NBER WORKING PAPER SERIES

DATA PRIVACY AND ALGORITHMIC INEQUALITY

Zhuang Liu
Michael Sockin
Wei Xiong

Working Paper 31250
http://www.nber.org/papers/w31250

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2023, Revised September 2025

We are grateful to Roland Benabou, Dirk Bergmann, Long Chen, Will Cong, Shaowei Ke, Andrey Ordin, Laura Veldkamp, Liyan Yang, Leeat Yariv, Lixin Ye, and seminar participants at American Finance Association Meetings, Cambridge, Cornell Johnson, CUHK Shenzhen, Luohan Academy, Northwestern, Princeton, Stanford, Toulouse, UCL, UT-Dallas, and the Virtual Finance and Paris Dauphine Digital Activities Seminars for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Data Privacy and Algorithmic Inequality
Zhuang Liu, Michael Sockin, and Wei Xiong
NBER Working Paper No. 31250
May 2023, Revised September 2025
JEL No. D0, E0

ABSTRACT

This paper develops a foundation for consumer privacy preferences by linking them to the desire to conceal behavioral vulnerabilities. Although data sharing with digital platforms improves matching efficiency for products and services, it also exposes individuals with self-control issues to predatory lending practices, creating a new form of inequality in the digital era—algorithmic inequality. Privacy regulations empower consumers to opt out of data sharing, but cannot fully protect vulnerable individuals because of data-sharing externalities. Moreover, coordination frictions among consumers may generate multiple equilibria with drastically different levels of data sharing, amplifying both efficiency gains and inequality risks.

Zhuang Liu
The University of Hong Kong (HKU)
liuz@hku.hk

Michael Sockin
University of Texas at Austin
Department of Economics
michael.sockin@mccombs.utexas.edu

Wei Xiong
Princeton University
Department of Economics
and NBER
wxiong@princeton.edu

The rise of big data has brought major benefits for consumers. By pooling together multidimensional user data and applying advanced analytics, platforms such as Google, Amazon, Facebook, and ChatGPT can provide personalized products and services that would otherwise be prohibitively costly. In finance, big data has enabled fintech lenders partnering with these platforms to expand credit access to underserved borrowers (Allen, Gu, and Jagtiani 2021; Berg, Fuster, and Puri 2022). Liu, Lu, and Xiong (2023) show how big tech lenders use unique user data to extend loans to small businesses, while Parlour, Rajan, and Zhu (2022) and He, Huang, and Zhou (2023) analyze how open banking's data-sharing rules affect bank–fintech competition.

Yet the same data practices raise unprecedented challenges for consumer protection and privacy. While personalization can improve services, it can also exploit behavioral biases, encouraging overspending—for example, by bundling credit with online purchases. Reports by the Stigler Committee (2019), Helberger et al. (2021), OECD (2021), and the FTC (2022) warn that digital platforms are especially effective at exploiting individual vulnerabilities. The rapid growth of unsecured credit via "buy now, pay later" (BNPL) lenders illustrates this concern. Di Maggio, Katz, and William (2022) find BNPL access substantially boosts spending beyond standard substitution effects. Using data on 10.6 million U.S. consumers, DeHaan et al. (2024) show BNPL adoption leads to spikes in overdraft charges and credit card fees, while Larrimore et al. (2024) find that financially vulnerable consumers overextend themselves when BNPL is offered.

A related concern is the prevalence of "dark patterns" that manipulate consumers into unwanted purchases or debt. A 2019 survey found that 24% of 1,760 websites employed such practices (OECD 2021). Mathur et al. (2019) identify dark patterns on 11.1% of 11,000 shopping sites and document third-party providers that design them. Johannesson (2021) found that two leading Swedish BNPL lenders used at least ten deceptive designs to encourage borrowers to take on debt with credit installments to increase profits. The FTC fined Epic Games $520 million in 2022 for using dark patterns to induce unwanted Fortnite purchases.

Protecting consumer data privacy is central to addressing these risks. As Zarsky (2019) and Spencer (2020) note, the problem lies less in manipulation per se than in the internet's capacity for intensive data collection, personalization, and real-time execution. Platforms like TikTok tailor content feeds to maximize engagement, a design linked to addiction and mental health issues among youth (BBC 2022).[1] In response, governments have enacted sweeping privacy regulations, including the EU's General Data Protection Regulation (GDPR, 2018) and California's Consumer Privacy Act (CCPA, 2020), both emphasizing informed consent and data safeguards.

Motivated by these observations, we develop a model of digital platforms to study how on-demand credit and data sharing affect both aggregate and cross-sectional consumer welfare. Platforms use shared data to match consumers with desirable goods and financing, but the same data expose vulnerable consumers to impulse spending and predatory lending. Our analysis shows that the ability of platforms, intermediaries, and vendors to profit from behavioral weaknesses creates costs of data sharing that fall unevenly across consumers. The resulting divide—between those who benefit and those who are exploited—gives rise to "algorithmic inequality," a new form of inequality in the digital age.[2]

To anchor our analysis, we focus on limited self-control as a key vulnerability shaping consumers' data-sharing decisions. Consumers with self-control problems may struggle to resist impulse purchases promoted by digital platforms through BNPL credit schemes. These schemes, often targeted at credit-constrained consumers, encourage borrowing up to the limit, downplay the costs of debt, and obscure late fees and interest charges (World Bank Group 2021). Ultimately, they may impose harm through fees, higher interest, and financial distress (DeHaan et al. 2024). Recognizing their vulnerability, such consumers may opt to withhold data to avoid being targeted, even at the cost of losing access to desirable goods and services.

We adopt the temptation utility framework of Gul and Pesendorfer (2001) to analyze the data-sharing choices of vulnerable consumers. Weak-willed consumers incur mental costs from resisting predatory financial products and therefore prefer smaller menus that exclude such options. Their data sharing, together with sellers' advertising, shapes these menus and, in turn, their privacy preferences on digital platforms. Unlike prior work that treats the cost of privacy as exogenous (e.g., Jones and Tonetti 2020; Cong, Xie, and Zhang 2020; He, Huang, and Zhou 2023), we microfound it. This approach not only allows consumers to distinguish between beneficial and harmful data sharing—absent in earlier models—but also

[1] https://www.bbc.com/news/uk-wales-62720657.

[2] Our concept of "algorithmic inequality" differs from concerns about algorithmic bias arising from statistical discrimination (e.g., Cowgill and Stevenson 2020; Cowgill and Tucker 2020). For instance, Luguri and Strahilevitz (2021) show that dark patterns disproportionately affect less educated consumers, leading to adverse distributive outcomes.

generates novel data-sharing externalities through the endogeneity of privacy costs.

Our model considers an online platform—such as Google, TikTok, or Facebook—that collects consumer data and shares it with sellers. The platform hosts N normal goods, which provide consumers with utility, and J predatory goods, whose sellers exploit vulnerable consumers through dark patterns to induce impulse purchases. Because such purchases are excessive, they require financing provided by a fintech lender partnering with the platform via a BNPL scheme. This financing exposes consumers to default risk through late fees and interest on missed payments. Both normal and predatory sellers can advertise to potential consumers, but doing so entails a convex cost, capturing the increasing difficulty of reaching a wider audience.

There are two types of consumers: strong-willed, who always resist predatory goods, and weak-willed, who may succumb to them. Both types derive utility from normal goods, but only weak-willed consumers are vulnerable to purchasing predatory goods financed through BNPL schemes. Such purchases expose them to the risk of financial distress.

While consumers might prefer their data to be shared only with sellers of normal goods, in practice the same data can also be accessed by predatory sellers. Although data's non-rivalry is often celebrated as a strength of the digital economy, we emphasize a problematic dimension: the lack of exclusivity in its use. This arises both because it is difficult to draft narrowly tailored data authorization agreements and because consumers fail to internalize the externalities of their data-sharing decisions.

Strong-willed consumers, who can resist predatory goods, are indifferent to such advertising and thus favor broader product menus and data sharing that enhances targeting of normal goods. Weak-willed consumers, by contrast, face a sharper trade-off: while data sharing improves targeting of normal goods, it also heightens their exposure to predatory marketing. This tension is central to our model. Recognizing their vulnerability, weak-willed consumers may withhold data to protect themselves, even at the cost of losing beneficial targeting.

We analyze three data-sharing schemes.

- **No data sharing**: Resembling traditional advertising, sellers face a dark pool of consumers and can only reach a random subset. This shields weak-willed consumers from predatory goods but also limits access to normal goods for everyone.

- **Full data sharing:** Sellers observe each consumer's type and target ads precisely.

This improves access to normal goods but heightens weak-willed consumers' exposure to predatory goods. The welfare gap between strong- and weak-willed consumers—our measure of algorithmic inequality—is always larger under full data sharing. If the harm from predatory goods is severe, full data sharing may also reduce total welfare relative to no data sharing.

- **Opt-in/opt-out (GDPR-style):** Strong-willed consumers always opt in, while weak-willed consumers weigh the benefits of better targeting against the risk of predation, following a cutoff rule: the most tempted opt out, others opt in. Their decisions shape the composition of opt-in and opt-out pools, which in turn affect sellers' targeting.

One might expect opt-in/opt-out to dominate the other two schemes, since consumers choose optimally. But data-sharing externalities complicate the comparison: when one consumer opts in, sellers also learn about others. As a result, full data sharing maximizes welfare if weak-willed consumers' self-control problem is mild; no data sharing dominates if it is severe; and opt-in/opt-out is best when the problem is intermediate. Regardless of the ranking, full data sharing always produces the greatest algorithmic inequality, since it offers the least protection to weak-willed consumers.

Because the cost of privacy is endogenous, the opt-in/opt-out regime can generate multiple equilibria. In particular, if all consumers opt in, even the most tempted weak-willed consumers lose the protection of opting out and are effectively forced to share data, collapsing the scheme into full data sharing. Such multiplicity implies that small changes in the environment can trigger sharp shifts in data-sharing outcomes, leading to extreme forms of algorithmic inequality.

We calibrate the model using 2024 e-commerce revenues, advertising costs, and evidence on GDPR efficacy to assess the severity of self-control problems on online platforms. The calibration yields three equilibria: one with full opt-in by weak-willed consumers, one with partial opt-in, and one with minimal opt-in. Among these, the full data-sharing equilibrium delivers the highest utilitarian welfare—16.6% above the minimal data-sharing equilibrium—but also widens algorithmic inequality between strong- and weak-willed consumers by 11.3%. Thus, while data sharing generates substantial social benefits, it does so at significant cost to vulnerable consumers. A dynamic extension shows that these problems intensify over time, as today's data sharing both improves the quality of normal goods and erodes consumers' future ability to resist predatory goods.

Recent advances in big data analytics and information technologies have profound implications for financing on digital platforms. Our model shows that data sharing affects not only the credit access of individual consumers but also the welfare of others. This perspective complements existing work: He, Huang, and Zhou (2023) examine how open banking data sharing shapes competition between banks and fintechs; Parlour, Rajan, and Zhu (2022) analyze fintech competition in payment services when banks use payment data to infer credit quality; and Berg et al. (2025) show how BNPL allows merchants to price discriminate through zero-interest loans. None of these studies, however, address the potential harm to vulnerable consumers, which is the focus of our analysis.

Our paper connects data privacy to the broader literature on exploiting consumer vulnerabilities in financial markets. Evidence shows that vulnerable consumers are often targeted through overpriced credit card debt and underpriced investment products such as health club memberships (DellaVigna and Malmendier 2004, 2006), payday loans (Bertrand and Morse 2011; Melzer 2011), add-on pricing (Gabaix and Laibson 2006), and overdraft fees (Stango and Zinman 2014). Digital technologies amplify these practices: for example, fintech lenders exploit borrowers' self-control problems using payment data (Di Maggio and Yao 2020).

The externalities we highlight build on the concept of social data (Acemoglu et al. 2019; Bergemann, Bonatti, and Gan 2019; Easley et al. 2019). These studies show that one consumer's data can reveal information about others, depressing data prices and encouraging excessive sharing. Galperti, Levkun, and Perego (2022) extend this logic to e-commerce platforms using a mechanism design approach. We differ by focusing on a new form of externality: when some consumers opt in, their data reveal the vulnerability of those who opt out. Moreover, unlike models with symmetric consumers, ours distinguishes between those with and without self-control problems. This allows us to show how data sharing can simultaneously raise aggregate welfare and impose disproportionate costs on the vulnerable, producing algorithmic inequality and motivating a policy role for data privacy regulation.

Finally, our analysis contributes to the macroeconomic literature on consumer data. Prior work emphasizes the non-rivalry of data and its increasing returns (Jones and Tonetti 2020; Cong, Xie, and Zhang 2020; Cong et al. 2020; Cong and Mayer 2022) or its role in long-run capital accumulation (Farboodi and Veldkamp 2020). These studies typically treat the cost of data sharing as exogenous. Abis and Veldkamp (2021) provide an exception, showing that AI adoption may reduce labor's income share by 5%. Our contribution is to emphasize that data sharing not only matters for growth but also generates inequality across consumers.

1 A Model of Data Sharing

We examine a digital platform where consumers purchase goods from sellers using BNPL credit provided by a fintech lender. The process unfolds in two stages. First, consumers join the platform and decide whether to share their data. Second, over multiple years, the platform collects digital histories—such as searches and purchases—subject to consent and shares them with sellers. Using this information, sellers target consumers with offers, and consumers can make purchases accordingly.

There are $N > 1$ normal goods, desirable to all consumers, and J predatory goods, representing excessive or unnecessary expenditures, such as impulse purchases.[3] Consumers are liquidity constrained and must finance purchases with short-term BNPL loans from the platform's affiliated fintech lender. Importantly, predatory goods may lead to default, reflecting their excessive nature: late fees and high interest on missed payments can trigger financial distress. We index the N normal goods by $n \in \{1, .., N\}$, and the J predatory goods by $j \in \{N + 1, ..., N + J\}$. While normal goods can also be BNPL-financed, we focus on predatory goods to highlight how data sharing amplifies self-control problems—overspending and debt accumulation.

There are two types of consumers: strong-willed, without self-control problems, and weak-willed, who are subject to them. Following Ichihashi (2020), we assume each consumer desires only one normal good, and each weak-willed consumer is tempted by a single predatory good. Sellers therefore face a nontrivial matching problem, and consumer data sharing improves matching efficiency.

Data is non-rival: the platform can share it with both normal and predatory sellers and cannot commit to restricting access to the former. This creates a cost of data sharing. Unlike prior work that treats this cost as exogenous (e.g., Jones and Tonetti 2020; Cong, Xie, and Zhang 2020; Cong et al. 2020), we model it as endogenous and heterogeneous. For strong-willed consumers the cost is zero, while for weak-willed consumers it is positive and rises with the severity of their self-control problem. This captures the reality that consumers

[3] While our model focuses on goods, it also applies to ecosystems of online services. For instance, normal services can represent conveniences include Google's free search or X's free news feeds, while predatory goods capture harms from data sharing such as cryptocurrency scams or addictive content.

cannot control how data is used—sharing data for beneficial purposes does not prevent its exploitation for harmful ones.

1.1 Consumers

There is a total of one unit of consumers, divided into strong-willed (π_S) and weak-willed (π_W) types, with $\pi_S + \pi_W = 1$. We assume $\pi_S > \frac{1}{2}$, consistent with survey evidence (e.g., Ameriks et al., 2007, Toussaert, 2018). Strong-willed consumers resist predatory goods, while weak-willed consumers may not. Each consumer desires one normal good n, while each weak-willed consumer is also tempted by one predatory good j. Let S_n denote the strong-willed consumers who want normal good n, and W_{nj} the weak-willed consumers who want normal good n and are tempted by predatory good j.

To buy good x, consumer i pays seller $p_i(x)$, which may be tailored to that consumer. For predatory goods, $p_i(j)$ also reflects the severity of self-control problems among the weak-willed. To focus on the role of BNPL credit in excessive expenditures, we assume that consumers always have sufficient wealth to repay loans for normal goods. By contrast, weak-willed consumers may face financial distress when repaying BNPL loans for predatory goods.[4] We treat normal and predatory purchases as separate accounts, consistent with evidence on mental accounting (Gelman and Roussanov 2024) that consumers view credit lines as nonfungible. Hence, each consumer chooses independently across advertised goods.[5]

Heterogeneity among weak-willed consumers is central to our model, as it provides camouflage under the opt-in/opt-out data-sharing scheme analyzed later. We focus on a symmetric equilibrium with uniformly distributed preferences: each consumer is equally likely to prefer any of the N normal goods, and each weak-willed consumer is equally likely to be tempted by any of the J predatory goods. Both types consume according to their preferences and the advertisements they receive.

To microfound data-sharing costs, we adopt the temptation utility framework of Gul

[4]Self-control problems are not confined to disadvantaged consumers. For example, Jike Zhang, a celebrated Chinese athlete and 2012 Olympic table tennis champion, reportedly accrued tens of millions of dollars in gambling debt despite earning around \$10 million in 2017. Pressured by creditors, he even resorted to releasing private photos of his former movie-star girlfriend, causing severe reputational and commercial losses. This illustrates that debt-related self-control issues can afflict even wealthy individuals.

[5]Adding a budget constraint would introduce an additional distortion: predatory goods might crowd out normal consumption, further worsening algorithmic inequality by incentivizing normal-good sellers to favor strong-willed consumers. We abstract from this to focus on the direct impact of data sharing. When a consumer shares data, she gains better targeting of normal goods but also risks exposure to predatory ones.

and Pesendorfer (2001), which axiomatizes self-control problems. Building on Kreps (1979), this framework models preferences over menus. Standard utility theory predicts that rational consumers strictly prefer larger menus, which expand their choice set and potential utility. By contrast, the temptation utility framework captures that a larger menu including tempting but predatory options may reduce welfare for consumers with self-control issues.

The consumer's preference for a menu $\mathcal{M}$ is given by:

$$\max_{x \in \mathcal{M}} \ [u(x) + v(x) - p_i(x)] - \max_{x' \in \mathcal{M}} \ v(x'), \tag{1}$$

where x is a choice from menu $\mathcal{M}$, and $u(x)$, $v(x)$, and $p_i(x)$ denote the commitment utility, temptation utility, and price, respectively.[6] The consumer's actual choice is determined by the first maximization in Equation (1):

$$x_* = \arg\max_{x \in \mathcal{M}} \ [u(x) + v(x) - p_i(x)], \tag{2}$$

which compromises between commitment and temptation utilities. If $x_* \neq \arg\max_{x' \in \mathcal{M}} v(x')$, the consumer exercises self-control. The cost of self-control arises from the gap $\max_{x' \in} v(x') - v(x_*)$: even unchosen tempting goods impose a welfare loss. If $v(x) = 0$, the model collapses to standard Von Neumann–Morgenstern utility.

The menu $\mathcal{M}$ depends on sellers' advertising strategies, themselves shaped by the platform's data-sharing scheme and the consumer's data-sharing choice.[7] A consumer's ex ante utility is thus the expected utility over all possible menus.

Temptation utility specification A consumer of type $\tau_i \in \{S_n, W_{nj}\}$ has commitment and temptation utilities from consuming normal good n, alternative normal goods $n' \neq n$,

[6] Following Gul and Pesendorfer (2004), we exclude prices from temptation utilities, though one could argue that higher prices and default risk reduce temptation. A modified formulation to capture this is:

$$\max_{x \in M} [u(x) + v(x) - 2p_i(x)] - \max_{x' \in M} [v(x') - p_i(x')],$$

without qualitatively impacting our key insights. For expositional simplicity, we use the first specification. We thank Shaowei Ke for this observation.

