Cognitive Inequality and Big Data

Indira Puri[1] and Laura Veldkamp[2]

[1]New York University

[2]Columbia University *

March 15, 2024

Abstract

We combine insights from medical and big data literature to propose a novel model, which suggests that the expansion of big data exacerbates cognitive inequality. While individuals with high cognitive abilities may benefit from the targeting and customization facilitated by big data, those with lower cognitive abilities—and even children—may suffer adverse effects. Data from political discourse supports our predictions. The findings introduce a new consideration to the debate on big data regulation and emphasize the necessity of addressing cognitive inequality.

The regulation of data has become a focal point in contemporary discourse. Scholars such as Seim et al. (2022) underscore three primary rationales for regulatory intervention: the absence of monetary compensation for users who generate data, the exacerbation of market power through data utilization, and the profit-centered exploitation of data by firms. Our study introduces a novel dimension to the discourse on data regulation by examining its disparate impact on individuals of varying

*Preliminary and incomplete. Please send comments to indira.puri@nyu.edu. Thanks to Angelica Munoz Rodriguez for her outstanding research assistance. All errors are our own.

cognitive abilities. Unlike previous approaches that focus on economic or legal aspects, our analysis delves into the cognitive implications of big data utilization. We argue that the differential effects of data on individuals. We use theory and data to argue that predictive data is likely exacerbating inequality in cognitive ability. We find evidence of this cognitive divergence in the complexity of online political communication. This raises the concern that the free exchange of ideas, central to a democracy, may face yet another hurdle.

Our conclusions stem from two fundamental assumptions, for which we document empirical support, and which serve as the cornerstone of our model. The first assumption posits that cognitive processes, particularly attention and information processing, can be likened to a muscle that requires regular exercise. We contend that engaging with complex information—paying attention—enhances cognitive abilities, whereas limited exposure to such stimuli may lead to cognitive decline.

The second assumption revolves around the unique capabilities of big data to tailor information complexity to individual consumers' cognitive profiles. Unlike traditional approaches where information dissemination is uniform, big data enables precise targeting based on users' cognitive abilities. Consequently, individuals with varying levels of cognitive aptitude are exposed to information tailored to their comprehension levels.

In this article, we integrate these two assumptions into a cohesive framework to elucidate the ramifications of big data utilization on cognitive inequality. Notably, our model does not rely on intricate mechanisms; rather, it derives its insights from the inherent disparities in information processing and dissemination facilitated by big data. By shedding light on the cognitive dimensions of data regulation, our study underscores the imperative for policymakers to consider the nuanced cognitive implications of data utilization in crafting effective regulatory frameworks.

Related literature First and foremost, our work is related to others who study the harms of the digital economy (Bergemann et al., 2023). Liu et al. (2021) is a closely related paper that examines a different harm of data, that it is used to exploit temptation. Data is used to predict which consumers who are susceptible to temptation so that firms can target them with ads and sell them goods, which ultimately lower their utility. The idea that data is used to target customers is similar, and is pervasive in the management literature (Matz, 2022). Moreover, Bergemann and Bonatti (2019) highlight the emergence of social externalities resulting from the widespread use of data. In legal scholarship, Ritter and Mayer (2017) advocate for treating data as a form of property and granting ownership rights to users, while Sokol and Comerford (2015) explore the intersection of big data and market power within the antitrust framework. However the focus on informational complexity is novel and has distinct predictions that matter for the future of labor and of democracy.

In the realm of finance, big data facilitates better prediction and parsing of market trends and investment opportunities (Goldstein et al., 2021). A special issue edited and summarized by Goldstein and colleagues provides comprehensive insights into the transformative potential of big data in financial analytics, showcasing how it enables more accurate predictions and enhances decision-making processes within the financial sector.

The study of complexity arises in behavioral finance, decision theory, and experimental economics (for example, Puri (2022), Bernheim and Sprenger (2020), and Moffatt et al. (2015)), where subjects are shown to be averse to content they deem too complex. This is related to the study of attention and rational attention (Sims, 2003; Mackowiak and Wiederholt, 2009; Caplin and Dean, 2015). What this work adds is a theory of where attention, or a tolerance for complexity, comes from.

