Nonlinear Pricing with Under-Utilization: A Theory of Multi-Part Tariffs

Roberto Corrao* — MIT
Joel P. Flynn† — MIT
Karthik A. Sastry‡ — MIT

February 10, 2022

Abstract

We study the optimal nonlinear pricing of goods whose usage generates revenue but cannot be contracted upon. The optimal price schedule is a multi-part tariff, featuring tiers of usage within which buyers pay a price of zero on the margin. We apply our model to digital goods markets, in which advertising makes usage valuable, but monitoring legitimate usage is infeasible. Our results rationalize common pricing schemes including premium plans, free trials, and free products. Under partial contractibility of usage, there is an endogenous separation between "users," who face a multi-part tariff, and "workers," who are paid on the margin.

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: rcorrao@mit.edu.
†MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: jpflynn@mit.edu.
‡MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: ksastry@mit.edu.
This paper previously circulated as "Screening with Under-Utilization: Tiers, Add-ons, and Welfare Implications." We are grateful to Daron Acemoglu, George-Marios Angeletos, Charles Angelucci, Ian Ball, Dirk Bergemann, Alessandro Bonatti, Oğuzhan Çelebi, Glenn Ellison, Drew Fudenberg, Robert Gibbons, Giacomo Lanzani, Lucas Mayer, Stephen Morris, Marco Ottaviani, Alessandro Pavan, Nicola Rosaia, Jean Tirole, Michael Whinston, and Alexander Wolitzky, as well as participants in the MIT Theory Lunch, the MIT Industrial Organization Lunch, and the MIT Sloan Organizational Economics Lunch for helpful comments. First posted version: April 1, 2021.

1 Introduction

Digital goods are often priced according to multi-part tariffs—that is, they are sold in tiers of zero marginal prices, potentially interspersed with regions of positive marginal prices. For example, making Google searches or browsing Facebook posts is always free; streaming certain movies from Amazon Prime's library is free after paying for a subscription, while additional movies can be streamed for a price; and reading *Wall Street Journal* articles is free on the margin for anyone up to a trial limit, and free for paid subscribers in unlimited quantities. A common theme in each of these cases is that the seller can monetize the buyer's time and attention via other channels, like serving advertisements or collecting valuable data. This indirect revenue model is big business for the internet's largest players—for example, Google, Facebook, and Amazon respectively made 98%, 108%, and 60% of their net profit in Fiscal Year 2020 from advertisement.[1]

But the fact that indirect revenue *can* make zero-marginal-price units profitable does not explain why they are optimal. The classical theory of nonlinear pricing as screening (Mussa and Rosen, 1978; Maskin and Riley, 1984; Wilson, 1993) predicts that sellers should use smoothly varying marginal prices, instead of tiers of zero marginal prices, to extract maximal profits.[2] Understanding multi-part tariffs requires an alternative economic mechanism.

In this paper, we introduce a nonlinear pricing model with the following form of non-contractibility: once the buyer purchases the right to use a good, the seller cannot fully enforce the good's utilization. The scope for under-utilizing digital goods—and its potential influence on digital markets—is perhaps best illustrated by the historical failure of "pay-to-click" businesses that try to incentivize valuable usage (e.g., to pay people to see advertisements), only to be defrauded by users' simple cheating strategies (e.g., having a computer script click through a website).[3] Our model captures this key issue: providers cannot contract upon legitimate, valuable usage, or otherwise prevent fraudulent, valueless usage.

Our main result is that the optimal price schedule in the presence of under-utilization and usage-derived revenue is a multi-part tariff. Intuitively, non-contractibility of usage prevents sellers from charging negative marginal prices, which they would like to do to encourage valuable usage. Thus, they instead charge a price of zero on the margin. The remainder of

[1] Net advertisement revenue figures are analyst estimates by eMarketer (eMarketer Insider Intelligence, 2020), and net income is from financial statements as collected by the *Wall Street Journal*.

[2] As we clarify in Section 3, pooling of many buyer types, as is standard under "ironed solutions," is unrelated to the issue of zero marginal pricing: under pooling, many buyers purchase the same amount of the good for the same price, but additional units of the good still have a strictly positive marginal price. Models with discrete buyer types or bang-bang solutions can also generate *weakly optimal* multi-part tariffs, but make the counterfactual prediction that no buyer ever consumes in the region with zero marginal prices.

[3] In Section 2.2, we discuss the case study of the pay-to-click AllAdvantage.com, and how modern legal infrastructure is designed to prevent platforms from compelling users to engage with advertisements.

our analysis explores the structure of multi-part tariffs, their welfare implications, and the richer pricing structures that are possible when usage is partially contractible.

Model. As in the classical nonlinear pricing framework, buyers differ in their demand for the product, represented by a scalar, privately known type, and have quasilinear utility in money. Higher types correspond with higher demand for the product, embedded in the familiar assumptions that preferences are convex and satisfy single-crossing. The seller values transfers as well as usage-derived revenues from advertisement, data generation, platform externalities, and/or user addiction. To model non-contractibility of usage, we give buyers the ability to use less than they are allocated—for instance, if a buyer purchases the right to spend y hours on an online platform, they may choose to spend $x \leq y$ hours.

The seller chooses an arbitrary price schedule that assigns a price to each level of purchases. Buyers first decide how much to purchase, and then what to use, within the scope of what is permitted by the contract. We study the problem of how to design the price schedule optimally, taking both of the buyers' decisions into account.

Pricing with Non-Contractible Usage. We first characterize the seller's optimal pricing when usage is completely non-contractible, or when buyers can freely dispose of what they purchase. The induced levels of buyer usage in the optimum cap the seller's preferred usage (i.e., the optimum for the seller were usage fully contractible) with buyers' bliss points (i.e., what buyers would use were the product free).[4] The corresponding nonlinear price must be flat whenever buyers consume their bliss points, since the marginal value of additional usage to the buyer is zero. Thus, sellers price according to *multi-part tariffs*.

We next explore the structure of multi-part tariffs. We show zero marginal pricing applies to more units of the good when there are greater marginal revenues from usage (e.g., from advertising) and smaller marginal information rents, or costs of screening. Intuitively, usage-derived revenue makes higher usage more attractive, while information rents distort down the amount of usage for all but the highest type.

The *shape* of these competing effects determines where in the price schedule zero-marginal-price regions emerge. Figure 1 previews four of the pricing schemes that our model can generate. *Free pricing*, in which all units have zero marginal price (e.g., Google search or Facebook), occurs when marginal revenues from usage globally dominate marginal information rents. *Freemium pricing* or *free trial pricing*, in which initial units of the good have zero marginal price and all subsequent "add-on" units have positive marginal price (e.g., the mobile game "Candy Crush Saga"), occurs when usage-derived revenues are highly concave

[4]Formally, we show this under the technical conditions that virtual surplus is single-crossing in usage and agents' types, and that virtual surplus is strictly quasiconcave in usage. In Appendix C.1, we relax the first of these assumptions.

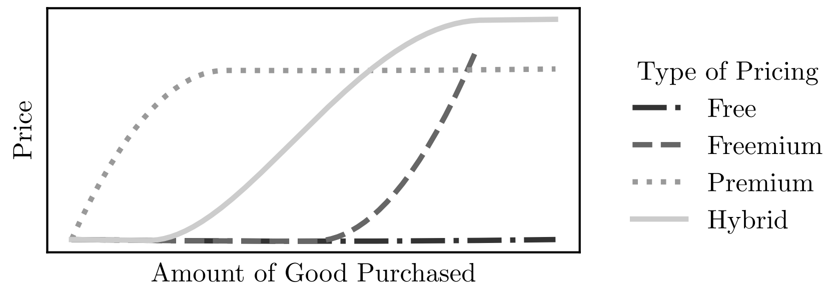

Figure 1: Example multi-part tariffs, which are derived in Section 3.

(e.g., because unique users generate valuable data) and overwhelm information rents for only low-usage buyers. *Premium pricing*, in which initial units of the good are sold for a strictly positive marginal price and subsequent units are sold for zero marginal price (e.g., Amazon Prime Video), occurs when usage-derived revenue dominates information rents for only high-usage buyers. Indeed, as information rents vanish for the highest-usage buyers, globally positive marginal revenue from usage always generates a "premium tier." Thus, *hybrid pricing* schemes like the combination of a free trial and premium plan (e.g., *The Wall Street Journal*) can occur with sufficiently concave and increasing revenues from usage.

We next use our results to understand the effects on buyer welfare of both contractible usage and changes in the structure of demand and revenue. First, while non-contractibility improves buyer welfare under a fixed price schedule, all buyers would obtain greater welfare with perfectly contractible usage under the corresponding optimal price schedule. The intuition for this result is that the lack of contractibility of usage, which enables buyers to escape being forced to use the good, also prevents them from being paid compensating differentials for usage, which leads to forgone gains from trade. Second, in the absence of contractible usage, buyer welfare is less sensitive to changes in marginal usage-derived revenue. These results echo popular claims, made especially about social media products, that users are not fairly remunerated for "being the product, not the consumer."[5] As we now argue, such remuneration is intimately linked to the nature of contractibility of usage.

Pricing with Partially Contractible Usage. Finally, we generalize this analysis to allow for different levels of contractibility corresponding to different levels of usage. We characterize implementation in this environment, use this result to study realistic forms of partial contractibility, and find that they can induce the combination of both negative and zero marginal prices. For example, contractibility of high levels of usage rationalizes the

[5]As one example, Apple co-founder Steve Wozniak had the following to say about why he deleted his personal Facebook account: "[Facebook's] profits are all based on the user's info, but the users get none of the profits back [. . . .] As they say, with Facebook, you are the product" (Guynn and McCoy, 2018).

combination of free services for low-intensity users with payments to high-intensity users. This mirrors the contract terms of YouTube's *Partner Program* and TikTok's *Creator Fund*, which pay a (nonlinear) wage per video view only to video creators with a sufficient level of following and activity. Instead, contractibility of low levels of usage rationalizes payments for sign-ups with zero marginal pricing for higher usage, like the Bing Rewards program that pays users a small amount for the first few searches they make per day. In both cases, contractibility of usage allows for platforms to separate "users" of a service from "workers" of the platform. The former face a multi-part tariff, whereas the latter face a negative marginal price that captures the information-rent-adjusted compensating differential of exerted effort in generating revenue for the principal.

Related Literature. The closest theoretical analysis to ours is by Grubb (2009), who demonstrates the optimality of three-part tariffs (or *freemium pricing*) when selling to overconfident consumers who can freely dispose of the purchased good.[6] In Online Appendix D.4, we show how our formal analysis nests and generalizes this model without necessarily appealing to behavioral foundations. By considering a richer class of external revenue functions and more general extents of contractibility, our analysis also endogenizes a richer set of pricing schemes and is applicable to a larger number of settings.

Our analysis relates to a literature on pricing under various constraints. Sundararajan (2004) studies non-linear pricing of information goods when the principal faces a "transaction cost" of measuring provision of the good, and derives an optimal tiered pricing schedule. Our analysis, by contrast, endogenizes multi-part tariffs as an optimal strategy in light of under-utilization. Amelio and Jullien (2012) and Choi and Jeon (2021) study markets with constraints for non-negative linear pricing and how bundling products across markets can effectively subvert such constraints. We complement this line of research by studying the *non-linear* pricing problem, albeit in a single market.

Our results fit into a theoretical literature on mechanism design with moral hazard (e.g., Forges, 1986; Laffont and Tirole, 1986; Fudenberg and Tirole, 1990; Carbajal and Ely, 2013; Strausz, 2017; Gershkov, Moldovanu, Strack, and Zhang, 2021). However, given its focus on the possibility of under-utilization, our model of *ex post* moral hazard has a specific structure that admits tractable analysis and, at the same time, has not previously been analyzed.

Outline. The rest of the paper proceeds as follows. Section 2 introduces our model. Section 3 solves for optimal contracts with non-contractible utilization and establishes their properties. Section 4 generalizes this to allow for partial contractibility. Section 5 concludes.

[6] Free disposal is also relevant for the sale of information goods, where agents may choose to optimally disregard information (see, e.g., Bergemann, Bonatti, and Smolin, 2018).

2 Model

2.1 Consumer Demand

There is a single good that can be bought and consumed in amounts x in the interval $X = [0, \overline{x}] \subset \mathbb{R}$. There is a unit measure of consumers with privately known type $\theta \in \Theta = [0, 1]$ that parameterizes their demand. The type distribution $F \in \Delta(\Theta)$ admits a density f that is bounded away from zero on Θ. For example, x might be the time that an agent spends on an online platform, and θ shifts how much they enjoy this activity.

Consumers' type-specific preferences over consumption are represented by a twice continuously differentiable utility function $u : X \times \Theta \to \mathbb{R}$. We assume that higher types value consumption more and that all types have single-peaked preferences over consumption with the following three conditions: (i) $u(\cdot)$ satisfies strict single-crossing in (x, θ); (ii) for each $x \in X$, $u(x, \cdot)$ is monotone increasing over Θ; and (iii) for each $\theta \in \Theta$, $u(\cdot, \theta)$ is strictly quasiconcave over X. All consumer types value zero consumption the same as their outside option payoff, which we normalize to zero, or $u(0, \theta) = 0$ for all types $\theta \in \Theta$. Agents have quasilinear preferences over consumption and money $t \in \mathbb{R}$, so their transfer-inclusive payoff is $u(x, \theta) - t$.

2.2 Under-Utilization

The primary departure of our analysis from traditional non-linear pricing is the ability of consumers to under-utilize what they buy at zero cost. To model under-utilization we define a *disposal correspondence* $D : X \rightrightarrows X$, which maps what consumers purchase $y \in X$ to a *disposal set* $D(y) \subseteq [0, y]$. A consumer buying y can consume any $x \in D(y)$. The extent of feasible under-utilization is determined by the scope of enforceable contracts to condition on ultimate consumption of the good, as described mathematically by the structure of D. For example, the standard screening model with perfectly contractible consumption is recovered in this setting with $\underline{D}(y) \equiv \{y\}$, which imposes that goods cannot be under-utilized. By contrast, we are primarily interested in the case of unrestricted under-utilization, or *free disposal*, which we model as $\overline{D}(y) \equiv [0, y]$.

We argue free disposal describes feasible contracting in digital goods markets, our primary application. For example, the *Wall Street Journal* can measure if a given consumer loads an article, but not if they actually read it. Likewise, Google can register that a search has been made, but not that this is done by a human as opposed to a bot.

Perhaps the most direct evidence for the non-contractibility of consumption is provided by the failure of "pay-to-click" internet businesses. One case study is the rise and fall of

AllAdvantage.com, a venture that paid users to view a permanent banner ad when browsing the internet. *The New York Times*, who interviewed the company's founder as well as eager customers, asked "Can it Pay to Surf the Web?" in a July 1, 1999, headline (Guernsey, 1999). But AllAdvantage.com was quickly bogged down by users' finding simple ways to automate web surfing. This was concisely summarized in the headline of a *Wired* magazine article from July 10, 2000, which (unintentionally) answers the *Times*' original question: "It Pays to Cheat, Not to Surf" (Kang, 2000). In our model's language, consumers could purchase y hours of time browsing the internet with the AllAdvantage banner, but then under-utilize to consume $x \leq y$ hours, with the residual $y - x$ hours handled by bots.

Other settings of intermediate contractibility of utilization are natural and arise, for example, when platforms such as YouTube and TikTok monitor high levels of content production. We return to these cases in Section 4.

2.3 Production and Revenues

The seller's revenue derives from two sources. The first is the sum of monetary payments $t \in \mathbb{R}$ from consumers to the seller. The second is revenue that derives from consumers' usage of the good, net of production costs. This second source is represented by a continuously differentiable $\pi : X \times \Theta \to \mathbb{R}$.[7] The seller values zero consumption the same as their outside option revenue from not selling the product at all, which we normalize to zero, or $\pi(0, \theta) = 0$ for all $\theta \in \Theta$.

We have four primary justifications of usage-dependent revenue in the digital goods case. In practice, we think of our model's π as a reduced-form representation of the combination of potentially all four mechanisms that best characterizes a given market.

Advertisement. Digital goods are commonly bundled with revenue-generating advertisements. For example, Google search results, Facebook social feeds, and *Wall Street Journal* articles all include advertisements. In these and other online settings, advertisers can directly measure both the number of times an advertisement is loaded (*impressions*) and the number of times an advertisement is clicked. Payments from the advertiser to the platform commonly depend *ex post* on both impressions and clicks per impression (*click through rate*).