[7] Our analysis builds on the random temptation utility framework of Stovall (2010), a special case of random Strotz (1955) utility as characterized by Bénabou and Pycia (2002) and Dekel and Lipman (2012).

predatory good j, and alternative predatory goods $j' \neq j$:

	strong-willed		weak-willed	
x	$u_{S_n}(x)$	$v_{S_n}(x)$	$u_{W_{nj}}(x)$	$v_{W_{nj}}(x)$
n	$\tilde{u}_n > 0$	0	$\tilde{u}_n > 0$	0
$n' \neq n$	0	0	0	0
j	$u_B < 0$	0	$u_B < 0$	$\gamma_i \bar{v} - u_B > 0$
$j' \neq j$	$u_B < 0$	0	$u_B < 0$	0

(3)

Here, both types draw a random utility $\tilde{u}_n \sim U[0, \bar{u}]$ from a uniform distribution with $\bar{u} > 0$ as the maximal utility. This random utility can be interpreted as a transient taste for the normal good—for example, a temporary desire for clothing or a durable good on a given day.

Predatory goods, financed through BNPL, entail possible financial distress from late fees and interest. Although consumers may enjoy some immediate utility $u_i > 0$, with probability δ (independent across consumers) they default, incurring a loss C. Following Berg et al. (2025), we treat δ as exogenous.[8] We define expected commitment utility as $u_B = u_i - \delta C < 0$, reflecting the net harm of predatory goods.

Strong-willed consumers derive no temptation utility from predatory goods ($v_{S_n}(j) = 0$) and never use them. Weak-willed consumers, however, receive temptation utility $\gamma_i \bar{v} - u_B$, where $\bar{v} > 0$ measures the overall temptation strength, and $\gamma_i \sim U[0, 1]$ captures heterogeneity in susceptibility.

The weak-willed consumer's choice problem reduces to:

$$\max_{x \in \{j, \emptyset\}} \left[u_{W_{nj}}(x) + v_{W_{nj}}(x) - p_i(x)\right] = \max\{\gamma_i \bar{v} - p_i(j), 0\}.$$

Thus, consumer i buys predatory good j if $\gamma_i \geq p_i(j)/\bar{v}$. We interpret γ_i as a behavioral weakness making consumers susceptible to marketing and credit offers they may later regret.

Note that temptation is persistent (γ_i is consumer-specific), whereas commitment utility for normal goods is random. This asymmetry prevents discrimination by normal-good sellers,

[8]Empirically, DeHaan et al. (2024) show that BNPL users face significantly higher overdraft charges, credit card interest, and late fees. Industry reports suggest similar risks: four in ten BNPL users report late payments, and most suffer credit-score declines (https://www.icba.org/newsroom/blogs/main-street-matters/2022/07/11/bnpl-and-the-illogical-argument-for-credit-card-interchange-fee-regulation).

even with complete information, but allows predatory sellers to target and price-discriminate effectively. This structure lets us focus on how data access amplifies temptation-driven exploitation of weak-willed consumers, rather than on normal-good price discrimination, which is well studied in prior work.

Menu preference The menu $\mathcal{M}$ that a consumer faces is determined by the advertisements she receives from sellers. A menu may contain none, one, or multiple normal and/or predatory goods. Each consumer has additive utilities from normal and predatory good consumption and may choose any combination of goods. We denote the menu for a normal product n as $\mathcal{M}_i^n \in \{\{n, \emptyset\}, \emptyset\}$, where $\emptyset$ means the good is not advertised and $\{n, \emptyset\}$ means it is. Similarly, $\mathcal{M}_i^j \in \{\{j, \emptyset\}, \emptyset\}$ is the menu for a predatory good j.

Given the framework in Equation (1), the choices of a consumer of type $\tau_i \in \{S_n, W_{nj}\}$ from menus $\mathcal{M}_i^{n'}$ and $\mathcal{M}_i^{j'}$ are:

$$x_{\tau_i}\left(\mathcal{M}_i^{n'}\right) = \arg\max_{x \in \mathcal{M}_i^{n'}} \left[\tilde{u}_{\tau_i}(x) - p_i(x)\right], \tag{4}$$

$$z_{\tau_i}\left(\mathcal{M}_i^{j'}\right) = \arg\max_{z \in \mathcal{M}_i^{j'}} \left[u_{\tau_i}(z) + v_{\tau_i}(z) - p_i(z)\right], \tag{5}$$

where $p_i(x)$ and $p_i(z)$ may reflect price discrimination, depending on whether the seller knows the consumer's type. Consumers take sellers' advertising and pricing policies as given.

The consumer's ex ante preference for the full menu is then

$$\begin{aligned} & U_{\tau_i}\left(\{\mathcal{M}_i^n\}_{n=1}^N, \{\mathcal{M}_i^j\}_{j=N+1}^{N+J}\right) \\ = & \sum_{n=1}^N \tilde{u}_{\tau_i}\left(x_{\tau_i}\left(\mathcal{M}_i^n\right)\right) - p_i\left(i, x_{\tau_i}\left(\mathcal{M}_i^n\right)\right) \\ & + \sum_{j=N+1}^{N+J} u_{\tau_i}\left(z_{\tau_i}\left(\mathcal{M}_i^j\right)\right) + v_{\tau_i}\left(z_{\tau_i}\left(\mathcal{M}_i^j\right)\right) - p_i\left(i, z_{\tau_i}\left(\mathcal{M}_i^j\right)\right) - \max_{z' \in \mathcal{M}_i^j} v_{\tau_i}(z'). \end{aligned} \tag{6}$$

This formulation allows us to study how data sharing shapes consumer menus via sellers' advertising.

Our use of temptation utility to model self-control problems parallels an alternative approach based on present bias (Laibson 1997; O'Donoghue and Rabin 1999; DellaVigna and Malmendier 2004). Hyperbolic discounting leads consumers to overweight the immediate gratification from predatory goods while undervaluing future repayment costs. Sophisticated consumers, anticipating this conflict, may prefer to keep predatory goods off their menus. Benabou and Pycia (2002) show that temptation utility is equivalent to a multiple-selves formulation, making it consistent with sophisticated present-biased consumers. We adopt temptation utility for its simplicity, as it captures menu preference without requiring a dynamic setup, even if the present-bias approach offers sharper positive predictions.

In our model, weak-willed but sophisticated consumers make fully rational data-sharing decisions, despite their lack of self-control in consumption. This provides a solid foundation for normative analysis of data-sharing schemes and privacy regulation (e.g., Attanasio and Weber 2010). We abstract from naïve present-biased consumers, who are unaware of their self-control problem and therefore fail to use opt-out options to protect themselves. While sophisticated consumers' choices may affect the welfare of naïve ones, incorporating the latter group poses a conceptual challenge: doing so requires paternalistic welfare judgments that conflict with their expressed preferences.

1.2 Sellers

There are $N + J$ sellers on the platform, one representative seller n for each normal good and one seller j for each predatory good. For simplicity, we assume each seller faces zero marginal production cost but incurs a convex advertising cost of $-(1-\phi)\, c \log(1-y)$ to reach a measure y of consumers, where $(1-\phi)\, c > 0$. As in Grossman and Shapiro (1984), the convexity reflects the increasing difficulty of reaching a broader audience.[9] Limited consumer attention implies advertisers avoid flooding users with ads, requiring progressively higher fees to expand reach.[10]

Normal-good sellers aim to target both strong- and weak-willed consumers who prefer their good, while predatory-good sellers target only weak-willed consumers tempted by their product. Their ability to discriminate depends on the platform's data-sharing scheme and

[9]This cost can be microfounded as follows. Similar to Grossman and Shapiro (1984), suppose each advertisement is seen by a consumer with probability η. If a seller sends out q ads, a consumer sees them with probability $1-(1-\eta)^q$, so by the Weak LLN, exactly a fraction $y_k = 1-(1-\eta)^q$ of consumers see the ad. Let $\xi = -\log(1-\eta)$. To reach y_k measure of consumers, the seller must buy $q = -(1/\xi)\log(1-y_k)$ ads. If the platform charges f per ad, the cost is $fq = -(f/\xi)\log(1-y_k) = -(1-\phi)\, c \log(1-y_k)$, giving an effective cost parameter $(1-\phi)\, c$.

[10]See Chen (2022) for a model of targeted advertising under limited attention and Roussanov et al. (2021) for evidence of its importance in mutual fund selection.

consumers' data-sharing choices. Sellers provide BNPL through the affiliated fintech lender by charging a discount equal to a fraction $\phi \in (0,1)$ of the purchase price.

Because consumers are risk-neutral and ultimately receive seller profits, sellers maximize expected profit. Seller $k \in \{1, ..., N+J\}$ chooses its advertising set Y_k and consumer-specific prices $p_i(k)_{i \in Y_k}$ to solve:

$$\Pi_k = \sup_{\{p_i(k), Y_k\}} (1-\phi)\, \mathbb{E}\left[\int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di + c \log(1-y_k) \,\middle|\, \mathcal{I}^k \right], \tag{7}$$

where y_k is the measure of consumers in Y_k, $\mathbf{1}_{\{x(i)=k\}}$ indicates whether consumer i uses good k, and $\mathcal{I}^k$ is seller k's information set. If a consumer is not advertised to, the good does not appear on her menu.

Sellers are strategic but lack commitment: they can condition advertising and pricing only on their information sets and must find it optimal to carry out their announced strategies after consumers choose data sharing. This will matter under the opt-in/opt-out arrangement, as it rules out self-confirming equilibria supported only by off-equilibrium beliefs (as in Perfect Bayesian equilibria).

Sellers face participation constraints:

$$p_i(n) \leq \bar{u}, \;\; p_i(j) \leq \bar{v}, \tag{8}$$

since charging above these thresholds would result in no sales.

1.3 Fintech Lender

There is a BNPL fintech lender affiliated with the digital platform, similar to Klarna or Afterpay. It provides short-term loans to consumers to facilitate purchases of both normal and predatory goods. Consistent with practice, the BNPL lender assumes all borrower default risk on the platform and is compensated through a merchant fee: the merchant remits a discounted amount of the purchase price to the lender (OCC, 2023).

The lender charges sellers a merchant fee equal to a fraction ϕ of the consumer's purchase price. In practice, BNPL lenders typically conduct only a soft or no credit check before issuing loans for online purchases. Their primary source of revenue is merchant fees, which scale with purchase size, supplemented by late fees and interest on unpaid balances. If a consumer defaults—which, in our model, occurs only for predatory goods purchases—with probability δ, the lender recovers only a fraction $1-\xi$ of the loan balance. This recovery reflects late fees, partial repayments, and proceeds from delinquent loans to collection agencies. By the Weak Law of Large Numbers, exactly a fraction δ of weak-willed consumers will default.

The fintech lender is ultimately owned by households and faces a cost of capital R for its lending activities. Its expected profits are:

$$\Pi_F = (1 - R(1-\phi))\, \mathbb{E}\left[\sum_{k=1}^{N+J} \int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di\right] - \delta\xi(1-\phi)\, \mathbb{E}\left[\sum_{k=N+1}^{N+J} \int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di\right], \tag{9}$$

which equals the effective fee revenue from all purchases minus expected losses from default. The lender is willing to provide loans on the platform as long as

$$\phi \geq 1 - \frac{\mathbb{E}\left[\sum_{k=1}^{N+J} \int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di\right]}{R\mathbb{E}\left[\sum_{k=1}^{N+J} \int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di\right] + \delta\xi\mathbb{E}\left[\sum_{k=N+1}^{N+J} \int_{i \in Y_k} p_i(k)\, \mathbf{1}_{\{x(i)=k\}} di\right]}.$$

In what follows, we assume $\phi \geq 1 - \frac{1}{R+\delta\xi}$, the fee that would be required if all purchases carried default risk, so that this participation constraint is always satisfied. In practice, BNPL merchant fees are roughly double those charged by credit cards for equivalent services, and late fees can reach up to 25% of the transaction value, which explains BNPL lenders' greater tolerance for consumer default.[11] As such, the lender's incentives are aligned with sellers in promoting larger consumer purchases, even when default risk is significant. Although we focus data sharing between platform and sellers for convenience, BNPL lenders in practice also collect consumer data to maximize incremental sales and the lifetime revenue that they can extract through tailored product offerings, marketing campaigns, and product experiences, and to induces some consumers to borrow more than they can repay (CFPB, 2022).

[11]See https://www.icba.org/newsroom/blogs/main-street-matters/2022/07/11/bnpl-and-the-illogical-argument-for-credit-card-interchange-fee-regulation.

1.4 Sequential Rational Expectations Equilibrium

We examine the effects of different data-sharing schemes on consumers and sellers, abstracting from the platform's own incentives. Implicitly, we assume that the platform shares all consumer data with sellers, subject to each consumer's sharing preferences. In Section 2, we analyze two benchmark data-sharing regimes: (i) no sharing and (ii) full sharing. In both cases, consumers have no individual choice over whether their data is shared. In Section 2.3, we consider a GDPR-inspired scheme that allows each consumer to decide whether to share their data with the platform, which then passes on the authorized data to sellers.

Under each of these data-sharing schemes, an equilibrium consists of:

- Consumer optimization: Given each seller's advertising and pricing policies, each consumer i chooses an optimal data-sharing choice s_i and then follows a purchasing policy $\left\{\{x_\tau\left(\mathcal{M}_i^n\right)\}_{n=1}^{N}, \left\{y_\tau\left(\mathcal{M}_i^j\right)\right\}_{j=N+1}^{N+J}\right\}$ over a menu set $\{\mathcal{M}_i^n, \mathcal{M}_i^j\}$.
- Seller optimization: Given consumers' optimal choices, each seller k selects an advertising policy Y_k and a pricing policy $p_i\left(k\right)_{i\in Y_k}$ for its good. These must be sequentially rational when consumers authorize data sharing.

For welfare analysis, we assume sellers pay the platform for advertising services. Hence advertising costs and fintech merchant fees are treated as zero-sum transfers between sellers, the platform, and the fintech lender—all ultimately owned by consumers. Because preferences are quasi-linear in expenditures, we can aggregate consumer utilities and seller, lender, and platform profits to obtain utilitarian welfare:

$$\begin{aligned} W \;=\; & \frac{1}{N}\sum_{n=1}^{N}\int \tilde{u}_n\left(\pi_S \mathbf{1}_{\left\{n\in\mathcal{M}_{S_n}^n\,\cap\, x_{S_n}=n\right\}} + \pi_W \mathbf{1}_{\left\{n\in\mathcal{M}_{W_n}^n\,\cap\, x_{Wn}=n\right\}}\right) dH\left(\tilde{u}_n\right) \\ & +\frac{\pi_W}{J}\sum_{j=N+1}^{N+J}\int\left(u_B \mathbf{1}_{\left\{j\in\mathcal{M}_{W_j}^j\,\cap\, x_{W_j}=j\right\}} + \left(u_B-\gamma_i\bar{v}\right)\mathbf{1}_{\left\{j\in\mathcal{M}_{W_j}^j\,\cap\, x_{W_j}=\emptyset\right\}}\right) dG\left(\gamma_i\right), \end{aligned} \tag{10}$$

The first term reflects the commitment utility of both strong- and weak-willed consumers from consuming normal goods. The second captures the social impact of predatory goods: weak-willed consumers who succumb incur negative commitment utility u_B, while those who resist suffer the mental cost $u_B - \gamma_i\bar{v}$.

As noted in Equation (1), when a weak-willed consumer buys a predatory good, temptation utility is fully offset, leaving zero net temptation utility. Interest payments do not affect social welfare because they are transfers to sellers. Welfare losses instead arise from (i) the negative commitment utility u_B when weak-willed consumers borrow to buy predatory goods, and (ii) the disutility of resisting temptation when such goods appear on menus but are not consumed.

Equation (10) highlights a central trade-off in data sharing. On one hand, sharing improves matching for normal goods, raising welfare; on the other, it exposes weak-willed consumers to predatory goods, reducing welfare. Unlike standard data privacy models, which focus on how data increases total surplus through better matching and redistributes it between consumers and sellers, our framework shows that weak-willed consumers value privacy not to secure better prices but to avoid exploitation and the harms of overspending—amplified by BNPL-style credit that fosters indebtedness.

Our model thus underscores that the costs and benefits of data sharing are distributed unevenly. Strong-willed consumers gain from better access to normal goods, while weak-willed consumers face heightened risks due to behavioral vulnerabilities. A representative-agent cost function misses this heterogeneity and, therefore, the critical role of privacy regulation in addressing algorithmic inequality.[12]

We measure this inequality as the welfare gap Δ, defined as the difference between the average utility of strong- and weak-willed consumers after accounting for the costs of purchasing normal and predatory goods:

$$\begin{aligned} \Delta \;=\; & \frac{1}{N}\sum_{n=1}^{N}\int\left(\tilde{u}_n - p_{\tilde{u}_n}(n)\right)\left(\pi_S \mathbf{1}_{\left\{n\in\mathcal{M}^n_{S_n}\,\cap\, x_{S_n}=n\right\}} - \pi_W \mathbf{1}_{\left\{n\in\mathcal{M}^n_{W_n}\,\cap\, x_{Wn}=n\right\}}\right) dH\left(\tilde{u}_n\right) \\ & -\frac{\pi_W}{J}\sum_{j=N+1}^{N+J}\int\left(\left(u_B - p_i(j)\right)\mathbf{1}_{\left\{j\in\mathcal{M}^j_{W_{nj}}\,\cap\, x_{Wnj}=j\right\}} + \left(u_B - \gamma_i\bar{v}\right)\mathbf{1}_{\left\{j\in\mathcal{M}^j_{W_{nj}}\,\cap\, x_{Wnj}=\emptyset\right\}}\right) dG\left(\gamma_i\right), \end{aligned} \tag{11}$$

where $p_{\tilde{u}_n}(n)$ indicates that for normal good n, consumers type is the random utility draw $\tilde{u}_n$.

To anchor the welfare analysis of different data-sharing schemes, we first consider the

[12] Related work, such as Ali and Benabou (2020) and Jann and Schottmüller (2020), emphasizes privacy as protection against social discrimination: public observability of actions allows inferences about private traits, which fosters conformity at the expense of individual preferences. Tirole (2021) further warns that absent privacy protections, governments could use social rating systems—linking political stances with social networks—to control society without overt repression. By contrast, in our setting, vulnerable consumers value privacy primarily as a shield against exploitation of their behavioral weaknesses rather than against discrimination.

planner's problem of maximizing social welfare in (10). Since advertising entails no social cost, the planner would want sellers of normal goods to advertise to all consumers, strong- and weak-willed alike. In contrast, predatory advertisements impose costs on weak-willed consumers, whether they succumb to temptation or resist it. Thus, the planner prefers that predatory sellers do not advertise to anyone. In this ideal outcome, there is no algorithmic inequality.[13]

Proposition 1 *In the first-best outcome, sellers of normal goods advertise to all consumers, while sellers of predatory goods advertise to none.*

The first-best outcome might suggest banning predatory financial products altogether. However, such products are often intertwined with desirable goods and services. For example, Larrimore et al. (2024) show that BNPL is used not only by financially vulnerable consumers prone to overextension but also by more financially secure consumers who use it to smooth payments and avoid interest charges. A product like BNPL may therefore be welfare-improving for some consumers but harmful for others, depending on what it finances.[14] These overlaps highlight why defining "predatory lending" in legal or regulatory terms is difficult, and why banning either use or advertising may be infeasible.[15]

2 Data-Sharing Equilibrium

In this section, we analyze the data-sharing equilibrium. To frame the main scheme in which consumers can opt in or out of sharing, we first consider two benchmarks: no sharing and full sharing. We then evaluate consumer welfare across these schemes and provide a calibrated analysis using realistic parameters. Proofs of the key propositions appear in the Appendix, with additional proofs in the Internet Appendix.