Many of our ideas are similar to ideas about commitment and addiction. The idea that cognitive work is unpleasant in the short run, beneficial in the long run and the consumers avoid doing it is simliar to DellaVigna and Malmendier (2006)'s analysis of physical exercise. Our theory of this lack of commitment draws from the seminal work on rational addiction by Becker and Murphy (1988). However, the argument that the consumption of complex information falls in this category is the key novel idea of this paper and has important consequences that are not relevant for exercise or drug addiction problems previously studied.

1 Motivating Evidence

Our model is based on two key assumptions: that attention is like exercise and that data is used to target people with different attributes with content matched to their attributes. We notice that each claim has been researched extensively, the first in a medical literature, and the second in the big data literature. We summarize the evidence across these literatures for each fact. Our contribution is to bring these disparate literatures together to develop a model with new predictions that provides a novel perspective on the heated debate surrounding the benefits and costs of big data. We present evidence of those two claims first, before we proceed to describe the model.

Mental Exercise. Our first key assumption is that the ability to process complex information is improved with practice. The more one reads, hears or processes complex information, the more capable a person is of performing this task. In other words, complex thinking is like a muscle that needs to be exercised. The medical literature finds evidence in favor of this key assumption: mental exercise increases cognitive ability. For example Pillai et al. (2011) and Verghese et al. (2003) find that doing crossword puzzles reduces cognitive aging. Devanand et al. (2022) document cognitive improvements for individuals with mild cognitive impairment when they train with crosswords. Sevinc et al. (2021) show that mindfulness training improves cognitive performance in older cognitively functioning adults. Ng et al. (2021) document that computerized cognitive training increase cognitive ability across adults of different ages. Petrella et al. (2023) finds weak support for crosswords helping slow the progression of Alzheimer's disease. And Jonaitis et al. (2013) notes that game-playing preserves cognitive strengths later in life.

This is a non-exhaustive list of medical studies showing a link between cognitive exercise and cognitive ability across a range of cognitive types and ages. While some studies, like Aartsen et al. (2002), find no link between cognitive ability and mental exercise, review articles conclude that the pre-ponderance of the evidence favors mental exercise helping with a variety of cognitive outcomes (Jak et al., 2013).

Our model also assumes that people who have a low ability to process complex information tend to avoid such content. This assumption is supported by experimental evidence. Puri (2022) finds that those with higher cognitive ability exhibit less complexity aversion.

Big Data and Targeting. Our second main assumption is that the rise in the abundance of data and the ability to harness it with big data technologies, like AI, has allowed providers to digital content to more accurately match content to user As highlighted by Veldkamp (2022), the utilization of big data has significantly enhanced predictive capabilities across diverse domains. One of the quintessential applications of big data lies in predicting human traits, a task pivotal for targeted marketing strategies, political campaigns, and financial analyses.

Predictive analytics leveraging large datasets play a pivotal role in tailoring products, campaigns, information, and advertisements to specific consumer segments (Matz, 2022). For instance, Nickerson and Rogers (2014) provide an extensive analysis of the integration of big data in political campaigns, illustrating how it enables more precise targeting of individuals for various forms of engagement, including donations, volunteering, and expressing support for candidates or issues.

Furthermore, the emergence of information firms, as documented by Dolfsma and Van Der Eijk (2018), underscores the commercial value of consumer data. Firms specialize in collecting and selling consumer data to companies seeking insights into target audiences for their products or services. Such data-driven insights empower companies to optimize their marketing strategies by precisely identifying and engaging with potential consumers who are most likely to be receptive.

2 Model

Our simple model is designed to show how the combination of the two assumptions, supported above, leads to the logical conclusion that growth in big data will exacerbate cognitive inequality. We use as simple a structure as possible to describe that argument. After setting up the model and extracting its ideas, we examine its predictions and consider the consequences of its predictions for democracy and the economy.

