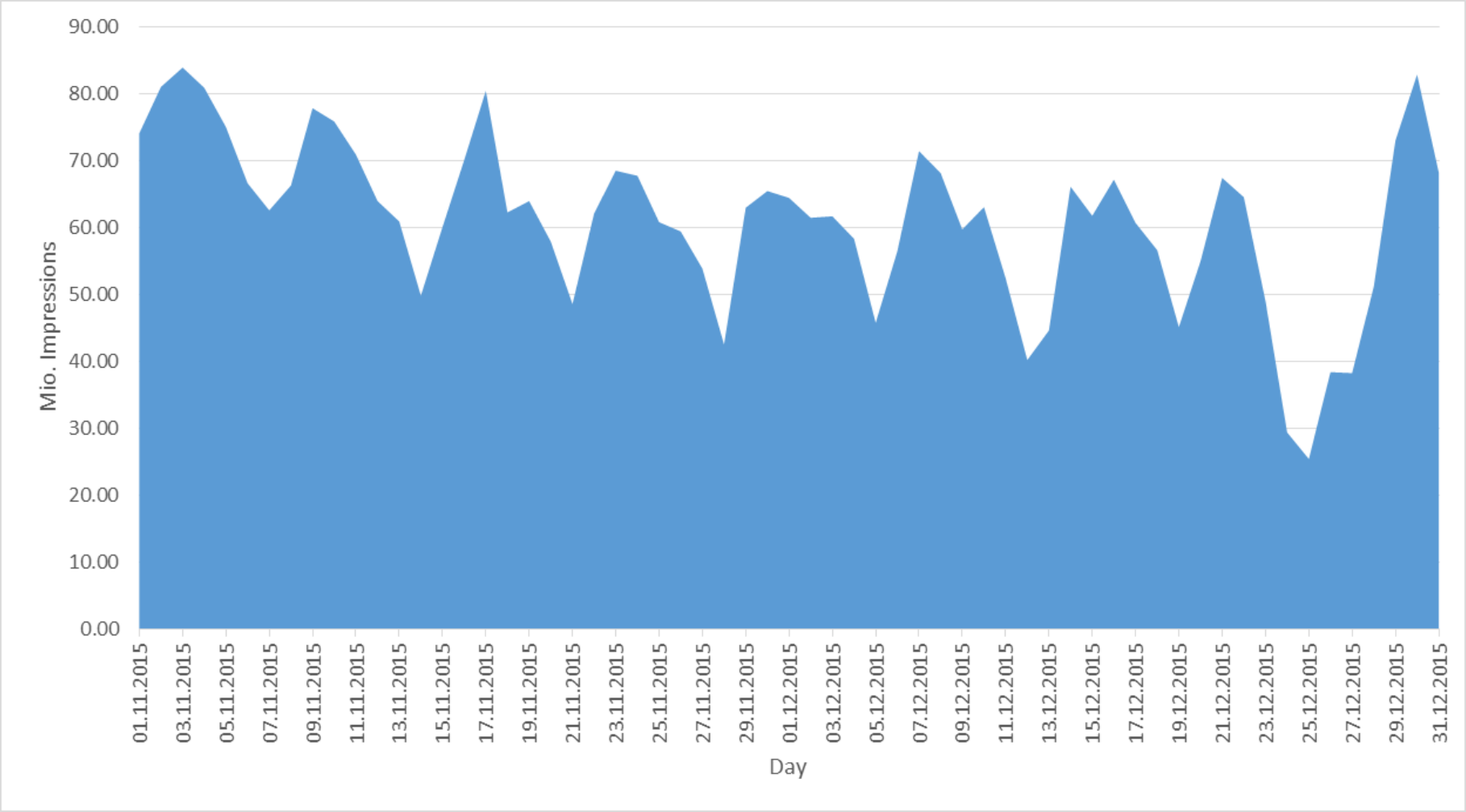
Redirect:
If browser is redirected to URL of ad server, ad server gets access to its cookies, if placed, or possibility to place a cookie

Ad Tag

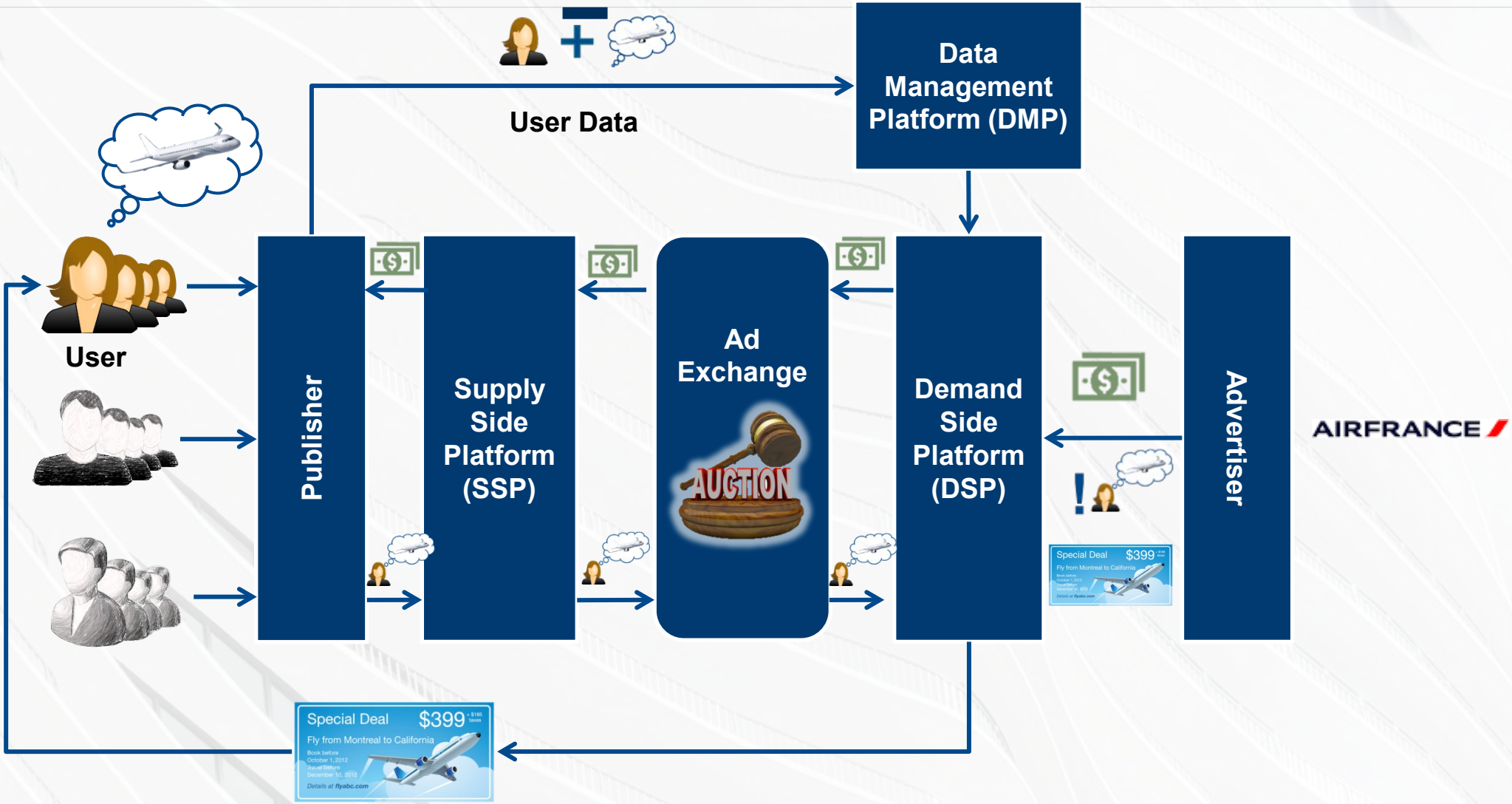
```
<script> src=„ib.adnxs.com/ttj?id=„123" type=text/javascript</script>
```

Laub, Miller, Skiera (2022) - When the Cookie Crumbles? The Impact of Online Tracking Restrictions on the Online Advertising Market

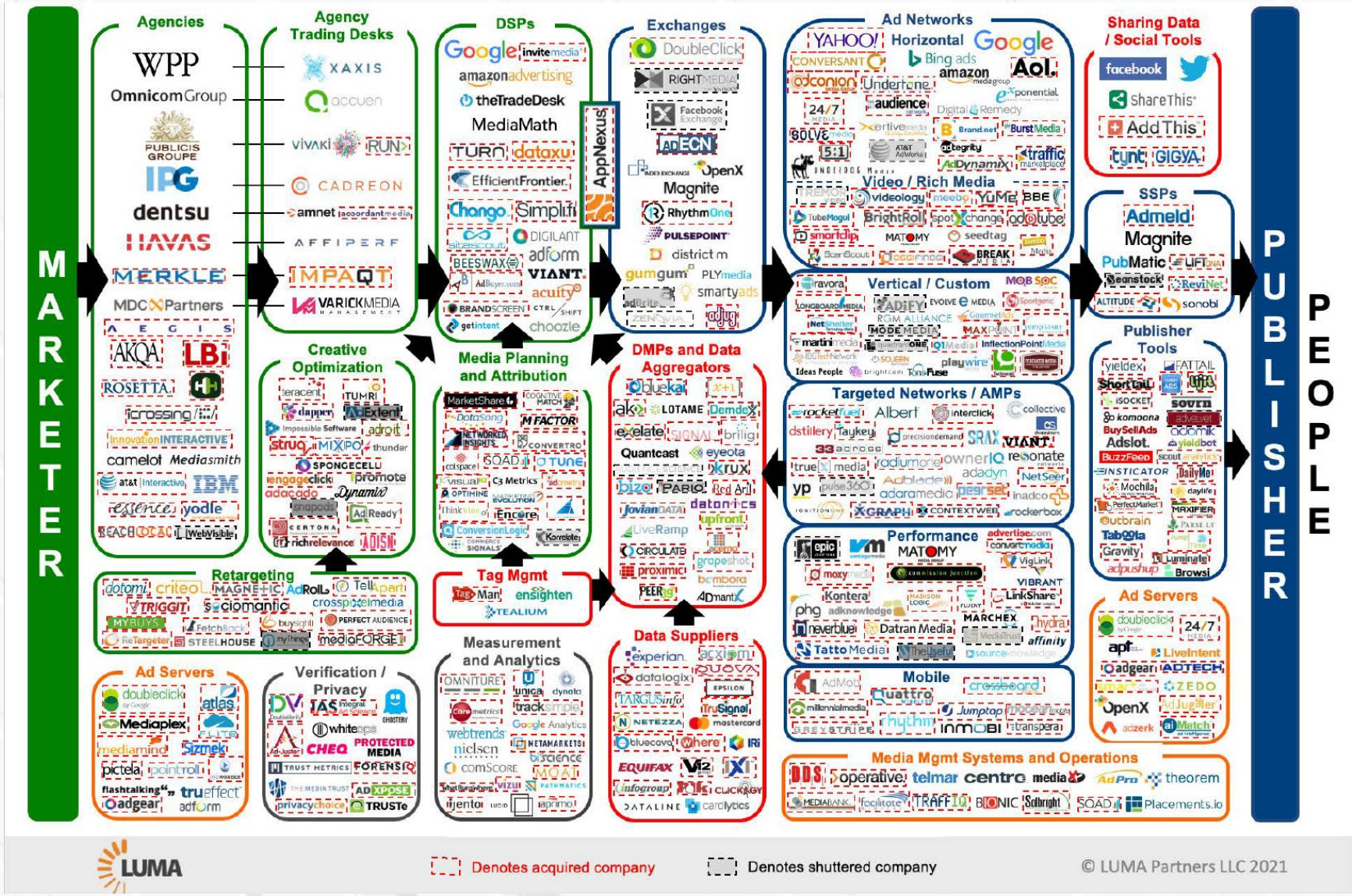
Daily Number of Auctions to Sell Online Display Ads



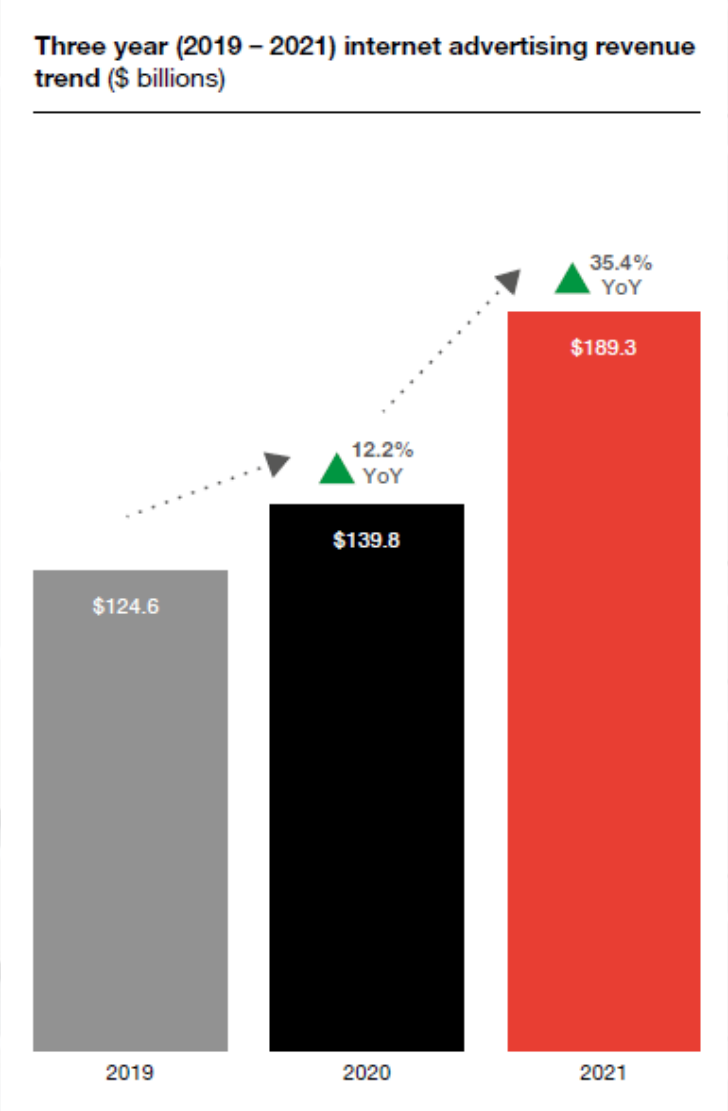
Different Parties in Online Advertising



Display Luma Landscape



US Internet Advertising Revenues



Source: IAB Internet Advertising Revenue Report 2021

Impact of Privacy Regulation on Online Advertising Market

Online Advertising Market Before GDPR (General Data Protection Regulation)

Publishers

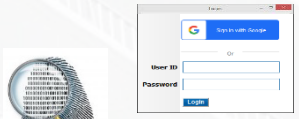
Contact & Data

Free Content

Contact & Data

Ad Revenue

Tracking



Users

Targeting

Purchase

Advertisers

Categorization of User Tracking Technologies

Identify a User on	Identify a User on: own Website	Identify a User on: other Website(s)
single Device	First-Party Cookie Digital Fingerprinting Advertising Identifiers Local Storage Tracking Pixel	Third-Party Cookie Digital Fingerprinting Advertising Identifiers Local Storage Tracking Pixel
multiple Devices	Login	Login (SSO)

