



Center for Technology, Innovation and Competition



# Big Data and Competition Law: Lessons from Innovation Markets

---

Christopher S. Yoo

John H. Chestnut Professor of Law, Communication, and Computer & Information  
Science; Founding Director of the Center for Technology, Innovation and Competition

University of Pennsylvania

December 1, 2023

---

# CONCENTRATION OF DATA AS A POTENTIAL INDEPENDENT CONCERN FOR COMPETITION LAW

---

- Calls to look at concentration in data as a separate consideration
    - U.S. Draft Merger Guidelines (2023)
    - Possible monopolization claims, e.g., ongoing U.S. case vs. Google
  - Insights from looking at past efforts to analyze concentration in input markets: research and development (R&D)
    - Look at the theoretical and empirical literature assessing the connection between R&D and consumer welfare
    - Examine Gilbert & Sunshine's (1995) seminal discussion of "innovation markets"
    - Analyze the limitations and critiques of their proposal
    - Assess the lessons this debate has for treating data as a separate market
-

# GILBERT & SUNSHINE'S "INNOVATION MARKET" PROPOSAL

---

- Gilbert & Sunshine (1995) proposed focusing on concentration in R&D
    - Fact pattern: merger of firms that do not compete in any current or foreseeable product market but are major investors in R&D
    - Concern: merger may reduce R&D spending/inhibit unforeseen future products
  - Innovation markets have not been widely used post *Genzyme* (2004)
    - 1992 U.S. Merger Guidelines discussed R&D only as a potential efficiency
    - 1995 and 2017 U.S. IP Licensing Guidelines recognized possible market for R&D
    - 2010 U.S. Merger Guidelines discussed R&D as a potential competitive harm
    - 2004 EU Horizontal Merger Guidelines regard mergers between innovators as a “potential special circumstance”
    - 2017 Dow-DuPont did not require connection to a concrete product market
-

# CONCEPTUAL CHALLENGES FOR INNOVATION MARKETS

---

- Key concern: impact of R&D spending on dynamic efficiency
  - Ambiguous relationship between scale/concentration and R&D spending
    - Theoretical and empirical literature fail to support monotonic relationship
    - Outcome depends on firm- and industry-specific factors: technological opportunity, appropriability, degree of market segmentation
  - Ambiguous relationship between R&D spending and innovation
    - Potential for excessive/redundant R&D spending (e.g., patent races)
    - Potential efficiencies: scale economies, access to complementary inputs
  - Similar concerns for relationship between scale/concentration and data-based innovation
-

# OPERATIONAL CHALLENGES FOR INNOVATION MARKETS: MARKET DEFINITION

---

- Problems of uncertainty, exacerbated by longer time frames
    - Different innovative modalities (e.g., process vs. product innovation, bus. models)
    - Innovation from unexpected sources
    - Riskiness/unpredictability of the innovative process
  - Danger of treating all R&D as fungible (not everyone is a competitor)
  - Proposed five-step rough guide for evaluating innovation markets
    - Key device: focus on particular product markets and specialized assets
    - Effect of limiting innovation markets to innovations in advanced stages
  - Similar problems for big data (n.b. different types, alternative sources)
-

# GILBERT & SUNSHINE'S ROUGH GUIDE FOR ANALYZING INNOVATION MARKETS IN MERGER ANALYSIS

---

1. Identify the overlapping R&D activities of the merging firms
  2. Identify alternative sources of R&D
  3. Evaluate actual and potential competition from downstream products
  4. Assess the increase in concentration in R&D and competitive effects on investment in R&D
  5. Assess R&D efficiencies
-

# STEP 1: IDENTIFY THE OVERLAPPING R&D ACTIVITIES OF THE MERGING FIRMS

---

- Gilbert & Sunshine's analysis and caveats
    - Difficulty in determining overlap: nonsubstitutability of R&D, variation in firm capabilities, unpredictability of R&D
    - Limit to R&D that may lead to improved products or processes
    - Limit to R&D that can have a significant impact on a relevant downstream market
    - Focus on specialized/specific assets
  - Commentary
    - Echo unpredictability of R&D
    - Propose limiting to products in advanced clinical trials in pharma
    - Question if must tie to product market, what is the benefit from adding innovation markets to analysis of current markets and potential competition
-