Model assumptions. There is a high type consumer H and a low type consumer L, where $L, H \in (0,1)$. The high type has a higher cognitive level c_H than the low type's cognitive level c_L.

There are two periods, 1 and 2. In period 1, each consumer is served information, $A_1^{(L)}$ and $A_1^{(H)}$ respectively, by a firm F. Information, which we call ads for short, could be news, tweets, blogs or other forms of information provision. We represent the complexity of the information provided by a firm as $A_1^{(i)}$, $i \in \{L, H\}$. This complexity measure lies in $(0,1)$.

The firm's has a single product, and its goal is to maximize the probability of a sale, a user engagement, or a click. The probability that consumer of type i buys a product is decreasing in the distance of the ad's complexity from the consumer's type $(A - i)$. Ads that are more complex than the consumer can process are additionally unlikely to result in a purchase:

$$P^{i,A} = P(\text{buy} \mid \text{consumer } i, \text{ ad } A) = \mathbb{1}\{A \leq i\} g(i - A) + \mathbb{1}\{A > i\} \theta g(A - i), \quad (1)$$

where $g : (0,1) \to (0,1)$ is a decreasing, strictly concave function (a lesser distance from the ad to consumer type increases the likelihood of purchase).

Consumers in this model are passive. They do not choose whether to see the ad or which ad to see; it is forced on them. We could add a consumer choice. However, in reality, much of this information is conveyed in the form of ads, which one is forced to consume, in order to reach some website with desired content.

The complexity of the ad the consumer sees affects the quality of their decision in the second period. The intuition is that processing complexity is a muscle that needs to be exercised. For now, we write

$$U_2^{(i)} = f(c(A_1^{(i)})). \quad (2)$$

That is, the utility in the second period is an increasing function of the complexity of the period 1 advertisement. Consumers are aware of this - they are not time inconsistent - but have no choices in our framework; the time-1 ad is delivered to them, in the process of their online activity, without their knowledge or explicit consent.

We conceptualize a world without big data as a pooling equilibrium, where both consumer L and consumer H are served the same ad. A world with big data allows for targeting, so the firm can tailor each advertisement to each consumer.

Equilibrium and welfare. We consider two different economies, with different data regimes. In the economy without big data, the firm chooses the ad complexity to maximize $P^{L,A} + P^{H,A}$. In the second economy with big data, the firm can run targeted advertisements, so they choose the complexity of ads A^L, A^H to maximize $P^{L,A^L} + P^{H,A^H}$.

Consumer welfare is given by period 2 utility, $U_2^{(i)}$.

3 Model Predictions

Our main point is to show how the assumptions of cognitive exercise and content targeting deliver cognitive divergence. Then, we consider how children might be represented in this setting and what effect data and its accompanying targeted content would have on them.

3.1 Information Targeting and Cognitive Divergence

Our first result simply describes the firms' optimal information complexity choices, in each of the two economies – the one without data targeting and the one with data-driven targeting.

Proposition 1. *Without big data, the single ad's complexity lies between L and H. With big data, each consumer's ad complexity is exactly their cognitive level.*

Proof. Since the probability that consumer of type i buys the product is given by $P^{i,A} = P(\text{buy — consumer } i, \text{ ad } A) = \mathbb{1}\{A \leq i\}g(i-A) + \mathbb{1}\{A > i\}\theta g(A-i)$, where

$g : (0, 1) \to (0, 1)$ is a decreasing, concave function, $P^{i,A}$ is maximized at $A = i$. This means that when ads can be perfectly targeted, each ad will match the consumer's cognitive level.

When ads cannot be targeted, the firm solves:

$$\begin{aligned}\max_A &\mathbb{1}\{A \leq H\}g(H - A) + \mathbb{1}\{A > H\}g(A - H) \\ &+ \mathbb{1}\{A \leq L\}g(L - A) + \mathbb{1}\{A > H\}g(A - L)\end{aligned}$$

There are three cases to consider:

1. $A \geq H$. Then the probability of sale is given by $g(A - H) + g(A - L)$. Since g is a strictly decreasing function, in this region, the constrained optimum is $A = H$.

2. $A \in (L, H)$.