Impact of GDPR on Online Advertising

Publishers

Contact & Data

Free Content

Ad Revenue

Contact & Data

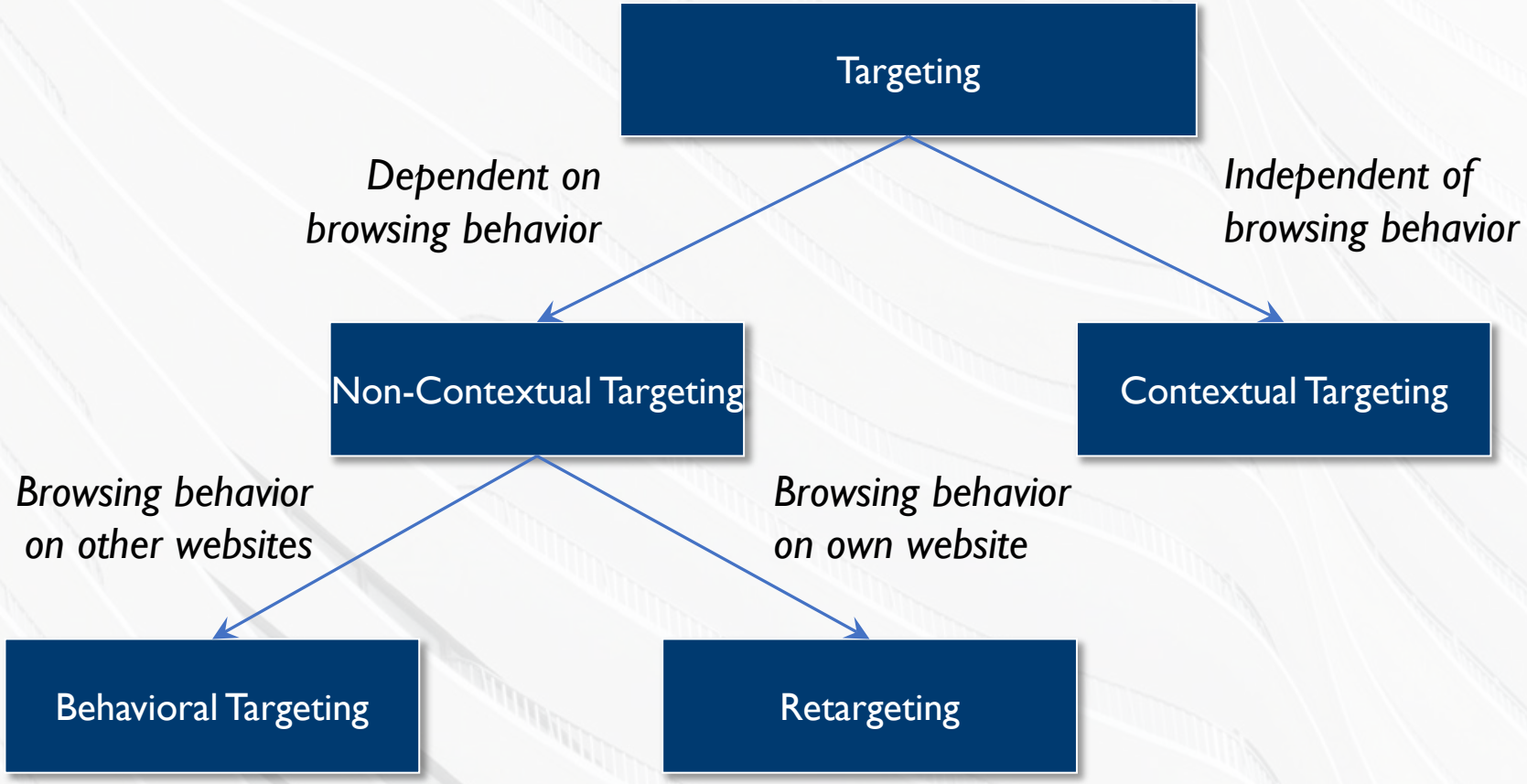
Users

Targeting

Purchase

Advertisers

Targeting with Display Advertisements



Overview of GDPR

General Data Protection Regulation (GDPR)

- **European online privacy law introduced in May 2018**

- **Aim**
  - Increase consumer privacy by
    - strengthening consumers' control over their personal data
    - harmonizing EU member states' privacy laws

- **Broad definition of personal data**
    - *„Any information that might be linked to an identifiable individual"*

- **Broad Applicability**
  - All European firms
  - All non-European firms processing European consumers' personal data

GDPR Fines

- **Data controller (e.g., a publisher)**
  - Firm that has obligation of deciding why and how to process the personal data (GDPR Article 4 point (7))

- **Data processor (e.g., an advertiser)**
  - Firm that processes personal data on behalf of a data controller (GDPR Article 4 point (8))

Definition of Personal Data

- **The GDPR defines personal data as follows (Article 4):**

  *[…] any information relating to an **identified** or **identifiable** natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a **name**, an **identification number**, **location data**, an **online identifier** or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*

- **Data directly identifying a consumer**
  - E.g., name, address, birth date, or social security number

- **Data indirectly identifying a consumer**
  - E.g., online identifies (cookies, digital fingerprints, …)

Definition of Personal Data

- **GDPR applies only to pseudonymous data, but not anonymous data**

- Personal data are considered **pseudonymous** if they do not directly identify a user but can be used to identify a user indirectly. For example, a customer number (e.g., “123456789”) does not, in itself, directly identify a user. However, when combined with information relating the customer number to an individual user (e.g., “123456789” represents user “X”), the customer number is considered pseudonymous data.

- Personal data are only considered **anonymized** if they do not identify a user at all (e.g., it is unknown which user the customer number “123456789” represents).

- Thus, firms that collect data from consumers and seek to avoid the GDPR cannot suffice with pseudonymizing user information but rather must anonymize it, which may be impractical and costly.

Definition of Data Controller and Data Processor

- **Data controller (e.g., a publisher)**
  - Firm that has obligation of deciding why and how to process the personal data (GDPR Article 4 point (7))

- **Data processor (e.g., an advertiser)**
  - Firm that processes personal data on behalf of a data controller (GDPR Article 4 point (8))

Obligations for Data Controller and Data Processor

Obligation	Data Controller	Data Processor
Processing Personal Data Based on a Legal Basis and Compliance with Associated Requirements	Yes	Yes
Documentation of Personal Data Processing	Yes	Yes
Implementation of Technical and Organizational Measures (Safeguarding Privacy by Design & Default)	Yes	Yes
Provision of Notification in Case of Data Breach	Yes	Yes
Selection of Purposes of Personal Data Processing	Yes	No
Justification of Relevance of Personal Data	Yes	No
Assurance of Data Processor's GDPR Compliance	Yes	No

Six Legal Basis to Process Personal Data under GDPR

- **Not applicable for online advertising industry**
  - Vital interest
  - Public interest
  - Contract fulfillment
  - Legal obligation

- **Applicable for online advertising industry**
  - Consent
    - for “one or more specific purposes” (Art. 6 (1) (a), GDPR)
  - Legitimate interest
    - “necessary for the purposes of the legitimate interest pursued by a data controller or a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject” (Art. 6 (1) (f), GDPR)
      - User has right to object

Comparison of Legal Bases for Tracking in Major Privacy Laws Worldwide

Privacy Law	Possible Legal Bases	Possible Legal Bases for Tracking
GDPR (EU)	• Consent • Legitimate Interest • Legal Obligation • Contract Fulfillment • Public Interest • Vital Interest	• Consent (opt-in approach) • Legitimate Interest
CCPA (California)	• Consent	• Consent (opt-out approach)
LGPD (Brazil)	• Consent • Legitimate Interest • Legal Obligation • Regular Exercise of Rights • Contract Fulfillment • Credit Protection • Public Interest • Vital Interest • Health Protection • Research Activities	• Consent (opt-in approach) • Legitimate Interest
PDPB (India)	• Consent • State Performance • Legal Obligation to State • Legal Obligation to Court • Vital Interest (Medical Services only) • Health Service (in Epidemics) • Individual Safety (for Disasters) • Employment Purposes • Legitimate Interest (limited to specific situations)	• Consent (opt-in approach)
PDPA (Thailand)	• Consent • Legitimate Interest • Legal Obligation • Contract Fulfillment • Public Interest • Vital Interest	• Consent (opt-in approach) • Legitimate Interest
PIPL (China)	• Consent • Legal Obligation • Contract Fulfillment • Public Interest • Vital Interest • Disclosed Information (disclosed by data subjects themselves or otherwise lawfully disclosed)	• Consent (opt-in approach) • Consent (opt-in separately for sensitive information)

Development of the Approaches to Ask Users for Consent

Before GDPR | Early Stage | Now

May 2018

- GDPR
- Clarifying consent

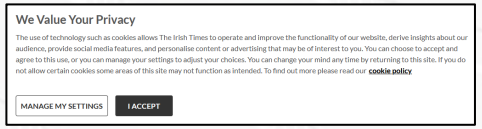
Over time

- Guidelines: equal prominence of refuse-option
- Regulators accept Pay-or-Consent Walls

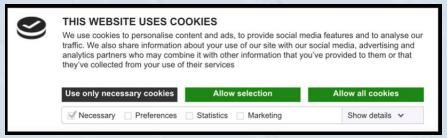
Notice-only

No consent banner

Notice-only

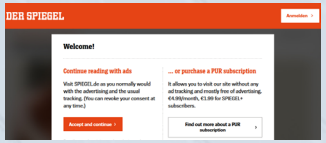


Consent & Settings



Consent & Free Refuse (CFR)

= *Alternative 1 (Status-quo)*



Pay-or-Consent Wall (PoC)

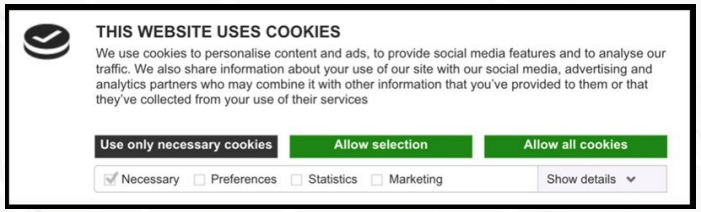
= *Alternative 2 (Spreading Fast)*

Publishers now decide between two alternatives to receive user's consent

Publishers Increasingly Make Users Pay for Privacy

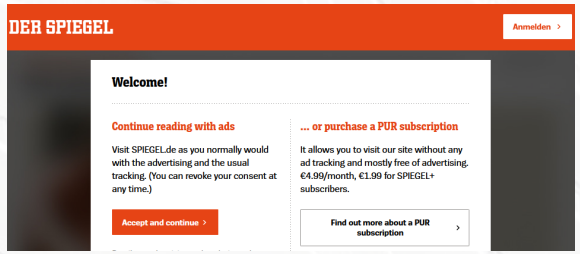
Consent & Free Refuse (CFR)

- Users can freely choose privacy
- Status-quo



Pay-or-Consent Wall (PoC)

- Users must pay if they want privacy
- Spreading fast

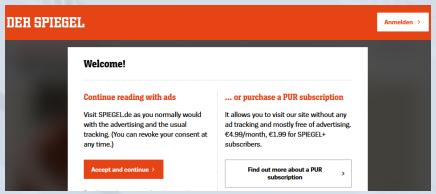


May 2018

Austrian "Der Standard" introduces PoC

February 2020

German "Der Spiegel" adopts PoC



Today

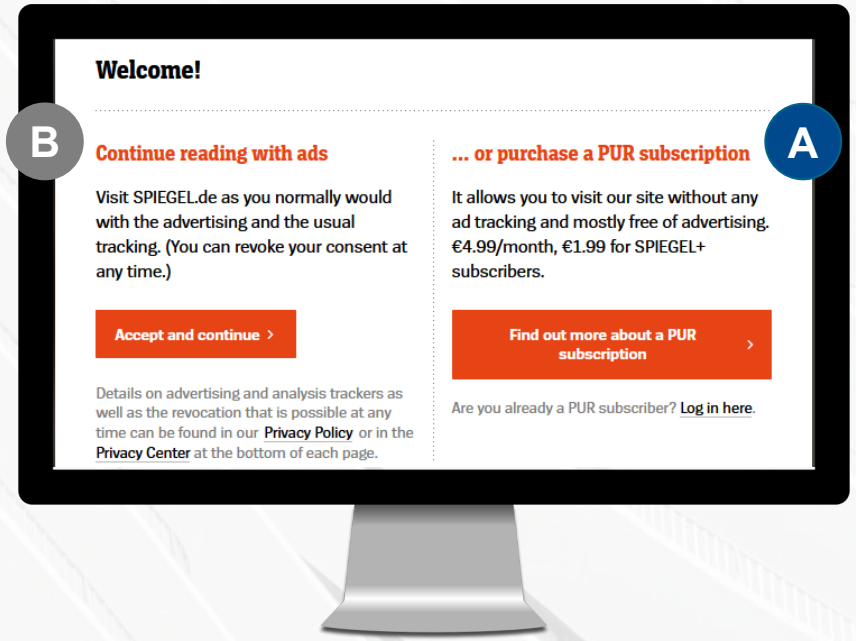
- Many publishers (e.g. Germany>55)
- Reach: Millions of users in Austria, Germany and France

e.g.