# STEP 1'S IMPLICATIONS FOR BIG DATA: DIFFICULTY DETERMINING OVERLAP

---

- Data for different business models
    - E-commerce: past purchase behavior to inform purchase recommendations
    - Search: past research behavior to inform relevant results
  - Structured vs. unstructured data
    - Structured – collected intentionally to inform a specific model (column-row)
    - Unstructured – collected incidentally and used to inform emergent models (photos, social media feeds, video, sensor data), most valuable and least used
  - The role of alternative dimensions in defining overlap (multiple Vs)
  - Lack of presence of specialized assets
  - Timing for understanding relevance of different types of data
-



## STEP 2: IDENTIFY ALTERNATIVE SOURCES OF R&D

---

- Gilbert & Sunshine's analysis and caveats
    - Parallel to market definition
    - Identification of both existing and potential sources of R&D (demand- and supply-side substitution), including existing firms and new entrants
    - Focus on specialized assets required to conduct R&D to establish limits
    - Recognition that even if R&D requires specific assets, firms that possess them may not be identifiable
  - Commentary
    - Difficulty in understanding future products that may compete > 2 years out
    - Risk that breadth will make all R&D markets competitive
-

## STEP 2'S IMPLICATIONS FOR BIG DATA: DIFFICULTY IDENTIFYING ALTERNATIVE SOURCES OF DATA

---

- Lack of specialized assets in big data
  - Presence of alternative sources of supply
    - Data brokers
    - Existing industry, esp. unstructured (financial services, retail, insurance,
    - Self-provisioning (esp. because data is nonrival)
    - Wide availability of alternative sources of unstructured data
  - Potential limits: impact of network effects on data collection (Android)
    - Based on oversimplified theories that posit inexhaustible returns to scale
    - Ignores features that dissipate winner-take-all dynamics: rapid growth, consumer heterogeneity, leapfrogging, large customers, multihoming, gateways
-

## STEP 3: EVALUATE ACTUAL & POTENTIAL COMPETITION FROM DOWNSTREAM PRODUCTS

---

- Gilbert & Sunshine's analysis and caveats
    - Downstream competition would make reductions in R&D unprofitable
    - Potential competition in downstream markets can also exert discipline
  - Commentary
    - Emphasis on competition in of downstream market means innovation market analysis may not add much to traditional antitrust
-

## STEP 3'S IMPLICATIONS FOR BIG DATA: WHAT IS THE DOWNSTREAM MARKET?

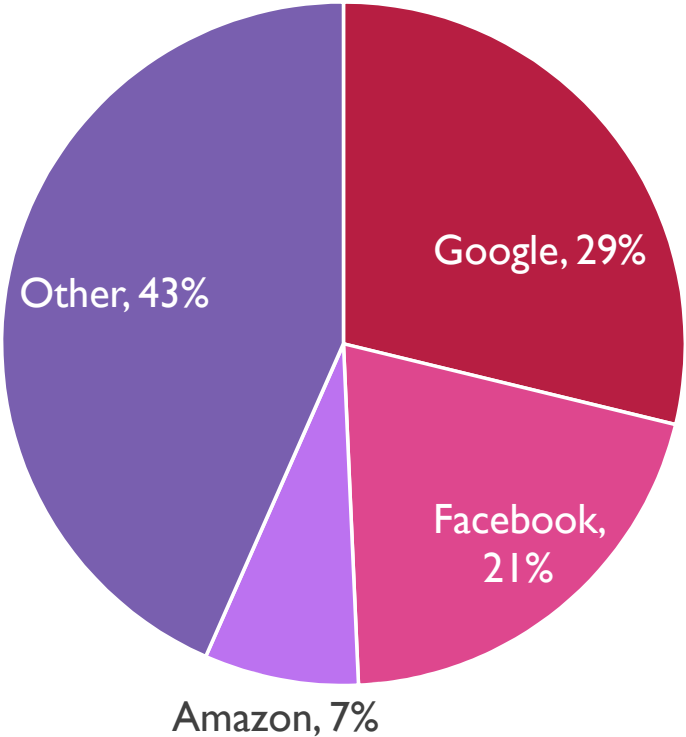
---

- Online services: many are competitive (travel, e-commerce)
  - Advertising
    - Online and offline ads are substitutes (Goldfarb & Tucker 2011a, 2011b; Zentner 2012; He, Lopez & Liu 2017)
    - Lack of proof of claims that different types of online ads are not substitutes (Ratliff & Rubinfeld 2011)
-