3. $A \leq L$. Since g is a strictly decreasing function, in this region, the constrained optimum within this region is $A = L$.

The firm is better off choosing $A \in (L, H)$ than $A = L$ when $g(H - L) + g(0) < g(H - A) + g(A - L)$. Since g is continuous, it is sufficient to show that this inequality holds for some $A \in H, L$. Take $A = (H + L)/2$. The inequality then reduces to $g(H - L) + g(0) < 2g(\frac{H-L}{2})$. This is always satisfied by g being strictly concave.

Similarly, the firm is better off choosing $A \in (L, H)$ than $A = H$ when the same inequality holds; we have proven that it holds by using the concavity of g.

$\square$

This model solution lays the foundation for our main result, which proves divergence in utility between high and low types. The divergence in utility arises because the data-driven targeting of ads exacerbates the divergence in cognitive ability.

Proposition 2. *The low type suffers from the move to the data economy, while the high type benefits.*

Proof. Let $A_T^{(i)}$ denote the complexity of the ad shown to type i under targeting, and $A_{NT}^{(i)}$ the complexity without targeting.

Since utility in period 2 is the welfare criterion, and is an increasing function of the advertisement complexity in the first period, and from Proposition 1 $A_T^{(L)} < A_{NT}^{(L)}$, then $U_2^{(L)}(A_T^{(L)}) < A_{NT}^{(L)}$.

Since, from Proposition 1 $A_T^{(H)} > A_{NT}^{(H)}$, then $U_2^{(H)}(A_T^{(H)}) < A_{NT}^{(H)}$. $\square$

Type uncertainty. In this simple model, firms without data know nothing about their consumers. Firms with information know their types perfectly. That's an extreme assumption. In reality, data allows firms to predict consumers' types with greater confidence or small error.

Uncertainty lowers the firm's optimal choice of complexity. The reason is that the risk of overly-complex information exceed the costs of overly-simple information. This assumption is supported by the experimental aversion to overly-complex information in Puri (2022). This difference creates concavity in the firm's objective function. With a concave objective function in complexity, uncertainty lowers the optimal choice of complexity. The firm chooses to hedge the greater risk. The greater risk is that of too much complexity. Therefore, an uncertain firm chooses a simpler message.

3.2 Adding Children

We model children as individuals whose cognitive capabilities are not yet fully developed. This means children have the same cognitive capability distribution, but shifted down: they are of type L and type $L' = L - (H - L)$. Since their spending capacity is lower than adults', the firm places ϵ small weight on them in the absence of big data, when deciding which (single) ad to run. With big data, the firm can target ads to children costlessly.

Proposition 3. *Children suffer from the introduction of targeted advertising.*

Proof. Since g is continuous, for ϵ sufficiently small, the proof of Proposition 1 continues to hold, e.g. that the single ad's complexity is greater than L in the absence of targeting.

With targeting, that complexity reduces to the child's cognitive type.

Since welfare is measured by period 2 utility, this means that $U_2(A_{NT}) > U_2(A_T^{(i)})$ for each child type i. $\square$

The reason kids are hurt is that they are served less-sophisticated material than they would be without data targeting. This robs them of the opportunity to challenge themselves and improve their cognitive functioning.

Of course, many children have cognitive capacity that is lower than most adults. Perhaps they would be challenged by L-type information. In this simple model, they are still hurt because their skills would be improved more with the higher-complexity information. See the discussion of overly-complex information in the next Section.

3.3 Complexity vs. Transparency

A potential criticism of this result is the following: Language that is too complex is undecipherable. It might be so confusing to the reader that its meaning and/or educational value are lost. We consider two potential downsides to complexity.

Costs of overly-complex information. Perhaps the best information, for the reader's well-being, challenges them, but not too much. Even in this case, the point still stands that the short-run information of the firm serving the ad is not to challenge the recipient's cognitive capacity, but to maximize the probability of sale or engagement. That is best accomplished by targeting at or below their current cognitive level.