- France: lemonde.fr, gala.fr
- Germany: bild.de, zeit.de
- Austria: krone.at

Tracking in Practice: Combination of Various Legal Basis to Track Users Online

Exemplary Pay-or-Consent Wall of spiegel.de



Pay-or-Consent Walls block the access to content until the user decides between “pay” or “consent”

Three Options:

A **“Pay”**

- User pays for not being tracked
- No personal data + no advertising

B **“Consent”**

- User consents to data collection
- Data used for advertising

C **“Leave”**

- User chooses none of the above and leaves

Session II:

Empirical Results

Overview of Empirical Results on the Impact of Privacy Regulation on the Online Advertising Market

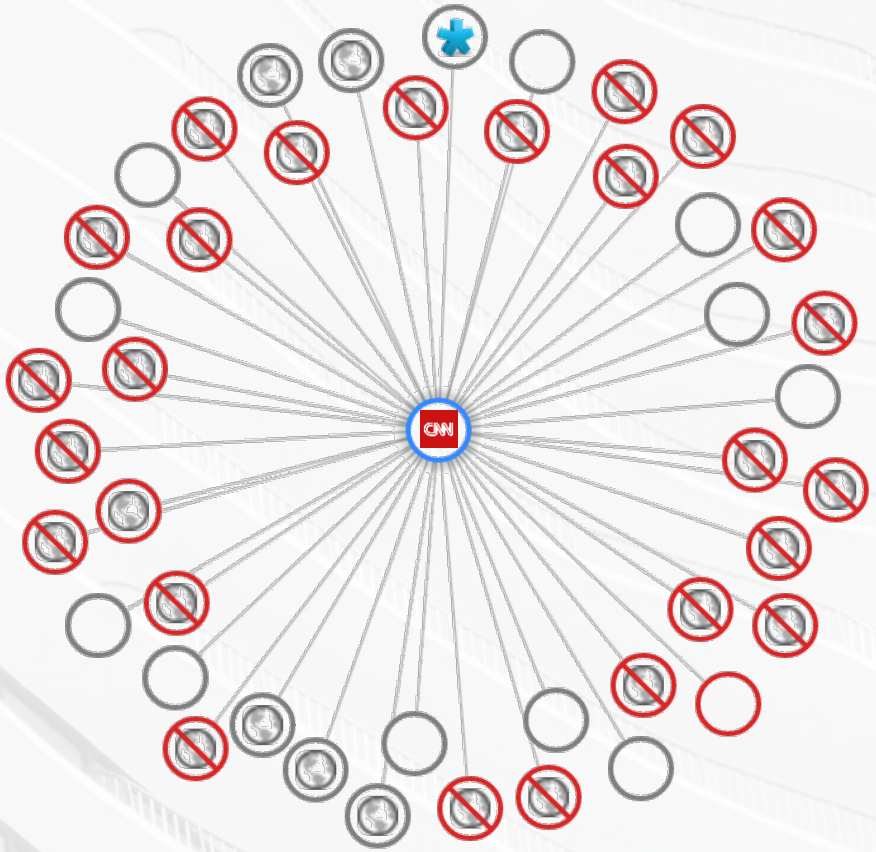
- Impact of the GDPR on User Tracking
- Quality of User Tracking Data
- Impact of the GDPR on the Value of User Tracking

Impact of General Data Protection Regulation (GDPR) on Online Tracking

Karlo Lukic, Klaus M. Miller, Bernd Skiera

HEC PARIS

Motivating Example:
CNN's Usage of Trackers in July 2022



cnn.com
used 36 trackers

Visualized with Disconnect browser extension for Firefox: https://disconnect.me/disconnect

cnn.com accessed on July 1st, 2022

Research Aim, Question and Relevance

- **Research aim**
  - Did GDPR increase EU users' online privacy?

- **Research questions**
  - Did GDPR decrease:
    - Quantity of online tracking (e.g., number of trackers)
    - Intensity of online tracking (e.g., share of tracked pages)

Today's focus

- **Relevance:**
  - Users
    - Did GDPR increase privacy?
  - Regulators
    - Did GDPR achieve its aim?
    - Insights for upcoming privacy regulation(s)?

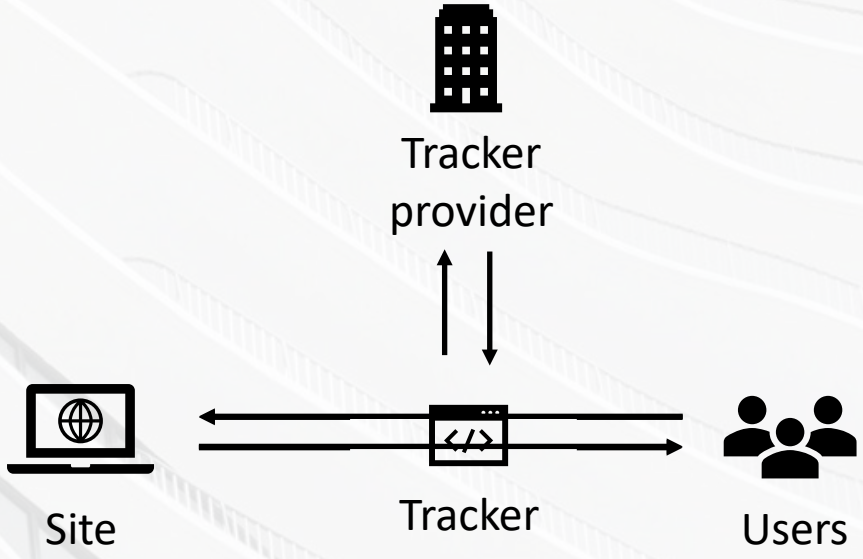
Description of Tracker Market Structure

Description of Online Tracking, Tracker and Tracker Provider

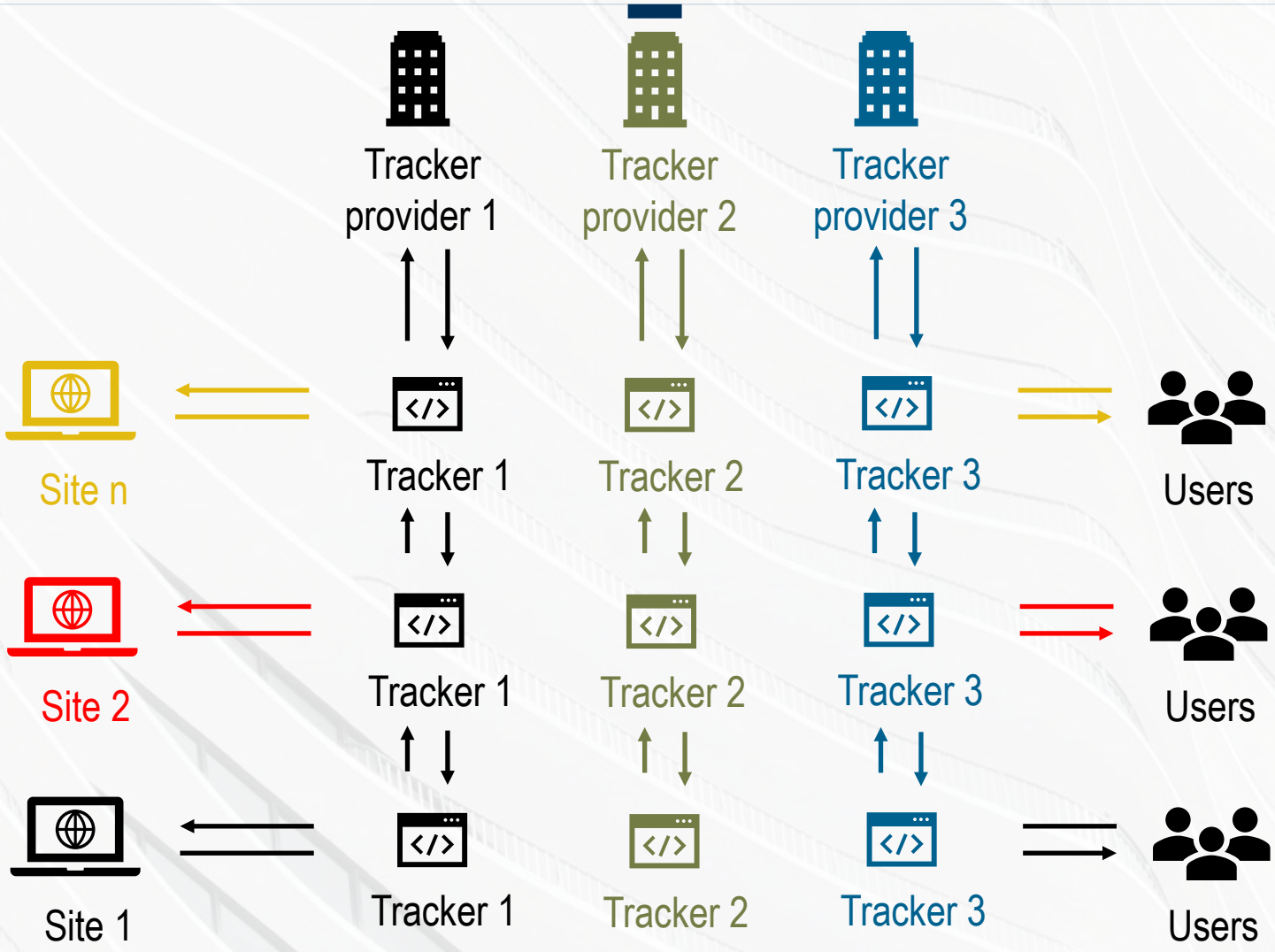
- **Online tracking**
  - Observing what users do online over time
- **Tracker**
  - Software installed on single or multiple pages of a site to track users
- **Tracker provider**
  - Company that develops tracker(s)
- **Site**
  - Company that uses tracker(s)

1. Example of Tracker Market: Single Site and Tracker

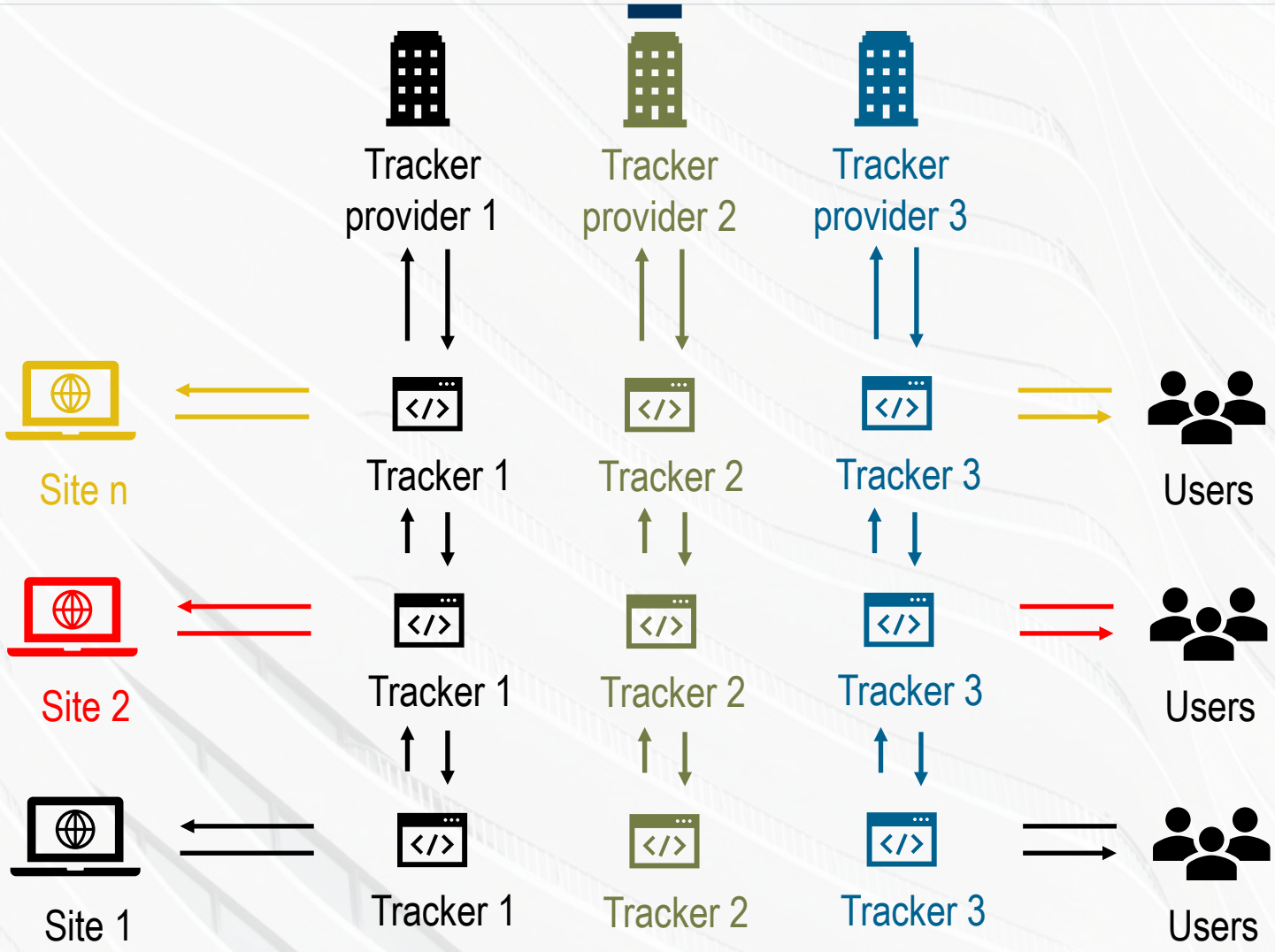
- 2 benefits of trackers for sites:
  - Use data for more targeted advertising
  - Use data for better user experience



2. Example of Tracker Market: Multiple sites and tracker providers



2. Example of Tracker Market: Multiple sites and tracker providers

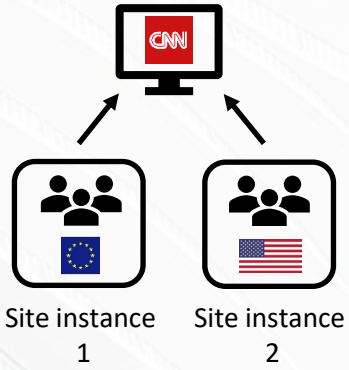


Setup of Empirical Study

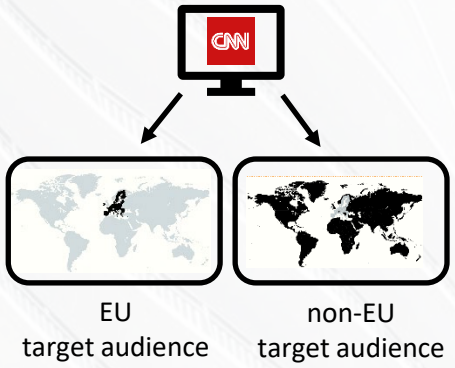
Description of Data

1. **WhoTracks.me**
   - Information on tracking measures (quantity and intensity)
   - 717 sites accessed by:
     - EU
     - US users

     } 2 site instances per site (1,434 site instances)
   - Each month from April 2018 to December 2019 (T = 21 months)
   - We use balanced panel (N = 30,114 = 717 * 2 * 21 observations)



2. **SimilarWeb**
   - Complements main data set to find site's target audience
   - Scraped data (September 2021)
     - Share of traffic from (non)-EU users for 717 sites
   - Acquired data (April 2018 – December 2019)
     - Share of traffic from (non)-EU users consistent over time for 368 of 717 sites (51%)



Description of Dependent Variable

- Dependent variable:
- Distinct number of trackers per site instance

- Illustrative example:

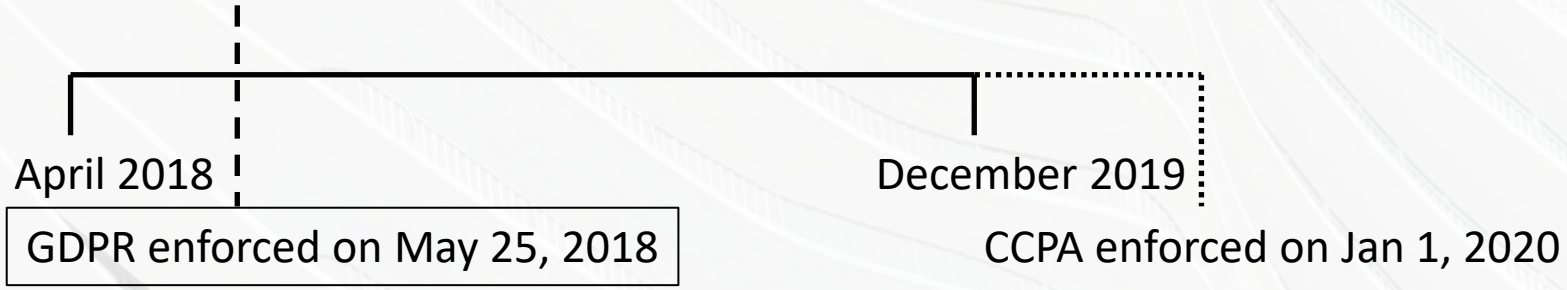
Site	Base of User	Site Instance	Page	Number of Distinct Trackers Per Page	**Distinct Number of Trackers Per Site Instance**
cnn.com	EU	1	https://edition.cnn.com/	44	**46 (or higher)**
			https://edition.cnn.com/world	46	
			…	…	
cnn.com	US	2	https://edition.cnn.com/	41	**48 (or higher)**
			https://edition.cnn.com/world	48	
			…	…	