We model these payments via our π as functions of platform consumption x, rather than purchases y—a *Wall Street Journal* user must load the article containing an advertisement, and perhaps click on it, to register a payment. We argue that the metaphor also applies when human and automated usage of online platforms may be substituted for one

[7]It may be reasonable to assume that production costs depend on the *allocation* rather than consumption of the good. But if costs are monotone, it is straightforward to argue that the principal will never produce more than is consumed and "waste" the product, leading to a representation of costs in terms of usage.

another. The aforementioned AllAdvantage.com failed because advertisers refused to pay out for inauthentic, bot-derived clicks. Modern advertisement contracts, having internalized the mistakes of the AllAdvantage era, tie payouts explicitly to "valid," human-derived clicks and impressions. As one example, the terms and conditions of Google AdSense, a popular service for adding advertisements to a webpage, define invalid activity as that "solicited or generated by payment of money, false representation, or requests for end-users to click on Ads or take other actions" (Google AdSense, 2020). For the erstwhile AllAdvantage.com, or any modern website monetized via AdSense, only human consumption x translates into revenue, while any bot-derived residual $y - x$ does not.

The function π is a possibly type-dependent mapping from usage to advertisement revenue. This subsumes details of consumer behavior and the advertisement contract, such as the rate with which consumers click advertisements and the payment per click. We further justify this mapping in the Online Appendix, where we explicitly model bargaining between advertisers and the platform (D.1) and the vendor's choosing advertising intensity to balance external revenue with depressed consumer demand (C.3).

Data Collection. A trend in digital advertising over the last decade is the rise of *targeted advertisements* tuned toward individuals' interests as revealed by their online activity.[8] This phenomenon has helped open up a "data economy" in which producers profit from collecting information about consumers, either directly via selling data to marketing intermediaries or indirectly via applying data toward internal advertisement. As in the advertisement case, we use the function π to model the reduced-form "usage to revenue" schedule subsuming the translation of usage into data and the valuation of that data.

Platform Externalities. Social media platforms, content-streaming platforms with social rating systems (e.g., Netflix and Hulu), matching services (e.g., Tinder and Match.com) and online games (e.g., Fortnite and Candy Crush Saga) rely on active use to boost the appeal of their product. In Appendix D.2, we describe how a simple model of platform externalities can micro-found an external revenue function π by affecting all agents' willingness to pay to participate in the platform. The framework can accommodate locally positive social externalities, as suggested by the previous examples, as well as negative externalities, due for instance to crowd-out or congestion.

Addiction. Conventional wisdom and recent empirical evidence (Allcott, Gentzkow, and Song, 2021) suggest that addicted users are a major source of demand for phone apps and social media services. In particular, assume that the quantity $x' \in X$ purchased tomorrow by

[8] A report by IHS Markit estimated that, in Europe, advertising that used behavioral data comprised 86% of all programmatic digital advertising (IHS Markit, 2017)

each consumer is increasing in their "addictive" consumption today $x \in X$, establishing an indirect link between current consumption x and future payments t'. Our revenue function π captures this effect in reduced form. In Appendix D.3, we show how selling to myopic consumers with habit formation gives rise to an identical nonlinear pricing problem to the one we study, where π is the future revenue obtained by addicting agents today.

2.4 The Nonlinear Pricing Problem

The seller's problem is to design a total revenue maximizing price schedule $T : X \to \overline{\mathbb{R}}$, where $T(y)$ is the monetary payment from a consumer purchasing $y \in X$. Following the choice of T, each type $\theta \in \Theta$ chooses whether to buy anything, how much to purchase $\xi(\theta) \in X$, and how much to ultimately consume $\phi(\theta) \in D(\xi(\theta))$. As is standard, we assume that the purchase and consumption functions $\xi : \Theta \to X$ and $\phi : \Theta \to X$ are the revenue-maximizing selections from the buyers' demand correspondence. Hence, the seller's problem can be formulated as:

$$
\begin{aligned}
\sup_{\phi,\xi,T} \quad & \int_{\Theta} \left(\pi(\phi(\theta),\theta) + T(\xi(\theta))\right) \mathrm{d}F(\theta) \\
\text{s.t.} \quad & \phi(\theta) \in \arg\max_{x \in D(\xi(\theta))} u(x,\theta) \quad \text{for all} \quad \theta \in \Theta && \text{(O)} \\
& \xi(\theta) \in \arg\max_{y \in X} \left\{ \max_{x \in D(y)} u(x,\theta) - T(y) \right\} \quad \text{for all} \quad \theta \in \Theta && \text{(IC)} \\
& u(\phi(\theta),\theta) - T(\xi(\theta)) \geq 0 \quad \text{for all} \quad \theta \in \Theta && \text{(IR)}
\end{aligned}
\tag{1}
$$

The first constraint (O) in Problem 1, or Obedience, establishes that each consumer θ chooses their optimal level of consumption $\phi(\theta)$ by optimally under-utilizing their initial purchase $\xi(\theta)$ to the extent allowed by the disposal correspondence D.[9] The second constraint (IC), or Incentive Compatibility, embodies the consumers' optimal purchase $\xi(\theta)$, taking into account their subsequent ability to under-utilize. The final constraint (IR), or Individual Rationality, ensures that all consumers are willing to participate.[10]

3 Optimal Pricing with Free Disposal

We now characterize optimal pricing under free disposal, $D(x) = \overline{D}(x) = [0, x]$. This models our primary application to pricing digital goods for the reasons described in Section 2.2.[11]

[9] We use the word "obedience" in the sense of Myerson (1982).

[10] Ensuring participation of all types is without loss of optimality for the seller owing to their ability to sell nothing and charge a price of zero, given the outside option of zero.

[11] Free disposal also serves as robust benchmark in the presence of partial contractibility of utilization as it corresponds to the worst possible case of partial contractibility for the seller.

We provide conditions under which the optimal price schedule endogenously takes the form of a multi-part tariff with discrete tiers of provision. We use this structure to provide sufficient conditions for optimal pricing to reduce to the freemium, premium and fixed cost pricing plans often observed in practice. Finally, we study the implications of free disposal and changes in external revenues for consumer and producer welfare.

3.1 The Optimal Price Schedule

We first define some important objects in which the optimal pricing schedule will be expressed. First, we define the consumer-optimal consumption of type θ as the function $\phi^A : \Theta \to X$:

$$\phi^A(\theta) = \arg\max_{x \in X} u(x, \theta) \tag{2}$$

which is unique and strictly increasing because of the strict quasiconcavity of $u(\cdot, \theta)$ in x for all $\theta \in \Theta$ and strict single-crossing of u in (x, θ). For example, $\phi^A(\theta)$ is the amount of time that type θ would optimally spend on the digital platform if it were freely available.

Second, we define the virtual surplus function $J : X \times \Theta \to \mathbb{R}$,

$$J(x, \theta) = \pi(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta) \tag{3}$$

which is simply the total surplus $\pi + u$, net of information rents, in the form of discounts to consumers, required to ensure local incentive compatibility. For the remaining analysis, we will assume that the function J satisfies *single-crossing* in (x, θ) and is *strictly quasiconcave* in x. These standard technical assumptions guarantee that virtual surplus has a unique maximum and is maximized pointwise under the optimal contract, thereby ruling out cases with bunching, or multiple agent types' consuming the same bundle.[12] Therefore, the producer-optimal consumption level $\phi^P : \Theta \to X$:

$$\phi^P(\theta) = \arg\max_{x \in X} J(x, \theta) \tag{4}$$

is unique and strictly increasing. In a counterfactual world with perfect contractibility, the revenue-maximizing seller would induce a type-θ consumer to spend $\phi^P(\theta)$ time on the platform.

With these definitions in hand, we state our result describing optimal pricing:

[12] In Appendix C.1, we relax the single-crossing assumption on J, strengthen strict quasiconcavity to strict concavity, and show how to adapt our analysis to settings which feature canonical bunching.

Proposition 1 (Optimal Pricing). *In any optimal contract, consumption is given by:*

$$\phi^* = \min\{\phi^P, \phi^A\} \tag{5}$$

The minimal optimal price schedule is monotone and given by:

$$T^*(x) = u(x_*, 0) + \left[\int_{x_*}^{\min\{x, x^*\}} u_x(s, {\phi^*}^{-1}(s))\, \mathrm{d}s \right]^+ \tag{6}$$

where $x_* = \phi^*(0)$, $x^* = \phi^*(1)$, and $[\cdot]^+$ denotes the positive part. Moreover, purchases are part of an optimal contract if and only if they are a selection from the correspondence $\Xi_{\phi^*} : \Theta \rightrightarrows X$:[13]

$$\Xi_{\phi^*}(\theta) = \begin{cases} \{\phi^*(\theta)\} & \text{if } \phi^*(\theta) < \phi^A(\theta) \\ \left[\phi^A(\theta), \inf_{\theta' \in [\theta,1]} \left\{\phi^*(\theta') : \phi^*(\theta') < \phi^A(\theta')\right\}\right] & \text{if } \phi^*(\theta) = \phi^A(\theta) \end{cases} \tag{7}$$

Thus, consumption in any optimal contract is simply the producer-optimal consumption level capped by the consumer-optimal consumption level. The remainder of the result derives corresponding optimal price schedules and levels of purchases from this optimal consumption function. We analyze the properties of this contract in the following sections.

We prove Proposition 1, as part of a more general result (Proposition 7), in three steps in Appendix A. In Appendix A.1, we extend the Taxation Principle (Guesnerie and Laffont, 1984) to our environment with disposal correspondences, obtaining an additional monotonicity restriction on the price schedule. In Appendix A.2, we use this *monotone taxation principle* to establish that a consumption function ϕ is implementable if and only if it is monotone and capped by ϕ^A. In Appendix A.3, we reduce the monopolist's problem to the control problem of maximizing virtual surplus (3) under the constraint $\phi \leq \phi^A$, show that the pointwise monotone solution is $\phi^* = \min\{\phi^A, \phi^P\}$, and show that the optimal price schedule is uniquely identified over the offered menu by the envelope formula in Equation 6. Finally, to characterize the corresponding purchases, we observe that whenever $\phi^*(\theta) = \phi^A(\theta)$, the monopolist can offer to buyer θ any amount up to

$$\overline{\xi}(\theta) = \inf\left\{\phi^P(\theta') : \theta' \in [\theta, 1], \phi^P(\theta') < \phi^A(\theta')\right\} \tag{8}$$

while still satisfying the incentive constraints in Problem 1.

[13] With the convention that if the set over which the infimum is taken is empty, then the infimum is equal to the supremum of the codomain of the relevant objective function. For example, ∞ for $\mathbb{R}$, and 1 for $[0, 1]$.

Importantly, a possible optimal solution for the seller is to offer only the tiered levels of consumption in the set $\overline{\xi}(\Theta)$ at the corresponding prices given by T^*. By construction, each consumer θ will purchase the corresponding tier $\overline{\xi}(\theta)$ and then consume only up to $\phi^*(\theta)$. This selection $\overline{\xi}$ can be always justified in the presence of arbitrarily small menu costs for the seller and will later help us rationalize the presence of several discrete tiers of service.

We provide a more specialized intuition for the form of the optimal contract in an example:

Example 1 (Digital Platform with Advertisement)**.** A digital platform sells access time $x \in [0, 1]$. Consumers have quadratic payoffs

$$u(x, \theta) = \theta x - \frac{x^2}{2} \tag{9}$$

where θ is uniformly distributed on $\Theta = [0, 1]$. A consumer who spends time x on the platform clicks on $x(k - cx)$ advertisements, for $k > 0$ and $c \in [0, k]$. The assumption $c \geq 0$ embodies user fatigue from seeing the same advertisements repeatedly, and $c \leq k$ ensures that total clicks remain positive. Each click yields revenue $p > 0$ for the seller. The principal's revenue function is therefore

$$\pi(x, \theta) = \frac{\text{Revenue}}{\text{Click}} \cdot \frac{\text{Clicks}}{\text{Time}} \cdot \text{Time} = p \cdot (k - cx) \cdot x = pkx - pcx^2 =: \alpha x - \frac{\beta}{2} x^2 \tag{10}$$

where, in the last equality, we define $\alpha = pk$ and $\beta = 2pc$ for simplicity.

We can use Proposition 1 to solve for the seller's optimal pricing in this setting. The seller-preferred consumption, which maximizes virtual surplus, is

$$\phi^P(\theta) = \max\left\{0, \min\left\{1, \frac{\alpha + 2\theta - 1}{\beta + 1}\right\}\right\} \tag{11}$$

The consumer-preferred demand is $\phi^A(\theta) = \theta$. Therefore the consumption implemented in the optimal contract is

$$\phi^*(\theta) = \min\left\{\phi^A(\theta), \phi^P(\theta)\right\} = \begin{cases} \phi^P(\theta) & \text{if } \theta < \frac{1-\alpha}{1-\beta} \\ \phi^A(\theta) & \text{if } \theta \geq \frac{1-\alpha}{1-\beta} \end{cases} \tag{12}$$

In the limit case of $\alpha = pk > 1$, which can be guaranteed by a sufficiently high revenue per click or intercept of the clicks-per-time function, the previous two expressions reduce to $\phi^*(\theta) = \phi^A(\theta)$ and $T(x) \equiv 0$. In words, the platform is free and consumers spend as much time as they wish on it. At the other extreme, if $\alpha < \beta$ or $k < 2c$ (corresponding to a low click-per-time intercept and/or high fatigue rate), $\phi^* = \phi^P$ and the tariff is always

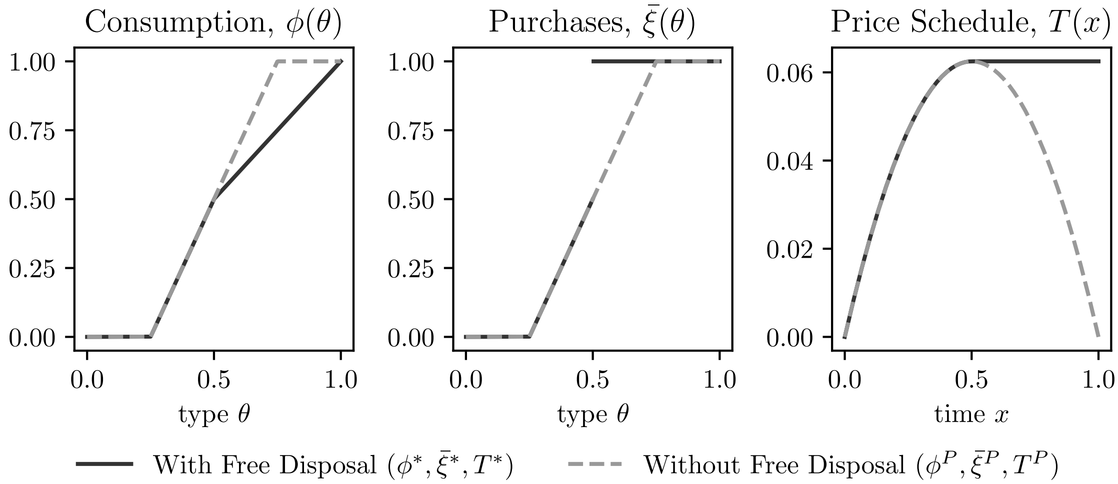

Figure 2: Optimal contracts with and without free disposal ($\alpha = \frac{1}{2}, \beta = 0$)

increasing. This is a "conventional" pricing scheme with positive marginal prices determined by the usual considerations of screening. Finally, all other cases are a hybrid between the two extremes: the tariff increases to the point $x = \frac{1-\alpha}{1-\beta}$ via the "conventional" formula and is flat thereafter, meaning that marginal units $x \in \left(\frac{1-\alpha}{1-\beta}, 1\right]$ have a zero marginal price.

We illustrate such a hybrid case in Figure 2 with $\alpha = \frac{1}{2}$ and $\beta = 0$. We show $\phi^*(\theta)$ in the leftmost panel and $T^*(x)$ in the right-most panel with a solid line. In the middle panel, we illustrate the "maximum purchase schedule" defined in Equation 8, which provides the good in metered quantities up to $x = \frac{1-\alpha}{1-\beta} = \frac{1}{2}$ and then in a maximum quantity $x = 1$ at a fixed price $T^{\max} = T^*\left(\frac{1}{2}\right) = \frac{1}{16}$. The region of zero marginal pricing can be interpreted as a "premium tier" of unlimited usage, an idea we will expand upon more generally in later analysis (Section 3.3).