Complex language vs. complex ideas There is a long-standing debate about the merits of using complex language. Defenders argue that complex language expresses complex ideas with more nuance and precision than simple language could. Detractors argue that complex language reflects a writer's lack of effort to communicate clearly. Neither our model nor our evidence will resolve this debate.

Our argument is not that one form of writing is better or worse, but that complex language is more likely to indicate complex thought. It is a proxy. Complex thought and language both challenge the human brain to exert effort and grow stronger. Even if complex language is the result of poor writing and is ineffective at persuasion, it may still challenge the cognition of the reader or listener. It is this cognitive capacity which is our focus.

4 Evidence of Cognitive Divergence

The testable hypothesis from the theory is that complexity of information that is served to viewers, who might be targeted, is diverging. In other words, the difference between the least complex and most complex messages is growing, as more data and better data technologies are deployed.

4.1 Measuring Complexity

There are many ways to measure language complexity. The Flesch-Kinkaid measure is probably the most well-known measure. It relies on the number of words per sentence and the number of syllables in each word. This metric is simplistic because it mis-classifies complex or less-phonetic, monosyllabic words like “fjord” or ”bough,” as well as simple, common words with many syllables, like ”anybody” or ”congratulations.”

More recently, two other measures of language complexity emerged to solve these problems. The Dale–Chall formula uses a list of about 3,000 words familiar to American fourth graders. The score estimates the U.S. grade level required to understand the text, based on how many words are on that fourth-grade list. The score also factors in the average sentence length.

The Gunning-Fog Index scores a text based on the average sentence length and the percentage of complex words. Complex words are judged to be those with three or more syllables, that are not proper nouns, familiar jargon, or compound words.

Our results make use of both the Dale-Chall and the Gunning-Fog complexity metrics.

4.2 Divergence of Complexity in Political News

To test our theory, we examine online communications of the major US political parties. Measuring divergence in political news is an important endeavor, especially in an era characterized by increasing polarization and the proliferation of partisan media outlets. This exercise supports the theory and shows that it operates in a realm of great importance to society.

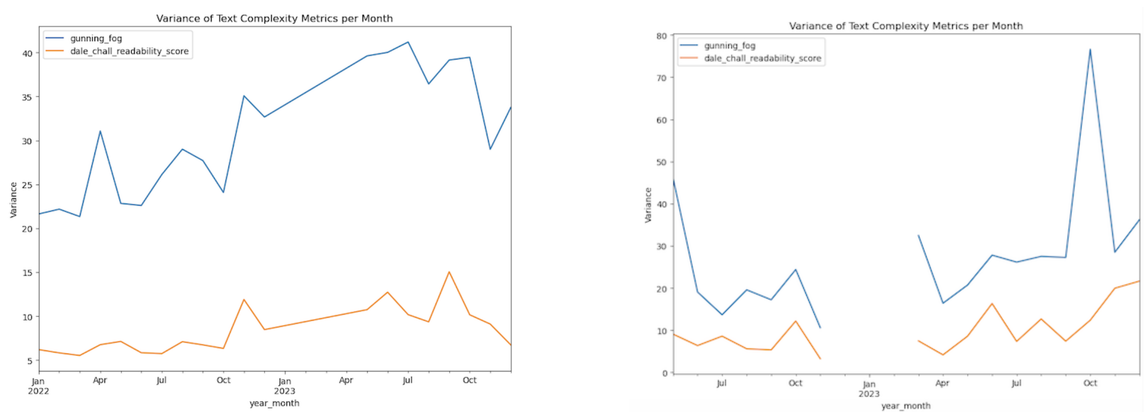
The hypothesis is that targeting of information causes the communications to differ more in their complexity. In other words, we are looking for dispersion in linguistic complexity, across posts of the same source. If data and algorithms are getting more precise at targeting recipients with communications designed for their cognitive level, then this disperson should rise over time.

In our analysis, we focus on the Facebook pages of the Democratic Party (DNC) and the Republican Party (RNC). Specifically, we examine the the Democratic Party page, which boasts 1.6 million followers, and the National Republican Congressional Committee page, which commands 1 million followers.[1] These figures are similar for RNC/Republican Party.