- Increase in online privacy = decrease in distinct number of trackers

Basic Idea of Analysis

- Difference in differences (DiD) design:
- 2 periods: before and after GDPR's enforcement
- 2 groups: treatment and control

- Time frame:



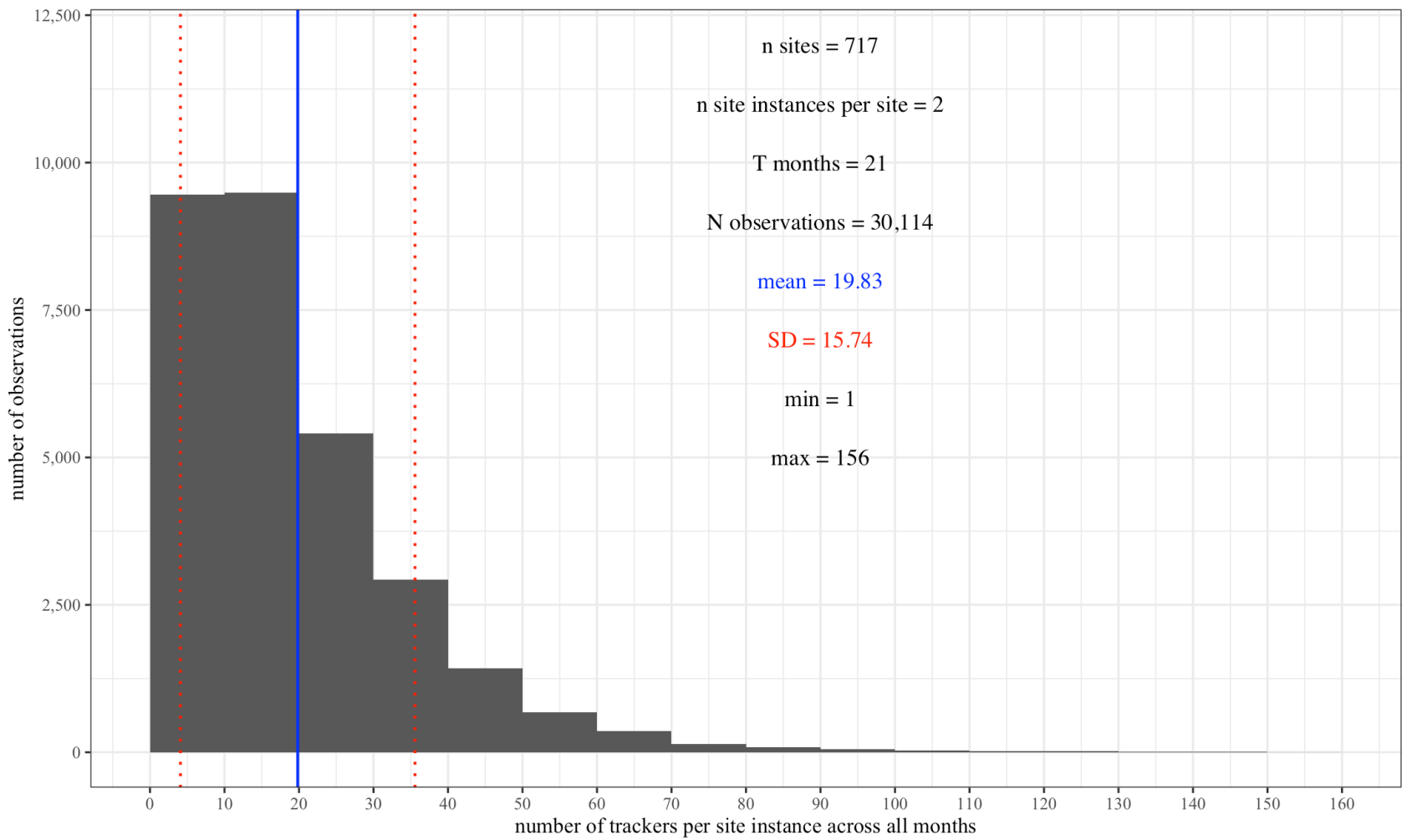
- Treatment assignment framework:

	Base of User	
Website Target Audience	**EU**	**non-EU**
EU	GDPR applies	GDPR applies
non-EU	GDPR applies	GDPR does not apply

Notes: The cell belonging to the control group is colored in blue, the cells belonging to the treatment group are colored in orange.

Results of Empirical Study

Distribution of Number of Trackers per Site Instance Across all Months



Notes: Multiplying the number of sites (717) with the number of site instances (2) and the number of months (21) yields the number of observations (30,114).

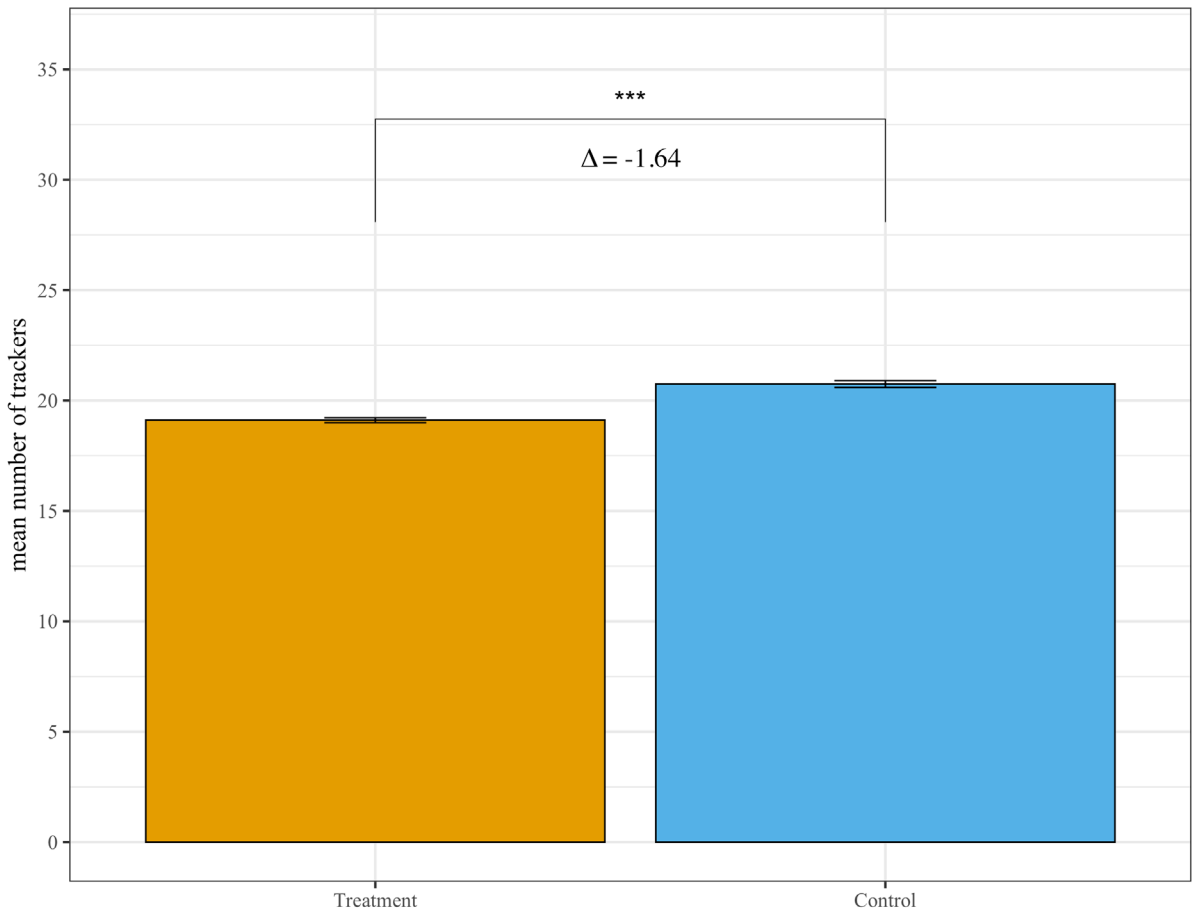
Distribution of Site Instances across Website Target Audience and Base of User

	Base of User	
Website Target Audience	**EU**	**US**
EU[1]	1,869 (6%)	1,869 (6%)
non-EU[2]	13,188 (44%)	13,188 (44%)
∑	15,057 (50%)	15,057 (50%)

[1]The website targets EU users if (1) the website uses an EU top-level domain (e.g., .de) or (2) the website received more traffic from the EU than the non-EU users in at least one period. [2]The website targets non-EU users if (1) the website uses a non-EU top-level domain (e.g., .com) and (2) the website received more traffic from the non-EU users than the EU users in at least one period.

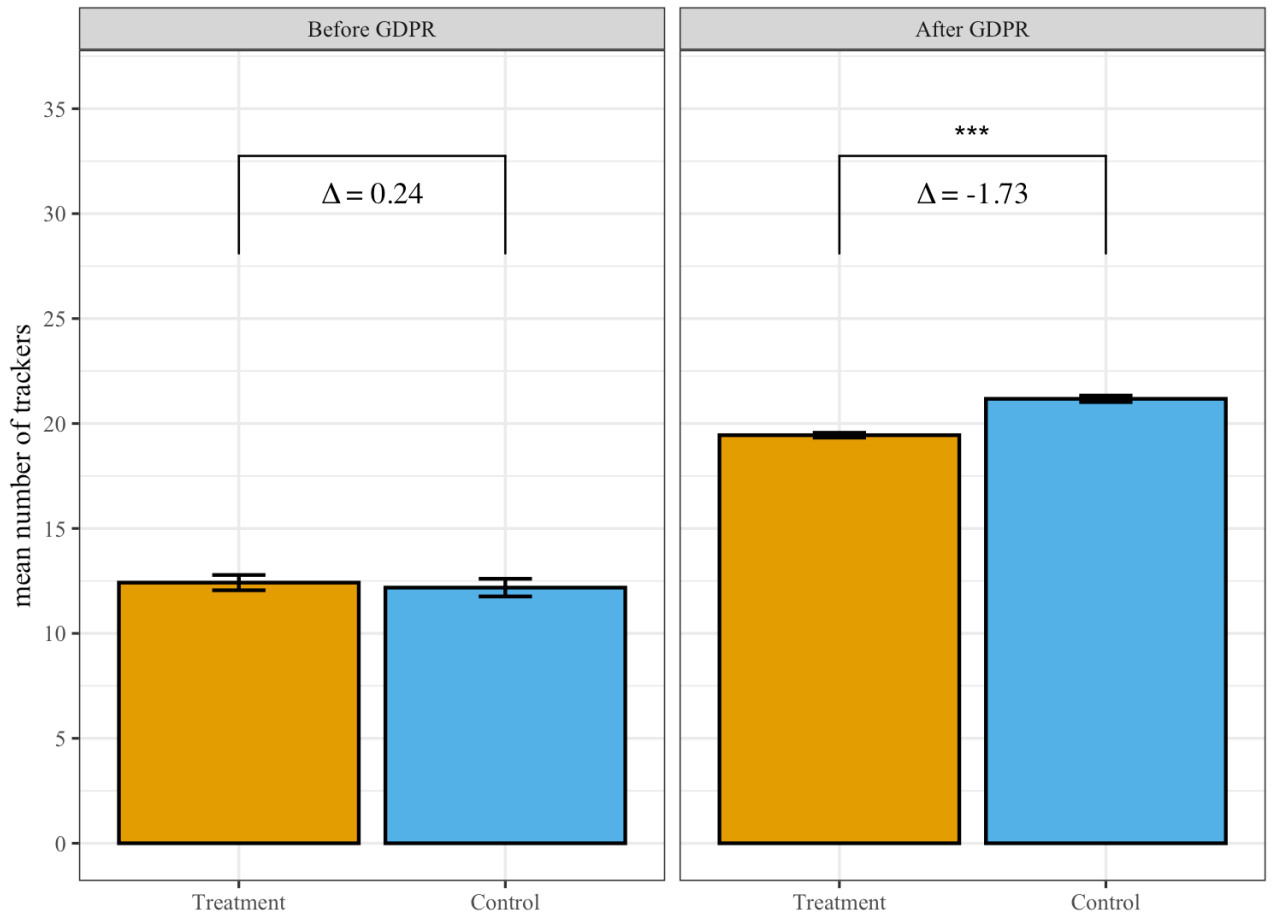
Notes: This table shows the number and the percentage of observations in our sample per cell. The cell belonging to the control group is colored in blue, the cells belonging to the treatment group are colored in orange. In total, 56% (N = 16,926) of all observations (N = 30,114) make the treatment group and 44% (N = 13,188) the control group.

Difference in Number of Trackers between Groups across all Months



Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001
Notes: Independent t-test comparisons between group means across all periods.

Difference in Number of Trackers between Groups before and after GDPR



Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001
Notes: Independent t-test comparisons between group means in periods before and after GDPR.