We can contrast these predictions with those in a variant market with the same demand and external revenues, but no potential for free disposal (i.e., the legitimate time spent on the platform is perfectly contractible). The seller implements a contract in which agents consume $\phi^*(\theta) = \phi^P(\theta)$. The corresponding price schedule may entail negative marginal and/or average prices depending on (α, β). We illustrate, in the same case of $\alpha = \frac{1}{2}$ and $\beta = 0$, the consumption, purchases, and prices in the dashed lines of Figure 2. Due to the non-monotonicity of prices, $x = 0$ and $x = 1$ are both priced for free. This starkly demonstrates the incentives for sellers in digital markets to use negative prices to induce valuable usage, which would be enforceable absent free disposal. $\triangle$

3.2 Rationalizing Multi-Part Tariffs

We now generally characterize when the optimal tariff is a multi-part tariff. We first formally define the concepts of *flatness*, *flat intervals*, and *tiers*. To this end, we define the space of consumed outcomes $X^* = \phi^*(\Theta) \subseteq X$ as the image of ϕ^*.[14] In what follows, with an abuse of notation, we let $T^* : X^* \to \mathbb{R}$ denote the restriction of the optimal price schedule over the space of consumed outcomes X^*, and call it the price schedule.[15]

Definition 1 (Flatness, Flat Intervals, and Tiers)**.** *A price schedule $T : X^* \to \mathbb{R}$ is flat at $x \in X^*$ if there exists an $\varepsilon > 0$ such that $T(x') = T(x)$ for all $x' \in (x - \varepsilon, x + \varepsilon) \cap X^*$. The interval $I \subseteq X^*$ is a flat interval if the price schedule is flat at all points $x \in I$. The interval $R \subseteq X^*$ is a tier if it is a maximal flat interval.*

Flatness of prices is more demanding than zero slope, or $T'(x) = 0$ wherever T' is defined, as it must apply on a positive measure interval around x.[16] Tiers are defined in terms of the price schedule, which is the relevant observable outcome of interest in our analysis.

To characterize when optimal price schedules are flat, we first define the constrained marginal revenue function $H : X^* \to \mathbb{R}$ that maps outcomes to the marginal revenue of the principal in the type who most prefers that outcome:[17]

$$H(x) = J_x\left(x, (\phi^A)^{-1}(x)\right) \tag{13}$$

To interpret H, imagine that the principal had to offer the product for free. The sign of $H(x)$ determines whether the principal would profit from more ($H(x) > 0$) or less ($H(x) < 0$) consumption from the type whose favorite consumption is x. The following result links flat pricing with this trade off.[18]

Proposition 2 (Flatness Characterization)**.** *If $H(x) > 0$, then the optimal price schedule T^* is flat at $x \in X^*$. Conversely, if the optimal price schedule T^* is flat at $x \in X^*$, then $H(x) \geq 0$.*

Proof. See Appendix A.4. □

[14] Given continuity of ϕ^*, the set X^* is an interval.

[15] Even if T^* is defined over the whole X, the outcomes $x \in X \setminus X^*$ are never consumed in the optimal allocation. Therefore, our restriction is without loss for exploring pricing and welfare implications.

[16] Thus, zero marginal transfers or tariffs "at the top," meaning a specific maximal point in the type or action space, do not imply flatness by our definition—no unit of size $\epsilon > 0$ is sold for zero price.

[17] Given an increasing function $\phi : \Theta \to X$, we denote with ϕ^{-1} its generalized inverse $\phi^{-1}(x) = \inf\{\theta \in \Theta : \phi(\theta) \geq x\}$.

[18] It is not possible to strengthen the claim that if optimal price schedule T is flat at $x \in X^*$, then $H(x) \geq 0$ and make the inequality strict. The proof of Proposition 2 provides an explicit counterexample.

To understand this result, we observe that the condition $H(x) > 0$ corresponds to the following comparison of two terms evaluated at x and $\theta = (\phi^A)^{-1}(x)$:

$$\underbrace{f(\theta)\,\pi_x(x,\theta)}_{\text{Marginal Revenues}} > \underbrace{(1-F(\theta))\,u_{x\theta}(x,\theta)}_{\text{Marginal Information Rents}} \tag{14}$$

Condition 14 holds when the marginal revenues from increasing x, multiplied by the density of consumers, exceed the increase in information rents paid to all higher-type consumers. Steepening the revenue function boosts the former force. Reducing complementarity or increasing $F(\cdot)$ in the hazard rate order reduces the latter one. Without positive marginal revenues, the second force would always win and the price schedule would never be flat.

This result allows us to understand how changing advertising revenues matter for the structure of zero marginal pricing. To illustrate this, we revisit Example 1:

Example 1 (continuing from p. 11)**.** The constrained marginal revenue function is $H(x) = (\alpha - \beta x) - (1 - x)$, where the first term is the marginal revenue from additional usage and the second is marginal increase of information rents. Zero marginal pricing therefore occurs on the set of consumption levels where $H > 0$. Globally increasing marginal revenues via higher α and/or lower β increases the size of this set. These comparative statics amount to increasing the revenue per click or decreasing the speed at which click-through-rate decreases per session (i.e., user fatigue). In this way, a more lucrative advertising business causes more units to be optimally priced at zero marginal price. △

Next, we use our characterization to determine how many parts the optimal multi-part tariff has. We say that a price schedule with N tiers is a generalized $N+2$-*part tariff*. This builds on the conventional definition of a *three-part tariff* in which there is a fixed cost, an initial allotment of zero-marginal-price goods, and then additional units with marginal positive price.[19] Our more general definition can account for zero-marginal-pricing away from the "bottom" of the contract. We observe, building on the characterization of *locally* flat pricing in Proposition 2, that the number of tiers depends on the number of times that H changes sign. Formally, we define the number of times H changes sign from below as:

$$k(H) = |\{x \in X^* : \exists \epsilon > 0, \forall y \in (x-\epsilon, x), \forall y' \in (x, x+\epsilon), H(y) < 0, H(y') \geq 0\}| \tag{15}$$

where this definition simply rules out cases where H hits 0 at a single point only (which does not lead to a flat tariff) and restricts to the earliest x at which a sign change happens

[19] A *two-part tariff*, via the conventional definition, combines fixed costs with positive marginal costs. This of course can be accommodated in the conventional non-linear pricing framework, with zero tiers, as the "intercept" of the tariff is a free parameter.

to prevent double-counting of tiers.

Corollary 1 (Multi-Part Tariffs)**.** *If $H(\min X^*) > 0$, respectively $H(\min X^*) < 0$, then the price schedule T is a $3 + \left\lfloor \frac{1}{2}k(H) \right\rfloor$, respectively $2 + \left\lceil \frac{1}{2}k(H) \right\rceil$, multi-part tariff.*

Proof. See Online Appendix B.1. □

Our model, depending on primitives and their effects on the function H, can therefore rationalize a diversity of price schedules combining tiers with marginally priced "add-ons" as strictly optimal for the principal.

Remark 1 (Standard models do not generate multi-part tariffs)**.** Strictly optimal multi-part tariffs are not possible in the canonical Mussa and Rosen (1978) and Wilson (1993) screening models with a (convex) continuum of agent types.[20] A *weakly optimal* multi-part tariff is possible, for instance, in a specialization with a discrete number of types, by extending the domain of the offered menu with a constant tariff. But we argue this is not economically meaningful, or relevant for our applications, for three reasons. First, no type would ever consume any outcome in the extended menu that does not lie in the initial menu, therefore ruling out variability of consumption within a pricing tier. This is clearly counterfactual—within the single tier of a zero-price product like Facebook, there is large variability in time spent on the platform. Second, there would be as many parts to the tariff as unique consumer types, which is an arbitrary choice of the modeler. This prevents meaningful comparative statics for the number of observed tiers as a function of primitives. Third, there is no principled reason for arguing that the multi-part tariff is the "right" selection from the set of optimal price schedules which are extended off-menu.

Remark 2 (Bunching is unrelated to multi-part tariffs)**.** As a technical point, we emphasize that *optimal bunching* in standard screening models is a different phenomenon from optimal multi-part tariffs. Intuitively, bunching is a feature that occurs in the type space Θ, where many different types buy the same amount. However, all units of the good are still sold at a strictly positive marginal price and so the optimal tariff is never flat. In Appendix C.1, we solve for the optimal contract with bunching when J does not satisfy single-crossing in (x, θ) and is strictly concave in x by adapting the method of Nöldeke and Samuelson (2007). Example 1 provides an explicit illustration of how bunching is unrelated to the issue of zero marginal pricing: multiple buyers bunch on buying nothing, but there are still strictly positive marginal prices for the first marginal units of the good (see Figure 2).

[20] Even if the virtual surplus is not strictly-concave and a multi-part tariff is weakly optimal, the same argument of this remark would apply.

3.3 The Structure of Multi-Part Tariffs

We now use our characterization to study where in the price schedule zero marginal pricing is optimal. In the process, we rationalize a number of commonly used price schedules.

3.3.1 Cut-off Price Schedules

We first restrict to a simple setting in which H crosses zero at most once. In this setting, there are four exhaustive possibilities for the global shape of the price schedule, based on cut-offs at which zero marginal pricing switches on or off.

Corollary 2 (Cut-off Pricing)**.** *The following statements are true:*

1. *If $H(x) \leq 0$ for all $x \in X^*$, then T is strictly increasing everywhere.*
2. *If $H(x) \geq 0$ if and only if $x \leq \hat{x} \in int(X^*)$, then T has a single tier $[\min X^*, \hat{x})$, and is strictly increasing on $[\hat{x}, \max X^*]$.*
3. *If $H(x) < 0$ if and only if $x < \hat{x} \in int(X^*)$, then T has a single tier $[\hat{x}, \max X^*]$ and is strictly increasing on $[\min X^*, \hat{x})$.*
4. *If $H(x) \geq 0$ for all $x \in X^*$, then T has a single tier X^*.*

Proof. See Online Appendix B.2. □

The first case corresponds to a standard, strictly positive marginal price. The second corresponds to a classic three-part tariff: initial units are provided at a fixed price plus zero marginal price, and subsequent units are provided at a positive marginal price. A common specialization of this strategy in digital goods contexts, under the added restriction that the fixed price is zero, is a *freemium* ("free plus premium") or *free trial* pricing strategy. The third case is an inverted three-part tariff, where initial units are provided at a positive marginal price and larger quantities are bundled together at a fixed price. While such schemes are rare for standard physical goods, they are common in digital goods contexts as *unlimited subscriptions*. The final case is a completely free product. Such products are of course ubiquitous online, mainly as free, ad-supported services in search (e.g., Google) and social media (e.g., Facebook, Twitter, and Instagram).

We can use Example 1 to illustrate these four pricing cases numerically:

Example 1 (continuing from p. 11)**.** The function $H(x) = (\alpha - \beta x) - (1 - x)$ satisfies single-crossing. The four cases of Corollary 2 exhaustively summarize the possibilities in the model.[21] In Figure 3, we illustrate each of these cases as parameterized by different choices of (α, β). We name the cases based on the interpretation sketched above. △

[21] Case 1 arises if $\alpha \leq 1$ and $\alpha \leq \beta$; case 2 if $\alpha > 1$ and $\alpha \leq \beta$; case 3 if $\alpha \leq 1$ and $\alpha > \beta$; and case 4 if $\alpha > 1$ and $\alpha > \beta$.

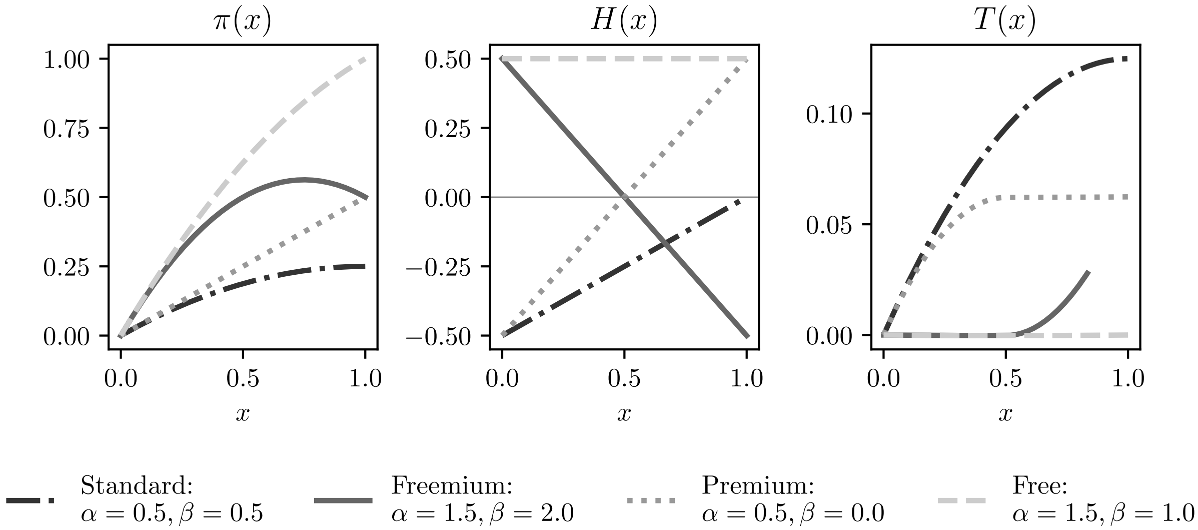

Figure 3: Cut-off price schedules in Example 1

3.3.2 General Price Schedules: Premium and Trial Tiers

We next derive sufficient conditions for zero marginal pricing at the bottom and top of the menu. We say a tariff T features a *premium tier* if it is flat at $\max X^*$, the highest level of consumption under the optimal contract, and a *trial tier* if it is flat at $\min X^*$, the lowest level of consumption under the optimal contract.

Corollary 3 (Premium and Trial Tiers)**.** *The price schedule has a premium tier if*

$$\pi_x(\phi^A(1), 1) > 0 \tag{16}$$

The price schedule has a trial tier if

$$f(0)\,\pi_x(\phi^A(0), 0) > u_{x\theta}(\phi^A(0), 0) \tag{17}$$

Proof. See Online Appendix B.3. □

The first part shows that the premium tier emerges as optimal any time that marginal external revenues are globally positive. The lack of a countervailing force from marginal information rents reflects the fact that the principal does not distort allocations of the highest-type agents away from the first-best surplus-maximizing allocation.

This prediction matches standard pricing for online news and content streaming (e.g., *Wall Street Journal* or Netflix subscriptions), which rarely cap usage in the highest-paid tier. The prediction also understandably breaks down for the "analog" version of each product, which necessarily entails production costs and therefore non-positive marginal revenues beyond a point—a print *Wall Street Journal* subscription does not promise access to unlimited paper copies of the *Journal* and all possible supplements, and the original DVD-by-post incarnation of Netflix had limits on videos per customer.

The second part shows that trial tiers require a more stringent condition, whereby marginal revenue (multiplied by the mass of low-type agents) exceeds the marginal information rent paid to *all* higher types. With zero or negative marginal revenues, this condition would never hold. Otherwise, it is more likely to hold in environments with high marginal revenues stemming from *low* consumption. This comparative static may be illustrated by comparing digital video streaming and cable television. Platforms for the former, like Netflix and Hulu likely have a large "extensive margin" of advertising and/or data collection revenue from small levels of usage. Consistent with the model, they regularly offer free trials for their main services. Cable television providers might have a similar incentive to sign up users, but face a large fixed cost to install a physical connection in the user's home. Cable TV pricing correspondingly has an installation fee in contrast to a free trial, combined often with a tiered structure on the "intensive margin" of channel selection.

Another relevant example comes from phone apps, particularly games (e.g., "Candy Crush Saga"). Such apps can directly generate revenue from merely being opened, as they can scrape data points like the user's location and sell them to advertising firms. They are also likely to be addictive and have concave effects of past usage on future willingness-to-pay (i.e., the first minute of the game hooks the user more than the hundredth).[22] In practice, these products often are priced with free basic versions and paid add-ons.

We conclude by using an example with randomly occurring impressions to numerically illustrate how premium and trial tiers can coexist in the optimal contract:

Example 2 (Optimal Pricing with Premium and Trial Tiers)**.** Consumer preferences, the outcome space, and the type distribution are identical to those in Example 1. Impressions are again generated at a constant rate, normalized to one, and advertisers pay $\alpha > 0$ per click. But consumer behavior toward advertisements has the following different structure. Consumers "notice" an advertisement according to an exponential process with hazard rate $\lambda > 0$, click on the first advertisement they notice, and ignore all subsequent advertisements. The principal therefore receives the following payoff in expectation:

$$\pi(x, \theta) = \frac{\text{Cost}}{\text{Click}} \cdot \text{Expected Clicks} = \alpha \cdot (1 - e^{-\lambda x}) \tag{18}$$

[22]The micro-foundation of revenue via addiction in Online Appendix D.3 defines the relevant notion of concavity more precisely. Allcott, Gentzkow, and Song (2021) show that a model of total demand for popular phone apps (Facebook, Instagram, Twitter, Snapchat, web browsing, and YouTube), calibrated to empirical evidence, is consistent with habit formation, myopia, and diminishing returns.