[13] If instead $u_B > 0$, predatory sellers would advertise to all weak-willed consumers in the first best. This contrasts with standard price-discrimination models, where reducing search frictions through data sharing always improves social surplus. Here, data sharing improves matching for normal goods but can reduce welfare when it facilitates exposure to predatory goods.

[14] Our model can be adapted so that each predatory good is a variant of a normal good—tempting only for specific weak-willed consumers, leading to negative commitment utility for them. While this complicates consumer inference, the core mechanism is unchanged.

[15] An alternative policy is to impose strict borrowing limits to protect vulnerable consumers. Yet BNPL schemes are deliberately designed to bypass traditional safeguards, limiting recourse for borrowers and using alternative data to extend credit to consumers who would not qualify under conventional underwriting.

2.1 Consumer Choice

We begin by analyzing individual consumer choices from a given menu of goods. A strong-willed consumer who prefers normal good n will purchase it if the seller's price is below the consumer's reservation value, and always reject other goods. A weak-willed consumer who prefers normal good n but also desires predatory good j behaves similarly for n: they buy if the price is below their reservation value. In addition, they may buy predatory good j if their temptation coefficient γ_i is sufficiently high relative to the price. The following proposition summarizes these choices.

Proposition 2 *A strong-willed consumer with commitment utility $\tilde{u}_n$ for normal good n purchases it if $p_{\tilde{u}_n}(n) \leq \tilde{u}_n$, and rejects all other products. A weak-willed consumer with commitment utility $\tilde{u}_n$ for normal good n and temptation coefficient γ_i for predatory good j purchases good n if $p_{\tilde{u}_n}(n) \leq \tilde{u}_n$, and purchases product j if $\gamma_i \geq \frac{p_i(j)}{\bar{v}}$.*

This proposition shows that both strong- and weak-willed consumers might reject good n if their random utility draw falls below the posted price. Consequently, seller of good n cannot perfectly price discriminate.[16] As a result, all strong- and weak-willed consumers prefer to receive advertisements for good n, as this allows them to benefit when their realized utility is high. This creates a motive for consumers to share their data with the platform.

2.2 Two Benchmarks

To analyze the equilibrium with individual data-sharing choices, we first consider two benchmark schemes: one with no data sharing and one with full data sharing.

In the no-sharing (NS) benchmark, the platform neither collects nor shares consumer data. Sellers thus lack information about consumer types and face a "dark pool" for advertising, with each advertisement reaching its target consumers only with the unconditional probability. This setting reflects market practices prior to the era of big data. The equilibrium is characterized as follows:

[16] Much of the literature on data sharing emphasizes its role in enabling price discrimination. For example, Taylor (2004) and Acquisti and Varian (2005) show that using past purchase data is optimal for sellers when consumers are naive about how information is used, but not when they are sophisticated. Ali, Lewis, and Vasserman (2019) find that disclosure choices can intensify competition and lower prices, while Ichihashi (2020) shows that multi-product sellers may prefer not to use data for pricing, allowing consumers to reveal preferences and improve matching. A recent review is provided by Goldfarb and Tucker (2019). Our analysis departs from this literature by focusing on data privacy as protection against consumer vulnerabilities rather than as a constraint on price discrimination.

Proposition 3 *With no data sharing (NS), there exists a unique equilibrium with the following properties:*

1. *Normal good seller n randomly advertises to a mass*

$$y_n^{NS} = \max\left\{1 - 4N\frac{c}{\bar{u}}, 0\right\}, \tag{12}$$

at a uniform price: $p^{NS}(n) = \frac{1}{2}\bar{u}$.

2. *Predatory good seller j randomly advertises to a mass*

$$y_j^{NS} = \max\left\{1 - 4\frac{J}{\pi_W}\frac{c}{\bar{v}}, 0\right\}, \tag{13}$$

at a uniform price: $p^{NS}(n) = \frac{1}{2}\bar{v}$.

In this regime, undirected advertising generates inefficiency. Normal good sellers restrict advertising to only a fraction of potential buyers. As Equation (12) shows, seller n's advertising intensity y_n falls with the cost parameter c, and rises with $\frac{1}{N}$ (the share of intended consumers in the population) and $\bar{u}$ (which determines the good's price). Since sellers fail to advertise to all consumers who value the good—contrary to the first-best—matching is inefficient, motivating demand for greater data sharing.

At the same time, Equation (13) shows that anonymity discourages predatory sellers from targeting all weak-willed consumers, which creates a source of welfare gain. Lacking information about reservation values, both normal and predatory sellers set uniform prices: $p^{NS}(n) = \frac{1}{2}\bar{u}$ and $p^{NS}(j) = \frac{1}{2}\bar{v}$, which means their advertisements are accepted by intended consumers only half of the time.

The second benchmark represents the opposite extreme, where the platform collects consumer data and can fully infer each consumer's type. This setting mirrors the practices of many digital platforms prior to the introduction of privacy regulations. For simplicity, we assume the platform can identify whether a consumer is strong- or weak-willed and, in the latter case, observe the consumer's temptation coefficient γ_i. While this may overstate current capabilities, rapid advances in data analytics make this an informative limiting case. By sharing data with sellers, the platform enables them to tailor advertising and pricing strategies to specific consumer types, particularly vulnerable ones (Nadler and McGuigan

2018; Stigler Committee Report 2019).[17]

Since strong- and weak-willed consumers behave identically with respect to normal good n, seller n need not differentiate them. Proposition 4 characterizes the measure of consumers targeted, y_n^{FS}, and the equilibrium price $p^{FS}(n)$. Data sharing improves efficiency by allowing seller n to avoid advertising to consumers who would never purchase, so $y_n^{FS} \geq y_n^{NS}$. Thus, both strong- and weak-willed consumers face a higher probability of being reached. As seller n still cannot observe reservation values, it charges a uniform price$p^{FS}(n) = \frac{1}{2}\bar{u}$.

By contrast, access to consumer data enables predatory seller j to perfectly target weak-willed consumers and price discriminate based on temptation severity. Each consumer is charged their full reservation value, $p_{\gamma_i}(j) = \gamma_i \bar{v}$. This motivates the seller to concentrate its advertising on the most tempted consumers, i.e., those with $\gamma_i \geq \hat{\gamma}^{FS}$. As a result, full data sharing leads to both precise targeting and perfect price discrimination against vulnerable consumers.

Proposition 4 *With full data sharing (FS), there exists a unique equilibrium with the following properties:*

1. *Normal good seller n advertises to y_n^{FS} measure of strong- and weak-willed consumers that desire the good:*
$$y_n^{FS} = \min\left\{\max\left\{1 - 4\frac{c}{\bar{u}}, 0\right\}, \frac{1}{N}\right\} \tag{14}$$
and charges a uniform price $p^{FS}(n) = \frac{1}{2}\bar{u}$.

2. *Predatory good seller j advertises to all weak-willed consumers that desire the good with $\gamma_i \geq \hat{\gamma}^{FS} = 1 - \frac{J}{\pi_W} y_j^{FS}(\bar{v})$, where $y_j^{FS}(\bar{v})$ is the total advertising by seller j:*
$$y_j^{FS}(\bar{v}) = \begin{cases} \frac{1+\pi_W/J}{2} - \sqrt{\left(\frac{1-\pi_W/J}{2}\right)^2 + \frac{\pi_W c}{J\bar{v}}} & \text{if } \bar{v} > c \\ 0 & \text{if } \bar{v} \leq c \end{cases}, \tag{15}$$
and sets consumer-specific prices $p_{\gamma_i}(j) = \gamma_i \bar{v}$.

[17] Our model abstracts from the full complexity of how platforms target vulnerabilities in practice. In reality, platforms use real-time analytics to infer not only cognitive and affective states but also the timing when consumers are most susceptible. Some strategies even deplete willpower—through aggressive advertising or continuous engagement—thus proactively inducing vulnerability. Social media platforms, for example, customize nudges and content to maximize engagement, occasionally bordering on addictive behavior (Allcott et al. 2020).

Full data sharing clearly benefits strong-willed consumers by improving access to normal goods. For weak-willed consumers, however, it creates a trade-off: better access to normal goods but increased exposure to predatory goods. This makes the welfare comparison between no sharing and full sharing non-trivial. As Proposition 6 shows, when weak-willed consumers' vulnerability is sufficiently severe (i.e., u_B is sufficiently negative), full data sharing reduces overall social welfare relative to no sharing.

Policymakers have long worried about these risks. A 2013 U.S. Senate review of the data broker industry revealed that firms were selling products explicitly designed to identify financially vulnerable consumers, exposing them to high-cost loans and other risky financial products.[18] Such risks have only intensified on digital platforms, whose access to consumer data is unprecedented.

2.3 Opt-in/Opt-out

We now analyze the main data-sharing scheme, modeled after the European Union's General Data Privacy Regulation (GDPR) and the California Consumer Privacy Act (CCPA). These regulations safeguard consumer privacy by granting individuals the right to decide whether to share their data with digital platforms,[19] thus offering the potential to achieve a Pareto efficient outcome as each consumer can make the most suitable choice for herself. Under such a scheme, which we refer to as GDPR, strong-willed consumers can opt-in and benefit from data sharing, while severely tempted consumers can opt-out to avoid exposure to predatory goods. [20]

Strong-willed consumers benefit from sharing their data with normal goods sellers and have no concern about predatory goods sellers. Hence, they all opt in. Weak-willed consumers, however, face a trade-off: opting in improves access to normal goods but also in-

[18]The report specifically states on page ii: "*One company reviewed sells a marketing tool that helps to "identify and more effectively market to under-banked consumers" that the company describes as individuals including "widows" and "consumers with transitory lifestyles, such as military personnel" who annually spend millions on payday loans and other "non-traditional" financial products. The names, descriptions and characterizations in such products likely appeal to companies that sell high-cost loans and other financially risky products to iii populations more likely to need quick cash, and the sale and use of these consumer profiles merits close review.*" See https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577 for the report.

[19]Similar laws—such as Virginia's Consumer Data Protection Act (VCDPA), Colorado's Privacy Act (CPA), and Utah's Consumer Privacy Act (UCPA)—follow the same framework.

[20]These regulations also have measurable effects in practice. Goldberg et al. (2019), for example, find that spending and visits by EU consumers fell by up to 7% in 2018 relative to 2017 following GDPR implementation, with the decline especially pronounced for smaller sellers.

creases exposure to predatory goods. As a result, severely tempted consumers tend to opt out, while less tempted consumers opt in. To simplify notation, we use *in* subscripts for consumers in the opt-in pool and *out* for those in the opt-out pool. For example, $p_{in}^{GDPR}(n)$ denotes the price changed by normal good seller n to opt-in consumers, while $p_{out}^{GDPR}(j)$ denotes the price charged by temptation good seller j to the opt-out pool.

Opting out does not fully shield vulnerable consumers. The degree of protection depends on the diversity of the opt-out pool, which creates "noise" that dilutes targeting by predatory sellers. Thus, each weak-willed consumer's decision to opt in or out depends on others' choices. We conjecture the existence of a cutoff temptation level γ^*: consumers with with $\gamma_i \geq \gamma^*$ opt out, while those with $\gamma_i < \gamma^*$ opt in. This threshold is symmetric across all temptation types.

Conditional on this cutoff γ^*, the utility of a weak-willed consumer with γ_i who opts in is:

$$\begin{aligned} U_{W,in}^{GDPR}(\gamma_i) &= \frac{y_{n,in}^{GDPR} N}{\pi_S + \pi_W \gamma^*} \int_0^{\bar{u}} \max\left\{\tilde{u}_{nt} - p_{in}^{GDPR}(n), 0\right\} dH(\tilde{u}_n) \\ &+ \frac{y_{j,in}^{GDPR}(\gamma_i) J}{\pi_W \gamma^*} \left(u_B - p_{\gamma_i,in}^{GDPR}(j) \mathbf{1}_{\left\{\gamma_i \geq \frac{p_{\gamma_i,in}^{GDPR}(j)}{\bar{v}}\right\}} - \gamma_i \bar{v} \mathbf{1}_{\left\{\gamma_i < \frac{p_{\gamma_i,in}^{GDPR}(j)}{\bar{v}}\right\}}\right), \end{aligned} \tag{16}$$

where $y_{n,in}^{GDPR}$ is the total advertising of seller n to the opt-in pool at price $p_{in}^{GDPR}(n)$, while $y_{j,in}^{GDPR}(\gamma_i)$ is the advertising intensity of seller j targeting opt-in consumers with temptation γ_i at price $p_{\gamma_i,in}^{GDPR}(j)$. The consumer's utility reflects the conditional probability of being targeted by both sellers, $\frac{y_{nin}^{GDPR}}{(\pi_S+\pi_W\gamma^*)/N}$ and $\frac{y_{j,in}^{GDPR}(\gamma_i)}{\pi_W/J}$.

By contrast, the utility of opting out is:

$$\begin{aligned} U_{W,out}^{GDPR}(\gamma_i) &= \frac{y_{n,out}^{GDPR}}{\pi_W(1-\gamma^*)} \int_0^{\bar{u}} \max\left\{\tilde{u}_n - p_{out}^{GDPR}(n), 0\right\} dH(\tilde{u}_n) \\ &+ \frac{y_{j,out}^{GDPR}}{\pi_W(1-\gamma^*)} \left(u_B - p_{out}^{GDPR}(j) \mathbf{1}_{\left\{\gamma_i \geq \frac{R_{out}^{GDPR}(j)}{\bar{v}}\right\}} - \gamma_i \bar{v} \mathbf{1}_{\left\{\gamma_i < \frac{p_{out}^{GDPR}(j)}{\bar{v}}\right\}}\right), \end{aligned} \tag{17}$$

where $y_{n,out}^{GDPR}$ is the total advertising by normal good seller n to the opt-out pool at price $p_{out}^{GDPR}(n)$, and $y_{j,out}^{GDPR}$ is the total advertising by predatory good seller j to the opt-out pool at price $p_{out}^{GDPR}(j)$.

A weak-willed consumer opts in if and only if

$$U_{W,in}^{GDPR}(\gamma_i) \geq U_{W,out}^{GDPR}(\gamma_i), \tag{18}$$

with equality holding for the marginal consumer with $\gamma_i = \gamma^*$ who is indifferent between the two choices.

The following proposition characterizes the equilibrium under the GDPR. Define

$$\underline{\gamma} \equiv \frac{1}{2}\left(1 - \sqrt{1 - \frac{1}{J}}\right), \tag{19}$$

the lowest feasible cutoff value for γ^*. If γ^* were below $\underline{\gamma}$, predatory sellers would not target the opt-in pool, rendering such a cutoff invalid.

Proposition 5 *There exists an equilibrium under the GDPR with the following properties:*

1. ***Consumer choice.*** *All strong-willed consumers opt in. A weak-willed consumer opts in if $\gamma_i < \gamma^*$ and opts out if $\gamma_i \geq \gamma^*$.*

2. ***Normal goods.*** *Seller n charges the same price in both opt-in and opt-out pools: $p_{in}^{GDPR}(n) = p_{out}^{GDPR}(n) = \frac{1}{2}\bar{u}$, and follows a water-filling advertising strategy that prioritizes the opt-in pool:*

$$y_{n,in}^{GDPR} = \min\left\{1 - 4\frac{c}{\bar{u}}, \frac{1 - \pi_W(1-\gamma^*)}{N}\right\}, \tag{20}$$

$$y_{n,out}^{GDPR} = \min\{\max\{\pi_W(1-\gamma^*) - 4N\frac{c}{\bar{u}}, 0\}, \pi_W(1-\gamma^*)\}. \tag{21}$$

3. ***Predatory goods.*** *Seller j also uses a water-filling strategy. If $\gamma^* > \underline{\gamma}$, it prioritizes the opt-in pool, targeting a measure $y_{j,in}^{GDPR}$ (Equation (A.56)) of the more-tempted consumers and charging their reservation utility: $p_{\gamma_i,in}^{GDPR}(j) = \gamma_i \bar{v}$. It may then extend advertising to a measure $y_{j,out}^{GDPR}$ (Equation (A.57)) of the opt-out pool at a fixed price: $p_{out}^{GDPR}(j) = \max\left\{\frac{1}{2}, \gamma^*\right\}\bar{v}$.*[21]

[21] In the knife-edge case that $\gamma^* = \underline{\gamma}$, it prioritizes the opt-out pool by targeting a measure $y_{j,out}^{GDPR}$, given in Equation (A.54), of the consumers in the opt-out pool by charging $p_{out}^{GDPR}(j) = \frac{1}{2}\bar{v}$, and it may also target an additional measure $y_{j,in}^{GDPR}$, given in Equation (A.55), of the more-tempted consumers in the opt-in pool and charging them: $p_{\gamma_i,in}^{GDPR}(j) = \gamma_i \bar{v}$.

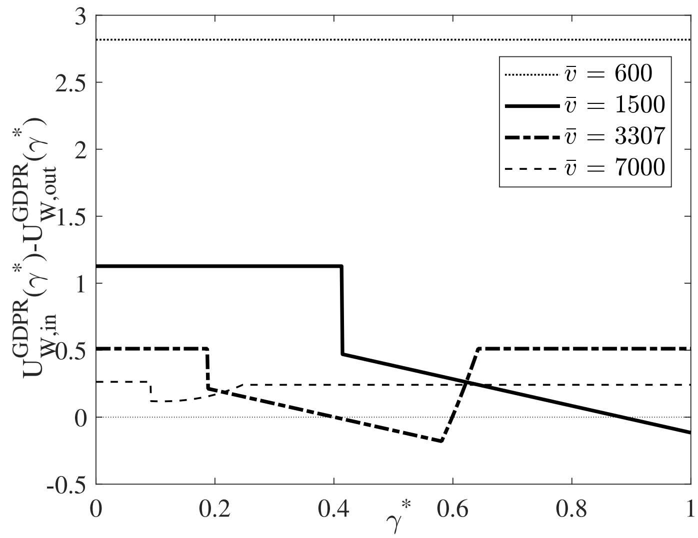

Figure 1: Illustration of the relative benefit for the marginal weak-willed consumer with temptation index γ^* to opt-in, $U^{GDPR}_{W,in}(\gamma^*) - U^{GDPR}_{W,out}(\gamma^*)$, for four values of $\bar{v}$. The parameters are $N = 10, J = 3, \pi_W = 0.25$, and the rest are listed in Table 1.

4. ***Equilibrium cutoff.*** *The cutoff $\gamma^* \geq \underline{\gamma}$ has the following properties:*

   (a) If $\bar{v} < \frac{\bar{u}}{8} \min\left\{\left(1 - \frac{4c}{\bar{u}}\right) N, 1\right\} + u_B$ or $\bar{v} > \left(1 - \frac{\pi_W}{J} + \frac{\pi_W}{J^2}\right)^{-1} Jc$, there is a full data-sharing equilibrium in which all weak-willed consumers opt in ($\gamma^* = 1$).

   (b) For $\bar{v} < v_{***}$ (Equation (A.71)) and $c > \frac{\bar{u}}{N}\pi_W \frac{1-\underline{\gamma}}{4}$, there exists an interior cutoff $\gamma^* \in \left(\underline{\gamma}, 1\right)$, solving Equation (A.69). If, in addition, $\bar{v} > \left(1 - \frac{\pi_W}{J} + \frac{\pi_W}{J^2}\right)^{-1} Jc$, multiple equilibria exist, including the full data-sharing equilibrium $\gamma^* = 1$.

Proposition 5 shows that weak-willed consumers follow a cutoff strategy in deciding whether to share their data. Predatory sellers exploit this by perfectly targeting and price discriminating against the more tempted opt-in consumers, while only imperfectly targeting the opt-out pool. The equilibrium cutoff γ^* is therefore shaped by the relative incentives of predatory sellers to search across the two pools, as detailed in part 4.