Results of DiD Analysis for Number of Trackers

- Cross-table for number of trackers:

Group	Pre-GDPR	Post-GDPR	Difference
Treatment	12.42	19.44	7.02
Control	12.18	21.18	9.00
Difference	0.24	-1.73	-1.97

- DiD coefficient for number of trackers:

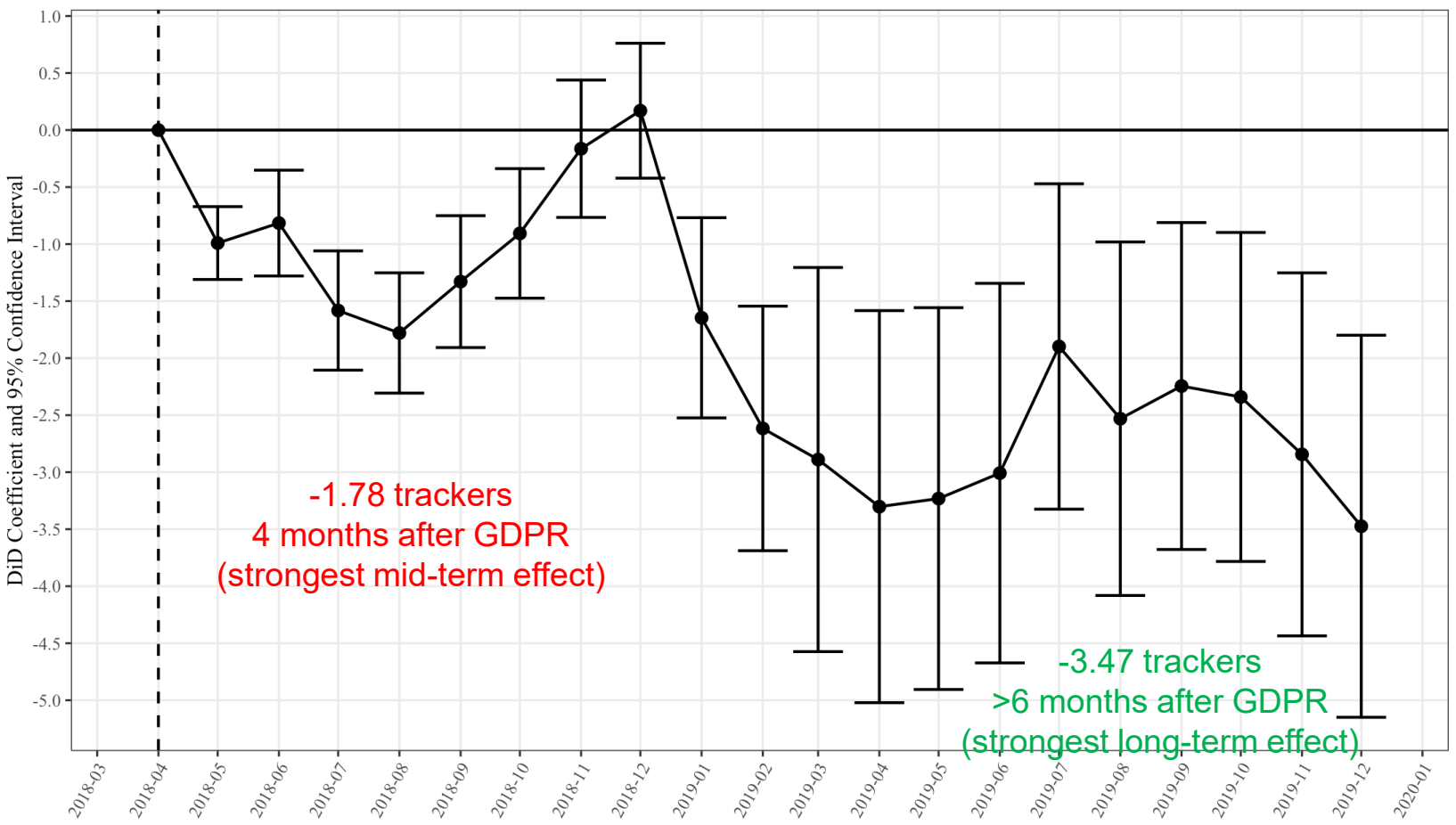
Dependent Variable:	Number of trackers per site instance and month
Model:	(1)
Treatment x Post	-1.97*** [-2.90; -1.05]
Site Instance FE	✓
Period FE	✓
N Observations	30,114
R^2	0.78

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

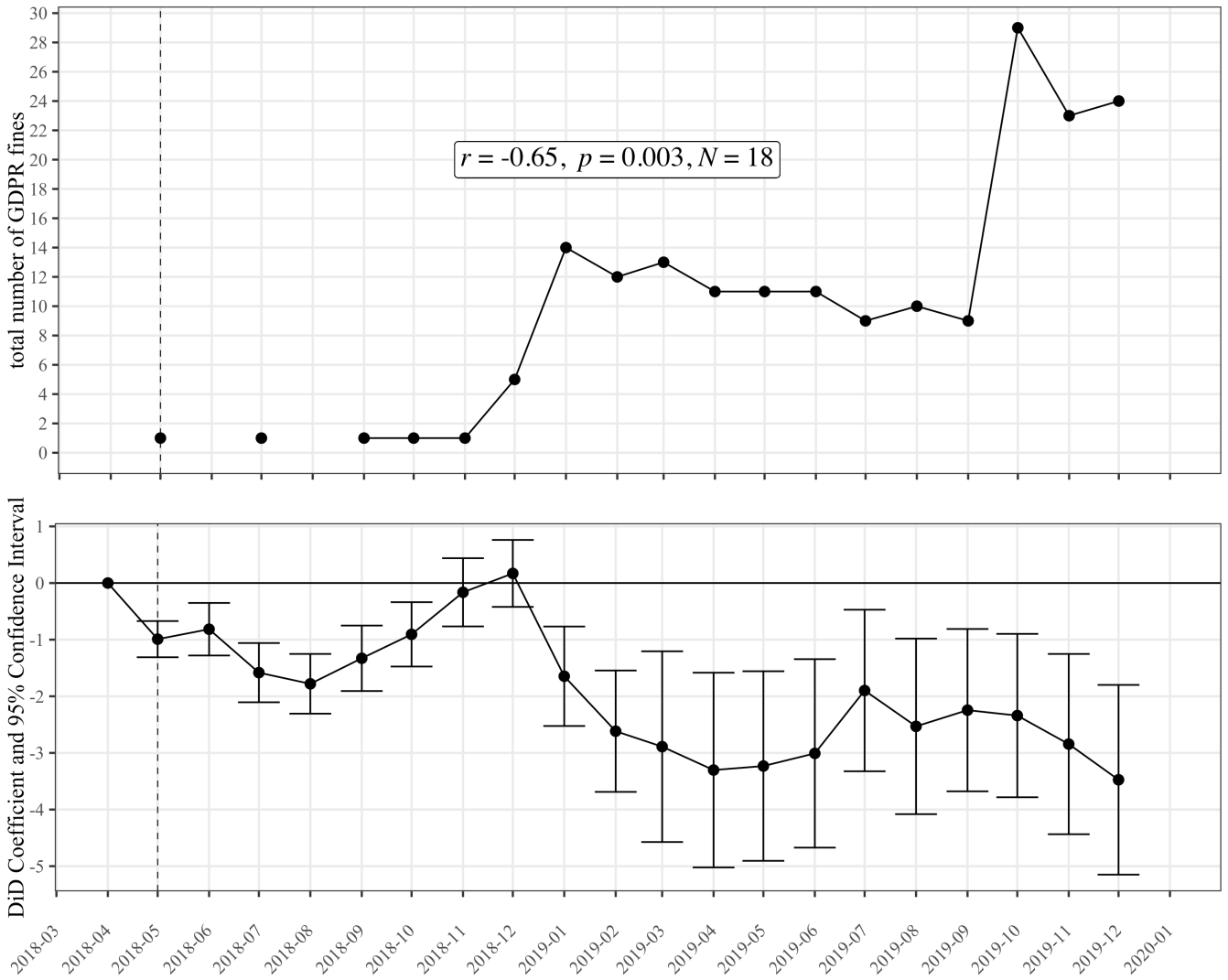
Notes: Multiplying the number of sites (717) with the number of site instances (2) and the number of months (21) yields the number of observations (30,114).

Development of DiD coefficients for Number of Trackers



Notes: The model includes site instance and period fixed effects. One-way standard errors are clustered at the site instance level.
The first period (April 2018) before GDPR enforcement was excluded from the estimation as a reference period.

Correlation between Monthly DiD Coefficients and Number of GDPR Fines



Notes: The model includes site instance and period fixed effects. One-way standard errors are clustered at the site instance level. The first period (April 2018) before GDPR enforcement was excluded from the estimation as a reference period.

Difference in Effect of GDPR on Number of Trackers across Site Industries and Categories of Trackers

Across Website Industries using the Number of Trackers as DV[1]			Across Categories of Trackers using the Number of Trackers as DV		
Website Industry	**DiD Coefficient**	**95% CI**	**Tracker Category**	**DiD Coefficient**	**95% CI**
Adult	-0.94	[-1.89; 0.02]	Advertising	-1.07	[-1.86; -0.29]
Business	-1.37	[-2.66; -0.07]	Analytics	-0.23	[-0.42; -0.04]
Entertainment	-2.04	[-3.67; -0.42]	Comments	0.01	[-0.01; 0.04]
News & Portals	-3.90	[-7.67; -0.13]	Essential	-0.52	[-0.89; -0.16]
Recreation	2.29	[-5.05; 9.63]	Other	-0.04	[-0.14; 0.05]
Reference	-2.37	[-4.84; 0.1]	Social Media	0.01	[-0.07; 0.09]
Total[2]	-1.97			-1.97	

[1]We abbreviated the dependent variable to DV.; [2]The total row includes the overall effect of the GDPR on the number of trackers and the share of tracked pages, respectively.

One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

Notes: The model includes site instance and period fixed effects. We ordered the table alphabetically by website industry.

Insights from Robustness Tests

1. **Longer period before GDPR**
   - Usage of sample with longer period before GDPR (T = 31 months)
   - Evidence that parallel trends' assumption is not violated

2. **Different treatment assignments**
   - Site's server location, site's use of cookie banner
   - Similar results for 2 other treatment assignments

3. **Breakdown of effect on different treatment groups**
   - Similar results for "cleaner" comparison of treatment and control group

4. **Different measure for quantity of online tracking**
   - Similar results for number of tracker providers

Summary of Results

Summary of Results

1. **Quantity of online tracking increases over time**
   - Treatment group: 7.02 trackers
   - Control group: 9.00 trackers

2. **GDPR decreased:**
   - Quantity of online tracking by -1.97 trackers

3. **GDPR affected quantity of online tracking:**
   - Weaker in mid-term (-1.78 trackers; 4-6 months after GDPR)
   - Stronger in long-term (-3.47 trackers; >6 months after GDPR)

4. **GPDR decreased number of trackers on:**
   - Business- (-1.37 trackers), entertainment- (-2.04 trackers), and news-related (-3.90 trackers) sites

5. **GPDR decreased site's usage of**:

- Advertising (-1.25 trackers), analytics (-0.21 trackers), and essential (-0.46 trackers) trackers

Implications of Empirical Study

Implications of Empirical Study

1. **Regulators**
   - Likely cause for 2nd wave of effect:
     - Increase in number of GDPR fines throughout 2019, or
     - €50,000,000 penalty enforced on Google in January 2019
   - Stronger GDPR enforcement would likely help achieve its aim

2. **Users**
   - GDPR increased online privacy to a small extent
   - Justifiable privacy concerns:
     - Quantity of online tracking increases over time

Thank You for Your Attention!

Karlo Lukic

lukic@wiwi.uni-frankfurt.de

Klaus M. Miller

millerk@hec.fr

Bernd Skiera

skiera@wiwi.uni-frankfurt.de

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 833714).

Richness and Consistency of Consumer Profiles and their Development over Time

Lennart Kraft, Klaus M. Miller, Bernd Skiera

HEC PARIS

Aim and Setup of Empirical Study of Consumer Profiling

- **Aim:**
  - Quality of consumer profiling
  - Speed of consumer profiling

- **Setup of empirical study**
  - 2.7 years (867 days: February 17th, 2014 – October 30th, 2016)
  - Random selection of 170,000 consumer profiles
  - Analysis of 850,000,000 data points

- **Approach**
  - Development of dimensions and measures
    - Richness
      - Profiling Breadth
      - Profiling Depth
      - Profiling Density
    - Profiling Consistency

Example of the Kind of User Data

Date: Wednesday, June 15th, 2015

URL: ESPN.com

Interests: Sports (Soccer, Snowboarding), Music, Marketing

Income: > 80,000 $

Education: PhD

Age: 34-39

Martial Status: Single

Gender: Male

Date: Wednesday, June 15th, 2015

URL: hidden

Interests:n/a

Income: n/a

Education: n/a

Age: n/a

Martial Status: n/a

Gender: Female

Visualization of Profiling Activities

Figure 3: Visualization of Profiling Activities of 30 out of 53,896 Consumers (Study 1)

Notes: Points mark events.
Orange points mark the "birth" of a cookie, i.e., the identifier of a consumer.
Purple points mark the "death" of a cookie.
Green points represent an "addition of a data point" to the consumer's profile.
The difference between the orange and the purple points represents the lifetime of the cookie (identifier).
The difference between the first green point and the purple point represents the time throughout which profiling of the consumer occurred.

Detailed Description of Results

Table 5: Summary and Comparison of Results of all Empirical Studies.

Dimension	Perspective of Time	Study 1	Study 2	Study 3
Profiling Breadth	On the Sampling date	73.45%	73.13%	72.57%
	At the End of the lifetime	76.38%	76.38%	76.06%
	After 1 month	50.18%	51.22%	50.55%
	After 6 months	40.37%	41.23%	40.57%
	After 7.5 months	37.70%	38.50%	37.88%
	After 12 months	30.13%	30.36%	30.11%
	After 18 months	19.73%	19.06%	19.12%
Profiling Depth	Sampling date	32.15	30.96	31.02
	On the Sampling date	41.62	41.57	41.28
	After 1 month	18.70	18.42	18.43
	After 6 months	32.19	32.12	31.92
	After 7.5 months	34.02	33.93	33.81
	After 12 months	37.92	37.99	37.71
	After 18 months	40.53	40.48	40.22

Note: 1 month means 30 days, 6 months means 180 days, 7.5 months represents the data of the profiling density's maximum, 12 months means 365 days, 18 months means 550 days. Note that the standard errors are very low for the profiling breadth and profiling depth, (e.g., 0.19pp for profiling breadth and 0.1742 for profiling depth on the sampling date), and also low for the profiling density (e.g., 0.1633 after 7.5 months). Similarly, the standard errors for profiling consistency are low (e.g., 0.1063 on the sampling date). Given that the standard errors are similar across all studies, we omitted them from this table for ease of reading.

Dimension	Perspective of Time	Study 1	Study 2	Study 3
Profiling Density	After 1 month	15.99	15.72	15.76
	After 6 months	23.25	23.14	23.06
	After 7.5 months	23.53	23.43	23.38
	After 12 months	22.46	22.31	22.29
	After 18 months	17.32	16.39	16.52
Profiling Consistency (Average AC_1 Statistic)	On the Sampling date	0.45	0.44	0.45
	At the End of the lifetime	0.43	0.42	0.43
	After 1 month	0.51	0.50	0.51
	After 6 months	0.44	0.44	0.45
	After 7.5 months	0.44	0.43	0.44
	After 12 months	0.43	0.42	0.43
	After 18 months	0.42	0.42	0.43

Note: 1 month means 30 days, 6 months means 180 days, 7.5 months represents the data of the profiling density's maximum, 12 months means 365 days, 18 months means 550 days. Note that the standard errors are very low for the profiling breadth and profiling depth, (e.g., 0.19pp for profiling breadth and 0.1742 for profiling depth on the sampling date), and also low for the profiling density (e.g., 0.1633 after 7.5 months). Similarly, the standard errors for profiling consistency are low (e.g., 0.1063 on the sampling date). Given that the standard errors are similar across all studies, we omitted them from this table for ease of reading.

Summary of Results

- **Profiling Breadth**
  - 75% of consumer profiles are non-empty
  - Cookie deletion outweighs profile creation
- **Profiling Depth**
  - 42 attribute levels (=information) per consumer profile
  - Most information accumulates early
- **Profiling Density**
  - Usable information peaks at 24 after 7.5 months
  - Flat maximum from 6 – 9 months
- **Profiling Consistency**
  - Most attributes only weakly consistent
  - Very small decrease over time

Thank you for Your Attention!