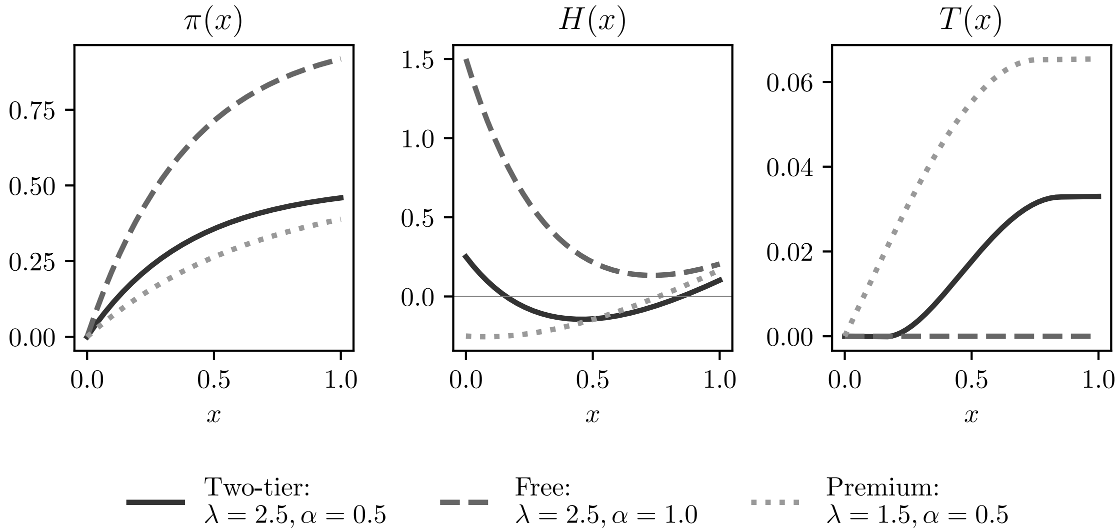

Figure 4: General price schedules in Example 2

Thus, $H(x) = \lambda\alpha e^{-\lambda x} - (1 - x)$, which has, depending on the values of (λ, α), either zero, one, or two tiers. The last case features both a free trial at the bottom and zero marginal pricing at the top while featuring positive marginal prices for intermediate units. In Figure 4, we illustrate these possibilities, including the two-tier scheme, for different parameter values. Intuitively, the concave revenue function makes usage at the bottom attractive to the principal, while the absence of information rents at the top coupled with some marginal revenues makes zero marginal pricing optimal. This combination of a trial and premium plan is characteristic of online newspaper and content streaming platforms. $\triangle$

3.4 Consumer and Producer Welfare Under Multi-Part Tariffs

We now study the impact of under-utilization on the levels of producer and consumer surplus. For an arbitrary price schedule T, denote the welfare of each type $\theta \in \Theta$ under optimal purchasing and utilization as $V(\theta; T)$ and the corresponding payoff of the seller as $\Pi(T)$.[23] With unlimited disposal $\overline{D}$ and no disposal $\underline{D}$ we write these payoffs, respectively, as $\overline{V}(\theta; T)$, $\underline{V}(\theta; T)$ and $\overline{\Pi}(T)$, $\underline{\Pi}(T)$. The following result summarizes how consumer and producer

[23] More formally:

$$V(\theta; T) = \sup_{y \in X, x \in D(y)} u(x, \theta) - T(y) \tag{19}$$

for all $\theta \in \Theta$, with seller-preferred $\xi(\theta; T)$ and agent level consumption $\phi(\theta; T)$ yielding:

$$\Pi(T) = \int_{\Theta} \left(\pi(\phi(\theta; T), \theta) + t(\xi(\theta; T))\right) \mathrm{d}F(\theta) \tag{20}$$

surplus are affected by the removal of contractibility of consumption.

Proposition 3 (Contractibility and Welfare)**.** *For any fixed price schedule T, $\overline{V}(\theta;T) \geq \underline{V}(\theta;T)$ for each $\theta \in \Theta$. However, in any corresponding optimal price schedule T^* and $T^{*'}$, $\overline{V}(\theta;T^*) \leq \underline{V}(\theta;T^{*'})$ for each $\theta \in \Theta$. Moreover, $\overline{\Pi}(T^*) \leq \underline{\Pi}(T^{*'})$.*

Proof. See Appendix A.5. □

All consumers lose out from increased contractibility for a fixed pricing policy, as the scope for payoff-increasing actions declines; but all consumers *gain* from increased contractibility under the seller's reoptimized menu. Intuitively, consumers would like to commit *ex ante* to avoid the possibility of moral hazard and have the seller pay them to take certain actions; increasing contractibility provides exactly this commitment device. Sellers gain from increased contractibility for the simple reason that it increases the set of implementable allocations.

This result allows us to speak to the benefits for consumers of advertising- and data-generated revenue. Technology companies have argued that advertising is consumer-friendly because it supports the provision of more free content.[24] Our results clarify how the constraints imposed by moral hazard and their observed consequence of zero-marginal-pricing skew the distribution of surplus toward producers.

A next natural question concerns how the gap between worlds with and without free disposal is itself affected by changes in the economic environment. Define the gaps in consumer and producer surplus in the optimum $\Delta_V(\theta) = \underline{V}(\theta;T^*) - \overline{V}(\theta;T^{*'})$ and $\Delta_\Pi = \underline{\Pi}(T^*) - \overline{\Pi}'(T^{*'})$. The following result shows that these gaps expand when the economic environment changes in ways that increase the principal's desire for more consumption:

Proposition 4 (Comparative Statics for Welfare)**.** *If either (i) $\pi_x(\cdot,\theta)$ pointwise increases for all $\theta \in \Theta$, or (ii) F decreases in the hazard rate order, then $\Delta_V(\theta)$ increases for all $\theta \in \Theta$ and Δ_Π increases.*[25]

Proof. See Appendix A.5. □

The intuition for this result is that with higher marginal value of consumption for the principal or smaller information rents, the principal-optimal consumption is larger. As a result, in the absence of contractibility, there are more types who end up consuming their bliss point rather than the principal's preferred amount. Thus, the principal gains more

[24]This sentiment is typified in a 2010 report by the Information Technology and Innovation Foundation, a US lobbying group, which claims that "data privacy regulations could sharply limit the principal funding mechanism for most of the free Internet enjoyed by consumers today and result in lost jobs, investment and Internet innovation." (Castro, 2010).

[25]That is, if the environment changes from $\mathcal{E} = (\pi, F)$ to $\tilde{\mathcal{E}} = (\tilde{\pi}, \tilde{F})$ and either (i) $\tilde{\pi}_x(\cdot,\theta) \geq \pi_x(\cdot,\theta)$ for all $\theta \in \Theta$ or (ii) $\tilde{F} \succeq_{HR} F$, where $\succeq_{HR}$ is the hazard rate order, then $\tilde{\Delta}_V(\cdot) \geq \Delta_V(\cdot)$ and $\tilde{\Delta}_\Pi \geq \Delta_\Pi$.

from contractibility when they would like the buyers to consume more, precisely because the disposal constraint is more often binding. Moreover, when incentives for the principal to allocate more are stronger, the introduction of contractibility is more valuable as a commitment device for the agents as they can be paid more for their now higher consumption.

This result allows us to shed light on the likely welfare effects of changes in advertising technology, such as the past decade's advent of more valuable *targeted* advertisements. This phenomenon translates in our model to an increase in the marginal value of usage, π_x. Our result implies that such improvements in technology will lead to welfare gains for all consumers, but that these gains would be larger if digital goods' usage were fully contractible. Moreover, government regulations that may, in the long run, diminish marginal advertising revenues, like the European Union's General Data Privacy Regulation, may *not* have large negative effects on prices and consumer surplus for digital products if sellers are far on the "interior" of the zero marginal pricing constraint.

Our focus on contractibility, and the technologies that would allow principals to monitor usage, contrasts with an influential perspective in the literature that focuses instead on the lack of collective bargaining for users of online products (Posner and Weyl, 2018). Taken to the extreme, our results imply a thorny "privacy paradox" for consumers—the only way to properly reap the benefits of the surplus generated by targeted advertising is to surrender *additional* privacy by enabling tools to more precisely monitor usage and attention.

4 Pricing with Partial Contractibility

While the extreme case of totally non-contractible consumption is relevant for many of our digital pricing applications, in some settings it appears that consumption is at least partially contractible. For example, search engine Bing monitors unique sign-ups and pays users for initial searches, while content-sharing platforms such as YouTube and TikTok monitor high-end users and share advertising revenues by paying them for their usage of the platform and the videos they create. In this section, we develop the theoretical apparatus required to consider partial contractibility via disposal correspondences, characterize implementable contracts, and simplify the optimal pricing problem. This enables us to fully characterize the optimal price in applications.

4.1 Modelling Partial Contractibility

Recall that a disposal correspondence $D : X \rightrightarrows X$ takes as input an allocated quantity $y \in X$ and returns a set of achievable outcomes $D(y) \subseteq X$. The interpretation of D as capturing under-utilization is ensured through the following five conditions:

1. *Reflexivity*: $y \in D(y)$ for every $y \in X$. It is possible for an agent to consume what they are allocated.
2. *No over-utilization*: for every $x, y \in X$, if $x \in D(y)$ then $x \leq y$. Disposal can allow the agent to consume or produce less but not more.
3. *Transitivity*: for every $x, y \in X$, if $x \in D(y)$ then $D(x) \subseteq D(y)$. If an agent can reach x by disposing from y, and z by disposing from x, then they can reach z by disposing directly from y.
4. *Monotonicity*: for every $x, y \in X$, if $x \leq y$ then $D(x) \leq_{SSO} D(y)$, where $\leq_{SSO}$ denotes the strong set order. When an agent disposes from a lower starting point, the set of things to which they can dispose is also lower.
5. *Closed-valuedness*: $D(y)$ is closed for all $y \in X$.

We say that a disposal correspondence D is *regular* whenever it satisfies all five properties listed above. Both the full-contractibility correspondence $\underline{D}(y)$ and the free-disposal correspondence $\overline{D}(y)$ are regular. Toward understanding the general nature of contractibility in our setting, we first characterize the structure of a regular disposal correspondence in terms of its minimum selection $\delta : X \to X$, defined as $\delta(y) = \min D(y)$ for all $y \in X$.

Lemma 1 (Characterization of Regular Disposal Correspondences)**.** *A disposal correspondence $D : X \rightrightarrows X$ is regular if and only if $D(y) = [\delta(y), y]$ for some function $\delta : X \to X$ that is increasing and such that, for every $x, y \in X$, $\delta(y) \leq y$, and,*

$$x \in [\delta(y), y] \implies \delta(x) = \delta(y) \tag{21}$$

Proof. See Online Appendix B.4. □

Intuitively, given any initial allocation $y \in X$, it is feasible to consume any $x \in [\delta(y), y]$. Graphically, each representing function δ can only switch between adjacent intervals $I \subseteq X$ and $I' \subseteq X$ such that: (i) there is no contractibility in the first interval and full contractibility in the second, that is $\delta(x) = \min I$ for all $x \in I$ and $\delta(x) = x$ for all $x \in I'$; (ii) there is full contractibility over the first interval and none in the second, that is $\delta(x) = x$ for all $x \in I$ and $\delta(x) = \min I'$ for all $x \in I'$; (iii) there is no contractibility over allocations within each interval, that is $\delta(x) = \min I$ for all $x \in I$ and $\delta(x) = \min I'$ for all $x \in I'$.

This characterization implies that three regular disposal correspondences form the building blocks of any other: *upper-threshold contractibility* D^u under which the seller can contract upon consumption above some threshold $c \in \mathbb{R}$; *lower-threshold contractibility* D^l under which the seller can contract upon consumption below c; and *grade-based contractibility* D^g the seller can contract upon whether or not consumption is above or below c. Hence:

$$D^u(y) = \begin{cases} [0, y] & \text{if } y < c \\ \{y\} & \text{if } y \geq c \end{cases} \quad D^l(y) = \begin{cases} \{y\} & \text{if } y < c \\ [c, y] & \text{if } y \geq c \end{cases} \quad D^g(y) = \begin{cases} [0, y] & \text{if } y < c \\ [c, y] & \text{if } y \geq c \end{cases} \tag{22}$$

In this section, we will explicitly apply upper-threshold and lower-threshold contractibility to digital goods pricing. In Online Appendix E.1, we apply grade-based contractibility to a problem of optimal monopoly regulation. However, we start with a general result characterizing implementability for arbitrary regular disposal correspondences.

4.2 Implementation with Partial Contractibility

We first define *implementable* consumption functions:

Definition 2 (Implementation). *A consumption function ϕ is implementable given D if there exists a purchase function ξ and a price schedule T such that (ϕ, ξ, T) jointly satisfy the constraints (O), (IC), and (IR) of Problem 1. In this case, we say that ϕ is supported by (ξ, T).*

We next characterize what is implementable with partial contractibility of utilization:

Proposition 5 (Characterization of Implementation). *Let D be a regular disposal correspondence with minimum selection δ. A consumption function ϕ is implementable given D if and only if ϕ is monotone increasing and such that for all $\theta \in \Theta$:*

$$\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\} \tag{23}$$

Moreover, ϕ is supported by $\xi = \phi$ and price schedule (with $C \leq 0$):

$$T(x) = C + \int_0^x u_x(s, \phi^{-1}(s)) \mathrm{d}s \tag{24}$$

Proof. See Appendix A.2. □

To understand this result, consider what the capping condition in Equation 23 requires. Intuitively, after being given any $y \in X$, the agent's favorite point is either y (if $y \leq \phi^A(\theta)$) or $\phi^A(\theta)$ (if $y > \phi^A(\theta)$). Thus, if $y \leq \phi^A(\theta)$, then the agent is content to consume y. However, if $\phi^A(\theta) < y$, then the agent's favorite feasible deviation is simply $\phi^A(\theta)$ (if $\phi^A(\theta) \in D(y)$) or the lowest possible feasible level of consumption $\delta(y)$ (if $\phi^A(\theta) < \delta(y) = \min D(y)$).

Thus, the agent optimally consumes no more than $\max\{\delta(y), \phi^A(\theta)\}$. Thus, it is clear that the capping requirement is necessary. The more involved part of this result shows that for any monotone ϕ, the capping requirement is sufficient. Standard arguments through the envelope formula and single-crossing rule out single deviations where a type θ would rather consume the allocation of any other type θ'. The main steps of the proof establish that there are also no double deviations where a type θ wishes to purchase $\phi(\theta')$ and consume some $x \in D(\phi(\theta'))$. Intuitively, single-crossing of agents' preferences remains sufficient to rule out any agent having the willingness-to-pay to profitably perform such a double deviation.

Beyond its use in characterizing optimal contracts, this result already informs us of interesting features of implementable contracts. In particular, the price schedule T must be increasing over the intervals $(\delta(y), y)$ for all $y \in X$. Thus, while the seller is constrained to offer positive or zero marginal prices on any such interval, they now have the potential to charge negative marginal prices across such intervals. This is an important possibility that we will shortly explore in an application to pricing content creation.

4.3 Optimal Pricing with Partial Contractibility

This implementation result allows us to express the optimal pricing problem (Problem 1) as a control problem, where the seller chooses a monotone and pointwise capped consumption function to maximize expected virtual surplus.

Lemma 2 (The Control Problem)**.** *Let D be a regular disposal correspondence with minimum selection δ. A consumption function is part of a solution to Problem 1 if any only if it solves*

$$\begin{aligned} \max_{\phi} \quad & \int_{\Theta} J(\phi(\theta), \theta)\, \mathrm{d}F(\theta) \\ s.t. \quad & \phi(\theta') \geq \phi(\theta), \quad \phi(\theta) \in [0, \phi^A(\theta)] \cup \{x \in X : \delta(x) = x\}, \quad \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{25}$$

Proof. See Appendix B.5 □

Thus, optimal consumption maximizes total virtual surplus subject to monotonicity and contractibiltiy constraints. The contractability constraint requires that consumption either lies below the bliss point or is at a fixed point of δ, which represents the lowest level of consumption in any interval of non-contractibility. This is as much as we can say about the optimum in general.