Figure 1 illustrates the equilibrium by plotting the equilibrium cutoff γ^* against the net benefit for the marginal consumer to opt-in, $U^{GDPR}_{W,in}(\gamma^*) - U^{GDPR}_{W,out}(\gamma^*)$, for different values of $\bar{v}$. An interior equilibrium arises when this difference equals zero. A full data-sharing equilibrium ($\gamma^* = 1$) occurs when the difference is positive at $\gamma^* = 1$, while a minimum data-sharing equilibrium ($\gamma^* = \underline{\gamma}$) occurs when this difference is negative at $\gamma^* = \underline{\gamma}$. The sustainability of full data sharing depends on whether predatory sellers find it optimal to search the opt-out pool, which in turn requires those opting out to be sufficiently tempted.

- **Small $\bar{v}$.** When $\bar{v}$ is low (e.g., the top dotted line in Figure 1 with $\bar{v} = 600$), the cost to weak-willed consumers of being targeted by predatory sellers is small and outweighed by the benefits of sharing data with normal goods sellers. Consequently, all consumers opt in, and predatory sellers target only the most tempted consumers in the opt-in pool.

- **Moderate $\bar{v}$.** When $\bar{v}$ is moderate (e.g., the solid black line in Figure 1 with $\bar{v} = 1500$), a unique interior equilibrium emerges where the curve intersects the x-axis. Here, predatory sellers prioritize advertising to the opt-in pool but do not fully cover the opt-out pool.

- **Large $\bar{v}$.** When $\bar{v}$ is large (e.g., the thick dot-and-dashed line in Figure 1 with $\bar{v} = 3307$, also used in our calibration exercise in Section 3), multiple equilibria exist: two interior solutions plus full data sharing. This multiplicity arises because the incentives of predatory sellers to search the opt-out pool are non-monotonic in γ^*: profits from targeting the most tempted in the opt-in pool rise with γ^*, but for sufficiently large γ^*, profits from targeting the opt-out pool also rise. The relative benefit curve thus crosses zero twice, producing three equilibria. This non-monotonicity generates a coordination problem among weak-willed consumers: the most tempted may still opt in because they fail to coordinate on one of the two lower interior cutoffs. Moreover, as $\bar{v}$ increases, the minimal and intermediate cutoffs behave differently: fewer weak-willed consumers opt in at the minimal cutoff, but more do so at the intermediate cutoff, reflecting the shape of the relative benefit curve.

- **Very large $\bar{v}$.** When $\bar{v}$ is extremely large (e.g., the dashed line in Figure 1 with $\bar{v} = 7000$), all consumers once again opt in. In this case, weak-willed consumers' temptation is so severe that even if they opt out, predatory sellers still target them in the opt-out pool. Thus, opting in becomes strictly preferable despite their vulnerability.[22]

[22] A subtle but important observation is that our full data-sharing equilibria do not rely on off-equilibrium beliefs that opting out would automatically expose consumers to predatory targeting with probability one, even when it is not in sellers' best interest. Instead, these equilibria arise because, in equilibrium, predatory sellers find it profitable to fully search the opt-out pool whenever any positive mass of weak-willed consumers opts out. Thus, full data sharing is not guaranteed: it emerges only when searching the opt-out pool is sufficiently profitable.

Our analysis highlights a crucial data-sharing externality. To avoid targeting by predatory sellers, the most tempted consumers may opt out and attempt to hide within the opt-out pool. Yet the opt-in decisions of others—such as strong-willed and moderately tempted consumers—undermine this protection. Their exit from the opt-out pool reduces its "camouflage," raising the probability that the remaining weak-willed consumers are targeted by predatory sellers. In this sense, the opt-in choices of strong- and moderately weak-willed consumers impose a negative externality on the most vulnerable, as they do not account for the harm their decisions create.

This mechanism echoes the concept of social data emphasized by Acemoglu et al. (2019), Bergemann, Bonatti, and Gan (2019), and Easley et al. (2019), who show that one individual's data can reveal information about others. In contrast, in our setting it is a consumer's behavior that reveals information about herself to the platform. Thus, the externality we identify is distinct: it stems not from cross-person inference but from how individual choices alter the effectiveness of concealment for others. The presence of this externality suggests that simply granting consumers the ability to opt in or out of data sharing may be insufficient to protect the most vulnerable.

The existence of multiple equilibria with different cutoffs γ^* is a sharp manifestation of this data-sharing externality. When $\bar{v}$ lies in an intermediate range, several cutoffs can be self-consistent with consumers' optimal data-sharing policies, including full data sharing. These multiple equilibria, and the associated coordination challenges, do not arise in models with reduced-form cost functions for data sharing (e.g., Jones and Tonetti, 2020) or in models of data markets (e.g., Acemoglu et al., 2019; Bergemann, Bonatti, and Gan, 2019; Easley et al., 2019).

The uneven impact of the data-sharing externality on strong- and weak-willed consumers underscores the presence of algorithmic inequality on the platform. While data sharing benefits strong-willed consumers, it can harm weak-willed consumers by eroding the camouflage provided by the opt-out pool and thereby intensifying their temptation problem.

Empirical evidence on how consumers make data-sharing choices remains limited because of the scarcity of detailed individual-level data. A central finding is the so-called data privacy paradox: although consumers often express concerns about privacy, they nevertheless share their data with sellers and digital platforms. Survey-based studies—including Gross and Acquisti (2005), Goldfarb and Tucker (2012), and Athey et al. (2017)—document this paradox, which the literature typically attributes to behavioral biases such as present bias (see Acquisti, Brandimarte, and Loewenstein 2020 for a review). Our model provides a novel explanation without invoking such biases: even severely tempted consumers may still choose to share data because the opt-out pool offers insufficient protection, making opting in the better choice.

Chen et al. (2021) survey Alipay users about their privacy concerns when sharing data with third-party mini-programs. Nearly half (49%) cite worries about seductive advertising and temptation consumption as reasons for hesitation. Paradoxically, however, users with stronger privacy concerns authorize data sharing with more mini-programs—a pattern the authors attribute to these users' greater demand for digital services, reflecting the core trade-off emphasized in our model. Moreover, the study finds that privacy-concerned users expand their data sharing over time. While our model is not designed to capture this specific behavior, the coordination problem it highlights—where each consumer's choice depends on the decisions of others—offers a rationale for why data sharing may increase even among privacy-concerned individuals.

Additional evidence shows that data sharing affects both firms and consumers. De Matos and Adjerid (2021), in a field experiment with a major telecommunications provider, find that new data authorizations improve targeting, leading to higher sales, more effective marketing, and stronger contractual lock-ins. Similarly, Aridor et al. (2020) show that GDPR's opt-in requirement reduced the number of observable consumers in the online travel industry by 12.5%, yet those who opted in could be tracked more persistently and targeted more efficiently.

2.4 Welfare Comparison

In this subsection, we compare the welfare consequences of the three data sharing schemes analyzed: no data sharing, full data sharing, and the opt-in/opt-out scheme. Social welfare is given by the aggregate utility of strong- and weak-willed consumers over their purchases (Equation (10)), while welfare gap is defined as the difference in the welfare between strong- and weak-willed consumers (Equation (11)).

Proposition 6 *The social ranking of full data sharing, no data sharing, and the GDPR-style opt-in/opt-out scheme has the following properties:*

- *Full data sharing yields the highest social welfare when the temptation problem is mild, i.e., when u_B is close to zero.*

- *No data sharing yields the highest social welfare when the temptation problem is severe, i.e., when u_B is sufficiently negative.*

- *For intermediate values of u_B, the opt-in/opt-out scheme may deliver the highest social welfare.*

- *Regardless of the ranking, no data sharing delivers the lowest welfare gap.*

Proposition 6 highlights the trade-off at the heart of data sharing: improving matching efficiency for normal goods versus protecting weak-willed consumers from predatory goods. Full data sharing maximizes matching efficiency but provides the least protection; no data sharing does the reverse. The opt-in/opt-out scheme lies between these extremes. Accordingly, full data sharing is optimal when temptation is minor ($u_B \approx 0$), while no data sharing is optimal when temptation is severe ($u_B << 0$). No data sharing also minimizes the welfare gap, since it reduces the harm suffered by weak-willed consumers from predatory goods.

The opt-in/opt-out scheme can dominate in an intermediate range of u_B, balancing the benefits and costs of data sharing. However, it may face coordination problems: if too many weak-willed consumers opt in, the protective value of the opt-out pool erodes. When multiple equilibria exist under the opt-in/opt-out scheme, the equilibrium on which weak-willed consumers coordinate determines the precise region of u_B where this scheme yields the highest welfare. This coordination issue affects the location of the intermediate region but not the overall welfare ranking of the three schemes.

3 A Calibration Exercise

We now present a calibration exercise to examine consumer welfare within the opt-in/opt-out data-sharing scheme. As shown in Proposition 5, the equilibrium cutoff γ^* may take one or three values depending on parameter choices. Our goal is to examine the possible equilibria using parameters calibrated to reflect a realistic economic environment. Given limited data availability on e-commerce and online consumer lending, we rely on aggregate statistics to discipline the model.

Parameter	Value	Target	Data	Model
π_W	0.25			
c	713.5	Advertising Revenue	945.52	945.52
$\bar{u}$	1557.7	Normal Goods Seller Revenue	3894.19	3894.19
$\bar{v}$	3808.3	Predatory Goods Seller Revenue	456.86	456.86
u_B	-432.8	Equilibrium Cutoff γ^*	0.40	0.40

Table 1: This table displays the parameters for our numerical experiment, the data moments to which they are targeted, and the model simulated values of these targets.

3.1 Parameters

We calibrate the model to annual online spending by U.S. consumers. Reflecting data-sharing practices in the U.S. in 2021—where regulation on sellers' use of consumer data was limited—we assume all consumers share their data. The calibrated parameters are reported in 1.

We set the number of normal goods to $N = 10$ and predatory goods to $J = 3$. The fraction of weak-willed consumers is set to $\pi_W = 0.25$, consistent with survey evidence: Ameriks et al. (2007) estimate that 10–30% of respondents exhibit self-control problems, while Toussaert (2018) reports 23-36%.

Three of the four remaining parameters, c, $\bar{u}$,and $\bar{v}$, are pinned down by targeting moments from 2021 U.S. e-commerce data. To align with the unit mass of consumers in our model, we normalize all quantities by the number of U.S. online shoppers, about 273.5 million (see Prose Media, 2025).[23]

- Advertising cost c. We target annual online advertising revenue in 2024, reported as $258.6 billion in IAB's 2025 Internet Advertising Revenue Report. Normalized per online shopper, this yields $945.52.

- Utility from normal goods $\bar{u}$ and maximal temptation $\bar{v}$. We use U.S. annual e-commerce revenue in 2024, reported by the U.S. Census Bureau at $1.19 trillion. Dividing by the number of U.S. online shoppers yields per capita annual spending of $4,351.05. Assuming 25% ($\pi_W$) of this expenditure comes from weak-willed consumers, strong-willed consumers account for $3,263.29 and weak-willed consumers for $1,087.76. We choose $\bar{v}$ so that predatory sellers capture 42% of weak-willed consumers'

[23]See, for instance, https://www.prosemedia.com/blog/ecommerce-shoppingstatistics-2025/.

Equil. Cutoff	Strong-Willed		Weak-Willed		Gap	Utilitarian
	aggregate	per consumer	aggregate	per consumer		
$\gamma^* = 1.0$	1460.3	1947.1	-54.8	-219.1	1515.1	5756.6
$\gamma^* = .60$	1460.3	1947.1	21.5	86.0	1374.3	5201.2
$\gamma^* = .40$	1460.3	1947.1	100.1	400.2	1360.2	4937.4

Table 2: This table displays strong- and weak-willed consumer welfare, their difference, and utilitarian welfare for the full, opt-in/opt-out, and no data-sharing schemes for the parameters listed in Table 1.

spending, consistent with 2019 survey evidence from Slickdeals.net showing that 42% of monthly discretionary spending exceeds budgeted amounts.[24] This implies predatory sellers earn \$456.86 annually from weak-willed consumers. We then set $\bar{u}$ to match revenue from normal goods, equal to \$3,263.29 from strong-willed consumers plus \$630.90 from weak-willed consumers, or \$3,894.19 in total.

- Utility cost of predatory goods u_B. We set u_B so that 15% of consumers opt out when the U.S. moves from full data sharing to the minimal opt-in/opt-out equilibrium (i.e., the equilibrium with the lowest γ^*). This matches the 15% reduction in web traffic observed after the GDPR's introduction, as reported by Congiu et al. (2022). We rely on this indirect measure because privacy preferences are malleable, making direct evidence difficult to interpret. This calibration implies a cutoff value of $\gamma^* = 0.40$ in the minimal data-sharing equilibrium.

With four parameters and four moments, we match the data perfectly.[25]

3.2 Consumer Welfare

Interestingly, under our calibrated parameters from Table 1, the opt-in/opt-out scheme admits three possible equilibria. As illustrated by the dashed line in Figure 1, these correspond to cutoffs of .40, .60, and 1.00 (i.e., full data sharing) for the marginal consumer. For each equilibrium, Table 2 compares the welfare of strong-willed and weak-willed consumers, both in the aggregate and on a per-consumer basis. Since utility is linear, these welfare values can be interpreted directly as dollar surpluses, measuring consumption-equivalent welfare.

[24] Slickdeals.net conducts annual surveys of unintended spending by polling 2,000 U.S. consumers.

[25] Additional analyses show that raising the number of predatory goods J mainly increases the calibrated value of $\bar{v}$, while increasing the number of normal goods N increases the calibrated cost of advertising c.

Across the three equilibrium, utilitarian welfare rises monotonically with the extent of data sharing: social welfare is highest under full data sharing ($\gamma^* = 1.00$) and lowest under the minimal data-sharing equilibrium ($\gamma^* = 0.40$), with a difference of 16.6% in consumption-equivalent welfare. This pattern echoes Jones and Tonetti (2020), who emphasize that because data are non-rival, greater sharing enhances social welfare. They estimate a 40% welfare loss between the no data-sharing equilibrium and the planner's optimum.

Despite this monotonic increase in social welfare, the distribution of gains is uneven:

- First, the welfare of strong-willed consumers remains constant across equilibria. Because the calibrated advertising cost c is relatively low, normal goods sellers always cover the entire opt-in pool, regardless of the cutoff.

- Second, the welfare of weak-willed consumers declines monotonically with data sharing. Their aggregate utility is highest (100.1) under the minimal data-sharing equilibrium ($\gamma^* = 0.40$) and lowest (-54.8) under full data sharing ($\gamma^* = 1.00$). Increased sharing exposes more weak-willed consumers to predatory goods, lowering their welfare. As a result, the welfare gap between strong- and weak-willed consumers is largest under full data-sharing equilibrium ($\Delta = 1515.1$) and smallest under minimal sharing ($\Delta = 1360.2$), a difference of 11.4%.

- Third, the upward-sloping social welfare pattern is driven largely by the rising gains of sellers. Utilitarian welfare also includes seller payoffs. Greater data sharing enables predatory sellers to target weak-willed consumers more effectively, increasing predatory sales even as those consumers' welfare declines.

- Finally, there is substantial heterogeneity among weak-willed consumers. Figure 2 illustrates how different equilibria affect consumers of varying temptation. Under full data sharing ($\gamma^* = 1.00$, dashed line), the least tempted weak-willed consumers benefit, but the most tempted suffer sharply, as reflected in the downward slope. Under minimal data sharing ($\gamma^* = 0.40$, solid line), welfare is improved for the most tempted: neither normal nor predatory sellers advertise to the opt-out pool, effectively setting a floor at zero utility. The intermediate data-sharing equilibrium ($\gamma^* = 0.60$, dotted line) favors mildly weak-willed consumers, who are less covered by normal sellers than under $\gamma^* = 0.40$, but harms the more severely tempted, who are now exposed to predatory targeting in the opt-out pool.

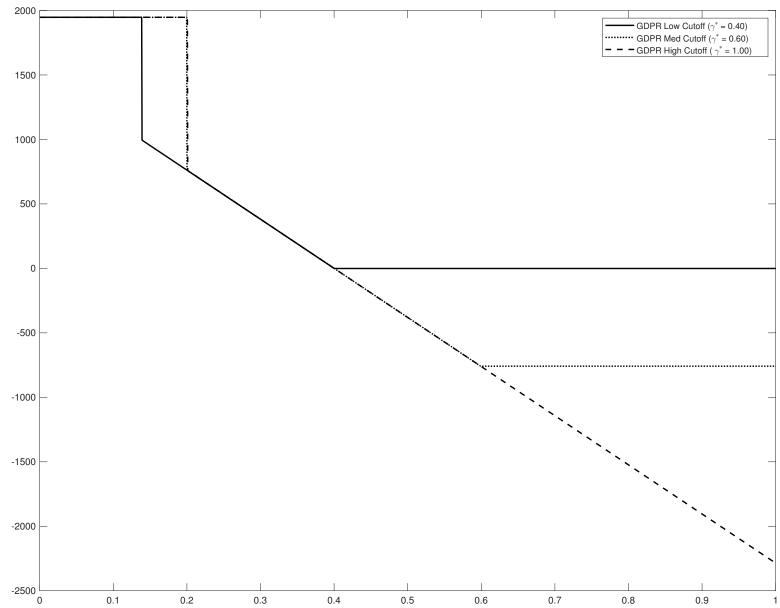

Figure 2: Heterogeneous welfare among weak-willed consumers under the three possible (High, Medium, and Low) GDPR cutoff equilibria for the parameters listed in Table 1.

As discussed in Section 2.3, these equilibria reflect a coordination problem among weak-willed consumers. Even without changes in the underlying environment, data sharing can generate significant welfare losses for the most vulnerable consumers, despite rising aggregate welfare. We refer to this unequal burden as "algorithmic inequality."

The multiplicity induced by coordination also underscores the importance of default options in privacy regulation. By guiding consumers toward either a minimal- or maximal-sharing equilibrium, default rules can serve as a coordination device. Thaler and Sunstein (2008) highlighted the powerful role of defaults in shaping welfare, and the Stigler Committee Report (2019) similarly emphasized their importance for protecting inattentive or biased consumers.

In Appendix B, we extend the model to a dynamic setting where platforms and sellers use accumulated data to improve their goods over time. In this dynamic framework, opt-in/opt-out decisions generate an intertemporal externality: today's data sharing affects the future quality of normal goods and the attractiveness of predatory goods. Simulations assuming coordination on the minimal-sharing equilibrium show that the opt-in/opt-out scheme reduces algorithmic inequality by 13.8% but lowers overall social welfare by 15.5% compared to full data sharing. This highlights a central trade-off of the digital economy: efficiency gains from greater data sharing versus the rising costs of algorithmic inequality.

4 Conclusion

This paper takes a novel approach to consumer privacy by emphasizing protection against individuals' own behavioral vulnerabilities. Data sharing with digital platforms improves matching efficiency for normal goods but also exposes weak-willed consumers to advertising that encourages excessive spending, often through BNPL and other potentially predatory products. Privacy regulations such as GDPR and CCPA provide opt-in and opt-out choices, yet they may fail to protect the most vulnerable because of data-sharing externalities and coordination problems that generate multiple equilibria with very different sharing outcomes. Greater data sharing can thus widen the welfare gap between strong- and weak-willed consumers—what we call algorithmic inequality—even as it raises overall welfare by improving goods quality over time. The result is a fundamental trade-off between efficiency and inequality in the digital economy.

Our analysis highlights how the data economy affects not only what goods and financing consumers receive but also the welfare of others. Externalities arise because platforms bundle data sharing with both normal and predatory sellers, making it costly for vulnerable consumers to opt out: the benefits of sharing with normal sellers often outweigh the risks of predatory targeting. In practice, platforms like Apple reinforce this bundling by offering free services such as messaging and mobile payments, further complicating efforts to regulate or unbundle data sharing.