Our methodology involves analyzing the textual content of each post on these pages since their inception. We exclude any URLs or links from our analysis, as these will be examined separately at a later stage. Within each month of a given year, we measure the dispersion in the complexity of the posts. This involves applying complexity measures to assess the readability and linguistic sophistication of the text. By analyzing the variance in grade level readability across posts within a month, we gain insights into how the complexity of messaging changes over time.

Using the Dale–Chall and Gunning-Fog complexity measures described above, we give each post a score. For each month, we plot the dispersion in complexity scores

[1]Note that the National Republican Congressional Committee, has more followers (1 million) than the Republican Party Page (21k). We therefore choose the page with more followers.



(A) Democratic Party Page

(B) National Republican Congressional Committee Page

FIGURE 1: Complexity of Facebook Posts Over Time

Note: This figure plots the variance within a given month of readability scores for posts made by the page in that month. The two readability scores we consider are Gunning-Fog and Dale-Chall. The NRCC page did not post from Dec 2022 - March 2023.

observed on a given website. This graphical representation, depicted in Figure 1, provides a visual understanding of how the complexity of posts on the DNC and RNC pages evolves over time. The figure depicted includes data only from 2022-2023; we are in the process of going back further in time. Table 1 depicts the point estimates for the variance of the readability scores on each page over time. While there a slight increase over time in this data, it is quite noisy, and we are in the process of going further back in time, because it is possible that use of big data for targeting complexity had already been incorporated by 2022.

As a continuation of our analysis, we extend the same methodology to analyze the text-based links included in the posts. This additional step allows us to assess the complexity of the content shared via external sources, providing a comprehensive understanding of the linguistic diversity present in the discourse on these Facebook pages. The results of this continuation analysis are coming soon.

Table 1: Variance in Readability Scores Over Time

Month	Page	Gunning-Fog	Dale-Chall
1/1/22	Democratic	21.63	6.19
2/1/22	Democratic	22.18	5.82
3/1/22	Democratic	21.34	5.53
4/1/22	Democratic	31.07	6.75
5/1/22	Democratic	22.83	7.12
6/1/22	Democratic	22.60	5.83
7/1/22	Democratic	26.12	5.74
8/1/22	Democratic	29.01	7.09
9/1/22	Democratic	27.71	6.73
10/1/22	Democratic	24.08	6.33
11/1/22	Democratic	35.08	11.90
12/1/22	Democratic	32.67	8.47
5/1/23	Republican	39.63	10.74
6/1/23	Republican	40.03	12.73
7/1/23	Republican	41.21	10.17
8/1/23	Republican	36.43	9.35
9/1/23	Republican	39.15	15.05
10/1/23	Republican	39.47	10.16
11/1/23	Republican	28.99	9.10
12/1/23	Republican	33.75	6.73
5/1/22	Republican	46.12	9.08
6/1/22	Republican	19.06	6.37
7/1/22	Republican	13.66	8.60
8/1/22	Republican	19.58	5.60
9/1/22	Republican	17.22	5.34
10/1/22	Republican	24.40	12.18
11/1/22	Republican	10.60	3.25
3/1/23	Republican	32.46	7.48
4/1/23	Republican	16.41	4.16
5/1/23	Republican	20.72	8.59
6/1/23	Republican	27.81	16.34
7/1/23	Republican	26.15	7.35
8/1/23	Republican	27.52	12.67
9/1/23	Republican	27.26	7.41
10/1/23	Republican	76.63	12.39
11/1/23	Republican	28.49	19.96
12/1/23	Republican	36.18	21.64

5 Conclusion: Potential Consequences of Cognitive Inequality

We introduce a novel dimension to the debate on data regulation by exploring its impact on the complexity of communication and cognitive abilities. We contend that if cognitive ability is like a muscle that needs to be exercised, and if data is used to target people with different characteristics and deliver them different content, that the natural outcome of those assumptions is that profit-maximizing firms will deliver content that exacerbates cognitive inequality. We find evidence of this divergence in the complexity of online political communication. This raises economic and politcal concerns.