Lennart Kraft

kraft@wiwi.uni-frankfurt.de

Klaus M. Miller

millerk@hec.fr

Bernd Skiera

skiera@wiwi.uni-frankfurt.de

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 833714).

When the Cookie Crumbles? The Impact of Online Tracking Restrictions on the Online Advertising Market

Rene Laub, Klaus M. Miller, Bernd Skiera

HEC PARIS

Increasing Restrictions on User Tracking

Online Advertising Industry

- **Online advertising industry tracks users' online behavior intensively, e.g.,**
  - Website visits
  - Ad exposure
  - Clicks, conversions
- **Advertisers use data to improve online advertising performance**
  - Targeting
  - User-level ad measurement

Users

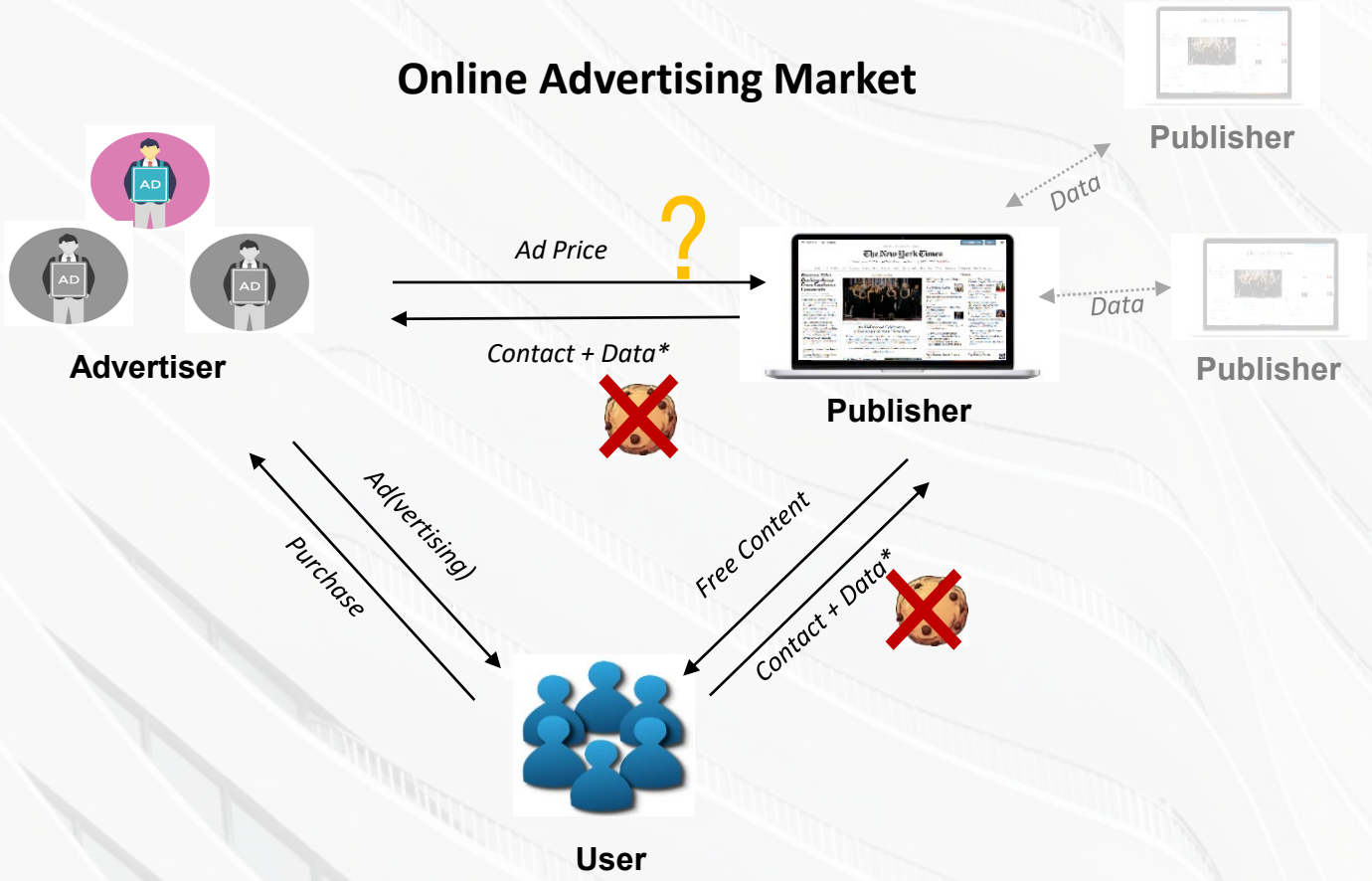
- **Users perceive user tracking as violation of their privacy** (e.g., PEW Research Center 2019)
- **Likely driven by users' perception of behaviorally targeted ads as "creepy" / privacy-intrusive** (e.g. Moore et al. 2015)

Introduction of restrictions on user tracking

- Privacy laws, e.g., General Data Protection Regulation, California Consumer Privacy Act, Personal Information Protection Law
- Industry initiatives, e.g., web browsers ban user tracking technologies (e.g., Apple, Google, Mozilla)

User tracking increasingly unavailable for online advertising industry

Relevance of User Tracking in Online Advertising Market



Research Project in a Nutshell

- **Research Aim:**
  - Determine value of user tracking for online advertising market (esp. publishers)

- **Research Questions:**
  - RQ1: How much value does user tracking generate for publishers?
  - RQ2: How does value of user tracking differ between publishers?
  - RQ3: How does value of user tracking differ between data categories?

- **Audience**
  - Policy makers
  - Online advertising industry
  - Users

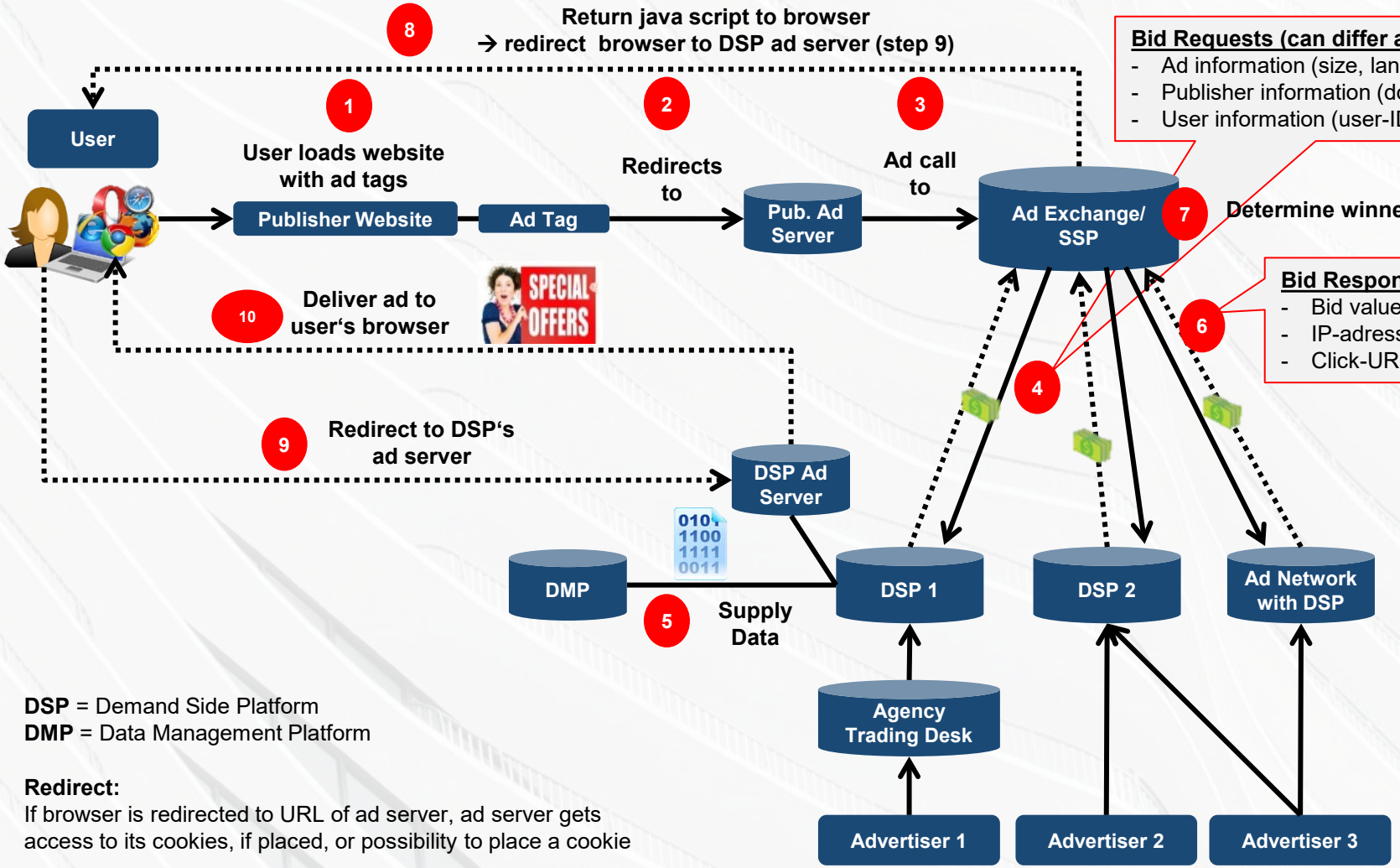
Description of User Tracking

Definition of User Tracking and Tracking Technologies

- **Definition of User Tracking**
  - User tracking is the collection of a user's data over time
  - Builds on the ability to (re-) identify the user on repeated visits to a website and within visits

- **Tracking Technologies**
  - Digital Fingerprinting
  - Cookies
    - Lifespan
      - Session cookies
      - **Persistent cookies**
    - Domain
      - First-party cookies
        (e.g. "Newspaper" web server stores info of "Newspaper" visit)
      - **Third-party cookies**
        (e.g., "Newspaper" web server stores info of visit at other websites)

Real-Time Bidding Auction Process



Bid Requests (can differ across bidders):
- Ad information (size, language)
- Publisher information (domain)
- User information (user-ID)

Bid Responses
- Bid value
- IP-adress of ad server
- Click-URL for ad

DSP = Demand Side Platform
DMP = Data Management Platform

Redirect:
If browser is redirected to URL of ad server, ad server gets access to its cookies, if placed, or possibility to place a cookie

Ad Tag

```
<script> src=„ib.adnxs.com/ttj?id=„123" type=text/javascript</script>
```

Bid Request Data in Online Advertising

Data Categories		Common Variables	Example
User data			
Ad slot data			

*Necessary to collect all other data categories from user tracking

Knowledge on the Value of User Tracking

Theoretical Results on Value of User Tracking

Negative Effect on Ad Prices (= decrease)

- **Advertising performance increasing practices no longer possible**

  -> Willingness-to-pay of advertisers decreases (e.g. Chen & Stallert 2014)

  -> Ad prices decrease

- **Niche advertisers exit a market without online user tracking**

  -> Competition decrease (Johnson 2013)

  -> Ad prices decrease

Positive Effect on Ad Prices (= increase)

- **User tracking creates narrow consumer segments and reduces competition among advertisers**

  -> Competition increases without user tracking (e.g. Levin & Milgrom 2010)

  -> Ad prices increase

Unclear effect of unavailability of user tracking on ad prices

Anecdotal Evidence from the Online Advertising Industry for Value of User Tracking

“The end of the cookie is going to harm […] publishers. “ – AdExchanger, 2020

Negative effect?

“**Behavioral targeting has been completely overhyped in its value for publishers from the day it was first invented.”**
- Michael Zimbalist (former CMO of The New York Times), WallStreetJournal, 2019

No Effect?

„The death of third-party cookies is the end of the free internet.“ - Wired, 2021

Strong negative effect?

“**The death of the cookie signals a new golden age for publishers.”-** WhatsNewInPublishing, 2020

Positive Effect?

Empirical Evidence for Value of User Tracking

Study	Marotta, Abhishek, Acquisti (2019, Working Paper)	Johnson, Shriver, Du (2020, Marketing Science)

Goal of Empirical Study: Determine value of user tracking for publishers

Setup of Empirical Study

Description of Data

- **Price of 42 million ad impressions from large European Ad-exchange**
  - User tracking via third-party-cookies
    - 85% of ad impressions with cookie (~1.4 Mio. cookie IDs)
    - 15% of ad impressions without cookie

- **100 publishers**

- **Covariates on ad impression characteristics, e.g.,**
  - Ad position
  - Ad size

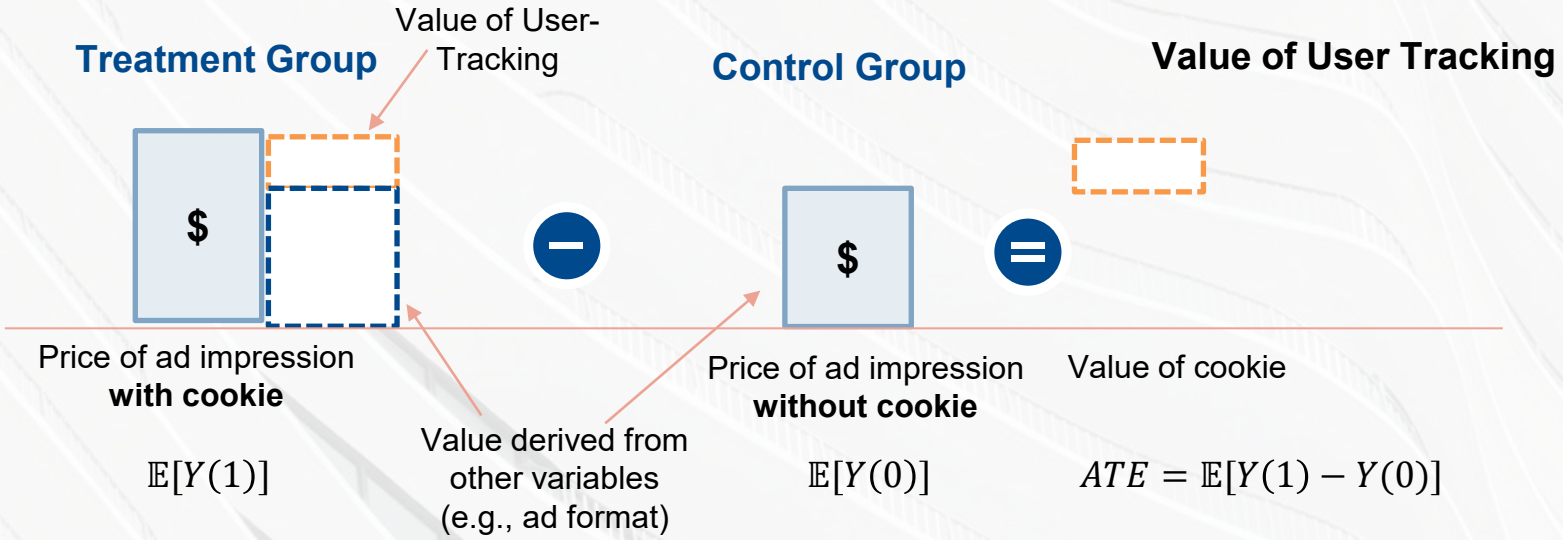
- **Covariates on user characteristics, e.g.,**
  - Device of user
  - Internet browser of user
  - …