Alternative remedies—banning predatory products, expanding legal recourse, or promoting platform competition—face serious limitations. Predatory products are difficult to define, as what harms one consumer may help another. Exploitation, unlike fraud or misrepresentation, is hard to measure and rarely policed under existing legal frameworks (Calo, 2013; Sunstein, 2015). And greater platform competition may worsen consumer outcomes by intensifying the race to design addictive content and exploitative practices (Stigler Committee, 2019; Ichihashi and Kim, 2021). Despite its imperfections, ex ante protection of data privacy remains the most effective safeguard for consumers in the digital marketplace. Such insights are relevant as states, such as New York, pass laws to regulate BNPL lenders.[26]

References

[26]For instance, see https://www.dfs.ny.gov/reports_and_publications/press_releases/pr20250509_2.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (2019), Too Much Data: Prices and Inefficiencies in Data Markets, National Bureau of Economic Research Working Paper 26296.

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein (2020), Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age, Journal of Consumer Psychology 30, 736–758.

Acquisti, Alessandro, and Hal R. Varian. (2005), Conditioning Prices on Purchase History, Marketing Science 24, 367–381.

Ali, S. Nageeb, and Roland Bénabou (2020), Image versus Information: Changing Societal Norms and Optimal Privacy, American Economic Journal, Microeconomics 12, 116–64.

Ali, S. Nageeb, Greg Lewis, and Shoshana Vasserman (2019), Voluntary Disclosure and Personalized Pricing, Working Paper.

Allen, Franklin, Xian Gu, and Julapa Jagtiani (2021), A Survey of Fintech Research and Policy Discussion, Review of Corporate Finance 1, 259-339.

Ameriks, John, Andrew Caplin, John Leahy, and Tom Tyler (2007), Measuring Self-Control, American Economic Review, 97, 966–972.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2020. The Economic Consequences of Data Privacy Regulation: Empirical evidence from GDPR, NBER Working Paper.

Athey, Susan, Christian Catalini, and Catherine Tucker (2017), The Digital Privacy Paradox: Small Money, Small Costs, Small Talk, NBER Working Paper No. 23488.

Attanasio, Orazio P., and Guglielmo Weber (2010), Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy, Journal of Economic Literature 48, 693–751.

Bénabou, Roland, and Marek Pycia (2002), Economic Letters 77, 419–424.

Berg, Tobias, Valentin Burg, Jan Keil, and Manju Puri (2025), The Economics of “Buy Now, Pay Later”: A Merchant’s Perspective, Journal of Financial Economics 171, https://doi.org/10.1016/j.jfineco.2025.104093.

Berg, Tobias, Andrea Fuster, and Manju Puri (2022), Fintech Lending, Annual Review of Financial Economics 14, 187-207.

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan (2019), The Economics of Social Data, Working Paper.

Bertrand, Marianne, and Adair Morse (2011), Information Disclosure, Cognitive Biases, and Payday Borrowing, Journal of Finance 66, 1865–1893.

Calo, R. (2013), Digital Market Manipulation, Geo. Wash. L. Rev. 82, 995.

Chen, Daneil (2022), The Market for Attention, Working Paper.

Chen, Long, Yadong Huang, Shumiao Ouyang, and Wei Xiong (2021), Data Privacy and Digital Demand, Working Paper.

Cong Will, Danxia Xie, and Longtian Zhang (2020), Knowledge Accumulation, Privacy, and Growth in a Data Economy, Management Science, forthcoming.

Cong Will, Wenshi Wei, Danxia Xie, and Longtian Zhang (2020), Endogenous Growth Under Multiple Uses of Data, Working Paper.

Cong, Will and Simon Mayer (2022), Antitrust, Regulation, and User Union in the Era of Digital Platforms and Big Data, Working Paper.

Congiu, Raffaele, Lorien Sabatino, and Geza Sapi (2022), The Impact of Privacy Regulation on Web Traffic: Evidence from the GDPR, Information Economics and Policy 61, forthcoming.

Consumer Financial Protection Bureau (2022), Buy Now, Pay Later: Market Trends and Consumer Impacts, Washington, DC, https://www.consumerfinance.gov/data-research/research-reports/buy-now-pay-later-market-trends-and-consumer-impacts/.

Cowgill, B. and M. T. Stevenson (2020), Algorithmic Social Engineering, AEA Papers and Proceedings, 110, 96–100.

Cowgill, B. and C. E. Tucker (2020), Algorithmic Fairness and Economics, Working Paper.

Deckel, Eddie, and Barrton L. Lipman (2012), Costly Self-Control and Random Self-Indulgence, Econometrica 80, 1271–1302.

De Matos, Miguel Godinho, Idris Adjerid (2021), Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. Management Science.

deHaan, Ed, Jungbae Kim, Ben Lourie, and Chenqi Zhu (2024), Buy Now Pay (Pain?) Later, Management Science, forthcoming.

DellaVigna, Stefano, and Ulrike Malmendier (2004), Contract Design and Self-control: Theory and Evidence, Quarterly Journal of Economics 119, 353–402.

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

DellaVigna, Stefano (2009), Psychology and Economics: Evidence from the Field, Journal of Economic Literature 47, 315–72.

Di Maggio, Marco, and Vincent Yao (2021), Fintech Borrowers: Lax Screening or Cream-Skimming?, Review of Financial Studies 34, 4565–4618.

Di Maggio, Marco, Emily Williams, and Justin Katz (2022), Buy Now, Pay Later Credit: User Characteristics and Effects on Spending Patterns, NBER Working Paper, w30508.

Easley, David, Shiyang Huang, Liyan Yang, and Zhuo Zhong (2019), The Economics of Data, Working Paper.

Farboodi, Maryam, and Laura Veldkamp (2020), A Model of the Data Economy, Working Paper.

FTC Bureau of Consumer Protection (2022), Bringing Dark Patterns to Light, FTC Staff Report.

Gabaix, Xavier, and David Laibson (2006), Shrouded Attributes, consumer Myopia, and Information Suppression in Competitive Markets, Quarterly Journal of Economics 121, 505–540.

Galperti, Simone, Aleksandr Levkun, and Jacopo Perego (2022), The Value of Data Records, Review of Economic Studies, forthcoming.

Gelman, Michael and Nikolai Roussanov (2024), Managing Mental Accounts: Payment Cards and Consumption Expenditures, Review of Financial Studies 37, 2586–2624.

Goldberg, Samuel, Garrett Johnson, and Scott Shriver (2019), Regulating Privacy Online: An Economic Evaluation of the GDPR,Working Paper.

Goldfarb, Avi, and Catherine Tucker (2012), Shifts in Privacy Concerns, American Economic Review 102, 349–53.

Goldfarb, Avi, and Catherine Tucker (2019), Digital Economics, Journal of Economic Literature 57, 3–43.

Gross, Ralph, and Alessandro Acquisti (2005), Information Revelation and Privacy in Online Social Networks, Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, 71–80.

Gul, Faruk, and Wolfgang Pesendorfer (2001), Temptation and Self-Control, Econometrica 69, 365–396.

– (2004), Self-Control and the Theory of Consumption, Econometrica 72, 119–158.

deHaan, Ed, Jungbae Kim, Ben Lourie, and Chenqi, Zhu (2024), Management Science, forthcoming.

He, Zhiguo, Jing Huang, and Jidong Zhou (2023), Open Banking: Credit Market Competition When Borrowers Own the Data, Journal of Financial Economics 147(2), 449-474.

Helberger, Natali, Orla Lynskey, Hans-W. Micklitz, Peter Rott, Marjin Sax, and Joanna Strycharz (2021), EU Consumer Protection 2.0: Structural Asymmetries in Digital Consumer Markets, BEUC Report.

Ichihashi, Shota (2020), Online Privacy and Information Disclosure by consumers, American Economic Review 110, 569–95.

Ichihashi, Shota, and Byung-Cheol Kim (2021), Addictive Platforms, SSRN 3853961.

Jann, Ole, and Christoph Schottmüller (2020), An Informational Theory of Privacy, Economic Journal 130, 93-124.

Johannesson, I. (2021), Dark Patterns in Digital Buy Now Pay Later Services, M.S. thesis. KTH Royal Institute of Technology.

Jones, Charles, and Chris Tonetti (2020), Nonrivalry and the Economics of Data, American Economic Review 110, 2819–58.

Kreps, David M (1979), A Representation Theorem for ”Preference for Flexibility”, Econometrica, 565–577.

Laibson, David (1997), Golden Eggs and Hyperbolic Discounting, Quarterly Journal of Economics 112, 443–478.

Larrimore, Jeff, Alicia Lloro, Zofsha Merchant, and Anna Tranfaglia (2024), “The Only Way I Could Afford It": Who Uses BNPL and Why, FEDS Notes.

Liu, Lei, Guangli Lu, and Wei Xiong (2023), The Big Tech Lending Model, Working Paper.

Luguri, Jamie and Lior Jacob Strahilevitz (2021), Shining a Light on Dark Patterns, Journal of Legal Analysis 13, 43–109

Mathur, Arunesh, Gunes Acar, Michael J. Friedman, Elena, Luchierini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan (2019), Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites, Proc. ACM Hum.-Comput. Interact. 3.

Melzer, Brian (2011), The Real Costs of Credit Access: Evidence From the Payday Lending Market, Quarterly Journal of Economics 126, 517–555.

Nadler, Anthony and Lee McGuigan (2018), An Impulse to Exploit: the Behavioral Turn in Data-driven Marketing, Critical Studies in Media Communication 35, 151-165.

O'Donoghue, Ted, and Matthew Rabin (1999), Doing It Now or Later, American Economic Review 89, 103–24.

OECD (2019), Challenges to consumer Policy in the Digital Age, `https://www.oecd.org/sti/consumer/challenges-to-consumer-policy-in-the-digital-age.pdf`.

OECD (2021), Roundtable on Dark Commercial Patterns Online: Summary of discussion, https://one.oecd.org/document/DSTI/CP/CPS(2020)23/FINAL/en/pdf.

Office of the Comptroller of the Currency (2023), Third-Party Relationships: Risk Management. OCC Bulletin 2023-37, https://www.occ.gov/news-issuances/bulletins/2023/bulletin-2023-37.html.

Parlour, Christine A., Uday Rajan, and Haoxiang Zhu (2022), When Fintech Competes for Payment Flows, Review of Financial Studies 35, 4985-5024.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei (2021), Marketing Mutual Funds, Review of Financial Studies 34, 3045–3094.

Spencer, S. B. (2020), The Problem of Online Manipulation, University of Illinois Law Review, 959–1005.

Stango, Victor, and Jonathan Zinman (2014), Limited and Varying consumer Attention: Evidence From Shocks to the Salience of Bank Overdraft Fees, Review of Financial Studies 27, 990–1030.

Stigler Committee (2019), Stigler Committee Report on Digital Platforms, University of Chicago.

Stovall, J. (2010), Multiple Temptations, Econometrica 78, 349–376.

Strotz, R. (1955), Myopia and Inconsistency in Dynamic Utility Maximization, Review of Economic Studies 23, 165–180.

Sunstein, C. R. (2015), Fifty Shades of Manipulation, Journal of Behavioral Marketing.

Tang, Huan (2019), The Value of Privacy: Evidence from Online Borrowers, Working Paper.

Taylor, Curtis R. (2004), consumer Privacy and the Market for Customer Information, RAND Journal of Economics 35, 631–650.

Thaler, Richard H., and Cass R. Sunstein (2008), Nudge: Improving Decision about Health, Wealth and Happiness, London: Penguin.

Tirole, John (2021), Digital Dystopia, American Economic Review 111(6), 2007–48.

Toussaert, Severine (2018), Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment, Econometrica 86, 859–889.

World Bank Group (2021), Consumer Risks in Fintech: New Manifestations of Consumer Risks and Emerging Regulatory Approaches, Policy Research Paper.

Zarsky, T. Z. (2019), Privacy and Manipulation in the Digital Age, Theoretical Inquiries in Law 20(1), 157–188.

Appendix

A Proofs of Propositions

A.1 Proof of Proposition 2

We first consider a strong-willed consumer, that is, $\tau(i) = S_n$, who has the following preferences over different menus:

$$U_S(\{n, \emptyset\}) = \max\{\tilde{u}_n - p_i(n), 0\}, \tag{A.1}$$
$$U_S(\{j, \emptyset\}) = 0. \tag{A.2}$$

Consequently, consumer S_n will buy good n if $\tilde{u}_n \geq p_i(n)$.

Consider now a weak-willed consumer, $\tau(i) = W_{nj}$, with the following preferences:

$$U_W(\{n, \emptyset\}) = \max\{\tilde{u}_n - p_i(n), 0\}, \tag{A.3}$$
$$U_W(\{j, \emptyset\}) = u_B + \max\{-p_i(j), -\gamma_i \bar{v}\}. \tag{A.4}$$

Choosing j from the menu $\{j, \emptyset\}$ is optimal if buying j delivers higher utility: $-p_i(j) > -\gamma_i \bar{v}$, which is equivalent to $\gamma_i > \frac{p_i(j)}{\bar{v}}$.

A.2 Proof of Proposition 3

Given the advertising and pricing strategies of normal good seller n, Proposition 2 implies that the quantity of goods sold that are financed by seller n is

$$Q_n^{NS} = \frac{1}{N} y_n^{NS} \left(1 - H\left(p^{NS}(n)/\bar{u}\right)\right), \tag{A.5}$$

and consequently the seller's profit net of the advertisement cost is

$$\Pi_n^{NS} = p^{NS}(n) \frac{1}{N} y_n^{NS} \left(1 - H\left(p^{NS}(n)/\bar{u}\right)\right) + c \log\left(1 - y_n^{NS}\right). \tag{A.6}$$

Similarly, the quantity of goods sold by the predatory good seller j is

$$Q_j^{NS} = \frac{\pi_W}{J} y_j^{NS} \left(1 - G\left(p^{NS}(j)/\bar{v}\right)\right), \tag{A.7}$$

and the net profit of seller j is

$$\Pi_j^{NS} = p_j^{NS} \frac{\pi_W}{J} y_j \left(1 - G\left(p^{NS}(j)/\bar{v}\right)\right) + c \log\left(1 - y_j^{NS}\right). \tag{A.8}$$

Technological feasibility requires that $y_n^{NS} \geq 0$ and $y_j^{NS} \geq 0$.

The first-order condition of Equation (A.6) with respect to y_n^{NS} is

$$p^{NS}(n) Q_n^{NS} = c \frac{y_n^{NS}}{1 - y_n^{NS}}. \tag{A.9}$$

Then, we have that

$$\Pi_n^{NS} = p^{NS}(n) Q_n^{NS} + c \log\left(1 - y_n^{NS}\right) = c \frac{y_n^{NS}}{1 - y_n^{NS}} + c \log\left(1 - y_n^{NS}\right). \tag{A.10}$$

Similarly, the first-order condition with respect to y_j^{NS} is

$$p^{NS}(j) Q_j^{NS} = c \frac{y_j^{NS}}{1 - y_j^{NS}}, \tag{A.11}$$

which further implies that

$$\Pi_j^{NS} = c \frac{y_j^{NS}}{1 - y_j^{NS}} + c \log\left(1 - y_j^{NS}\right). \tag{A.12}$$

The first-order conditions for the goods prices set by the two sellers are

$$Q_n^{NS} = \frac{p^{NS}(n)}{\bar{u}} \frac{1}{N} y_n^{NS} 1_{\{0 \leq p^{NS}(n) \leq \bar{u}\}}, \tag{A.13}$$

$$Q_j^{NS} = \frac{p^{NS}(j) \pi_W y_j^{NS}}{J \bar{v}} 1_{\{0 \leq p^{NS}(j) \leq \bar{v}\}}. \tag{A.14}$$

Note that the expected quantities sold by both sellers, Q_n^{NS} and Q_j^{NS}, are nonnegative, and the net profits with respect to prices are concave, since

$$\frac{d^2\Pi_n^{NS}}{d\left(p^{NS}(n)\right)^2} = -\frac{2}{\bar{u}}\frac{1}{N}y_n^{NS}h\left(p^{NS}(n)/\bar{u}\right)1_{\{0\leq p^{NS}(n)\leq\bar{u}\}}\leq 0, \tag{A.15}$$

$$\frac{d^2\Pi_j^{NS}}{d\left(p^{NS}(j)\right)^2} = -2\frac{\pi_W}{J}y_j^{NS}g\left(\gamma_*^{NS}\right)\frac{1}{\bar{v}}1_{\{0\leq p^{NS}(j)\leq\bar{v}\}}\leq 0. \tag{A.16}$$

It follows that the optimal price will always be nonnegative. Since

$$\frac{d^2\Pi_n^{NS}}{d\left(y_n^{NS}\right)^2} = -\frac{c}{\left(1-y_n^{NS}\right)^2} < 0, \tag{A.17}$$

and $\frac{d^2\Pi_n^{NS}}{dp^{NS}(n)dy_n^{NS}} = 0$, it follows that the Hessian for seller n's optimization with respect to $\left(p^{NS}(n), y_n^{NS}\right)$ is negative definite and that the FOCs are sufficient.

For strong-willed consumers, there are two possibilities: $p^{NS}(n) \in [0,\bar{u}]$ or $p^{NS}(n) \notin [0,\bar{u}]$. If $R^{NS}(n) \notin [0,\bar{u}]$, then either $p^{NS}(n) = 0$ or $p^{NS}(n) > \bar{u}$, neither of which generates revenue for seller n, and advertising is costly. Consequently, it must be the case that $p_n^{NS} \in [0,\bar{u}]$. Then, Equations (A.9) and (A.13) imply that $p^{NS}(n) = \frac{1}{2}\bar{u}$.

Similarly, for seller j, if $p^{NS}(j) \notin [0,\bar{v}]$, then either $p^{NS}(j) = 0$ or $p^{NS}(j) > \bar{v}$. Neither case generates any revenue, but advertising is costly. If $p^{NS}(j) \in [0,\bar{v}]$, then Equations (A.11) and (A.14) imply $p^{NS}(j) = \frac{1}{2}\bar{v}$.

From the FOCs for y_n^{NS} and y_j^{NS}, it then follows that

$$y_n^{NS} = 1 - \frac{4Nc}{\bar{u}}, \text{ and } y_j^{NS} = 1 - \frac{J}{\pi_W}\frac{4c}{\bar{v}}. \tag{A.18}$$

Thus, the equilibrium is unique. Note that if $y_n^{NS} \leq 0$, then seller n advertises to zero consumers. Similarly, if $y_j^{NS} \leq 0$, then seller j advertises to zero consumers.

A.3 Proof of Proposition 4

With full data sharing, sellers can now separately advertise to strong-willed and weak-willed consumers. We first consider the optimal advertisement and pricing policies of normal good seller n. It shall be clear that seller n target both strong-willed and weak-willed consumers that prefer good n. We denote y_n^{FS} as the measure of strong-willed and weak-willed consumers, to which seller n advertises, and p_n^{FS} as the price the seller sets.