However, with additional structure, it is possible to say a lot more about optimal solutions with partial contractibility using Lemma 2. In particular, under lower-threshold contractibility and with no additional assumptions, in Appendix Proposition 7 we use this result to solve pointwise for the optimal contract as:

$$\phi^{*l}(\theta) = \min\left\{\max\left\{c, \phi^A(\theta)\right\}, \phi^P(\theta)\right\} \tag{26}$$

This mirrors the solution with free disposal, but now with the agents' bliss points replaced with $\max\{c, \phi^A(\theta)\}$. Intuitively, if agents' bliss points are below c, as their consumption is contractible, they can be forced to consume up to the level of c. Moreover, whenever agents have $\phi^A(\theta) < c$, they are paid on the margin for their consumption. We argue that such a contractibility structure mirrors that of search engine Bing where initial sign-ups are contractible, but subsequent usage is not. In this context, our result rationalizes the pricing of Bing, where, under the Bing Rewards program, users are remunerated for a certain quantity of initial searches, but subsequently face a price of zero on the margin.

Without additional structure, the pointwise solutions to the problems with upper-threshold and grade-based contractibility are possibly non-monotone as the implementation constraint is not monotone in the strong set order. Nevertheless, when the pointwise solution is monotone, pointwise optimization of virtual surplus subject to the contractibility constraint is optimal. We now lever this structure to study the optimal pricing of content creation platforms, where, as we will argue, upper-threshold contractibility appears descriptively realistic. In Online Appendix E.1, we use this structure to solve for optimal quality regulation of a monopolist under a pass-fail, grade-based structure.

4.4 Application: Pricing Content Creation Platforms

Content creation platforms, websites on which users post their own videos, music, writings, or other shareable digital content, perhaps epitomize the internet's grey area between "users" and "workers." One important example is the video-sharing website YouTube. The parent company Alphabet (formerly, Google) reported in its 2019 annual earnings report that YouTube generates about $15 billion per year in advertising revenue (Alexander, 2020). As per the website's current Terms of Service, advertisements may appear on any YouTube video. But creators can share in that advertisement revenue only if they belong to the *YouTube Partner Program*. The Partner Program, introduced in December 2007, is available to content creators who meet a minimum threshold of activity, measured by "valid public watch hours" per year and subscriber count (YouTube Help, 2021). Initial approval and periodic verification of Partner status involve human review, and in particular the definition of "valid public watch hours" is intentionally qualitative to guard against cheats like watching one's own videos on loop. Approved creators earn revenues that scale with video views at variable rates. Popular video-sharing app TikTok has a similar *Creator Fund*, which like the YouTube Partner Program is available only to creators with sufficient "authentic followers" and "authentic video views" (TikTok, 2021).

We model YouTube's Partner Program (henceforth, the leading example of the two) as nonlinear pricing of video creation. The good x is video content measured, as described above, by valid public watch hours. Creators have satiated preferences over video creation—the process is intrinsically enjoyable up to some point, after which it requires costly effort. Type θ scales individuals' ability to produce a given amount of content, measured by the eventual extent to which it is watched, per unit of effort. The advertising revenue that YouTube receives from x hours of watching one content creator of type θ is $\pi(x, \theta)$.

The monitoring of quality production described by YouTube follows closely upper-threshold contractibility D^u, as defined in Equation 22. In particular, above the threshold c (as per current guidelines, 4000 valid public watch hours per year), there is perfect monitoring of content creation. More concretely, we think of under-utilization in this context as manipulation of watch-hours, for instance by writing a computer script to replay videos on loop or hiring bots as part of a "click farm." However, it is not possible to "fake" the relevant time spent on the platform and the content created beyond a certain level as YouTube employs extensive algorithmic and human reviewing to ensure legitimacy.

Toward solving for optimal pricing, we assume that YouTube, absent any imperfect contractibility, prefers that all sufficiently productive users produce more videos than they would for free. We do this through the condition that $\phi^P(\theta) \geq \phi^A(\theta)$ if and only if $\theta \geq \hat{\theta}$ for some $\hat{\theta} \in [0, 1]$. Moreover, we define the function $\Delta_c(\theta) = J(c, \theta) - J(\phi^A(\theta), \theta)$, which is positive whenever it is better to assign type θ to c rather than their bliss point, and negative otherwise. We can now use Lemma 2 to solve for the structure of the optimal contract:

Proposition 6 (Optimal Usage with Upper-Threshold Contractibility)**.** *If Δ_c crosses zero at most once from below, then usage in any optimal contract is given by:*[26]

$$\phi^*(\theta) = \begin{cases} \phi^P(\theta) & \text{if } \theta < \theta_0 \\ \phi^A(\theta) & \text{if } \theta \in [\theta_0, \theta_1) \\ c & \text{if } \theta \in [\theta_1, \theta_2) \\ \phi^P(\theta) & \text{if } \theta \geq \theta_2 \end{cases} \tag{27}$$

where $\theta_0 = \hat{\theta}$, θ_1 is the smallest solution to $\Delta_c(\theta_1) = 0$ if one exists and 1 otherwise, and θ_2 solves $\phi^P(\theta_2) = c$ if a solution exists and is 1 otherwise.

Proof. See Online Appendix B.6. □

[26] That is, $\Delta_c(\theta) \geq 0$ if and only if $\theta \geq \theta_1$ for some $\theta_1 \in [0, 1]$.

We now provide intuition for this solution and describe the implied optimal pricing. In the region with non-contractibility, the solution starts out by following the free-disposal solution of $\min\{\phi^A, \phi^P\}$. In this setting, where ϕ^P crosses ϕ^A once from below, the seller starts out by offering ϕ^P before transitioning to ϕ^A after they cross. Thus, the seller may start out with positive marginal pricing before switching to zero marginal pricing. Next, the seller is faced with choosing to allow production at the bliss-point level, or bumping up agents' production into the region with full contractibility. Bumping agents up is attractive exactly when $\Delta_c(\theta) \geq 0$. Thus, at this point, the optimal production level jumps from $\phi^A(\theta)$ to c.[27] Here, the seller makes a lump-sum payment to agents who produce at or above the threshold c. Finally, when ϕ^P reaches c, it is of course optimal to simply request the optimal level of production and the seller pays agents on the margin for additional production.

Example 3. To illustrate the properties of this contract, we solve for the optimal pricing under the same quadratic structure of preferences and advertisement revenue used in Example 1.[28] We use the revenue structure $\pi(x) = x$, or in the Example's language, $\alpha = 1$ and $\beta = 0$. This generates a free product in the absence of contractibility of production. We finally consider the upper threshold cut-off c at levels "∞" (i.e., the free-disposal case), 1, and 0, which increase subsets of X on which the principal can contract on production. In Figure 5, for each threshold, we plot the schedule of video creation as a function of type and the tariff or wage schedule in the optimal contract.

First, when $c \to \infty$, all agents consume their bliss point and the product is free. This models YouTube before the December 2007 introduction of the Partner Program.

Second, when $c = 1$, sufficiently high types are paid both in absolute terms and on the margin for content creation, as in the YouTube Partner Program.[29] The payment schedule beyond $x > c = 1$ is determined by the standard screening motives to provide incentives toward effort—in the mechanism design interpretation, truthful reporting of one's type. In particular, in our quadratic example with uniformly distributed types, the tariff is concave or the wage schedule is convex, consistent with the empirical observation that top creators receive higher rates per view (Alexander, 2020). The split of surplus between YouTube and creators, a contentious issue in practice, relates both to the location of the threshold cutoff

[27] This is why we restrict attention to cases in which Δ_c crosses zero at most once from below. When this fails, the pointwise solution becomes non-monotone as the seller wishes to move some agents down from consuming c to their bliss point $\phi^A(\theta)$. In Online Appendix B.6, we provide sufficient conditions on primitives for this condition to hold.

[28] In the earlier examples, there was a natural and relevant notion of a binding "maximum time" spent on a platform. For video creation, such a limit is less natural, so we choose a domain such that the upper limit is never binding, $X = [0, 2]$.

[29] This "free then payment" structure arises because marginal advertising revenues are sufficiently high or, in our language, $\phi^P > \phi^A$ globally.

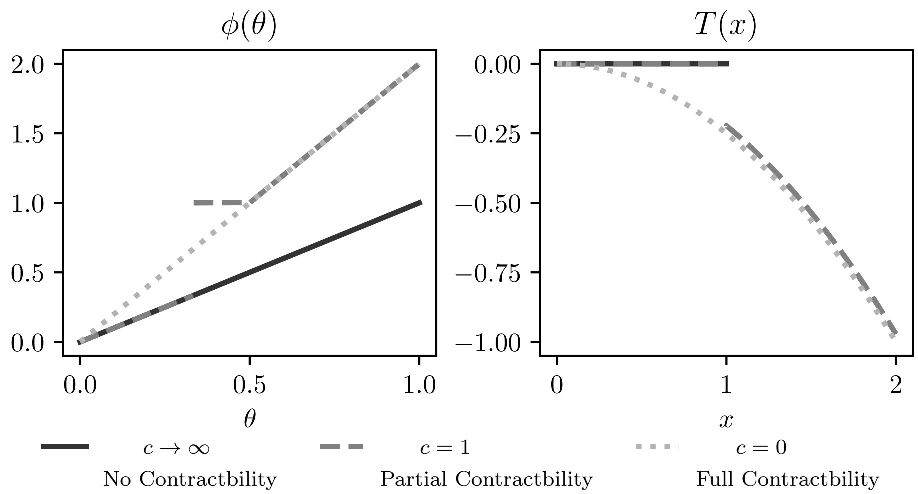

Figure 5: Optimal contracts under various levels of upper-threshold contracting.

c and the deep parameters determining the extent of information rents for better producers.

Third, when $c = 0$, all consumption is contractible and all types are paid. This is a counterfactual world in which the Partner Program is not exclusive only to high-productivity creators. In particular, switching from the observed world to this one not only creates surplus for low-type creators who were excluded from the program, but also for the highest-type creators who are paid larger information rents.[30] This switch does not, however, necessarily increase the aggregate amount of watch hours on YouTube, as a mass of creators who were previously bunched at producing exactly at the monitoring threshold $c = 1$ now produces at a lower level which is preferable to the platform. $\triangle$

5 Conclusion and Summary of Extensions

We study optimal nonlinear pricing in environments with feasible under-utilization and usage-derived revenue, features that are ubiquitous in the digital goods context. We show how the combination of these two forces rationalizes the occurrence of multi-part tariffs, or price schedules that include at least one tier of zero marginal prices. The key mechanism is that sellers have an incentive to pay users on the margin, due to usage-derived revenue (e.g., from advertisement or data collection), but non-contractibility prevents such arrangements from being enforceable. More succinctly, zero marginal pricing is the sellers' constrained optimum in a world in which "pay to click" is impossible.

[30] While creator welfare is not plotted, this comparative static is immediate from observing that the $c = 0$ tariff is pointwise lower than the $c = 1$ tariff—creators are paid more for any given level of activity.

We apply these results to study positive and normative features of digital goods markets. We show how different structures of external revenue translate into different familiar pricing schemes like free trials, premium tiers, and free products. We next show the scope for under-utilization reduces buyers' welfare and dampens the ability of buyers to reap the rewards from the revenue they generate. We finally show how partial contractibility predicts endogenous switching from agents being consumers, who either pay for a good or receive it for free, to being workers, who are paid for their usage of a platform.

We finish by discussing additional analysis contained within the Online Appendix.

Under-Utilization with (Perfect) Competition. It is of course natural to ask how our analysis in a monopoly setting would extend to markets with competition. In Online Appendix C.2, we solve for the equilibrium outcome of our screening model when the monopolist faces a perfectly competitive fringe of potential entrants. The equilibrium price schedule under perfect competition features zero marginal pricing more often than under monopoly pricing (Corollary 4 in Online Appendix C.2). The reason is simply that the perfectly competitive outcome is total, rather than virtual, surplus maximizing. Total surplus is maximized by a higher level of consumption owing to the absence of information rents, and so the constraint of under-utilization is more often binding. As a result, our model implies that multi-part tariffs are likely to be more prevalent in scenarios with fierce competition between sellers than under monopoly.

Pricing Meets Product Design. Our main analysis took the structure of the seller's revenues as exogenous. In practice, of course, the revenue function reflects platform design decisions. In Online Appendix D.1, we study an extended model in which the seller (e.g., Google) and an outside advertiser (e.g., the retailer that wants to show sponsored results) engage in Nash bargaining over advertising revenue generated from usage (e.g., Google searches). We show that the principal designs the optimal mechanism as if the external advertisers' revenue is their own. Providers also have latitude in determining the "extent" of usage-derived revenue (e.g., how many sponsored search results to show), which in turn may directly affect demand. We study this issue of optimal bundling in Online Appendix C.3 and show how this both provides a microfoundation for the particular example revenue functions we consider and allows us to explore how the extent of contractibility affects such bundling.

Additional Applications: Beyond Digital Goods. While we have so far discussed the implications of our results for digital goods pricing, we argue that our results are relevant beyond this context. We summarize these additional results in Online Appendix E. First, we study problems of government monopoly regulation as in Baron and Myerson (1982) (E.1). The government (principal) values production instrumentally due to concerns about consumer surplus, but can contract only on maximum permitted production; and the monopolist does not wish to produce beyond the point that maximizes profits. Our analysis shows how, in optimal government contracts with pass-fail quality regulation, the optimal contract features endogenous fines for failing to produce a sufficiently high quality. Second, we study optimal income taxation in the presence of human capital or aggregate demand externalities (E.2). In this case, with perfect contractibility, it may be optimal for the government to charge a negative marginal tax for sufficiently high income. Instead, we show how, under a partial contractibility scenario in which governments cannot detect over-reported income, zero marginal taxes may be optimal incentives.

Appendices

A Proofs of Main Results

In this appendix, we provide the proofs of the main results. To do this, we present and prove a series of more general results, from which we derive the results in the main text. First, we characterize implementation under regular disposal correspondences in terms of monotonicity of price schedules (A.1), before providing the proof of our implementation result for regular disposal correspondences (A.2). Second, we characterize optimal contracts (A.3) under lower-threshold contractibility and both zero marginal prices and multi-part tariffs in the case of free disposal (A.4). Finally, we derive comparative statics for welfare (A.5). The proofs of all additional results are postponed to Online Appendix B.

A.1 Toward Implementation: A Monotone Taxation Principle

To prove our implementation result, we first prove a novel extension of the taxation principle in our setting: a monotone taxation principle. Let $\bar{\mathbb{R}}$ denote the extended set of reals including $\{-\infty, +\infty\}$. In this section, we only assume that X and Θ are compact metric spaces. All the five properties defining regular disposal correspondences extend as written in Section 4.2. Given a regular disposal correspondence D, we say that $T : X \to \bar{\mathbb{R}}$ is monotone with respect to D if $T(x) \geq T(y)$ for all $x, y \in X$ such that $y \in D(x)$. The following intermediate result shows that monotonicity of the tariff with respect to D is necessary and sufficient for implementability (cf. Definition 2).

Lemma 3 (Monotone Taxation Principle)**.** *Fix a regular disposal correspondence D. A consumption function ϕ is implementable given D if and only if there exists a tariff $T : X \to$*

$\bar{\mathbb{R}}$ *that is monotone with respect to D and such that:*

$$\phi(\theta) \in \arg\max_{x \in X} \{u(x, \theta) - T(x)\} \tag{28}$$

and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. In this case, ϕ is supported by $\xi = \phi$ and T.

Proof. We begin by proving necessity of the existence of a monotone tariff. Suppose that ϕ is implementable. It follows that there exists (ξ, T) that support ϕ. In particular, observe that (O) implies that $\phi(\theta) \in D(\xi(\theta))$ for all $\theta \in \Theta$. Next define $\hat{T} : X \to \bar{\mathbb{R}}$ as:

$$\hat{T}(x) = \inf_{y \in X} \{T(y) : x \in D(y)\} \tag{29}$$

We next show that ϕ is also supported by $(\phi, \hat{T})$. By (O) of (ϕ, ξ, T), we have

$$u(\phi(\theta), \theta) \geq u(x, \theta) \tag{30}$$

for all $x \in D(\phi(\theta)) \subseteq D(\xi(\theta))$ and for all $\theta \in \Theta$, yielding (O) of $(\phi, \phi, \hat{T})$. By (IR) of (ϕ, ξ, T) and the definition of $\hat{T}$, we have

$$u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \tag{31}$$

for all $\theta \in \Theta$, yielding (IR) of $(\phi, \phi, \hat{T})$. Next, assume by contradiction that $(\phi, \phi, \hat{T})$ does not satisfy (IC), that is, there exists $\theta \in \Theta$ and $y \in X$ such that

$$\max_{x \in D(y)} u(x, \theta) - \hat{T}(y) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \tag{32}$$

With this, there exists $\hat{y} \in X$ such that $y \in D(\hat{y})$ and

$$\begin{aligned} \max_{x \in D(\hat{y})} u(x, \theta) - T(\hat{y}) \geq \max_{x \in D(y)} u(x, \theta) - T(\hat{y}) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \\ \geq u(\phi(\theta), \theta) - T(\xi(\theta)) = \max_{x \in D(\xi(\theta)} u(x, \theta) - T(\xi(\theta)) \end{aligned} \tag{33}$$

where we used Transitivity of D to argue that $D(y) \subseteq D(\hat{y})$. However, the previous equation yields a contradiction of (IC) of (ϕ, ξ, T), proving that $(\phi, \phi, \hat{T})$ satisfies (IC). This shows that $(\phi, \phi, \hat{T})$ is implementable, hence that Equation 28 holds and that $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. Finally, we argue that $\hat{T}$ is monotone with respect to D. Fix $x, y \in X$ such that $y \in D(x)$. By Transitivity of D we have

$$\{\hat{x} \in X : x \in D(\hat{x})\} \subseteq \{\hat{x} \in X : y \in D(\hat{x})\} \tag{34}$$

yielding that $\hat{T}(y) \leq \hat{T}(x)$, as desired.