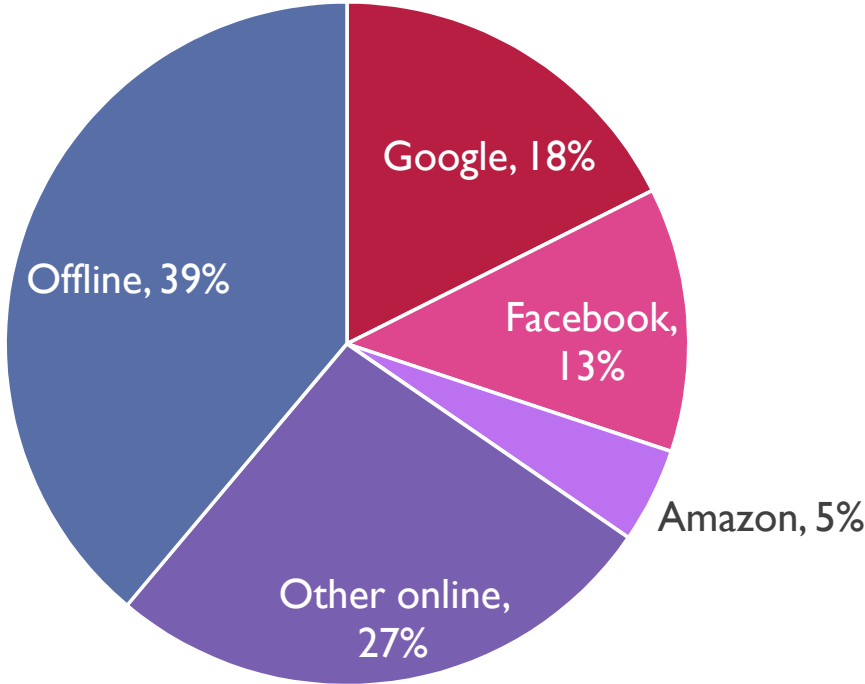
# MARKET SHARE (REVENUE) COMPARISON: WITH AND WITHOUT OFFLINE ADVERTISING (2022)

---

Market Share Excluding Offline Advertising



Market Share Including Offline Advertising



# THE ROLE OF COMPLEMENTARY INPUTS

---

- Variable proportions (McKenzie 1951; Vernon & Graham 1971)
    - Can respond to exclusion in one input by substituting complementary inputs
    - Substitution limits market power, but requires an inefficient input mix
  - Market power in complementary inputs (Teece 1986)
    - Even inputs with complete appropriability must combine with other inputs
    - If those inputs have market power, may have problems
    - Solutions can involve long-term contracts before investing in sunk costs
  - General purpose technologies (Bresnahan & Trajtenberg 1995)
    - Platforms create positive externalities for complementors
    - Solution may be to allow vertical integration to internalize more of these benefits
-

## STEP 4: ASSESS THE INCREASE IN CONCENTRATION & THE EFFECT ON R&D INVESTMENT

---

- Gilbert & Sunshine's analysis and caveats
    - Acknowledgement ways that concentration can promote R&D investment
      - Greater appropriability when intellectual property protection is incomplete
      - Rent dissipation/patent races
      - Better use of investments in complementary assets, firm-specific skills, private info
    - Theory and empirics have failed to resolve Schumpeter vs. Arrow conjectures
  - Commentary
    - Literature does not support presumption either way (Katz & Shelanski 2007)
    - Firm- & industry-specific factors make results "fragile" (Carlton & Gertner 2003)
    - Harm to consumers should be evaluated on facts of each case
-