Proposition 2 implies that strong-willed and weak-willed consumers use the same threshold $p_n^{FS}/\bar{u}$ in their random utility $\tilde{u}_n$ for purchasing good n. Thus, the sales by the seller associated with good n is

$$Q_n^{FS} = y_n^{FS}\left[1 - H\left(p^{FS}(n)/\bar{u}\right)\right], \tag{A.19}$$

and the net profit is

$$\Pi_n^{FS} = p^{FS}(n)\, y_n^{FS}\left[1 - H\left(p^{FS}(n)/\bar{u}\right)\right] + c\log\left(1 - y_n^{FS}\right). \tag{A.20}$$

Following the same proof for Proposition 3, it is optimal for seller n to set a price $p^{FS}(n) = \frac{1}{2}\bar{u}$. The first-order condition with respect to y_n^{FS} implies that

$$y_n^{FS} = 1 - 4\frac{c}{\bar{u}}. \tag{A.21}$$

Like before, if $1 - 4\frac{c}{\bar{u}} \leq 0$, it is optimal for the seller to advertise to no consumers. That is, $y_n^{FS} = 0$. Furthermore, if $1 - 4\frac{c}{\bar{u}} > \frac{1}{N}$, then $y_n^{FS} = \frac{1}{N}$.

We now consider the policies of the predatory good seller j. Seller j will advertise only to weak-willed consumers. Since seller j can discriminate by temptation types, it will exercise first-degree price discrimination by charging a weak-willed consumer his full reservation value: $p_{\gamma_i}^{FS}(j) = \gamma_i \bar{v}$. It can also make its advertising strategy y_j^{FS} dependent on γ_i. Since consumers with stronger temptation are willing to pay more, seller j optimally prioritizes strong temptation consumers:

$$d\hat{z}_j^{FS} c = \begin{cases} 0, & \text{if } \gamma_i < \hat{\gamma}^{FS} \\ \frac{\pi_W}{J} d\gamma_i, & \text{if } \gamma_i \in \left(\hat{\gamma}^{FS}, 1\right] \end{cases}. \tag{A.22}$$

Thus, seller j's profit is

$$\Pi_j^{FS} = \bar{v}\int_0^1 \gamma_i y_j^{FS}(d\gamma_i) + c\log\left(1 - y_j^{FS}\right) \quad \text{with} \quad y_j^{FS} = \int_0^1 y_j^{FS}(d\gamma_i) \in [0, \pi_W/J], \tag{A.23}$$

where $\int_0^1 \gamma_i y_j^{FS}(d\gamma_i)$ is understood as a Riemann-Stieljes integral.

Note that the expected revenue of seller j reduces to $\bar{v}\int_{\hat{\gamma}_{jt}^{FS}}^{1} \frac{\pi_W}{J}\gamma_i d\gamma_i = \bar{v}\frac{\pi_W}{J}\frac{1-\left(\hat{\gamma}_j^{FS}\right)^2}{2}$, where $\hat{\gamma}_j^{FS} = 1 - \frac{y_{jt}^{FS}}{\pi_W/J}$, since $y_j^{FS} \in [0, \pi_W/J]$. The expected revenue of seller j is then $\bar{v}y_j^{FS}\left(1-\frac{1}{2}\frac{y_{jt}^{FS}}{\pi_W/J}\right)$, which is determined by the seller's total advertising y_j. Consequently, we can rewrite seller j's maximization problem as choosing y_j^{FS}:

$$\Pi_j^{FS} = \bar{v}y_j^{FS}\left(1-\frac{1}{2}\frac{y_j^{FS}}{\pi_W/J}\right) + c\log\left(1-y_j^{FS}\right) \quad \text{with } \ y_j^{FS} \in [0, \pi_W/J]. \tag{A.24}$$

The first-order condition for y_j^{FS} is

$$\left(1-\frac{y_j^{FS}}{\pi_W/J}\right)\bar{v} - \frac{c}{1-y_j^{FS}} \leq 0, \tag{A.25}$$

which has an interior solution when $\bar{v} > c$. This leads to a quadratic equation:

$$\left(y_j^{FS}\right)^2 - (1+\pi_W/J)\,y_j^{FS} + \frac{\pi_W}{J}\left(1-\frac{c}{\bar{v}}\right) = 0, \tag{A.26}$$

which has the following solutions:

$$y_j^{FS} = \frac{1+\pi_W/J}{2} \pm \sqrt{\left(\frac{1-\pi_W/J}{2}\right)^2 + \frac{\pi_W c}{J\bar{v}}}. \tag{A.27}$$

We select the negative root because to a first-order approximation the positive root is greater than 1:

$$y_j^{FS} = \frac{1+\pi_W/J}{2} + \sqrt{\left(\frac{1-\pi_W/J}{2}\right)^2 + \frac{\pi_W c}{J\bar{v}}} \approx 1 + \frac{\pi_W/J}{1-\pi_W/J}\frac{c}{\bar{v}} > 1. \tag{A.28}$$

Consequently, we have that

$$y_j^{FS} = \frac{1+\pi_W/J}{2} - \sqrt{\left(\frac{1-\pi_W/J}{2}\right)^2 + \frac{\pi_W c}{J\bar{v}}}. \tag{A.29}$$

Again, if this solution to the first-order condition moves outside the feasible range $[0, \pi_W/J]$, it is optimal for the seller to advertise at the corner value. Consequently, the equilibrium is again unique.

Letting $p_n^{FS} = p_n^{FS}(\bar{u})$, $p_j^{FS} = p_j^{FS}(\bar{v})$, $y_n^{FS} = y_n^{FS}(\bar{u})$ and $y_j^{FS} = y_j^{FS}(\bar{v})$, we arrive at the statement of the proposition.

A.4 Proof of Proposition 5

In what follows, we search for a symmetric opt-in/opt-out strategy in which all weak-willed follow the same cutoff opt-in/opt-out strategy. Specifically, we conjecture that all weak-willed consumers will opt-in if their temptation index γ_i is less than some critical γ^*, and opt-out otherwise.

To avoid an uninteresting problem, we assume sellers for normal good do not have the capacity to advertise to all consumers even with data-sharing. Otherwise, there is no trade-off to opting-out for weak-willed consumers.

Sellers: We first characterize the optimal strategies of sellers of both normal and predatory goods taking the opt-in cutoff of weak-willed consumers γ^* as given. We start with the optimal strategy of normal good seller n. Suppose that seller n advertises to $y_{n,in}^{GDPR}$ measure of strong-willed and weak-willed consumers in the opt-in pool at price $p_{in}^{GDPR}(n)$ and $y_{n,out}^{GDPR}$ measure of consumers in the opt-out pool at price $p_{out}^{GDPR}(n)$. Then, the seller's expected profit, by the law of large numbers, is given by

$$\begin{aligned}\Pi_n \;=\; & \frac{1}{N} p_{out}^{GDPR}(n)\, y_{n,out}^{GDPR} \left(1 - \frac{p_{out}^{GDPR}(n)}{\bar{u}}\right) + p_{in}^{GDPR}(n)\, y_{n,in}^{GDPR} \left(1 - \frac{p_{in}^{GDPR}(n)}{\bar{u}}\right) \\ & + c \log\left(1 - y_{n,out}^{GDPR} - y_{n,in}^{GDPR}\right), \end{aligned} \tag{A.30}$$

where $y_{n,out}^{GDPR} \in [0, (1-\gamma^*)\,\pi_W]$ and $y_{n,in}^{GDPR} \in [0, \pi_S + \gamma^* \pi_W]/N$. Note that an advertisement to the opt-in pool reaches a strong or weak-willed consumer who desires the good with perfect precision, while one to the opt-out pool reaches a weak-willed consumer (who desires the good) at a probability of $1/N$.

If $y_{n,in}^{GDPR} > 0$ and $y_{n,out}^{GDPR} > 0$, the FOCs for $p_{in}^{GDPR}(n)$ and $p_{out}^{GDPR}(n)$ reveal that

$$p_{in}^{GDPR}(n) = p_{out}^{GDPR}(n) = \frac{1}{2}\bar{u}. \tag{A.31}$$

Then, the seller's profit becomes

$$\Pi_n = \frac{\bar{u}}{4N} y_{n,out}^{GDPR} + \frac{\bar{u}}{4} y_{n,in}^{GDPR} + c \log\left(1 - y_{n,out}^{GDPR} - y_{n,in}^{GDPR}\right). \tag{A.32}$$

The marginal profit from $y_{n,in}^{GDPR}$ is strictly higher than that from $y_{n,out}^{GDPR}$, as the advertising efficiency to the opt-in pool is higher. Thus, seller n gives higher priority to the opt-in pool.

The first-order condition with respect to $y_{n,in}^{GDPR}$ gives

$$\frac{\bar{u}}{4} - c\frac{1}{1 - y_{n,out}^{GDPR} - y_{nn,in}^{GDPR}} \begin{cases} < 0 & \text{if } y_{n,in}^{GDPR} = 0 \\ = 0 & \text{if } y_{n,in}^{GDPR} \in (0, \pi_S + \gamma^* \pi_W)/N \\ > 0 & \text{if } y_{n,in}^{GDPR} = (\pi_S + \gamma^* \pi_W)/N \end{cases} \tag{A.33}$$

The parameter restriction $c < \frac{\bar{u}}{4}$ ensures that $y_{n,in}^{GDPR} > 0$. As $y_{n,in}^{GDPR}$ has higher priority than $y_{n,out}^{GDPR}$, we have

$$y_{n,in}^{GDPR} = \min\left\{1 - 4\frac{c}{\bar{u}}, (\pi_S + \gamma^* \pi_W)/N\right\}. \tag{A.34}$$

If $y_{n,in}^{GDPR} = (\pi_S + \gamma^* \pi_W)/N$, the seller may have capacity to cover the opt-out pool. The first-order condition for $y_{n,out}^{GDPR}$ in this scenario gives

$$y_{n,out}^{GDPR} = \min\{\max\{\pi_W (1 - \gamma^*) - 4N\frac{c}{\bar{u}}, 0\}, \pi_W (1 - \gamma^*)\}. \tag{A.35}$$

Since seller n gives a higher priority in advertising to the opt-in pool, we can directly prove that each strong-willed consumer would prefer opt-in to opt-out. For simplicity, we skip the proof here.

We now analyze the optimal advertising strategy of predatory seller j. Suppose that seller j advertises with intensity $y_{j,in}^{GDPR}(\gamma_i)$ to weak-willed consumers in the opt-in pool at price $p_{\gamma_i,in}^{GDPR}(j) = \gamma_i \bar{v}$ and $y_{j,out}^{GDPR}$ measure of consumers in the opt-out pool at price $p_{out}^{GDPR}(j)$. Note that an advertisement to the opt-out pool reaches, with probability of $\frac{1}{J}$, a weak-willed consumer, who desires the good, and whether this weak-willed consumer buys the good or not depends on whether his temptation coefficient γ_i is above $p_{j,out}^{GDPR}/\bar{v}$. A further complication is that only weak-willed consumers with γ_i above γ^* are in the opt-out pool. Thus, the seller's profit is

$$\begin{aligned}
\Pi_j &= c\log\left(1-y_{j,out}^{GDPR}-y_{j,in}^{GDPR}\right)+\bar{v}_j\int_0^{\gamma^*}\gamma_i dy_{j,in}^{GDPR}\left(\gamma_i\right) \\
&\quad +\frac{1}{\left(1-\gamma^*\right)J}y_{j,out}^{GDPR}p_{out}^{GDPR}\left(j\right) \\
&\quad \cdot\left[\left(1-p_{out}^{GDPR}\left(j\right)/\bar{v}\right)1_{\left\{R_{j,out}^{GDPR}\geq\gamma^*\bar{v}\right\}}+\left(1-\gamma^*\right)1_{\left\{R_{j,out}^{GDPR}<\gamma^*\bar{v}\right\}}\right], \qquad \text{(A.36)}
\end{aligned}$$

where $y_{j,out}^{GDPR}\in\left[0,\left(1-\gamma^*\right)\pi_W\right]$ and $y_{j,in}^{GDPR}=\int_0^{\gamma^*}y_{j,in}^{GDPR}\left(d\gamma_i\right)\in\left[0,\gamma^*\pi_W/J\right]$ is the total advertisement to the opt-in pool.

If $y_{j,out}^{GDPR}>0$, then the first-order condition for $p_{out}^{GDPR}\left(j\right)$ gives the following:

$$\begin{aligned}
\text{If } \gamma^* &\leq \frac{1}{2},\ \left(1-2p_{out}^{GDPR}\left(j\right)/\bar{v}\right)1_{\left\{p_{out}^{GDPR}(j)\geq\gamma^*\bar{v}\right\}}=0, \\
\text{If } \gamma^* &> \frac{1}{2},\ p_{out}^{GDPR}\left(j\right)=\gamma^*\bar{v}. \qquad \text{(A.37)}
\end{aligned}$$

Thus, the optimal price satisfies

$$p_{out}^{GDPR}\left(j\right)=\begin{cases}\frac{1}{2}\bar{v} & \text{if } \gamma^*\leq\frac{1}{2}\\ \\ \gamma^*\bar{v} & \text{If } \gamma^*>\frac{1}{2}\end{cases}=\max\left\{\frac{1}{2},\gamma^*\right\}\bar{v}. \qquad \text{(A.38)}$$

Since consumers with stronger temptation are willing to pay more with $p_{\gamma_i,in}^{GDPR}\left(j\right)=\gamma_i\bar{v}$, it is optimal for seller j to prioritize consumers with higher γ_i:

$$dy_{j,in}^{GDPR}\left(\gamma_i\right)=\begin{cases}0 & \text{if } \gamma_i<\hat{\gamma}^{GDPR}\\ \frac{\pi_W}{J}d\gamma_i & \text{if } \gamma_i\in\left(\hat{\gamma}^{GDPR},\gamma^*\right]\end{cases}. \qquad \text{(A.39)}$$

Therefore, the expected revenue of seller j from the opt-in pool reduces to $\bar{v}\int_{\hat{\gamma}^{GDPR}}^{\gamma^*}\frac{\pi_W}{J}\gamma_i d\gamma_i=\bar{v}\frac{\pi_W}{J}\frac{(\gamma^*)^2-\left(\hat{\gamma}^{GDPR}\right)^2}{2}$. As $\hat{\gamma}^{GDPR}=\gamma^*-\frac{y_{j,in}^{GDPR}}{\pi_W/J}$ by definition, the expected revenue of seller j from advertising to the opt-in pool is determined by the seller's total advertising to the opt-in pool $y_{j,in}^{GDPR}$: $\bar{v}y_{j,in}^{GDPR}\left(\gamma^{*}-\frac{1}{2}\frac{y_{j,in}^{GDPR}}{\pi_W/J}\right)$. Thus, the expected profit of seller j becomes

$$\begin{aligned}\Pi_j &= \frac{1}{(1-\gamma^{*})J}\left[\frac{1}{4}-\left(\gamma^{*}-\frac{1}{2}\right)^{2}1_{\left\{\gamma^{*}>\frac{1}{2}\right\}}\right]\bar{v}y_{j,out}^{GDPR} \\ &\quad +y_{j,in}^{GDPR}\left(\gamma^{*}-\frac{1}{2}\frac{y_{jt,in}^{GDPR}}{\pi_W/J}\right)\bar{v}+c\log\left(1-y_{j,out}^{GDPR}-y_{j,in}^{GDPR}\right),\end{aligned} \tag{A.40}$$

and the seller's choice reduces to choosing $y_{j,in}^{GDPR}$ and $y_{j,out}^{GDPR}$.

Which pool has priority depends on which has higher marginal revenue. The marginal revenue from the opt-in pool $\bar{v}\left(\gamma^{*}-\frac{y_{j,in}^{GDPR}}{\pi_W/J}\right)$ is decreasing with $y_{j,in}^{GDPR}$ and has the highest value of $\bar{v}\gamma^{*}$ when $y_{j,in}^{GDPR}=0$. The marginal revenue from the opt-out pool is constant: $\frac{1}{(1-\gamma^{*})J}\left[\frac{1}{4}-\left(\gamma^{*}-\frac{1}{2}\right)^{2}1_{\left\{\gamma^{*}>\frac{1}{2}\right\}}\right]\bar{v}$.

The first-order condition for $y_{j,in}^{GDPR}$ is

$$\bar{v}\left(\gamma^{*}-\frac{y_{j,in}^{GDPR}}{\pi_W/J}\right)-\frac{c}{1-y_{j,out}^{GDPR}-y_{j,in}^{GDPR}}\begin{cases}<0 & \text{if } y_{j,in}^{GDPR}=0\\ =0 & \text{if } y_{j,in}^{GDPR}\in(0,\pi_W\gamma^{*}/J)\\ >0 & \text{if } y_{j,in}^{GDPR}=\pi_W\gamma^{*}/J\end{cases}, \tag{A.41}$$

and the first-order condition for $y_{j,out}^{GDPR}$ is

$$\frac{1}{(1-\gamma^{*})J}\left[\frac{1}{4}-\left(\gamma^{*}-\frac{1}{2}\right)^{2}1_{\left\{\gamma^{*}>\frac{1}{2}\right\}}\right]\bar{v}-\frac{c}{1-y_{j,out}^{GDPR}-y_{j,in}^{GDPR}}\begin{cases}<0 & \text{if } y_{j,out}^{GDPR}=0\\ =0 & \text{if } y_{j,out}^{GDPR}\in(0,(1-\gamma^{*})\pi_W)\\ >0 & \text{if } y_{j,out}^{GDPR}=(1-\gamma^{*})\pi_W\end{cases}. \tag{A.42}$$

When $\gamma^{*}<\frac{1}{2}$, the opt-in pool has priority whenever $\gamma^{*}>\frac{1}{4}\frac{1}{(1-\gamma^{*})J}$, which is equivalent to $\gamma^* \in \frac{1}{2}\left[1-\sqrt{1-\frac{1}{J}}, 1+\sqrt{1-\frac{1}{J}}\right]$, which exists and has its upper end above $\frac{1}{2}$. When $\gamma^* > \frac{1}{2}$, it is direct to verify that the opt-in pool has priority. Taken together, the opt-in pool has priority if and only if[27]

$$\gamma^* > \underline{\gamma} = \frac{1}{2}\left(1-\sqrt{1-\frac{1}{J}}\right). \tag{A.43}$$

If $\gamma^* \leq \underline{\gamma}$, the opt-out pool has priority. In this case, the seller first targets the consumers in the opt-out pool until it covers the full pool of $\pi_W (1-\gamma^*)$. Before it hits the corner, the interior choice is the first order condition in Equation (A.42) with $y_{j,in}^{GDPR} = 0$, which gives

$$y_{j,out}^{GDPR} = 1 - \frac{4cJ}{\bar{v}}(1-\gamma^*). \tag{A.44}$$

If $1 - \frac{4cJ}{\bar{v}}(1-\gamma^*) \geq \pi_W (1-\gamma^*)$, which is equivalent to $\gamma^* \geq 1 - \left(\frac{4cJ}{\bar{v}} + \pi_W\right)^{-1}$, then $y_{j,out}^{GDPR} = \bar{y}_{out} = \pi_W (1-\gamma^*)$ and $y_{j,in}^{GDPR}$ is given by the first order condition in Equation (A.41):

$$\left(\gamma^* - \frac{J}{\pi_W} y_{j,in}^{GDPR}\right)\left(1 - \bar{y}_{out} - y_{j,in}^{GDPR}\right) = \frac{c}{\bar{v}}. \tag{A.45}$$

This equation takes the advertising to the opt-out pool $y_{j,out}^{GDPR} = \bar{y}_{out}$ as given and solves for the advertising to the opt-in pool $y_{j,in}^{GDPR}$. Generically, we define $y_{in*}(x)$ as the solution to the following equation:

$$\left(\gamma^* - \frac{J}{\pi_W} y\right)(1-x-y) = \frac{c}{\bar{v}}, \tag{A.46}$$

which gives y the optimal amount of advertising to the opt-in pool for a given level of advertising x to the opt-out pool. This leads to

$$y^2 - \left(\frac{\pi_W}{J}\gamma^* + (1-x)\right) y + \frac{\pi_W}{J}\gamma^*(1-x) - \frac{\pi_W}{J}\frac{c}{\bar{v}} = 0. \tag{A.47}$$

[27] We also recognize that $\frac{c}{\bar{v}}$ is the minimum γ^* at which predatory goods sellers advertise a positive amount to the opt-in pool. This value is recovered by recognizing at zero advertising to the opt-in pool (i.e., $y_{j,in}^{GDPR} = 0$), the first-order condition is $\bar{v}\gamma^* - c$, which is nonpositive if $\gamma^* \leq \frac{c}{\bar{v}}$. Notice, however, the seller financing the predatory good also chooses zero advertising for the opt-out pool because $\gamma^* > \underline{\gamma}$, and the marginal revenue of the opt-in is always higher than that of the opt-out pool.