Divergence in cognitive capabilities among Americans carries significant economic and political ramifications, influencing various facets of society. Firstly, within the economic realm, cognitive disparities can exacerbate existing inequalities in income, wealth, and employment opportunities. Individuals with higher cognitive abilities are often better equipped to excel in complex and lucrative fields such as technology, finance, and management, thereby widening the income gap between cognitive elites and those with limited cognitive capacities. This economic stratification can perpetuate cycles of poverty and hinder social mobility, as individuals from disadvantaged backgrounds may face barriers to accessing high-quality education and opportunities for skill development.

Moreover, cognitive disparities can impact labor market dynamics, shaping patterns of occupational segregation and wage differentials. Industries that require advanced cognitive skills may experience heightened demand for talent, leading to increased competition and wage inflation for individuals possessing such capabilities. Conversely, sectors reliant on manual labor or routine tasks may face downward pressure on wages, as automation and technological advancements diminish the value of non-cognitive skills. This polarization of the labor market can contribute to social unrest and political polarization, as disenfranchised individuals perceive themselves as economically marginalized and excluded from the benefits of technological progress.

Furthermore, cognitive disparities may intensify social divisions and ideological polarization within the political landscape. Cognitive elites may gravitate towards political ideologies and movements that prioritize meritocracy, individualism, and market-driven solutions, advocating for policies that favor intellectual achievement and economic competitiveness. Conversely, individuals with lower cognitive abilities may be drawn to populist or authoritarian narratives that promise simplistic solutions to complex societal problems, such as nativism, protectionism, or anti-establishment rhetoric. This ideological polarization can undermine social cohesion, compromise the functioning of democratic institutions, and impede efforts to address pressing challenges such as inequality, climate change, and healthcare reform.

Of course, big data and new data algorithms offer many benefits, which were not the focus of this study. Firms that can predict demand and supply become more efficient, consumers find products they prefer and the pace of innovation may accelerate. Throughout history, similar advances in technology have lifted people out of poverty and improved lives. But such advances usually have some adverse effects that societies need to understand and to mitigate. One such adverse effect may be the reshaping of communication.

References

Aartsen, M. J., C. H. M. Smits, T. Van Tilburg, K. C. P. M. Knipscheer, and D. J. H. Deeg. 2002. "Activity in Older Adults: Cause or Consequence of Cognitive Functioning? A Longitudinal Study on Everyday Activities and Cognitive Performance in Older Adults." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 57: P153–P162.

Becker, Gary S, and Kevin M Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96: 675–700.

Bergemann, Dirk, and Alessandro Bonatti. 2019. "The Economics of Social Data." *SSRN Electronic Journal*.

Bergemann, Dirk, Jacques Crémer, David Dinielli, Carl-Christian Groh, Paul Heidhues, Maximilian Schäfer, Monika Schnitzer, Fiona Scott Morton, Katja Seim, and Michael Sullivan. 2023. "Market design for personal data."

Bernheim, B. Douglas, and Charles Sprenger. 2020. "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting." *Econometrica*, 88.

Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*, 105: 2183–2203.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96: 694–719.

Devanand, D. P., Terry E. Goldberg, Min Qian, Sara N. Rushia, Joel R. Sneed, Howard F. Andrews, Izael Nino, Julia Phillips, Sierra T. Pence, Alexandra R. Linares, Caroline A. Hellegers, Andrew M. Michael, Nancy A. Kerner, Jeffrey R. Petrella, and P. Murali Doraiswamy. 2022.

"Computerized Games versus Crosswords Training in Mild Cognitive Impairment." *NEJM Evidence*, 1.

Dolfsma, Wilfred, and Rene Van Der Eijk. 2018. "How info-firms use big data to target customers." *Journal of Business Strategy*, 39: 50–55.

Goldstein, Itay, Chester S Spatt, and Mao Ye. 2021. "Big Data in Finance." *The Review of Financial Studies*, 34: 3213–3225.

Jak, Amy J., Adriana M. Seelye, and Sarah M. Jurick. 2013. "Crosswords to Computers: A Critical Review of Popular Approaches to Cognitive Enhancement." *Neuropsychology Review*, 23: 13–26.