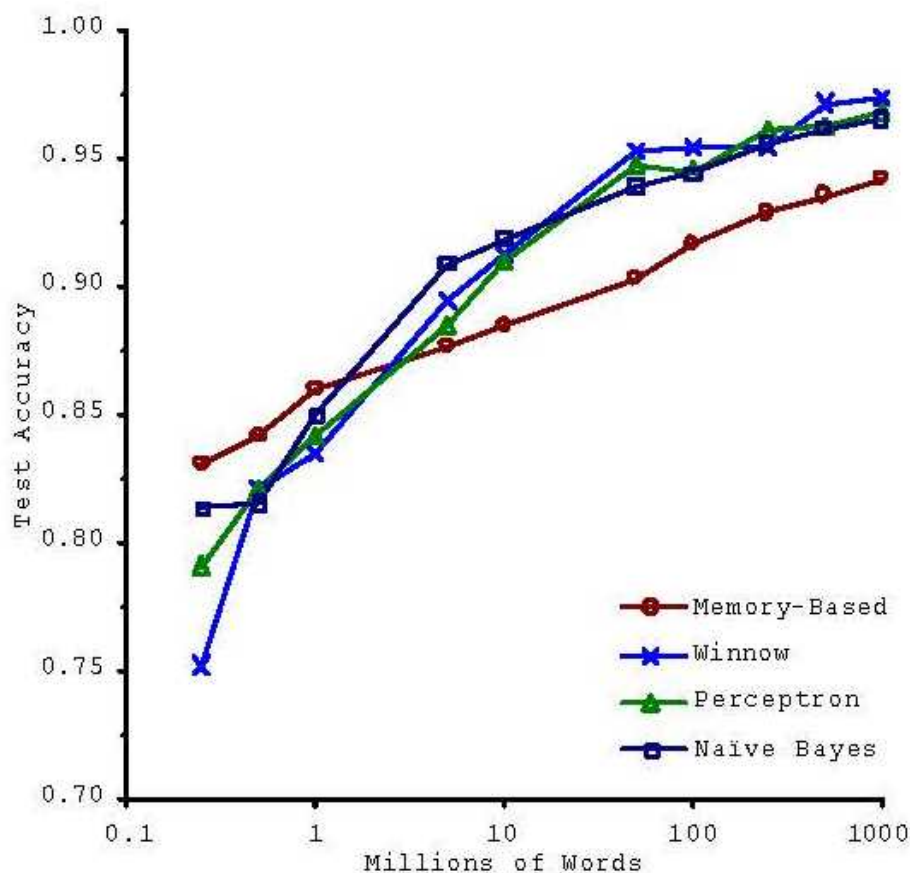
# STEP 4'S IMPLICATIONS FOR BIG DATA: AMBIGUITIES ABOUT THE IMPACT OF SCALE IN DATA

---

- Size of scale economies in data are unclear
    - Natural experiments show no scale economies for unstructured data (Chiou & Tucker 2017; Neuman, Whitfield & Tucker 2018)
    - Industry reports only use samples of data (Varian 2014; Bajari et al. 2018)
    - Differences between structured and unstructured data
    - Differences depending on business model
  - Relevance of aspects aside from quantity to of data
    - Potential interaction with features aside from scale (recency, variety, accuracy)
    - Algorithm quality as a source of value (Brynjolfsson & McElheran 2016)
-



# EMPIRICAL STUDY OF THE ROLE OF ALGORITHMS: BANKO & BRILL (2001)



- Size of training corpus on natural language disambiguation
- Findings
  - More data improved result quality, but may not be cost justified
  - Active learning and unsupervised learning may attain similar advantages
- All four exhibit diminishing returns
- Significance of the differences in performance depends on context

## STEP 5: ASSESS R&D EFFICIENCIES

---

- Gilbert & Sunshine's analysis and caveats
    - Several sources of R&D efficiencies
      - Scale economies in R&D
      - Better use of investments in complementary assets & firm-specific skills
      - Elimination of redundant activities
  - Commentary
    - Literature on optimal levels of innovation/excessive innovation
    - Difficulty proving efficiencies in conventional merger analysis
-

# STEP 5'S IMPLICATIONS FOR BIG DATA: POSSIBILITY OF DATA EFFICIENCIES

---

- Feasibility of similar efficiencies
    - Scale economies (for unstructured data)
    - Better use of algorithms
    - Rationalization of redundant activities
  - Other potential efficacies
    - Reduction in operating costs (run time, memory usage)
-

# A COMMENT ON SINGLE-FIRM CONDUCT

---

- Antitrust holds greater concern for mergers than for unilateral conduct
    - Combinations can more easily harm competition than single-firm conduct
    - Penalizing single-firm conduct that can be procompetitive
  - Extension of Gilbert & Sunshine to single-firm conduct requires even more justification
-

# ASSESSMENT OF IMPLICATIONS OF INNOVATION MARKETS FOR DATA

---

- Gilbert & Sunshine shows difficulties of assessing input markets
    - Caveats and preconditions imposed by Gilbert & Sunshine themselves
    - External critiques, even by sympathetic authors
    - Importance of limiting to specialized assets tied to specific products
  - Cautionary note/roadmap for treating markets for data as an independent consideration
-