As the larger root of this equation is larger than 1, we choose the smaller root:

$$y_{in*}(x) = \frac{1}{2}\left((1-x) + \frac{\pi_W}{J}\gamma^*\right) - \sqrt{\frac{1}{4}\left((1-x) - \frac{\pi_W}{J}\gamma^*\right)^2 + \frac{\pi_W}{J}\frac{c}{\bar{v}}}. \tag{A.48}$$

Thus, in this case, $y_{j,in}^{GDPR} = y_{in*}\left(\bar{y}_{out}\right)$.

We now consider the case $\gamma^* > \underline{\gamma}$. In this case, the opt-in pool has priority. Before the marginal revenue of the opt-in pool drops down to that of the opt-out pool, we have an interior solution for $y_{j,in}^{GDPR} = y_{in*}\left(0\right)$ with $y_{j,out}^{GDPR} = 0$.

These two marginal revenues will intersect at a unique level y_{in**} for $y_{j,in}^{GDPR}$, where

$$y_{in**}\left(\gamma^*\right) = \frac{\pi_W}{J}\gamma^* - \left(\frac{\pi_W}{J}\right)^2 \frac{\frac{1}{4} - \left(\gamma^* - \frac{1}{2}\right)^2 1_{\left\{\gamma^* > \frac{1}{2}\right\}}}{\left(1-\gamma^*\right)\pi_W}, \tag{A.49}$$

which is positive whenever $\gamma^* \geq \underline{\gamma}$. We can further simplify

$$y_{in**}\left(\gamma^*\right) = \begin{cases} \frac{\pi_W}{J}\left(\gamma^* - \frac{1}{4J}\frac{1}{1-\gamma^*}\right) & \text{if } \underline{\gamma} \leq \gamma^* \leq \frac{1}{2} \\ \\ \frac{\pi_W}{J}\gamma^*\left(1 - \frac{1}{J}\right) & \text{if } \gamma^* > \frac{1}{2} \end{cases}. \tag{A.50}$$

Note if $y_{in*}\left(0\right)$ rises above $y_{in**}\left(\gamma^*\right)$, it becomes profitable for the seller to target the opt-out pool together with the opt-in pool. In this situation, $y_{j,in}^{GDPR} = y_{in**}\left(\gamma^*\right)$, then the first-order condition in Equation (A.41) determines the interior level of $y_{j,out}^{GDPR} = y_{out*}(x)$ for a given level of $y_{j,in}^{GDPR} = x$ with

$$y_{out*}\left(x\right) = \min\left\{\max\left\{1 - \frac{c}{\bar{v}}\left(\gamma^* - \frac{x}{\pi_W/J}\right)^{-1} - x, 0\right\}, \pi_W\left(1-\gamma^*\right)\right\}. \tag{A.51}$$

In this expression, $y_{j,out}^{GDPR}$ is bounded from above by the size of the opt-out pool $\pi_W\left(1-\gamma^*\right)$. Consequently, substituting Equation (A.51) with Equation (A.50)

$$y_{out*}\left(y_{in**}\left(\gamma^*\right)\right) = \min\left\{\max\left\{z, 0\right\}, \pi_W\left(1-\gamma^*\right)\right\}. \tag{A.52}$$

where

$$z = \begin{cases} 1 - \frac{4Jc}{\bar{v}}\left(1-\gamma^*\right) - \frac{\pi_W}{J}\left(\gamma^* - \frac{1}{4J}\frac{1}{1-\gamma^*}\right) & \text{if } \underline{\gamma} \leq \gamma^* \leq \frac{1}{2} \\ \\ 1 - \frac{Jc}{\gamma^*\bar{v}} - \frac{\pi_W \gamma^*}{J}\left(1 - \frac{1}{J}\right) & \text{if } \gamma^* > \frac{1}{2} \end{cases}. \tag{A.53}$$

If $y_{j,out}^{GDPR}$ is constrained at its upper bound $\bar{y}_{out} = \pi_W\left(1-\gamma^*\right)$, then the first-order condition in Equation (A.41) gives an interior level of $y_{j,in}^{GDPR}$, with $y_{j,out}^{GDPR} = \bar{y}_{out}$. That is $y_{j,in}^{GDPR} = y_{in*}\left(\bar{y}_{out}\right)$.

Taken together, if $\gamma^* \leq \underline{\gamma}$, the opt-out pool has priority:

$$y_{j,out}^{GDPR} = \min\left\{1 - \frac{4cJ}{\bar{v}}\left(1-\gamma^*\right), \bar{y}_{out}\right\}, \tag{A.54}$$

and

$$y_{j,in}^{GDPR} = \begin{cases} 0 & \text{if } y_{j,out}^{GDPR} < \bar{y}_{out} \\ \\ y_{in*}\left(\bar{y}_{out}\right) & \text{if } y_{j,out}^{GDPR} = \bar{y}_{out}. \end{cases} \tag{A.55}$$

If $\gamma^* > \underline{\gamma}$, the opt-in pool has priority:

$$y_{j,in}^{GDPR} = \begin{cases} y_{in*}\left(0\right) & \text{if } y_{in*}\left(0\right) < y_{in**}\left(\gamma^*\right) \\ \\ y_{in**}\left(\gamma^*\right) & \text{if } y_{in*}\left(0\right) \geq y_{in**}\left(\gamma^*\right) \text{ and } y_{out*}\left(y_{in**}\left(\gamma^*\right)\right) < \bar{y}_{out} \\ \\ y_{in*}\left(\bar{y}_{out}\right) & \text{if } y_{out*}\left(y_{in**}\left(\gamma^*\right)\right) \geq \bar{y}_{out} \end{cases}, \tag{A.56}$$

and

$$y_{j,out}^{GDPR} = \begin{cases} 0 & \text{if } y_{in*}(0) < y_{in**}(\gamma^*) \\ \\ y_{out*}(y_{in**}(\gamma^*)) & \text{if } y_{in*}(0) \geq y_{in**}(\gamma^*) \text{ and } y_{out*}(y_{in**}(\gamma^*)) < \bar{y}_{out} \\ \\ \bar{y}_{out} & \text{if } y_{out*}(y_{in**}(\gamma^*)) \geq \bar{y}_{out} \end{cases} . \quad \text{(A.57)}$$

Weak-willed customers: We first verify if other weak-willed customers follow the conjectured cutoff strategy with cutoff γ^*, it is optimal for a weak-willed consumer with temptation γ_i to follow the same cutoff strategy. We then characterize the equilibrium cutoff γ^*.

Consider a weak-willed consumer with temptation index γ_i. Following Equation (16), his expected utility from opt-in is

$$U_{W,in}^{GDPR}(\gamma_i) = \frac{y_{n,in}^{GDPR} N}{\pi_S + \gamma^* \pi_W} \frac{\bar{u}}{8} + \frac{y_{j,in}^{GDPR}(\gamma_i) J}{\pi_W} (u_B - \gamma_i \bar{v}) . \quad \text{(A.58)}$$

This expression shows that $U_{W,in}^{GDPR}$ increases with $y_{n,in}^{GDPR}$ but decreases with $y_{j,in}^{GDPR}(\gamma_i)$. Following Equation (17), his expected utility from opt-out is

$$\begin{aligned} U_{W,out}^{GDPR}(\gamma_i) &= \frac{y_{n,out}^{GDPR}}{(1-\gamma^*)\pi_W} \frac{\bar{u}}{8} + \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W} u_B - \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W} \\ &\quad \cdot \left[\max\left\{\frac{1}{2}, \gamma^*\right\} \mathbf{1}_{\{\gamma_i > \max\{\frac{1}{2},\gamma^*\}\}} + \gamma_i \mathbf{1}_{\{\gamma_i \leq \max\{\frac{1}{2},\gamma^*\}\}} \right] \bar{v}, \end{aligned} \quad \text{(A.59)}$$

which increases with $y_{n,out}^{GDPR}$ and decreases with $y_{n,out}^{GDPR}$. Then,

$$\begin{aligned} V(\gamma_i) &= U_{W,in}^{GDPR}(\gamma_i) - U_{W,out}^{GDPR}(\gamma_i) \quad \text{(A.60)} \\ &= \left[\frac{y_{n,in}^{GDPR} N}{\pi_S + \gamma^* \pi_W} - \frac{y_{n,out}^{GDPR}}{(1-\gamma^*)\pi_W} \right] \frac{\bar{u}}{8} + \left(\frac{y_{j,in}^{GDPR}(\gamma_i)}{\pi_W / J} - \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W} \right) (u_B - \bar{v}\gamma_i) \\ &\quad - \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W} \bar{v} \left(\gamma_i - \max\left\{\frac{1}{2}, \gamma^*\right\} \right) \mathbf{1}_{\{\gamma_i > \max\{\frac{1}{2},\gamma^*\}\}} \end{aligned}$$

Note that $\frac{y_{n,in}^{GDPR} N}{\pi_S + \gamma^* \pi_W} \geq \frac{y_{n,out}^{GDPR}}{(1-\gamma^*)\pi_W}$ from our earlier analysis of seller n's strategy. Therefore, whether $U_{W,in}^{GDPR}(\gamma_i) - U_{W,out}^{GDPR}(\gamma_i)$ crosses zero depends on the second and third terms. In the second term, whether $\frac{y_{j,in}^{GDPR}(\gamma_i)}{\pi_W / J} - \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W}$ is positive or not depends on whether γ^* is higher or lower than $\underline{\gamma}$.

We first show that in equilibrium γ^* cannot be lower than $\underline{\gamma}$. We use contradiction. Suppose that $\gamma^* < \underline{\gamma}$. Then, $\frac{y_{j,in}^{GDPR}(\gamma_i)}{\pi_W / J} \leq \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W}$ because the predatory goods sellers give higher priority to opt-out pool. It is then clear that $V(\gamma_i) > 0$ for γ_i slightly above γ^*, implying that this consumer would choose opt-in. This contradicts with γ^* being the equilibrium threshold so that consumers with γ_i above γ^* all choose opt-out.

For $\gamma^* \geq \underline{\gamma}$, define the (adjusted) net benefit to opt-in for the marginal weak-willed consumer when she follows the conjectured cutoff strategy:

$$\begin{aligned} C(\gamma^*) &\equiv \frac{1}{\bar{v}} \left(U_{W,in}^{GDPR}(\gamma^*) - U_{W,out}^{GDPR}(\gamma^*) \right) \\ &= \frac{\bar{u}}{8\bar{v}} \left(\frac{y_{n,in}^{GDPR} N}{\pi_S + \gamma^* \pi_W} - \frac{y_{n,out}^{GDPR}}{(1-\gamma^*)\pi_W} \right) \\ &\quad + \left(\frac{u_B}{\bar{v}} - \gamma^* \right) \left(\frac{y_{j,in}^{GDPR}(\gamma^*)}{\pi_W / J} - \frac{y_{j,out}^{GDPR}}{(1-\gamma^*)\pi_W} \right). \end{aligned} \tag{A.61}$$

For γ^* to be the equilibrium cutoff, there are three possibilities:

$$\gamma^* \begin{cases} = \underline{\gamma} & \text{if } C(\underline{\gamma}) < 0 \\ \in (\underline{\gamma}, 1) & \text{if } C(\gamma^*) = 0 \\ = 1 & \text{if } C(1) > 0 \end{cases}. \tag{A.62}$$

Corner Solution for $\gamma^* = 1$: We consider this corner as a limiting case. There are two reasons why all consumers would choose opt-in.

First, the severity of temptation $\bar{v}$ may be so high that predatory goods sellers will search the opt-out pool on the margin if all weak-willed consumers opt-in. Suppose a fraction ε of weak-willed consumers opt-out, i.e., $\gamma^* = 1 - \varepsilon$, then substituting for $y_{n,in}^{GDPR}$, Equation

(A.61) reduces t

$$C(1-\varepsilon)=\frac{\bar{u}}{8\bar{v}}\left(\min\left\{\left(1-2\sqrt{\frac{c}{\bar{u}}}\right)\frac{N}{1-\varepsilon\pi_W},1\right\}-\frac{y_{n,out}^{GDPR}}{\varepsilon\pi_W}\right)+\left(\frac{u_B}{\bar{v}}+\varepsilon-1\right)\left(1-\frac{y_{j,out}^{GDPR}}{\varepsilon\pi_W}\right). \tag{A.63}$$

Recognizing that unless normal goods sellers can cover all consumers, they will eschew the opt-out pool ($y_{n,out}^{GDPR}=0$) with a $\varepsilon\pi_W$ mass of consumers because it has lower expected revenue. Even though predatory sellers give priority to the opt-in pool, a seller financing predatory good j may still cover the opt-out pool if $y_{in*}(0)$ rises above $y_{in**}(\gamma^*=1-\epsilon)$, which is given by Equation (A.50). This condition $\epsilon\to 0$ is equivalent to

$$\bar{v}\geq v_{**}\equiv\frac{c}{\left(1-\frac{\pi_W}{J}\right)\frac{1}{J}+\frac{\pi_W}{J^3}}. \tag{A.64}$$

Second, the severity of temptation $\bar{v}$ may be so low that all weak-willed consumers choose opt-in for the benefit of matching with normal goods sellers despite the cost of being targeted by predatory goods sellers. In this case, predatory goods sellers do not also advertise to the opt-out pool ($\frac{y_{j,out}^{GDPR}}{\varepsilon\pi_W}=0$) and $C(1-\varepsilon)$ reduces to

$$C(1-\varepsilon)=\frac{\bar{u}}{8\bar{v}}\min\left\{\left(1-4\frac{c}{\bar{u}}\right)\frac{N}{1-\varepsilon\pi_W},1\right\}+\frac{u_B}{\bar{v}}+\varepsilon-1. \tag{A.65}$$

Then $C(1-\varepsilon)$ is positive as $\epsilon\to 0$ if

$$\bar{v}<\frac{\bar{u}}{8}\min\left\{\left(1-\frac{4c}{\bar{u}}\right)N,1\right\}+u_B, \tag{A.66}$$

and again all weak-willed consumers opt-in. Similarly, if $\bar{v}<c$, then predatory goods sellers never advertise to the opt-in pool because the marginal revenue $\bar{v}$ is always less than the marginal cost, c. As such, all weak-willed consumers opt-in if

$$\bar{v}<\min c,\frac{\bar{u}}{8}\min\left\{\left(1-\frac{4c}{\bar{u}}\right)N,1\right\}+u_B \tag{A.67}$$

Note that the conditions $\bar{v}\leq v^*$ and Equation (A.67) may overlap. When this happens, there are two equilibria. In the equilibrium with $\gamma^*=1$, all weak-willed consumers choose opt-in to maximize matching with normal product sellers. In this case, the opt-out pool offers no protection for the severely tempted consumers. In the other equilibrium with $\gamma^* = \underline{\gamma}$, only a fraction of weak-willed consumers choose opt-in, and the opt-out pool provides substantial protection for the most tempted consumers. The complementarity in the consumers' opt-out decision contributes to the rise of multiple equilibria.

Interior Solution for $\gamma^* \in (\underline{\gamma}, 1)$: Note that $V(\gamma_i)$ in Equation (A.60) is monotonically decreasing with γ_i. Given that $V(\gamma^*) = C(\gamma^*) = 0$, consumers with $\gamma_i < \gamma^*$ want to opt-in and those with $\gamma_i > \gamma^*$ want to opt-out, confirming the optimality of the cutoff strategy for weak-willed consumers.

Given the non-linearity of $C(\gamma^*)$, there may be multiple values in $(\underline{\gamma}, 1)$ with $C(\gamma^*) = 0$, and consequently multiple equilibria. In this case, note that $\frac{y_{j,in}^{GDPR}(\gamma^*)}{\pi_W/J} = 1$ and $\frac{y_{j,out}^{GDPR}}{\pi_W(1-\gamma^*)} < 1$. The optimal advertising policy of seller j for the opt-in and opt-out pools is given by Equations (A.56) and (A.57). Substituting for $y_{n,in}^{GDPR}$ and $y_{n,out}^{GDPR}$ with Equations (A.34) and (A.35), we recognize

$$\begin{aligned} & \frac{y_{n,in}^{GDPR} N}{\pi_S + \gamma^* \pi_W} - \frac{y_{n,out}^{GDPR}}{\pi_W(1-\gamma^*)} \\ = \quad & \min\left\{\frac{1 - 4\frac{c}{\bar{u}}}{1 - \pi_W(1-\gamma^*)} N, 1\right\} - \max\left\{1 - \frac{4N\frac{c}{\bar{u}}}{\pi_W(1-\gamma^*)}, 0\right\}, \end{aligned} \tag{A.68}$$

and substituting this into Equation (A.61) gives

$$\begin{aligned} C(\gamma^*) \quad = \quad & \frac{\bar{u}}{8\bar{v}}\left(\min\left\{\frac{1 - 4\frac{c}{\bar{u}}}{1 - \pi_W(1-\gamma^*)} N, 1\right\} - \max\left\{1 - \frac{4N\frac{c}{\bar{u}}}{\pi_W(1-\gamma^*)}, 0\right\}\right) \\ & + \left(\frac{u_B}{\bar{v}} - \gamma^*\right)\left(1 - \frac{y_{j,out}^{GDPR}}{\pi_W(1-\gamma^*)}\right), \end{aligned} \tag{A.69}$$

where $y_{j,out}^{GDPR}$ is given by Equation (A.57).

The first term is continuous and positive on γ^*, and not equal to zero because $1 > 4\frac{c}{\bar{u}}$ by assumption. Whenever the max term is positive, the min term must be 1. The second term is continuous in γ^* because $\frac{y_{j,out}^{GDPR}}{\pi_W(1-\gamma^*)} \in [0, 1]$ is (piece-wise) continuous in γ^* from Equation (A.57). Consequently, $C(\gamma^*)$ is continuous in γ^* on $[\underline{\gamma}, 1]$.

Thus, by the Intermediate Value Theorem, there exists a $\gamma^* \in (\underline{\gamma}, 1)$ such that $C(\gamma^*) = 0$, and an interior equilibrium exists. Notice

$$C\left(\underline{\gamma}\right) > \frac{\bar{u}}{8\bar{v}}\left(\min\left\{\frac{1-4\frac{c}{\bar{u}}}{1-\pi_W\left(1-\underline{\gamma}\right)}N,1\right\} - \max\left\{1-\frac{4N\frac{c}{\bar{u}}}{\pi_W\left(1-\underline{\gamma}\right)},0\right\}\right) + \frac{u_B}{\bar{v}} - \underline{\gamma}. \tag{A.70}$$

It is sufficient that $c > \frac{\bar{u}}{N}\pi_W\frac{1-\underline{\gamma}}{4}$ (i.e., normal goods sellers ignore the opt-out pool and advertise only to the opt-in pool) and

$$\bar{v} < v_{***} \equiv \frac{u_B}{\underline{\gamma}} + \frac{\bar{u}}{\underline{\gamma}}\min\left\{\frac{N-\pi_W\left(1-\underline{\gamma}\right)}{1-\pi_W\left(1-\underline{\gamma}\right)},1\right\}, \tag{A.71}$$

for $C\left(\underline{\gamma}\right) > 0$. Consequently, it is sufficient that $c > \frac{\bar{u}}{N}\pi_W\frac{1-\underline{\gamma}}{4}$ and $\bar{v} < v_{***}$ to ensure there is an interior equilibrium.