Jonaitis, Erin, Asenath La Rue, Kimberly D. Mueller, Rebecca L. Koscik, Bruce Hermann, and Mark A. Sager. 2013. "Cognitive activities and cognitive performance in middle-aged adults at risk for Alzheimer's disease." *Psychology and Aging*, 28: 1004–1014.

Liu, Zhuang, Michael Sockin, and Wei Xiong. 2021. "Data Privacy and Temptation." Princeton University. Economics Department. Working Papers 2021-77.

Mackowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99: 769–803.

Matz, Sandra. 2022. *The Psychology of Technology: Social Science Research in the Age of Big Data.* . 1st ed., American Psychological Association.

Moffatt, Peter G., Stefania Sitzia, and Daniel John Zizzo. 2015. "Heterogeneity in preferences towards complexity." *Journal of Risk and Uncertainty*, 51: 147–170.

Ng, Nicole F., Allen M. Osman, Kelsey R. Kerlan, P. Murali Doraiswamy, and Robert J. Schafer. 2021. "Computerized Cognitive Training by Healthy Older and Younger Adults: Age Comparisons of Overall Efficacy and Selective Effects on Cognition." *Frontiers in Neurology*, 11: 564317.

Nickerson, David W., and Todd Rogers. 2014. "Political Campaigns and Big Data." *Journal of Economic Perspectives*, 28: 51–74.

Petrella, Jeffrey R., Andrew M. Michael, Min Qian, Adaora Nwosu, Joel Sneed, Terry E. Goldberg, Davangere P. Devanand, and P. Murali Doraiswamy. 2023. "Impact of Computerized Cognitive Training on Default Mode Network Connectivity in Subjects at Risk for Alzheimer's Disease: A 78-week Randomized Controlled Trial." *Journal of Alzheimer's Disease*, 91: 483–494.

Pillai, Jagan A., Charles B. Hall, Dennis W. Dickson, Herman Buschke, Richard B. Lipton, and Joe Verghese. 2011. "Association of Crossword Puzzle Participation with Memory Decline in Persons Who Develop Dementia." *Journal of the International Neuropsychological Society*, 17: 1006–1013.

Puri, Indira. 2022. "Simplicity and Risk." *Working Paper.*

Ritter, Jeffrey, and Anna Mayer. 2017. "Regulating Data as Property." *Duke Law and Technology Review*, 16: 220–277.

Seim, Katja, Dirk Bergemann, Jacques Cre, David Dinielli, Carl-Christian Groh, Paul Heidhues, Maximillian Schaefer, Monika Schnitzer, Ludwig-Maximilians-University Munich, Fiona M Scott Morton, and Michael Sullivan. 2022. "Market Design for Personal Data." *The Tobin Center at Yale University: Digital Regulation Project.*

Sevinc, Gunes, Johann Rusche, Bonnie Wong, Tanya Datta, Robert Kaufman, Sarah E. Gutz, Marissa Schneider, Nevyana Todorova, Christian Gaser, Götz Thomalla, Dorene Rentz, Bradford D. Dickerson, and Sara W. Lazar. 2021. "Mindfulness Training Improves Cognition and Strengthens Intrinsic Connectivity Between the Hippocampus and Posteromedial Cortex in Healthy Older Adults." *Frontiers in Aging Neuroscience*, 13: 702796.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics*, 50: 665–690.

Sokol, D. Daniel, and Roisin Comerford. 2015. "Antitrust and Regulating Big Data." *George Mason Law Review*, 23.

Veldkamp, Laura. 2022. "Valuing Data as an Asset." *Forthcoming, Review of Finance.*

Verghese, Joe, Richard B. Lipton, Mindy J. Katz, Charles B. Hall, Carol A. Derby, Gail Kuslansky, Anne F. Ambrose, Martin Sliwinski, and Herman Buschke. 2003. "Leisure Activities and the Risk of Dementia in the Elderly." *New England Journal of Medicine*, 348: 2508–2516.