- **Covariates on publisher characteristics, e.g.,**
  - Topic area of content
  - …

Measuring the Value of User Tracking

- **Aim of Empirical Study:** Estimate price difference between ad impressions with and without user tracking

- User tracking via Cookie
- Treatment = 1, if cookie is available (0 otherwise)



Identification Strategy (1/2)

1. Estimate price of each ad impression i :

- $\log(p_i) =$

$$\beta_1(User\ data\ generated\ by\ user\ tracking_i) +$$
$$\beta_2\ (User\ data\ not\ generated\ by\ user\ tracking_i) +$$
$$\beta_3\ (Ad\ slot\ data_i) +$$
$$\beta_4\ \mathrm{I}_i^{Cookied} +$$
$$\beta_5\ (User\ data\ not\ generated\ by\ user - tracking_i) * \mathrm{I}_i^{Cookied} +$$
$$\beta_6\ (Ad\ slot\ data_i) * \mathrm{I}_i^{Cookied} +$$
$$\beta_7 Advertiser_i +$$
$$\beta_8\left(Advertiser_i\right) * \mathrm{I}_i^{Cookied}$$
$$+ u_i$$

Identification Strategy (2/2)

2. Predict Counterfactual for each ad impression i:

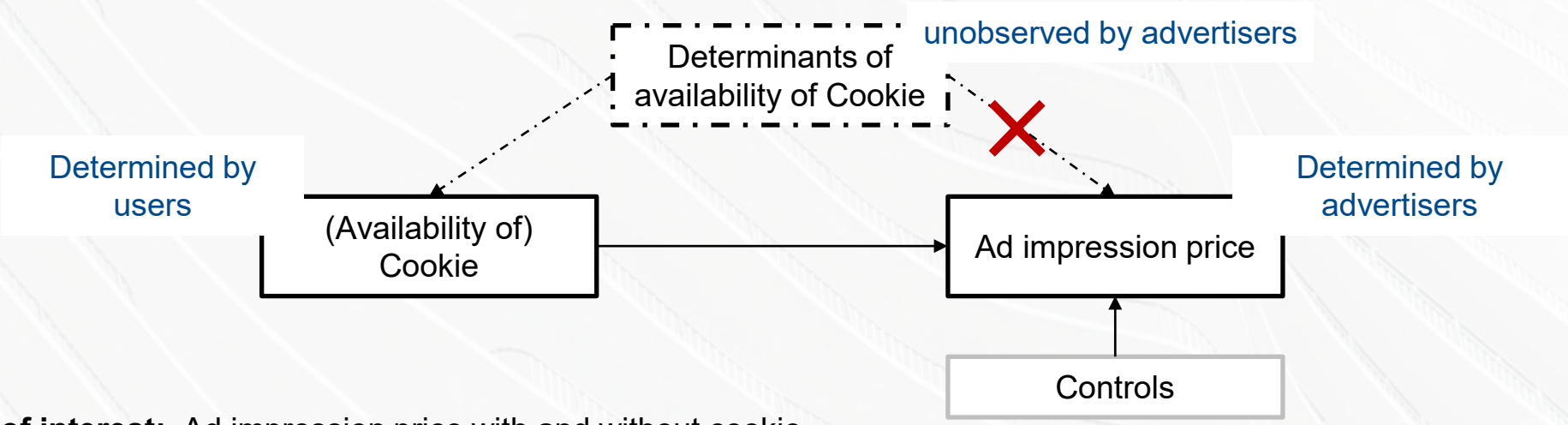
- Treatment group (= cookied users), $\log(p_i | No\ user\ tracking_i)$,
  if $\mathrm{I}_i^{Cookied}$= 0 and no data based on user tracking available
- Control group (= uncookied users), $\log(p_i | With\ user\ \mathrm{tracking})$,
  if $\mathrm{I}_i^{Cookied}$ = 1 and other data from user tracking set to zero*

- **3. Calculate Potential Outcomes (across all ad impressions):**
  - $\mathrm{E}(p | No\ user\ tracking)$
  - $\mathrm{E}(p | With\ user\ tracking)$

*Equivalent to a “new born” user

Selection into Treatment

- **Effect of interest:**



- **Outcomes of interest:** Ad impression price with and without cookie

- Availability of Cookie is not-random, however
  - Underlying determinants (e.g., age) unobserved by advertiser
  - We observe same variables as advertisers (e.g., proxies for age)

  -> selection on unobservable (self-selection) should not be a concern for estimation

- **Procedure: Estimate unobserved counterfactuals with regression analysis (RA) controlling for selection on observables** (e.g., ad format, browser)
- Augmented inverse probability weighting (AIPW) and Heckman Selection Model to account for remaining self-selection concerns

Robustness Check 1: Augmented Inverse Probability Weighting (AIPW)

•Step 1: Treatment Model Estimation (Controls for Selection into Treatment)

•Inverse Probability Weighting

- Estimate treatment probability $\hat{\pi}_i$ for each ad impression *i*
- Weight observations:
  - Treatment group (T_i=1): 1/$\hat{\pi}_i$
  - Control group (T_i=0): 1/(1-$\hat{\pi}_i$)

$$\widehat{ATE}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{\pi}_i} - \frac{(1-T_i)Y_i}{(1-\hat{\pi}_i)}\right\}$$

Y_i = price of ad impression i

Step 2: Outcome Model Estimation (Controls for Selection on Observables)

$

Regression Adjustment

- Estimate price of each ad impression *i* based on covariates Z_i (e.g. ad size, ad position)
- Separate estimation per group
- Estimate counterfactual price using regression estimates from opposite group

$$\widehat{ATE}_{RA} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\mathbb{E}}(Y_i|T_i = 1; Z_i) - \hat{\mathbb{E}}(Y_i|T_i = 0; Z_i)\}$$

Step 3: Estimate ATE

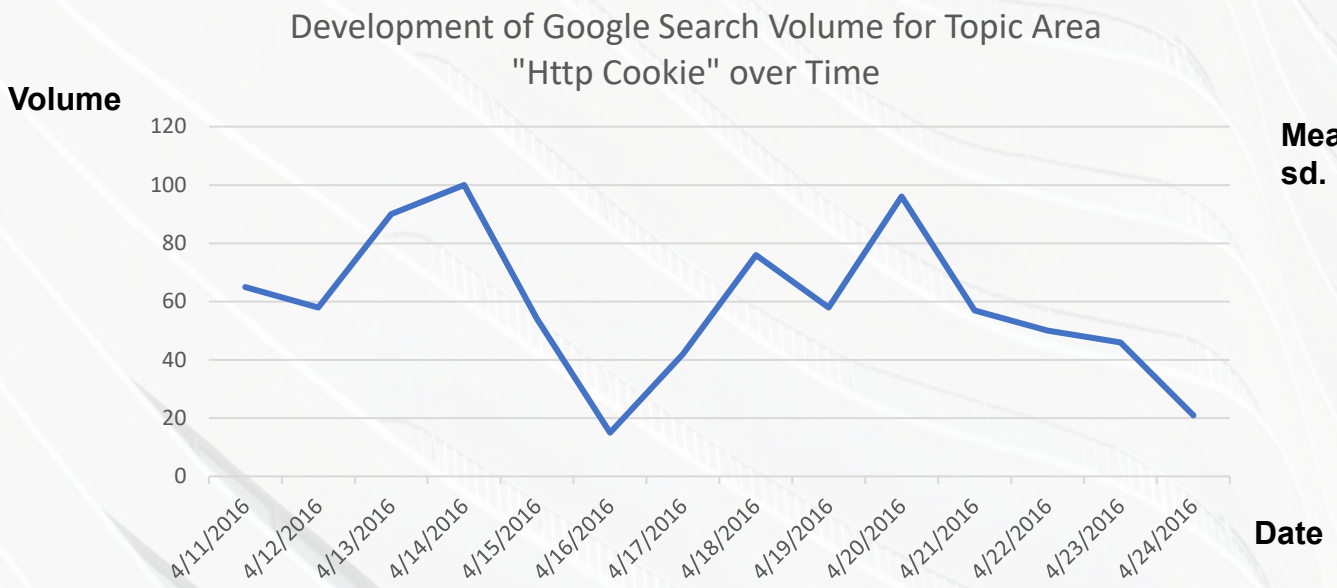
$$\widehat{ATE}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{T_i Y_i}{\hat{\pi}_i} - \frac{(1-T_i)Y_i}{(1-\hat{\pi}_i)}\right] - \frac{(T_i-\hat{\pi}_i)}{\hat{\pi}_i\,(1-\hat{\pi}_i)}\left[(1-\hat{\pi}_i)\hat{\mathbb{E}}(Y_i|T_i = 1; Z_i) + \hat{\pi}_i\hat{\mathbb{E}}(Y_i|T_i = 0; Z_i)\right]\right\}$$

- **"Double Robust" Estimator:**
  - Either treatment model or outcome model needs to be correctly specified.
  - Performs best* even under severe confounding (Glynn & Quinn 2010)

*Comparable or lower mean square error than propensity score matching, IPW or regression adjustment

Robustness Check 2: Heckman Selection Model

- **Exogenous variable:**
  - Google search volume for topic area „http Cookie“



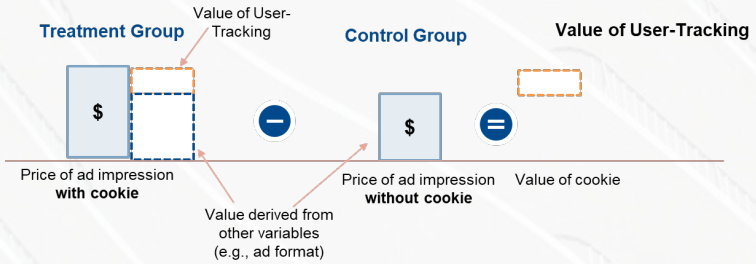
Mean = 59.4, sd. = 23.8

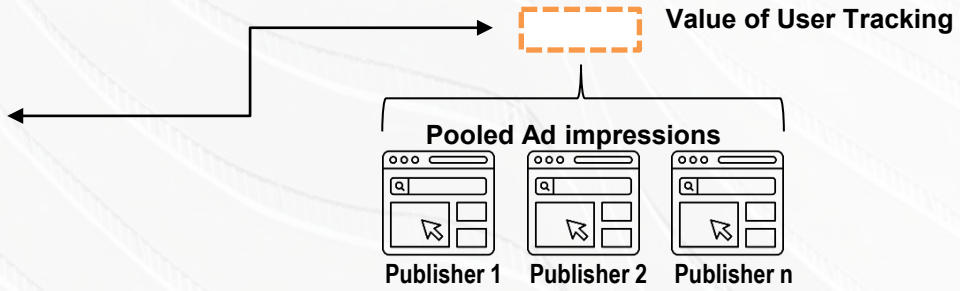
- **Significantly negative influence on probability of cookie availability**

Results of Empirical Study

Overview of Results of Empirical Study

RQ 1: What is the value of user tracking?

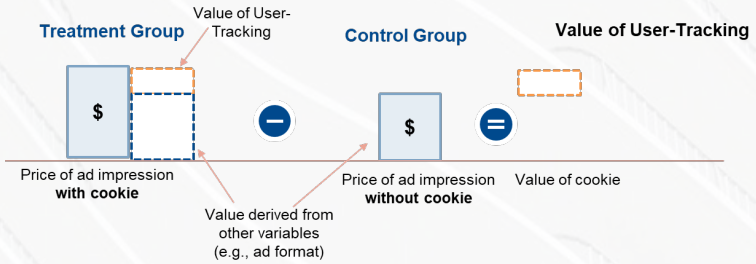


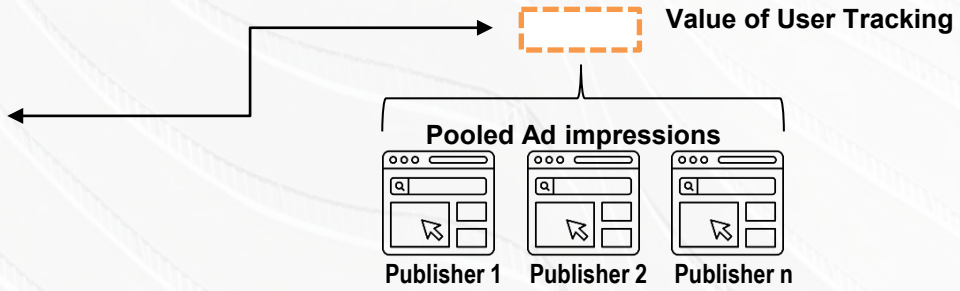


Results for
Value of User Tracking (RQ1)

Overview of Results of Empirical Study

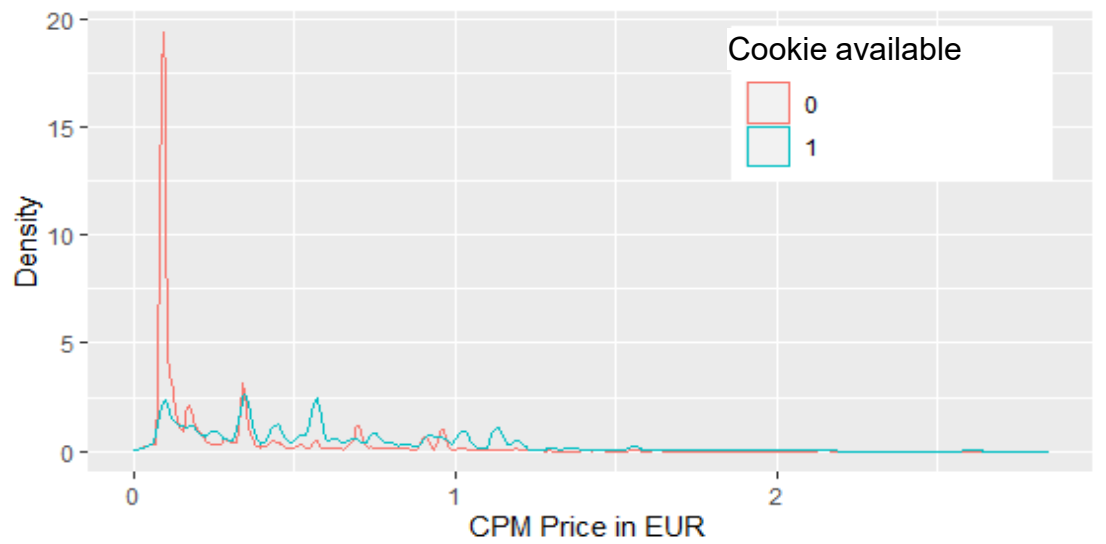
RQ 1: What is the value of user tracking?