When, in addition, $\bar{v} > v_{**}$, then there are multiple equilibrium in which full data-sharing is an equilibrium.

Internet Appendix for Data Privacy and Algorithmic Inequality

Internet Appendix A: Additional Proofs of Propositions and Results

Proof of Proposition 6

We first compare social welfare under three data sharing schemes: no data sharing, full data sharing, and the GDPR. Under a specific data-sharing scheme, social welfare is determined by the aggregate utility of strong- and weak-willed consumers, as indicated by Equation (10). This is based on the assumptions that the marginal cost of goods production is zero, and the price of goods and advertising costs are zero-sum transfers within the population. As the seller financing normal good n cannot price discriminate against its customers because of the consumers' random utility for normal products, it always sets a price of $\bar{u}/2$ for its good. Consequently, only half of the intended consumers with random utility above $\bar{u}/2$ consume the good. Thus, consumers' net utility gain from good n is $\frac{3}{8}\bar{u}\rho_n$, where ρ_n is the measure of strong-willed and weak-willed consumers receiving seller n's advertising. For a predatory good j, weak-willed consumers who borrow (with a measure of ρ_j) experience a negative utility of $u_B < 0$. Meanwhile, those who receive advertising from seller j but resist the temptation (marked in a set S_j) suffer a mental cost of $u_B - \gamma_i\bar{v}$. Note that ρ_n, ρ_j, and S_j are determined by the sellers' advertising and pricing strategies under each of the data sharing schemes.

Taken together, the social welfare is

$$W = \frac{3}{8}\sum_{n=1}^{N}\bar{u}\rho_n + \sum_{j=N+1}^{N+J}\left[u_B\rho_j + \int_{i\in S_j}\left(u_B - \gamma_i\bar{v}\right)dG\left(\gamma_i\right)\right]. \tag{IA.1}$$

Across the data sharing schemes, the key trade-off is between the first term (the benefit from normal goods) and the second and third terms (the cost from predatory goods). Note that these terms only account for the consumers' utility from the normal and predatory goods without including the price to sellers for the goods, which are transfers within the population.

No data sharing: The social welfare is

$$\begin{aligned} W^{NS} &= \sum_{n=1}^{N} \frac{1}{N} y_n^{NS} \int_{R_n}^{\bar{u}} u_n \frac{du_n}{\bar{u}} + \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} y_j^{NS} u_B \int_{p^{NS}(j)/\bar{v}}^{1} d\gamma_i \\ &\quad + \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} y_j^{NS} \int_0^{R^{NS}(j)/\bar{v}} (u_B - \gamma_i \bar{v})\, d\gamma_i \\ &= \frac{3}{8} \bar{u} y_n^{NS} + \pi_W y_j^{NS} \left(u_B - \frac{1}{8} \bar{v} \right) = \frac{3}{8} \bar{u} y_n^{NS} + \pi_W y_j^{NS} \left(u_B - \frac{\bar{v}}{8} \right). \end{aligned} \tag{IA.2}$$

Full data sharing: The social welfare is

$$W^{FS} = \frac{3}{8} \sum_{n=1}^{N} \bar{u} y_n^{FS} + \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} u_B \int_{\hat{\gamma}_j^{FS}}^{1} d\gamma_i = \frac{3}{8} N \bar{u} y_n^{FS} + J y_j^{FS} u_B. \tag{IA.3}$$

GDPR: From a social welfare perspective, we have

$$W^{GDPR} = \left(N y_{n,in}^{GDPR} + y_{n,out}^{GDPR} \right) \frac{3}{8} \bar{u} + \left(J y_{j,in}^{GDPR} + y_{j,out}^{GDPR} \right) u_B - y_{j,out}^{GDPR} \left(\frac{1}{8} - \frac{\gamma^{*2}}{2} \right) \bar{v} \mathbf{1}_{\gamma^* < \frac{1}{2}}, \tag{IA.4}$$

where the last term reflects that the least tempted weak-willed in the opt-out pool suffer a temptation cost if $\gamma^* < \frac{1}{2}$.

We first compare full data sharing to no data sharing:

$$W^{FS} - W^{NS} = < \pi_W \left(\frac{y_j^{FS}}{\pi_W / J} - y_j^{NS} \right) u_B + \frac{1}{8} \pi_W y_j^{NS} \bar{v} + \frac{3}{8} \bar{u} \left(N y_n^{FS} - y_n^{NS} \right) < 0, \tag{IA.5}$$

if $u_B < u_{B**}$, where

$$u_{B**} = - \frac{\frac{3}{\pi_W} \bar{u} \left(N y_n^{FS} - y_n^{NS} \right) + \bar{v} y_j^{NS}}{8 \left(\frac{y_j^{FS}}{\pi_W / J} - y_j^{NS} \right)}. \tag{IA.6}$$

That is, social welfare is lower with full data sharing than with no data sharing.

We now compare the GDPR to no data sharing:

$$\begin{aligned} W^{GDPR} - W^{NS} &= \left(N y_{n,in}^{GDPR} + y_{n,out}^{GDPR} - y_n^{NS} \right) \frac{3}{8} \bar{u} + \left(J y_{j,in}^{GDPR} + y_{j,out}^{GDPR} - y_j^{NS} \right) u_B \\ &\quad + \pi_W y_j^{NS} \frac{\bar{v}}{8} - y_{j,out}^{GDPR} \left(\frac{1}{8} - \frac{\gamma^{*2}}{2} \right) \bar{v} \mathbf{1}_{\gamma^* < \frac{1}{2}}, \end{aligned} \tag{IA.7}$$

where y_n^{NS} and y_j^{NS} are independent of u_B. The first term in $W^{GDPR} - W^{NS}$ is positive, representing the improved matching with normal goods sellers under opt-in/opt-out, while the second is negative, reflecting the increased exposure of weak-willed consumers to predatory goods sellers.

Notice when $u_B = 0$, it must be the case $W^{GDPR} > W^{NS}$ because of the improved matching with normal goods sellers. When $u_B < 0$ however, the most-tempted weak-willed consumers suffer from lack of camouflage because not only all strong-willed, but also the more-mildly tempted weak-willed, consumers opt-in. Because the social benefit of GDPR from increased matching with normal goods sellers is bounded from above by $\left(Ny_{n,in}^{GDPR} - y_n^{NS}\right)\frac{3}{8}\bar{u}$, it follows for sufficiently negative u_B that $W^{GDPR} < W^{NS}$. Since the objectives are continuous, there exist critical values of u_B, u_{B*}, such that $W^{GDPR} < W^{NS}$ when $u_B \leq u_{B*}$.

We now compare the GDPR with the full data sharing. The difference in the social welfare is given by

$$W^{GDPR} - W^{FS} = \left(Ny_{n,in}^{GDPR} + y_{n,out}^{GDPR} - Ny_n^{FS}\right)\frac{3}{8}\bar{u} + \left(Jy_{j,in}^{GDPR} + y_{j,out}^{GDPR} - Jy_{j,out}^{FS}\right)u_B - y_{j,out}^{GDPR}\left(\frac{1}{8} - \frac{\gamma^{*2}}{2}\right)\bar{v}\mathbf{1}_{\gamma^* < \frac{1}{2}}. \tag{IA.8}$$

Note that under full data sharing, normal goods sellers have higher advertising efficiency and therefore are able to better cover their intended consumers, that is, the first term is negative. It is further clear that total advertising by predatory goods sellers under opt-in/opt-out is less than that under full data sharing, $Jy_{j,in}^{GDPR} + y_{j,out}^{GDPR} - Jy_j^{FS} < 0$. Because predatory goods sellers are less efficient at targeting the most-tempted customers, the coefficient of u_B in the second term is negative, i.e., the second term is positive. Consequently, there may exist a critical u_{B***} such that $W^{GDPR} > W^{FS}$ if $u_B \leq u_{B***}$ (and $W^{GDPR} < W^{FS}$ otherwise).

Ranking the three schemes: Suppose u_B is sufficiently severe ($u_B < \min\{u_{B*}, u_{B**}\}$), then no data sharing delivers the highest social welfare. Further, if u_B is in an intermediate range ($u_B < u_{B***}$ and $u_B > u_{B*}$), GDPR delivers the highest social welfare.

Comparing the welfare gap: We now consider the welfare gap Δ. From Equation

(11), the welfare gap under no data sharing is given by:

$$\begin{aligned}\Delta_{NS} &= \sum_{n=1}^{N} \frac{\pi_S - \pi_W}{N} y_n^{NS} \int_{p(n)}^{\bar{u}} u_n \frac{du_n}{\bar{u}} - \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} y_j^{NS} \left(u_B - p^{NS}(j)\right) \int_{p^{NS}(j)/\bar{v}}^{1} d\gamma_i \\ &\quad - \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} y_j^{NS} \int_0^{R_j^{NS}/\bar{v}} \left(u_B - \gamma_i \bar{v}\right) d\gamma_i = \frac{1}{8}\left(\pi_S - \pi_W\right) \bar{u} y_n^{NS} - \pi_W y_j^{NS} \left(u_B - \frac{3}{8}\bar{v}\right). \end{aligned} \tag{IA.9}$$

The welfare gap under full data sharing is

$$\begin{aligned}\Delta_{FS} &= \frac{1}{8}\left(\pi_S - \pi_W\right) \sum_{n=1}^{N} \bar{u} y_n^{FS} - \sum_{j=N+1}^{N+J} \frac{\pi_W}{J} \int_{\hat{\gamma}_j^{FS}}^{1} \left(u_B - \gamma_i \bar{v}\right) d\gamma_i \\ &= \frac{1}{8} N \left(\pi_S - \pi_W\right) \bar{u} y_n^{FS} - J y_j^{FS} u_B + J \bar{v} y_j^{FS} \left(1 - \frac{1}{2} \frac{y_j^{FS}}{\pi_W / J}\right). \end{aligned} \tag{IA.10}$$

It then follows:

$$\begin{aligned}\Delta_{FS} - \Delta_{NS} &= \frac{1}{8}\left(\pi_S - \pi_W\right) \bar{u} \left(N y_n^{FS} - y_n^{NS}\right) - \pi_W \left(\frac{y_j^{FS}}{\pi_W / J} - y_j^{NS}\right) u_B \\ &\quad + \pi_W \left(\frac{y_j^{FS}}{\pi_W / J} \left(1 - \frac{1}{2} \frac{y_j^{FS}}{\pi_W / J}\right) - \frac{3}{8} y_j^{NS}\right) \bar{v} \geq 0, \end{aligned} \tag{IA.11}$$

and the welfare gap is always positive because the first two terms are nonnegative (recall $u_B < 0$) and the last term is strictly positive (recall the revenue of predatory product sellers is always higher with full data-sharing).

We now establish that the welfare gap is also higher under GDPR than no data-sharing. First, notice the welfare gap from Equation(11 is divided into two pieces: the difference in utility from normal goods Δ^n_{GDPR} and the drag on weak-willed consumer welfare from predatory goods Δ^j_{GDPR}. Because there are fewer weak-willed than strong-willed consumers, and some weak-willed consumers opt-out, strong-willed consumers differentially benefit more from improved access to normal goods by opting-in. As such, the first term in the welfare gap $\Delta^n_{GDPR} > 0$ is higher under GDPR than no data sharing, i.e., $\Delta^n_{GDPR} > \Delta^n_{NS}$. In addition, because predatory goods sellers can better target weak-willed consumers when a subset opts-in, $\Delta^j_{GDPR} < 0$ is also more negative under GDPR (i.e., the profits of predatory goods sellers are higher), or $\Delta^j_{GDPR} > \Delta^j_{NS}$. Consequently:

$$\Delta_{GDPR} = \Delta^n_{GDPR} + \Delta^j_{GDPR} > \Delta^n_{NS} + \Delta^j_{NS} = \Delta_{NS}. \tag{IA.12}$$

As such, the welfare gap is smallest under no data sharing.

Internet Appendix B: A Dynamic Model of Data Sharing

Motivated by the analysis of Jones and Tonetti (2020), Cong, Xie, and Zhang (2020), and Abis and Veldkamp (2021) on the long-term effects of data sharing on economic growth, we extend our model to a dynamic setting in this section. We demonstrate that data sharing not only helps to boost long-run growth but may also exacerbate algorithmic inequality. Specifically, we first highlight a dynamic externality of data sharing on the platform, in which today's data sharing by consumers impacts the quality of goods offered by sellers to future consumers. We then use a calibrated exercise to examine the long-run implications of data sharing.

Suppose now that time is discrete with $t = 0, 1, 2, ...$ In each period, there is a new generation of consumers that join the platform. There are two sub-periods in each date that correspond to the two stages in our static model. In the first sub-period, consumers join the platform and make their data-sharing decisions with the platform. In the second sub-period, consumers can borrow from sellers, which target their intended customers based on the data they receive from the platform about the consumers. Similar to our static model, we assume the platform shares the consumer data authorized by consumers with sellers.

A key feature of the dynamic model is that more data allow each seller to enhance its good over time. That is, normal goods can improve in quality over time, and predatory goods can be made more tempting. If a seller financing normal good n collects data on a mass d_{nt} of the strong-willed and weak-willed consumers who prefer product n at time t, then the firm linked to the seller increases the quality of its product $\bar{u}_t$ according to an AR(1) process

$$\log \bar{u}_{t+1} = (1-\theta)\log \bar{u} + \theta \log \bar{u}_t + d_{nt}, \tag{IA.13}$$

where $\theta \in [0, 1]$ is the rate of mean reversion. That the impact of data on good quality decays over time reflects the idea that old data becomes obsolete, as discussed in Jones and Tonetti (2020) and Abis and Veldkamp (2021).

Similarly, data enables predatory goods firms linked to predatory sellers to make their products more enticing by utilizing big-data analytics to identify and exploit the behavioral vulnerabilities of their customers. For example, by analyzing the attention and clicking patterns of weak-willed consumers on their platform, these firms and sellers can tailor their marketing strategies to cater to these tendencies, enhancing the effectiveness of their offerings in attracting such consumers. In other words, if the company gathers data on a mass d_{jt} of the weak-willed customers who desire product j at time t, it can improve not only the allure of its good $\bar{v}_t$, but also the potential harm $u_{B,t}$, according to the following AR(1) processes:

$$\log \bar{v}_{t+1} = (1-\theta)\log \bar{v} + \theta \log \bar{v}_t + d_{jt}, \tag{IA.14}$$

$$\log(-u_{B,t+1}) = (1-\theta)\log(-u_B) + \theta \log(-u_{B,t}) + d_{jt}. \tag{IA.15}$$

In each period, the equilibrium follows what is characterized in Proposition 5 for the opt-in/opt-out scheme with $\bar{u}, \bar{v}$,and u_B being replaced by $\bar{u}_t$, $\bar{v}_t$, and $u_{B,t}$. The fraction of weak-willed consumers that opt in γ_t^* evolves over time, and can exhibit path dependence when consumers must coordinate over multiple potential opt-in/opt-out equilibria.

Dynamic Data-sharing Externality

Through the enhancement of the quality of both normal and predatory goods, data sharing by one generation of consumers may impose both positive and negative externalities on future generations of consumers. While data sharing by consumers contributes to better normal goods tomorrow, data sharing by weak-willed consumers also contributes to more-tempting predatory goods tomorrow. Consumers do not internalize this feedback loop between their data-sharing decisions and the quality of both types of goods, which can result in more weak-willed consumers opting in tomorrow, exacerbating algorithmic inequality. This creates a virtuous cycle for consumers and normal goods sellers, and a vicious cycle for consumers and predatory goods sellers. Although data advances the technological frontier over time, as illustrated by Jones and Tonetti (2020) and Abis and Veldkamp (2021), such improvements are not necessarily beneficial for all consumers and may worsen algorithmic inequality.

This feedback loop also highlights a dynamic aspect of the non-rivalry of data. Because the platform cannot commit to withholding its data from predatory goods sellers, normal goods sellers subsidize the data accumulation of predatory products sellers through the voluntary data sharing of strong-willed and moderately tempted consumers. Conversely, for predatory goods impede the data accumulation of normal goods sellers because of the opt-out decisions of severely-tempted consumers.

		Full Sharing	Intermediate Sharing	Minimal Sharing
$\bar{u}_\infty$		20564.3	19888.0	19824.0
$\bar{v}_\infty$		4800.3	4293.9	4248.0
$u_{B,\infty}$		-534.2	-477.8	-472.7
γ^*_∞		1.0	0.52	.47
Strong-willed	aggregate	1927.9	1864.5	1858.5
	per consumer	2570.5	2486.0	2478.0
Weak-willed	aggregate	-54.3	126.4	150.2
	per consumer	-217.1	505.8	600.8
Gap	aggregate	1982.2	1738.0	1708.3
Utilitarian		7599.4	6514.2	6415.7

Table 3: This table displays the steady-state values for $\bar{u}_t$, $\bar{v}_t$, and $u_{B,t}$, strong- and weak-willed consumer welfare, their difference, and utilitarian welfare for the full, intermediate, and minimal data-sharing equilibria under the parameters in Table 1.

A Calibrated Assessment

Under the opt-in/opt-out data-sharing scheme, multiple equilibria with vastly different levels of data sharing by consumers may exist because of the coordination problem among consumers discussed in Section 2.3. We now conduct a numerical exercise using the model parameters calibrated earlier to evaluate how various equilibrium paths might impact data accumulation and, consequently, consumer welfare in the long-run.

We initialize our economy at $t = 0$ with the parameters from Table 1, and choose an AR(1) parameter θ of 0.64 based on the observation of Abis and Veldkamp (2020) that standard accounting practices amortize data warehouses over 36 months. We then simulate the economy under the opt-in/opt-out scheme until it converges to the steady state. If multiple equilibria emerge, we assume consumers coordinate on the same cutoff over time.

Table 3 presents the simulation results, showing significant differences in the steady-state across the three equilibrium paths, which includes full data sharing, intermediate data sharing and minimal data sharing, as determined by the three levels of the equilibrium cutoff of weak-willed consumers. As consumer data accumulates over time, good quality increases, particularly in the full data-sharing path where all consumers share their data. For example, normal goods are 3.7% more valuable in the full data-sharing path compared to the minimal sharing path (higher $\bar{u}_\infty$), while predatory goods are 13.0% more tempting (higher $\bar{v}_\infty$) and 13.0% more harmful (more negative $u_{B,\infty}$). Because of the dynamic good quality, 47% of weak-willed consumers opt-in in the steady-state of the minimal sharing path, compared to 40% in the static equilibrium analyzed earlier. In contrast, only 52% of weak-willed consumers opt-in in the steady-state of the intermediate data-sharing path compared to 60% in the static equilibrium because of the behavior of the intermediate cutoff discussed in Section 2.3. Strong-willed consumers fare worse under the intermediate and minimal sharing paths, as less data is shared, leading to lower normal good quality in the long-run compared to the full data-sharing path.

Interestingly, the minimal sharing path mitigates algorithmic inequality compared to the full sharing path by reducing the welfare gap between strong- and weak-willed consumers by 13.8%, even though full data sharing results in 18.4% higher overall welfare. Data accumulation raises both utilitarian welfare and the welfare gap by 32.0% and 30.8% under the full data-sharing path and 29.9% and 25.6% under the minimal data-sharing path compared to the static equilibrium reported in Table 2. These magnitudes are of a similar order of magnitude to what Jones and Tonetti (2020) find across data sharing schemes in a representative agent production economy. Thus, more data sharing, both within and across time, not only raises efficiency and social welfare, but also increases algorithmic inequality.