We now establish sufficiency. Suppose that there exists a monotone $T : X \to \bar{\mathbb{R}}$ such that Equation 28 holds and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. We next show that (ϕ, ϕ, T) is implementable. (IR) is immediately satisfied. Suppose toward a contradiction that (O) is not satisfied. That is, there exists $\theta \in \Theta$ and $x \in D(\phi(\theta))$ such that:

$$u(x, \theta) > u(\phi(\theta), \theta) \tag{35}$$

However, by monotonicity of T, we have

$$u(x, \theta) - T(x) > u(\phi(\theta), \theta) - T(\phi(\theta)) \tag{36}$$

yielding a contradiction. Finally, suppose toward a contradiction that (IC) is not satisfied. That is, there exist $\theta \in \Theta$, $y \in X$, and $x \in D(y)$ such that

$$u(x, \theta) - T(y) > \max_{\hat{x} \in D(\phi(\theta))} u(\hat{x}, \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\phi(\theta)) \tag{37}$$

But then, we have the following contradiction of monotonicity of T:

$$u(x, \theta) - T(y) > u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(x, \theta) - T(x) \tag{38}$$

where the second implication uses the fact that $\phi(\theta)$ solves the program in Equation 28. This proves sufficiency.

Finally, the fact that any implementable consumption function can be implemented as part of an allocation (ϕ, ϕ, T) follows by the construction in the necessity part of our proof.

□

A.2 Implementation (Proof of Proposition 5)

In this section, we go back to the standard assumptions of the main text. The proof of Proposition 5 follows.

Proof. Fix a regular disposal correspondence with minimum selection δ and a consumption function ϕ. We first prove that there exists a tariff T that is monotone with respect to D and such that (ϕ, T) satisfy Equation 28 and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$ if and only if ϕ is monotone increasing and satisfies $\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}$ for all $\theta \in \Theta$. In turn, by Lemma 3, this proves the first part of Proposition 5.

($\implies$) By the standard taxation principle with our assumptions, ϕ must be incentive compatible when the disposal correspondence is $D = \underline{D}$, that is, when disposal is not feasible. It is well-known that under our assumption that u is strictly single-crossing, a necessary condition for incentive compatibility is that $\phi(\theta)$ is monotone increasing in θ. By Lemma 3, (ϕ, ϕ, T) satisfies (O). Therefore, it is immediate by strict quasiconcavity of u and Lemma 1 that θ's optimal point in $D(\phi(\theta)) = [\delta(\phi(\theta)), \phi(\theta)]$ is given by $\max\{\delta(\phi(\theta)), \phi^A(\theta)\}$. Thus, it is necessary that $\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}$ for all $\theta \in \Theta$.

($\Longleftarrow$) Now suppose that $\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}$ for all $\theta \in \Theta$ and ϕ is monotone increasing. Define the function $t : \Theta \to \mathbb{R}$ as

$$t(\theta) = C + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \tag{39}$$

for some $C \leq 0$, and the tariff $T : X \to \overline{\mathbb{R}}$ as

$$T(x) = \inf_{\theta' \in \Theta} \{t(\theta') : x \in D(\phi(\theta'))\} \tag{40}$$

Fix $x, y \in X$ such that $y \in D(x)$. By Transitivity, for all $\theta \in \Theta$, if $x \in D(\phi(\theta))$, then $y \in D(\phi(\theta))$. This shows that

$$\{\theta' \in \Theta : x \in D(\phi(\theta'))\} \subseteq \{\theta' \in \Theta : y \in D(\phi(\theta'))\} \tag{41}$$

and therefore that $T(x) \geq T(y)$. We conclude that T is monotone with respect to D. Next, for every $\theta \in \Theta$, we have

$$u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - t(\theta) = \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s - C \tag{42}$$

Note that the right-hand side of this last equation is monotone increasing in θ since it is continuously differentiable with derivative $u_\theta(\phi(\theta), \theta) \geq 0$ for all $\theta \in \Theta$, owing to the fact that u is monotone increasing over Θ. Given that $C \leq 0$, we have that $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$.

We are left to prove that (ϕ, T) satisfy Equation 28. We first prove that, for all θ, θ', we have

$$u(\phi(\theta), \theta) - t(\theta) \geq \max_{x \in D(\phi(\theta'))} u(x, \theta) - t(\theta') \tag{43}$$

This is a variation of the standard reporting problem under consumption function ϕ and transfers t, where each agent, on top of misreporting their type, can also consume everything allowed by D. Violations of this condition can take two forms. First, an agent of type θ could report type θ' and consume $x = \phi(\theta')$. We call this a single deviation. Second, an agent of type θ could report type θ' and consume $x \in [\delta(\phi(\theta')), \phi(\theta')) = D(\phi(\theta')) \setminus \{\phi(\theta')\}$. We call this a double deviation. Under our construction of transfers t and monotonicity of ϕ, by a standard mechanism-design argument (e.g., Nöldeke and Samuelson (2007)), there is no strict gain to any agent of reporting θ' and consuming $x = \phi(\theta')$. Thus, there are no profitable single deviations under (ϕ, t).

We now must rule out double deviations. To this end, we first state an ancillary Lemma that transfers are monotone on any interval of types who can achieve a common lowest consumption. Its proof is relegated to the Online Appendix.

Lemma 4. *For any $\theta, \theta' \in \Theta$ such that $\theta \leq \theta'$ and $\delta(\phi(\theta)) = \delta(\phi(\theta')) \equiv \delta_0$, it holds that $t(s') \geq t(s)$ for all $s, s' \in [\theta, \theta']$ such that $s \leq s'$*

Proof. See Online Appendix B.7. □

We now use this Lemma to rule out double deviations. Define the value function $V : \Theta \to \mathbb{R}$ under ϕ and t as

$$V(\theta) = \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s - C \tag{44}$$

We separate the argument by various cases comparing $(\theta, \phi(\theta), \phi^A(\theta))$ and $(\theta', \phi(\theta'), \phi^A(\theta'))$:

1. $\theta' < \theta$: for there to be a strict double deviation, it must be that $\phi(\theta') < \phi(\theta)$. We have already argued that Obedience holds, so that $\phi(\theta')$ is optimal for type θ' when they could choose and $x \in [\delta(\phi(\theta')), \phi(\theta')]$. Moreover, by strict single-crossing of u in (x, θ) and strict quasiconcavity of $u(\cdot, \theta)$ in x, it is optimal for type θ to consume some $x \geq \phi(\theta')$. But, we know that $x \in [\delta(\phi(\theta')), \phi(\theta')]$, thus $x = \phi(\theta')$ is optimal. Hence, if there is a double deviation with $\theta' < \theta$, there is also a single deviation. This is a contradiction as we already showed that there are no strictly profitable single deviations.
2. $\theta' > \theta$ and $\phi^A(\theta) \geq \phi(\theta')$: the optimal choice of consumption for agent θ in $D(\phi(\theta'))$ is given by $\phi(\theta')$ by strict quasiconcavity of u, and there is a profitable single deviation, which is a contradiction.
3. $\theta' > \theta$ and $\phi^A(\theta) < \phi(\theta')$ and $\delta(\phi(\theta)) = \delta(\phi(\theta'))$: There are two sub-cases.
   (a) $\phi^A(\theta) < \delta(\phi(\theta')) = \delta(\phi(\theta))$: The optimal consumption choice for type θ in $D(\phi(\theta'))$ is given by $\delta(\phi(\theta')) = \delta(\phi(\theta)) = \phi(\theta)$ by strict quasiconcavity of u. Thus, there can be a strict double deviation only if $t(\theta') < t(\theta)$, which we have already shown to be impossible.
   (b) $\phi^A(\theta) \geq \delta(\phi(\theta')) = \delta(\phi(\theta))$: We know $x = \phi^A(\theta)$ is most attractive following any misreport θ'. Suppose that there exists some $\hat{\theta} \in (\theta, \theta']$ such that $\phi(\hat{\theta}) = \phi^A(\theta)$.

As t is monotone over the interval $[\theta, \theta']$, we know that a single deviation to $\hat{\theta}$ is weakly more attractive than a double deviation to $x \in D(\phi(\theta'))$. As no single deviations exist, this is a contradiction. It must then be that ϕ is discontinuous and no type receives $\phi^A(\theta)$. We know that the most attractive misreport is the smallest type θ' such that $\phi(\theta') \geq \phi^A(\theta)$. It follows that $\phi^A(\theta) \leq \phi(\theta') \leq \phi^A(\theta')$ and therefore that there exists some $\hat{\theta}$ such that $\phi^A(\hat{\theta}) = \phi(\theta')$, by continuity of ϕ^A.

We now show that if there exists a double deviation for type θ, there exists a single deviation for type $\hat{\theta}$. By the hypothesis of a double deviation for type θ:

$$u(\phi^A(\theta), \theta) - t(\theta') > u(\phi(\theta), \theta) - t(\theta) \tag{45}$$

Define for any type θ, the value of optimal autarkic consumption as $V^*(\theta) = u(\phi^A(\theta), \theta)$. We can write:

$$V^*(\theta) - V(\theta) > t(\theta') \tag{46}$$

As we have ruled out single deviations, we know that:

$$u(\phi^A(\hat{\theta}), \hat{\theta}) - t(\theta') \leq u(\phi(\hat{\theta}), \hat{\theta}) - t(\hat{\theta}) \tag{47}$$

Thus:

$$V^*(\hat{\theta}) - V(\hat{\theta}) \leq t(\theta') \tag{48}$$

Together, we then have that:

$$V(\hat{\theta}) - V(\theta) > V^*(\hat{\theta}) - V^*(\theta) \tag{49}$$

Or:

$$\int_{\theta}^{\hat{\theta}} u_\theta(\phi(s), s)\mathrm{d}s > \int_{\theta}^{\hat{\theta}} u_\theta(\phi^A(s), s)\mathrm{d}s \tag{50}$$

But we know that $\phi^A(s) \geq \phi(s)$ for all $s \in [\theta, \hat{\theta}]$ owing to being in this case, and this implies by single-crossing of u that $u_\theta(\phi^A(s), s) \geq u_\theta(\phi(s), s)$, which contradicts the inequality above.

4. $\theta' > \theta$ and $\phi^A(\theta) < \phi(\theta')$ and $\delta(\phi(\theta)) < \delta(\phi(\theta'))$: There are two sub-cases.

   (a) $\phi^A(\theta) < \delta(\phi(\theta'))$: We know that $x = \delta(\phi(\theta'))$ is the most attractive deviation. Suppose that there is some type s who receives $\phi(s) = \delta(\phi(\theta'))$, as transfers are monotone on each interval we know that $t(s) \leq t(\theta')$, and thus deviating to type s is more attractive. However, this is a single deviation, a contradiction. It must then be that for all types s such that $\delta(\phi(\theta')) \leq \phi(s) \leq \phi(\theta')$ that $\phi(s) > \delta(\phi(\theta'))$. In this case $\phi^A(\theta') \geq \phi(\theta')$ and we can apply a variation of the argument in 3(b) above. Namely, there exists $\hat{\theta} \in (\theta, \theta')$ such that $\phi^A(\hat{\theta}) = \phi(\theta')$. We have already shown that $\hat{\theta}$ has no strict incentive to imitate type θ'. Thus, by the same arguments as in case 3 (b):

$$V^*(\hat{\theta}) - V(\hat{\theta}) \leq t(\theta') \tag{51}$$

By the hypothesis that type θ has a strict incentive to imitate type θ' and consume $x = \delta(\phi(\theta'))$, we also have that:

$$u(\delta(\phi(\theta')), \theta) - t(\theta') > u(\phi(\theta), \theta) - t(\theta) \tag{52}$$

It follows from this that:

$$u(\delta(\phi(\theta')), \theta) - V(\theta) > t(\theta') \tag{53}$$

Combining above inequalities, it follows that:

$$V(\hat{\theta}) - V(\theta) > V^*(\hat{\theta}) - u(\delta(\phi(\theta')), \theta) \tag{54}$$

Which we can express as:

$$\int_{\theta}^{\hat{\theta}} u_\theta(\phi(s), s)\mathrm{d}s > u(\phi^A(\hat{\theta}), \hat{\theta}) - u(\delta(\phi(\theta')), \theta) \tag{55}$$

Observe further that there exists some $\theta'' \in (\theta, \hat{\theta})$ such that $\phi^A(\theta'') = \delta(\phi(\theta'))$. Therefore we can write:

$$u(\phi^A(\hat{\theta}), \hat{\theta}) = u(\delta(\phi(\theta')), \theta'') + \int_{\theta''}^{\hat{\theta}} u_\theta(\phi^A(s), s)\mathrm{d}s \tag{56}$$

Thus, we can rewrite inequality 55 as:

$$\int_{\theta''}^{\hat{\theta}} u_\theta(\phi(s), s)\mathrm{d}s + \int_{\theta}^{\theta''} u_\theta(\phi(s), s)\mathrm{d}s > \int_{\theta''}^{\hat{\theta}} u_\theta(\phi^A(s), s)\mathrm{d}s + \int_{\theta}^{\theta''} u_\theta(\delta(\phi(\theta')), s)\mathrm{d}s \tag{57}$$

However, we know that for all $s \in [\theta'', \hat{\theta}]$ that $\phi^A(s) \geq \delta(\phi(\theta')) = \phi^A(\theta'')$. It follows that $\phi(s) \le \phi^A(s)$ for all $s \in [\theta'', \hat{\theta}]$. Therefore, by strict single-crossing of u it follows that:

$$\int_{\theta''}^{\hat{\theta}} u_\theta(\phi(s), s)\mathrm{d}s < \int_{\theta''}^{\hat{\theta}} u_\theta(\phi^A(s), s)\mathrm{d}s \tag{58}$$

Moreover, we know that for all $s \in [\theta, \theta'']$ that $\phi^A(s) \le \delta(\phi(\theta')) = \phi^A(\theta'')$. It follows that $\phi(s) \le \delta(\phi(\theta'))$ for all $s \in [\theta, \theta'']$. Therefore, by strict single-crossing of u it follows that:

$$\int_{\theta}^{\theta''} u_\theta(\phi(s), s)\mathrm{d}s < \int_{\theta}^{\theta''} u_\theta(\delta(\phi(\theta')), s)\mathrm{d}s \tag{59}$$

The above two inequalities contradict inequality 55. That is, a double deviation would imply a single deviation, which is a contradiction.

(b) $\phi^A(\theta) \ge \delta(\phi(\theta'))$: We know that $x = \phi^A(\theta)$ is most attractive following a misreport θ'. By the same argument as in case 3(b), if any type in fact was allocated $\phi^A(\theta)$, there would also be a single deviation, a contradiction. If instead there was no such type, we apply the same argument as in 3(b), and reason similarly that there would be a single deviation for type $\hat{\theta}$.

We have ruled out double deviations in all cases and thereby completed the proof of the claim in Equation 43.

We next prove that Equation 43 implies that (ϕ, T) satisfy Equation 28. Indeed, for all $\theta \in \Theta$, we have

$$\begin{aligned} u(\phi(\theta), \theta) - T(\phi(\theta)) &\ge u(\phi(\theta), \theta) - t(\theta) \ge \sup_{\theta' \in \Theta} \max_{x \in D(\phi(\theta'))} u(x, \theta) - t(\theta') \\ &= \sup_{x \in X} \sup_{\theta' \in \Theta : x \in D(\phi(\theta'))} u(x, \theta) - t(\theta') = \sup_{x \in X} u(x, \theta) - T(x) \end{aligned} \tag{60}$$

yielding Equation 28. This concludes the proof of the first part of Proposition 5.

Fix an implementable ϕ. By the previous part of the proof and by Lemma 3, it follows that we can take $\xi = \phi$. By the construction of t and T in the previous part of the proof and Lemma 1 in Nöldeke and Samuelson (2007), we can take T as in Equation 24 with $C \le 0$. □

A.3 Optimal Contracts with Lower Threshold Contractibility (Proof of Proposition 1)

We first solve for optimal consumption under lower threshold contractibility.