Visualization of Distribution of Ad Prices With and Without User Tracking

Kernel-Density Plot of Ad Impression Prices



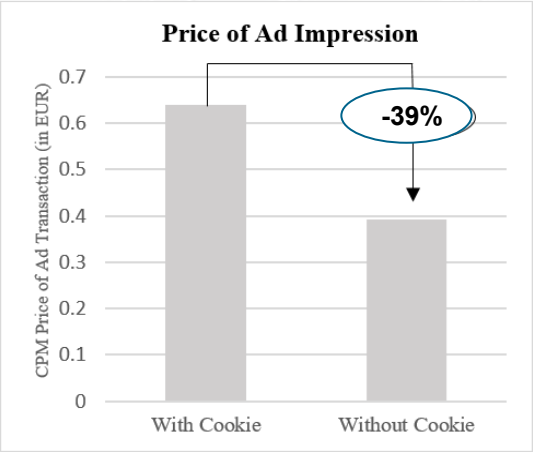
Summary Statistics

CPM Price in EUR	Mean	q95	q98	N (= Ad Impressions)
Cookied	0.691	1.429	2.843	36,024,873
Uncookied	0.276	0.954	1.077	6,397,826

- (Raw) mean price difference: 0.691 EUR - 0.276 EUR = 0.415 EUR
- Relative price difference: $\frac{(0.276 - 0.691)}{0.691} = \frac{-0.415}{0.691}$ = - 60%

Results for All Ad Impressions (~Average Treatment Effect)

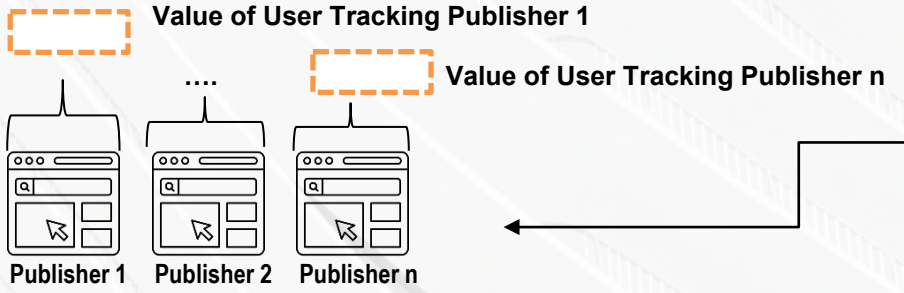
Potential Outcomes	RA	AIPW	Heckman Selection Model*
$\mathrm{E}(p \mid \text{With } user - tracking)$	0.65 EUR	0.64 EUR	0.65 EUR
$\mathrm{E}(p \mid No\ user - tracking)$	0.40 EUR	0.39 EUR	0.40 EUR
Price Difference	-0.25 EUR	-0.25 EUR	-0.25 EUR
Relative Price Difference	-39%	-39%	-39%
N	42,421,954		



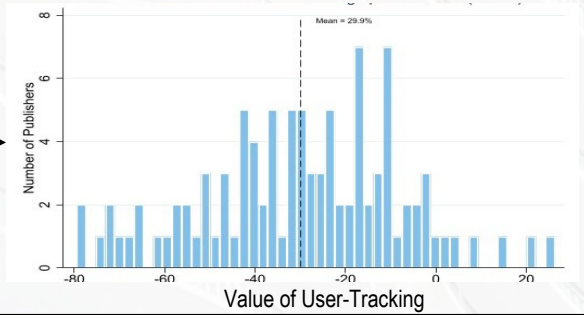
Results for
Heterogeneity Across Publishers (RQ2)

Overview of Results of Empirical Study

RQ 2: How does the value of user tracking differ across publishers?

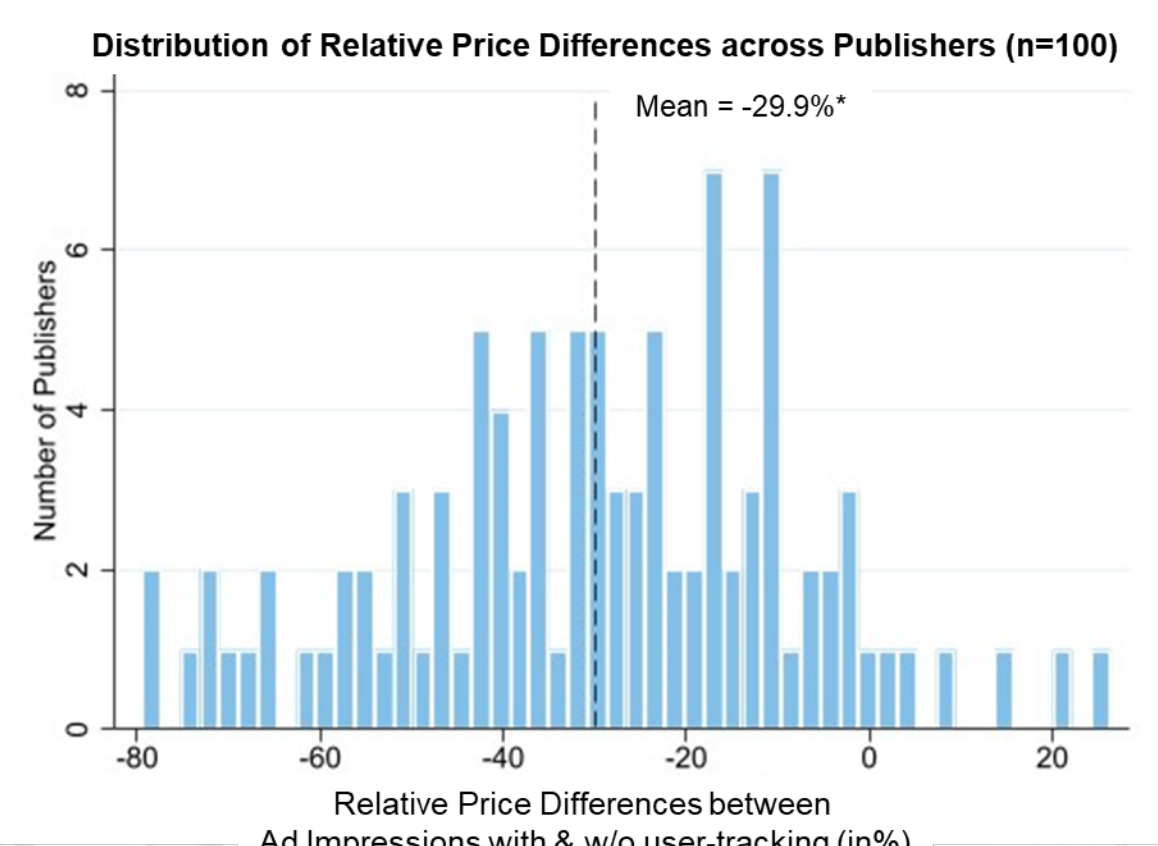


Distribution of Value of User Tracking



How does the Value of User Tracking Differ Across Publishers?

- **Repeat regression adjustment separately for each publisher**



- **Higher ad impression price with user tracking for ~ 93% of publishers.**

* Equivalent to a weighted mean of results from RQ1 with number of impressions per publisher as weights.

Regression Results for Publisher Heterogeneity

- **Aim:** Understand relationship between price change and publisher characteristics.

- **Publisher characteristics:**
  - Premium = quality of content
  - Niche = breadth of topic of content
  - Revenue concentration = concentration* of ad revenue across advertisers

- **Results:**
  - Premium publishers more reliant on cookie
  - Niche publishers less reliant on cookie
  - Higher revenue concentration associated with lower reliance on cookie

*based on Hirschman-Herfindahl-Index

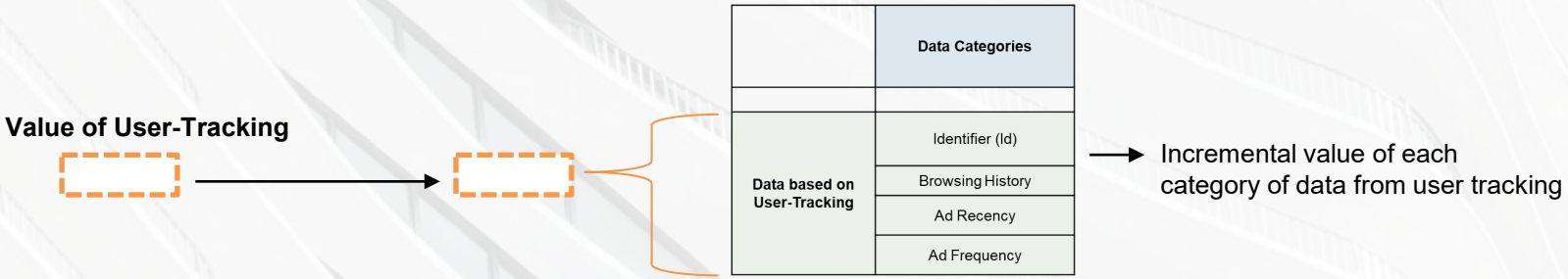
Dependent Variable:	Price Change (in %) Coefficient Estimate (SE)
Premium	-11.17 (4.34)***
Niche	9.56 (4.33)**
Revenue Concentration	0.007 (0.001)***
Controls	
Log(Reach)	1.42 (1.01)
Mobile	-30.36 (4.30)***
Publisher Category:	
Cars	-19.05 (3.83)***
Computer and Technology	20.65 (11.29)*
E-Mail Provider	-40.90 (8.05)***
Games	-22. (6.29)***
Health	-6.28 (5.42)
Lifestyle & Shopping	-3.45 (6.10)
Money, Finance & Real Estate	-5.99 (5.42)
Sport	11.43 (4.80)**
Women's Interests	-5.95 (3.92)
R^2	0.507
N	100

Note: *p<0.1; **p<0.05; ***p<0.01
Robust Standard Errors in Parentheses;
Inverse Probability Weighted Estimation.
Baseline Publisher Category: News & Information.

Results for
Heterogeneity Across Data Categories (RQ3)

Overview of Results of Empirical Study

RQ 3: How does the value of user tracking differ between categories of data?



How does the value of user tracking differ across categories of data?

- **Aim: Estimate incremental value of categories of data for value of user tracking**

- **Motivation:**
  - Data based on user tracking may vary in their
    - privacy intrusiveness for users
    - usefulness for the online advertising industry

Data Categories		Common Variables	Example
User data	Data generated by user tracking	Identifier (Id)*	Cookie- id “123-ABC-789”.
		Browsing History	The user visited www.sports.com already three times.
		Ad Recency	The user saw the ad “ABC” two minutes ago.
		Ad Frequency	The user saw the ad “ABC” already four times.
	Data not generated by user tracking	Device of user	The user is using an iPhone.
		Location of user	The user is in Berlin.
		Date and time	The time of the user’s visit is 2 pm on a Monday.
Ad slot data	Ad slot specific data	Position of ad slot	The ad slot is on the top of the website.
		Site of ad slot	The ad slot has a size of 728 x 90 pixels.
	Publisher specific data	Context of publisher	The ad slot is on a website offering financial news, e.g., www.financialtimes.com.

Ad price

Value added based on user tracking

Value added not based on user tracking

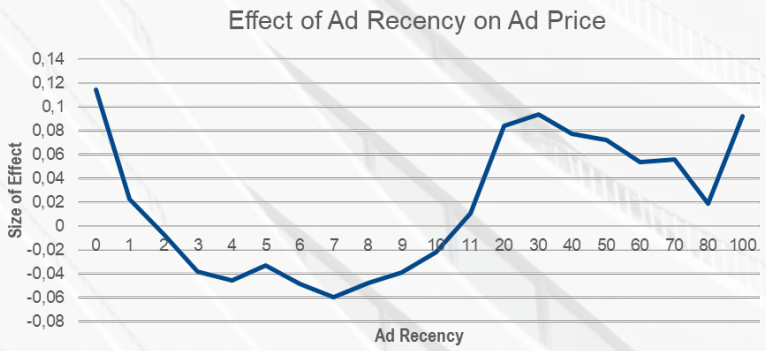
Estimation of Incremental Value of Categories of Data

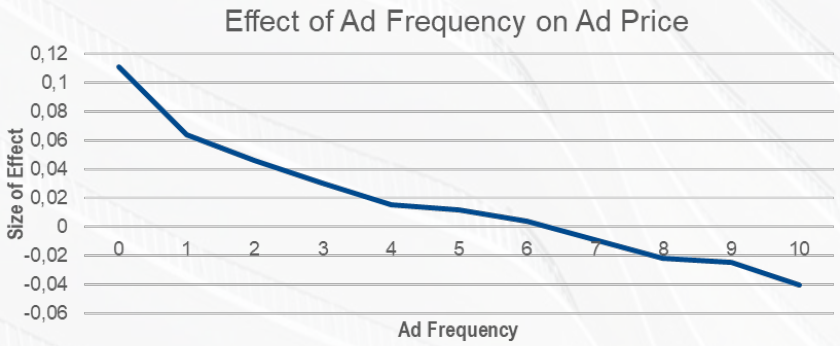
- **Prediction of p_i for 5 different scenarios for all cookied ad impressions:**

Scenario Number	Setting	Predicted price (Example)	Incremental Value of
Scenario 1	**No data** from user tracking	0.20 EUR	Contextual information: 0.20 EUR

Incremental Value of Different Categories of Data in Online Advertising

	Category of Data	Incremental Value…							
		Mean (in $)	Sd.	Quantiles					
				0%	50%	75%	90%	98%	100%
Without Cookie	Only Contextual	0.375	0.42	0.014	0.312	0.421	0.581	0.801	16.71





Summary and Implications

Summary and Implications

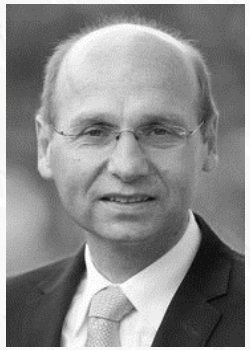
- **Ad prices are on average 39% lower without user tracking**
  - Threat of severe publisher revenue decline by restrictions on user tracking
  - Quantity and quality of free-content for users at risk
- **Almost all publishers (~93%) realize higher ad prices with user tracking**
  - Premium publishers, publishers with broader content, and publishers with less concentrated revenue on advertisers are stronger affected
- **Value of user tracking mainly driven by identifier, that enables ad performance measurement**
- **Browsing history does only generate considerable incremental value for a very small share of users (<2%)**
  - If users are primarily concerned about tracking of browsing history, then restrictions on user tracking could be limited to this type of data

DELETE COOKIES?!
gegen-den-strich.com
FERNANDEZ

Thank You for Your Attention!

Rene Laub
rlaub@wiwi.uni-frankfurt.de

Klaus Miller
millerk@hec.fr



Bernd Skiera
skiera@wiwi.uni-frankfurt.de

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 833714).

Summary of Tutorial

Summary of Tutorial

- **Online Advertising Market**
  - High-Tech
  - Extremely complex and networked
  - Big data
- **Economic Consequences of More Privacy (here: less tracking) can often be measured empirically**
- **Restrictions on Tracking**
  - Larger Impact on Publishers than Advertisers
- **Interesting Research Area**

The Impact of Privacy Regulation on the Online Advertising Market

gdpr-impact.com

Contact Information

Prof. Dr. Klaus M. Miller

HEC Paris

1 Rue de la Libération

78350 Jouy-en-Josas

Email: millerk@hec.fr

Web: fromdatatodecisions.com

@klausmiller

2022

2022