Proposition 7. *Let D be a lower threshold disposal correspondence with parameter c. Consumption in the optimal contract is given by:*

$$\phi^* = \min\left\{\max\left\{c, \phi^A\right\}, \phi^P\right\} \tag{61}$$

Proof. By Lemma 2, any optimal consumption function must solve:

$$\begin{aligned} &\max_{\phi} \int_{\Theta} J(\phi(\theta), \theta) \, \mathrm{d}F(\theta) \\ &\text{s.t} \quad \phi(\theta') \geq \phi(\theta), \quad \phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}, \quad \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{62}$$

Define the set of fixed points of δ as $\mathcal{C}_\delta = \{x \in X : \delta(x) = x\}$ and define the correspondence $S_\delta : \Theta \rightrightarrows X$ as $S_\delta(\theta) = [0, \phi^A(\theta)] \cup \mathcal{C}_\delta$. Observe that the constraint $\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}$ is equivalent to the constraint $\phi(\theta) \in S_\delta(\theta)$. Moreover, in the case of lower threshold contractibility we have that $\mathcal{C}_\delta = [0, c]$ so $S_\delta(\theta) = [0, \max\{\phi^A(\theta), c\}]$.

Consider now the family of problems for each $\theta \in \Theta$:

$$\max_{x \in S_\delta(\theta)} \{J(x, \theta)\} \tag{63}$$

As J is strictly quasiconcave in x, there is a unique maximum in this problem, which we call $\phi^*(\theta)$. Moreover, whenever $\phi^P(\theta) < \max\{\phi^A(\theta), c\}$, we know that $\phi^*(\theta) = \phi^P(\theta)$. Otherwise $\phi^*(\theta) = \max\{\phi^A(\theta), c\}$. Thus, the solution of this pointwise problem is:

$$\phi^*(\theta) = \min\left\{\max\left\{c, \phi^A(\theta)\right\}, \phi^P(\theta)\right\} \tag{64}$$

If in addition ϕ^* is monotone increasing, then ϕ^* is the unique solution to Problem 1. We now verify this monotonicity. We assumed that J satisfies single-crossing in (x, θ). Moreover, J is quasisupermodular in x by the fact that $x \in X \subset \mathbb{R}$; X is a lattice; Θ is partially ordered; and the correspondence $S_\delta(\theta) = [0, \min\{c, \phi^A(\theta)\}]$ is increasing. Thus, by Theorem 4′ in Milgrom and Shannon (1994), we have that $\phi^*(\theta)$ is increasing in θ. This completes the proof. □

We now prove Proposition 1 by explicitly constructing the claimed supporting tariffs and purchases.

Proof. Free disposal corresponds to lower threshold contractibility with $c = 0$. Thus, optimal consumption is given by:

$$\phi^* = \min\{\phi^A, \phi^P\} \tag{65}$$

By the envelope formula from Lemma 2, we then have that:

$$T^*(x) = u(x_*, 0) + \left[\int_{x_*}^{\min\{x, x^*\}} u_x(s, {\phi^*}^{-1}(s))\,\mathrm{d}s\right]^+ \tag{66}$$

where $x_* = \phi^*(0)$, $x_* = \phi^*(1)$, and $[\cdot]^+$ denotes the positive part. We now show that this optimal level of consumption is supported by any selection from Ξ_{ϕ^*} and only by selections from Ξ_{ϕ^*}. To this end, consider the selection $\overline{\xi} \in \Xi_{\phi^*}$ defined as $\overline{\xi} = \max \Xi_{\phi^*}$. We want to show that the triple $(\overline{\xi}, \phi, T^* \circ \overline{\xi})$ is incentive compatible for all price schedules schedule T^* defined as above, where we adopt the notation that $t = T^* \circ \overline{\xi}$. We verify that the three constraints in Problem 1 hold under this selection.

Consider the first constraint that for all $\theta \in \Theta$:

$$\phi(\theta) \in \arg \max_{x \in D(\xi(\theta))} u(x, \theta) \tag{67}$$

Observe that $\phi \in D(\overline{\xi})$ by construction of $\overline{\xi}$. Moreover, toward a contradiction, suppose that there exists $\theta \in \Theta$ and $x \in D(\overline{\xi}(\theta))$ such that $u(\phi(\theta), \theta) < u(x, \theta)$. There are two cases:

1. If $\phi(\theta) < \phi^A(\theta)$, then by construction $x \leq \xi(\theta) = \phi(\theta) < \phi^A(\theta)$ implying that $u(\phi(\theta), \theta) \geq u(x, \theta)$ by strict quasiconcavity of $u(\cdot, \theta)$, hence yielding a contradiction.
2. If $\phi(\theta) = \phi^A(\theta)$, then by construction $u(\phi(\theta), \theta) \geq u(x, \theta)$ yielding again a contradiction.

Given that $\theta \in \Theta$ was arbitrarily chosen, the proposed triple satisfies the constraint that it is optimal for any type θ to consume $\phi(\theta)$ when they purchase $\overline{\xi}(\theta)$.

Consider now the constraint that for all $\theta \in \Theta$:

$$\xi(\theta) \in \arg \max_{y \in X} \left\{ \max_{x \in D(y)} u(x, \theta) - T(y) \right\} \tag{68}$$

and define $g(y, \theta) = \max_{x \in D(y)} u(x, \theta)$. Toward a contradiction, suppose that there exist $\theta, \theta' \in \Theta$ such that:

$$g(\overline{\xi}(\theta), \theta) - t(\theta) < g(\overline{\xi}(\theta'), \theta) - t(\theta') \tag{69}$$

There are two cases to consider:

1. If $\phi(\theta') < \phi^A(\theta')$, then define $\theta'' = \theta'$ and note that

$$g(\phi(\theta), \theta) - t(\theta) = g(\overline{\xi}(\theta), \theta) - t(\theta) < g(\overline{\xi}(\theta'), \theta) - t(\theta') = g(\phi(\theta''), \theta) - t(\theta'') \tag{70}$$

where the last equality follows by construction of $\overline{\xi}$ and θ''.

2. If $\phi(\theta') = \phi^A(\theta')$, then define $\theta'' = \inf\left\{\hat{\theta} \in \Theta : \hat{\theta} \geq \theta', \phi(\hat{\theta}) < \phi^A(\hat{\theta})\right\}$. Note that, by monotonicity of ϕ and by construction of $\overline{\xi}$, we have $\phi(\theta'') = \overline{\xi}(\theta'') = \overline{\xi}(\theta')$. Moreover, by construction we necessarily have that $[\theta', \theta''] \subseteq \left\{\hat{\theta} \in \Theta : \phi(\hat{\theta}) = \phi^A(\hat{\theta})\right\}$. Therefore, by the previous claim on t, we have $t(\theta') = t(\theta'')$. Altogether, we have

$$g(\phi(\theta), \theta) - t(\theta) = g(\overline{\xi}(\theta), \theta) - t(\theta) < g(\overline{\xi}(\theta'), \theta) - t(\theta') = g(\phi(\theta''), \theta) - t(\theta'') \quad (71)$$

In both cases, there exists $\theta'' \in \theta$ such that

$$g(\phi(\theta), \theta) - t(\theta) < g(\phi(\theta''), \theta) - t(\theta'') \quad (72)$$

This contradicts the fact that (ϕ^*, ϕ^*, T^*) is implementable (see Proposition 5). Thus, the second constraint is satisfied.

The third constraint that for all $\theta \in \Theta$:

$$u(\phi^*(\theta), \theta) - T^*(\overline{\xi}(\theta)) \geq 0 \quad (73)$$

is unaffected as, by Lemma 4, T^* is monotone on all intervals where types receive their bliss points.

This proves that $(\phi^*, \overline{\xi}, T^*)$ is implementable and therefore optimal. Finally, note that for any other selection $\xi \in \Xi_{\phi^*}$, the triple (ϕ^*, ξ, T^*) is necessarily implementable and therefore optimal. Indeed, by way of contradiction, suppose that the latter is not implementable. It follows that $(\phi^*, \overline{\xi}, T^*)$ is not implementable either as all the feasible deviations under purchase function ξ are still feasible under $\overline{\xi}$. However, this contradicts our previous claim.

We finally show that if $\xi \notin \Xi_{\phi^*}$, then it is not part of an optimal contract. We will use the observation that all agents' payments to the seller are pinned down by the envelope formula for t. There are two cases to consider. First, suppose that there exists a $\theta \in \Theta$ such that $\xi(\theta) \neq \phi^*(\theta)$ and $\phi^*(\theta) < \phi^A(\theta)$. If $\xi(\theta) < \phi^*(\theta)$, then $\phi^*(\theta) \notin D(\xi(\theta))$, which makes ϕ^* infeasible. If $\xi(\theta) > \phi^*(\theta)$, then, as $\phi^*(\theta) < \phi^A(\theta)$, $t(\theta)$ is strictly increasing at θ. Thus $T(\xi(\theta)) > t(\theta)$, which is a contradiction. Second, suppose that there exists a $\theta \in \Theta$ such that $\xi(\theta) \notin [\phi^A(\theta), \inf_{\theta' \in [\theta, 1]} \left\{\phi^*(\theta') < \phi^A(\theta')\right\}$ and $\phi^*(\theta) = \phi^A(\theta)$. Once again if $\xi(\theta) < \phi^*(\theta)$, then $\phi^*(\theta) \notin D(\xi(\theta))$, which makes ϕ^* infeasible. If $\xi(\theta) > \inf_{\theta' \in [\theta, 1]} \left\{\phi^*(\theta') < \phi^A(\theta')\right\}$, then as before $T(\xi(\theta)) > t(\theta)$, which is a contradiction.

$\square$

A.4 Characterizing Zero Marginal Pricing (Proof of Proposition 2)

In this Appendix, we prove Proposition 2 as stated.

Proof. ($\Longrightarrow$) We first prove that $H(x) > 0$ implies that $T(x) = T_{\phi^*}(x)$ is flat at x, for any $x \in X^*$. Consider first any $x \in \text{Int}(X^*)$, where $\text{Int}(\cdot)$ denotes the interior of a set. Recall by Proposition 1 that $\phi^* = \min\{\phi^P, \phi^A\}$. Moreover, u and J are strictly single-crossing, so we have that ϕ^P and ϕ^A are strictly increasing over Θ. Thus, we have that ϕ^* is strictly increasing over $\theta \in \Theta$ and is therefore invertible at any $x \in X^* = \phi^*(\Theta)$. Suppose now that $H(x) = J_x(x, (\phi^A)^{-1}(x)) > 0$ and define $\theta(x) = (\phi^*)^{-1}(x)$. It is either the case that $x = \overline{x}$ (which is not in $\text{Int}(X^*)$), or $\phi^A(\theta(x)) < \phi^P(\theta(x))$. Thus, when $H(x) > 0$, $\phi^A(\theta(x)) < \phi^P(\theta(x))$, so $\phi^*(\theta(x)) = \phi^A(\theta(x))$. As ϕ^A and ϕ^P are continuous functions by the Theorem of the Maximum and invertible at x, we can find a neighborhood $O(x)$ of x, such that for all $x' \in O(x)$, and corresponding $\theta' = (\phi^*)^{-1}(x')$, we have that $\phi^*(\theta') = \phi^A(\theta')$. In Lemma 4, we showed that transfers must be constant on such an $O(x)$.

We now show that transfers are increasing over Θ. First, we know that transfers are a.e differentiable and that ϕ^* is a.e. differentiable as it is monotone. Thus, to show that t is monotone increasing it suffices to show that $t' \geq 0$ whenever it exists. Moreover, whenever t' exists, it is given by:

$$t'(\theta) = u_x(\phi^*(\theta), \theta)\phi^{*'}(\theta) \tag{74}$$

We know that $\phi^{*'} \geq 0$ as ϕ^* is monotone. We also know that $u_x(\phi^*(\theta), \theta) \geq 0$ for all $\theta \in \Theta$ as u is quasiconcave and $\phi^* \leq \phi^A$. Monotonicity of transfers, combined with the fact that ϕ^* is increasing over Θ and $x \in X^*$ implies that:

$$T(x) = T_{\phi^*}(x) = \inf_{\theta \in \Theta} \{t(\theta) : x \leq \phi(\theta)\} = t(\theta(x)) \tag{75}$$

As this is constant for all $x' \in O(x)$, we have shown that $T_{\phi^*}(x)$ is flat for all $x \in \text{Int}(X^*)$ whenever $H(x) > 0$.

It remains to consider all $x \notin \text{Int}(X^*)$. However, continuity of ϕ^* implies the result for the boundary points of X^*.

($\Longleftarrow$) We now prove that, for every $x \in X^*$, if $T(x) = T_{\phi^*}(x)$ is flat at x, then $H(x) \geq 0$. First, consider $x \in \text{Int}(X^*)$. From the first part of the proof, we have that $T_{\phi^*}(x) = t(\theta(x))$, where $\theta(x) = (\phi^*)^{-1}(x)$. As ϕ^A and ϕ^P are strictly increasing, we have that $\theta(x)$ is strictly increasing in x. Thus, if $T_{\phi^*}(x)$ is flat at x, there must exist a neighborhood of $O(x)$ of x such that $t(\theta(x')) = t(\theta(x))$ for all $x \in O(x)$. Toward a contradiction, suppose that $H(x) < 0$ at x. This implies that $\phi^P(\theta(x)) < \phi^A(\theta(x))$, and so $\theta(x) = (\phi^P)^{-1}(x)$. By continuity of

ϕ^P, there must exist some neighborhood $O'(x) \subseteq O(x)$ such that $\theta(x') = (\phi^P)^{-1}(x')$ for all $x' \in O'(x)$. But then we have that almost everywhere in $O'(x)$ that transfers are strictly increasing:

$$t'(\theta(x')) = u_x(\phi^P(\theta(x')), \theta)\phi^{P'}(\theta(x')) > 0 \tag{76}$$

where the above inequality follows as u is strictly quasiconcave and $\phi^P(\theta(x')) < \phi^A(\theta(x'))$, so $u_x(\phi^P(\theta(x')), \theta) > 0$, while the fact that ϕ^P is strictly increasing implies that $\phi^{P'}(\theta(x'))$. But we know that $t(\theta(x')) = t(\theta(x))$ for all $x \in O(x)$, so this is a contradiction. It follows that $H(x) \geq 0$. It remains to consider all $x \notin \mathrm{Int}(X^*)$. As before, continuity of H implies the result for the boundary points of X^*.

A careful reader may ask why it is not true that $T_{\phi^*}(x)$ is flat at any $x \in X$ implies $H(x) > 0$. Counterexamples arise to this claim when ϕ^P and ϕ^A do not strictly cross and instead coincide for positive Lebesgue measure of outcomes. In particular, suppose that $\phi^P \equiv \phi^A$; economically, this means the principal's unconstrained optimum coincides with the agents'. We have that $T_{\phi^*}(x)$ is flat everywhere but $H(x) \equiv 0$. □

A.5 Welfare (Proofs of Propositions 3 and 4)

We first prove Proposition 3 in the more general case with lower threshold monitoring.

Proof. We first prove the part for a fixed price schedule T with corresponding (ϕ, ξ). We can, for this part, in fact prove a more general result comparing disposal correspondences in the sense of how much contractibility they feature. We say that D features more contractibility than D', if we have that $D(y) \subseteq D'(y)$ for all $y \in X$. Consider the value to any agent of type $\theta \in \Theta$ of reporting any other type $\theta' \in \Theta$ and optimally disposing of $\xi(\theta')$:

$$V(\theta'; \theta, T, D) = \max_{x \in D(\xi(\theta'))} \{u(x, \theta)\} - t(\theta') \tag{77}$$

When D features more contractibility than D', we have that $D(\xi(\theta')) \subseteq D'(\xi(\theta'))$. It follows that $V(\theta'; \theta, a) \leq V'(\theta'; \theta, a)$ for all $\theta, \theta' \in \Theta, a \in \mathcal{A}$. Thus:

$$V(\theta; T) = \max_{\theta' \in \Theta} \{V(\theta'; \theta, T)\} \leq \max_{\theta' \in \Theta} \{V'(\theta'; \theta, T)\} = V'(\theta; T) \tag{78}$$

for all $\theta \in \Theta, T$. As $\underline{D}$ features more contractibility that $\overline{D}$, it immediately follows that $\overline{V}(\theta; T) \geq \underline{V}(\theta; T)$.

We now turn to comparing welfare under the principal optimal mechanism. With perfect contractibility, the optimal allocation is $\phi^*(\theta) = \phi^P(\theta)$. With lower threshold contractibility, the optimal allocation is $\phi^*(\theta) = \min\{\max\{\phi^A(\theta), c\}, \phi^P(\theta)\}$. Compare two regimes with lower threshold contractibility indexed by $c' \leq c$ with corresponding $\phi^{*'}$ and ϕ^*. It follows that $\phi^{*'}(\theta) \leq \phi^*(\theta)$ for all $\theta \in \Theta$. Using the formula for agent welfare under the optimal mechanism, we can then see that:

$$V(\theta; T^{*'}, c') = \int_0^\theta u_\theta(\phi^{*'}(s), s)\,\mathrm{d}s \leq \int_0^\theta u_\theta(\phi^*(s), s)\,\mathrm{d}s = V(\theta; T^*, c) \tag{79}$$

for all $\theta \in \Theta$, where the inequality follows as u is single-crossing in (x, θ) and $\phi^{*'} \leq \phi^*$. Setting $c' = 0$ and $c = 1$ yields the claim in the proposition.

For the principal, by Proposition 5 we proved that a consumption allocation ϕ is incentive compatible if and only if ϕ is monotone over Θ and $\phi(\theta) \leq \max\{\delta(\phi(\theta)), \phi^A(\theta)\}$ for all $\theta \in \Theta$. Moreover, if D features more contractibility than D', we have that $\delta \geq \delta'$. Thus, this constraint is weaker under D than D' and the set of incentive compatible allocations is larger. As the monopolist optimizes over a larger set under D than D', it is immediate that $\Pi(T^*) \geq \Pi(T^{*'})$. □

We now prove Proposition 4.

Proof. We first use the envelope formula for agent welfare under the optimal mechanism to derive:

$$\Delta_V(\theta) = \int_0^\theta \Big(u_\theta(\phi^*(s), s) - u_\theta(\phi^{*'}(s), s)\Big)\,\mathrm{d}s \tag{80}$$

for all $\theta \in \Theta$. Using the virtual surplus formulation of the principal's payoff, we moreover obtain that:

$$\Delta_\Pi = \int_0^1 \Big(J(\phi^*(\theta), \theta) - J(\phi^{*'}(\theta), \theta)\Big)\,\mathrm{d}F(\theta) \tag{81}$$

See moreover that:

$$J_x(x, \theta) = \pi_x(x, \theta) + u_x(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_{x\theta}(x, \theta) \tag{82}$$

Thus, if (i) $\pi_x(\cdot, \theta)$ pointwise increases for all $\theta \in \Theta$, (ii) F decreases in the hazard rate order, then we have that $J_x(\cdot, \theta)$ pointwise increases for all $\theta \in \Theta$. As J is strictly quasiconcave in x, we have that $\phi^P(\theta)$ increases for all θ under (i), or (ii). Moreover, π and F have no impact on ϕ^A. Finally, see that we can write:

$$J(x, \theta) = J(0, \theta) + \int_0^x J_x(s, \theta)\,\mathrm{d}s \tag{83}$$

As $J_x(\cdot, \theta)$ increases pointwise, we have that $J(\cdot, \theta)$ increases pointwise for all $\theta \in \Theta$ so long as $u(0, \theta)$ does not decrease under (i), or (ii) for any $\theta \in \Theta$. To see this, observe that

$u(0, \theta) = \pi(0, \theta) = 0$ by assumption. Thus, $u_\theta(0, \theta) = 0$ also. It then follows that:

$$J(0, \theta) = \pi(0, \theta) + u(0, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(0, \theta) = 0 \tag{84}$$

for all $\theta \in \Theta$ and does not change under (i), or (ii). Hence, any change according to (i), or (ii), leads ϕ^P and J to increase pointwise while ϕ^A is unchanged.

Denote the ϕ^P and J after the change in π_x or F as $\hat{\phi}^P$ and $\hat{J}$. To prove the given claims, we need to show that:

$$\begin{aligned}&\hat{\Delta}_V(\theta) - \Delta_V(\theta) = \\ &\int_0^\theta \left[\left(u_\theta(\hat{\phi}^*(s), s) - u_\theta(\hat{\phi}^{*'}(s), s)\right) - \left(u_\theta(\phi^*(s), s) - u_\theta(\phi^{*'}(s), s)\right)\right] \mathrm{d}s \geq 0\end{aligned} \tag{85}$$

and:

$$\begin{aligned}&\hat{\Delta}_\Pi - \Delta_\Pi = \\ &\int_0^1 \left[\left(\hat{J}(\hat{\phi}^*(\theta), \theta) - \hat{J}(\hat{\phi}^{*\prime}(\theta), \theta)\right) - (J(\phi^*(\theta), \theta) - J(\phi^{*\prime}(\theta), \theta))\right] \mathrm{d}F(\theta) \geq 0\end{aligned} \tag{86}$$

Consider first the welfare of consumers. In any case when we have $\underline{D}$, we have that $\phi^*(s) = \phi^P(s)$ and $\hat{\phi}^*(s) = \hat{\phi}^P(s)$. There are three possible cases for each $s \in \Theta$ to compute the integrand, which we address in sequence:

- i $\phi^P(s) < \phi^A(s)$ and $\hat{\phi}^P(s) < \phi^A(s)$, so $\phi^{*'}(s) = \phi^P(s)$ and $\hat{\phi^*}'(s) = \hat{\phi}^P(s)$. In this case, the value of the integrand is zero.
- ii $\phi^P(s) < \phi^A(s)$ and $\hat{\phi}^P(s) \geq \phi^A(s)$, so $\phi^{*'}(s) = \phi^P(s)$ and $\hat{\phi^*}'(s) = \phi^A(s)$. Thus, the value of the integrand is $u_\theta(\hat{\phi}^P(s), s) - u_\theta(\phi^A(s), s) \geq 0$ by single-crossing of u.
- iii $\phi^P(s) \geq \phi^A(s)$ and $\hat{\phi}^P(s) \geq \phi^A(s)$, so $\phi^{*'}(s) = \phi^A(s)$ and $\hat{\phi^*}'(s) = \phi^A(s)$. Thus, the value of the integrand is $u_\theta(\hat{\phi}^P(s), s) - u_\theta(\phi^P(s), s) \geq 0$ by single-crossing of u.

Thus, the integrand is positive for all $s \in \Theta$ and the claimed inequality holds.

Consider now the welfare of the principal. Again, we have that $\phi^*(s) = \phi^P(s)$ and $\hat{\phi}^*(s) = \hat{\phi}^P(s)$, and there are three cases:

- i $\phi^P(s) < \phi^A(s)$ and $\hat{\phi}^P(s) < \phi^A(s)$, so $\phi^{*'}(s) = \phi^P(s)$ and $\hat{\phi^*}'(s) = \hat{\phi}^P(s)$. In this case, the value of the integrand is zero.
- ii $\phi^P(s) < \phi^A(s)$ and $\hat{\phi}^P(s) \geq \phi^A(s)$, so $\phi^{*'}(s) = \phi^P(s)$ and $\hat{\phi^*}'(s) = \phi^A(s)$. In this case, the value of the integrand is $\hat{J}(\hat{\phi}^P(s), s) - \hat{J}(\phi^A(s), s) \geq 0$ as $\hat{\phi}^P$ is $\hat{J}$ maximal.
- iii $\phi^P(s) \geq \phi^A(s)$ and $\hat{\phi}^P(s) \geq \phi^A(s)$, so $\phi^{*'}(s) = \phi^A(s)$ and $\hat{\phi^*}'(s) = \phi^A(s)$. In this case, the value of the integrand is $\left(\hat{J}(\hat{\phi}^P(s), s) - \hat{J}(\phi^A(s), s)\right) - \left(J(\phi^P(s), s) - J(\phi^A(s), s)\right)$,

where we wish to show that this is positive. Now observe that we can write this inequality as:

$$\int_{\phi^P(s)}^{\hat{\phi}^P(s)} \hat{J}_x(z,s)\,\mathrm{d}z + \int_{\phi^A(s)}^{\phi^P(s)} \Big(\hat{J}_x(x,s) - J_x(z,s)\Big)\,\mathrm{d}z \geq 0 \tag{87}$$

As $\hat{J}$ is strictly quasiconcave and $\hat{\phi}^P$ is $\hat{J}$ maximal, we know that $\int_{\phi^P(s)}^{\hat{\phi}^P(s)} \hat{J}_x(z,s)\,\mathrm{d}z \geq 0$. Moreover, as $\hat{J}_x \geq J_x$, we have that $\int_{\phi^A(s)}^{\phi^P(s)} \Big(\hat{J}_x(x,s) - J_x(z,s)\Big)\,\mathrm{d}z \geq 0$. The claimed inequality follows.

We have shown that the integrand is positive for all $s \in \Theta$, and the initially claimed inequality follows. □

References

Alexander, J. (2020): "Creators finally know how much money YouTube makes, and they want more of it," *The Verge*, February 4. Retrieved from https://www.theverge.com/2020/2/4/21121370/youtube-advertising-revenue-creators-demonetization-earnings-google.

Allcott, H., M. Gentzkow, and L. Song (2021): "Digital Addiction," Working Paper 28936, National Bureau of Economic Research.

Amelio, A., and B. Jullien (2012): "Tying and freebies in two-sided markets," *International Journal of Industrial Organization*, 30(5), 436–446.

Badel, A., M. Huggett, and W. Luo (2020): "Taxing top earners: a human capital perspective," *The Economic Journal*, 130(629), 1200–1225.

Baron, D. P., and R. B. Myerson (1982): "Regulating a monopolist with unknown costs," *Econometrica*, pp. 911–930.

Becker, G. S., and K. M. Murphy (1988): "A theory of rational addiction," *Journal of Political Economy*, 96(4), 675–700.

Bergemann, D., A. Bonatti, and A. Smolin (2018): "The design and price of information," *American Economic Review*, 108(1), 1–48.

Carbajal, J. C., and J. C. Ely (2013): "Mechanism design without revenue equivalence," *Journal of Economic Theory*, 148(1), 104–133.

Castro, D. (2010): "Stricter Privacy Regulations for Online Advertising Will Harm the Free Internet," Policy brief, The Information Technology & Innovation Foundation, Retrieved from: https://itif.org/publications/2010/09/08/stricter-privacy-regulations-online-advertising-will-harm-free-internet.

Choi, J. P., and D.-S. Jeon (2021): "A Leverage Theory of Tying in Two-Sided Markets with Nonnegative Price Constraints," *American Economic Journal: Microeconomics*, 13(1), 283–337.

eMarketer Insider Intelligence (2020): "Google's US Ad Revenues to Drop for the First Time," June 22. Retrieved from: https://www.emarketer.com/newsroom/index.php/google-ad-revenues-to-drop-for-the-first-time/.

Forges, F. (1986): "An approach to communication equilibria," *Econometrica*, pp. 1375–1385.

Fudenberg, D., and J. Tirole (1990): "Moral hazard and renegotiation in agency contracts," *Econometrica*, pp. 1279–1319.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2021): "A theory of auctions with endogenous valuations," *Journal of Political Economy*, 129(4), 1011–1051.

Google AdSense (2020): "Online Terms of Service," Accessed November 22, 2020, from https://www.google.com/adsense/localized-terms.

Grubb, M. D. (2009): "Selling to overconfident consumers," *American Economic Review*, 99(5), 1770–1807.

Guernsey, L. (1999): "Can it Pay to Surf the Web?," *The New York Times*, November 22.

Guesnerie, R., and J.-J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Probems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25(3), 329–369.

Guynn, J., and K. McCoy (2018): "Apple co-founder Steve Wozniak says he's left Facebook over data collection," *USA Today*, April 8.

IHS Markit (2017): "The Economic Value of Behavioural Targeting in Digital Advertising," Research report, Retrieved from: https://iabeurope.eu/knowledgehub/policy/the-economic-value-of-data-driven-advertising/.

Kang, Y. P. (2000): "It Pays to Cheat, Not Surf," *Wired*, July 10.

Laffont, J.-J., and J. Tirole (1986): "Using cost observation to regulate firms," *Journal of Political Economy*, 94(3, Part 1), 614–641.

Lockwood, B. B., C. G. Nathanson, and E. G. Weyl (2017): "Taxation and the Allocation of Talent," *Journal of Political Economy*, 125(5), 1635–1682.

Maskin, E., and J. Riley (1984): "Monopoly with incomplete information," *The RAND Journal of Economics*, 15(2), 171–196.

Milgrom, P., and I. Segal (2002): "Envelope theorems for arbitrary choice sets," *Econometrica*, 70(2), 583–601.

MILGROM, P., AND C. SHANNON (1994): "Monotone comparative statics," *Econometrica*, pp. 157–180.

MIRRLEES, J. A. (1971): "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, 38(2), 175–208.

MUSSA, M., AND S. ROSEN (1978): "Monopoly and product quality," *Journal of Economic Theory*, 18(2), 301–317.

MYERSON, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics*, 10(1), 67–81.

NÖLDEKE, G., AND L. SAMUELSON (2007): "Optimal bunching without optimal control," *Journal of Economic Theory*, 134(1), 405–420.

POSNER, E. A., AND E. G. WEYL (2018): *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

ROTHSCHILD, C., AND F. SCHEUER (2016): "Optimal taxation with rent-seeking," *The Review of Economic Studies*, 83(3), 1225–1262.

STRAUSZ, R. (2017): "A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard," *American Economic Review*, 107(6), 1430–76.

SUNDARARAJAN, A. (2004): "Nonlinear pricing of information goods," *Management Science*, 50(12), 1660–1673.

TIKTOK (2021): "TikTok Creator Fund Terms," Accessed November 12, 2021, from https://www.tiktok.com/legal/tiktok-creator-fund-terms?lang=en.

WILSON, R. B. (1993): *Nonlinear Pricing*. Oxford University Press, New York.

YOUTUBE HELP (2021): "YouTube Partner Program overview & eligibility," Accessed October 30, 2021, from https://support.google.com/youtube/answer/72851.

Online Appendix

B Additional Proofs of Results in the Main Text

B.1 Proof of Corollary 1

Proof. If $H(\min X^*) > 0$, we observe that there is automatically a flat region for sufficiently low x. There is then one additional flat region for every two times the function $H(x)$ "strictly" crosses zero from below. Thus, the number of flat regions is $1 + \left\lfloor \frac{k(H)}{2} \right\rfloor$. To see this formally, first suppose that H is strictly increasing or decreasing at any x such that $H(x) = 0$. In this case, $k(H) = |\{x \in X^* : \exists \epsilon > 0, \forall y \in (x-\epsilon, x), \forall y' \in (x, x+\epsilon), H(y) < 0, H(y') > 0\}|$, and the result is immediate by Proposition 2. When H is neither strictly increasing nor decreasing at x such that $H(x) = 0$, there are three cases. First, $H(x) = 0$ and there exists $\varepsilon > 0$ such that $H(y) < 0$ for all $y \in (x-\varepsilon, x)$ and $H(y') \geq 0$ for all $y' \in (x, x+\varepsilon)$. These values of x imply the existence of a flat region by Proposition 2. Second, if $H(x) = 0$ but there exists $\varepsilon > 0$ such that $H(y) = 0$ for all $y \in (x-\varepsilon, x)$. These values of x occupy a flat region, but they do not index a tier as they are necessarily preceded by an x that falls into the first category. Third, $H(x) = 0$ but there exists $\varepsilon > 0$ such that $H(y) < 0$ for all $y \in (x-\varepsilon, x+\varepsilon)$. In this case, H hits 0 but the price schedule is not flat at x. Thus, we must "remove" all such from $k(H)$. Performing these removals yields the claimed $k(H)$.

Conversely, if $H(0) < 0$, there is one flat region the first time that $H(x)$ crosses zero, and a further flat region every two times the function $H(x)$ "strictly" crosses zero from below. By the same arguments, the number of flat regions is $\left\lceil \frac{k(H)}{2} \right\rceil$. $\square$

B.2 Proof of Corollary 2

Proof. Each part follows immediately from Proposition 2. $\square$

B.3 Proof of Corollary 3

Proof. By Proposition 2, if $H(x) > 0$ at $\phi^A(1)$, then the price schedule is flat at $\phi^*(1) = \phi^A(1)$. Moreover, at $x = \phi^A(1)$, we have $H(\phi^A(1)) = f(1)\pi_x(\phi^A(1), 1)$. It follows that when $\pi_x(\phi^A(1), 1) > 0$, that $H(\phi^A(1)) > 0$ and the price schedule features a premium tier. Likewise, if $H(x) > 0$ at $\phi^A(0)$, then the price schedule is flat at $\phi^*(0) = \phi^A(0)$. Moreover, at $x = \phi^A(0)$, we have that $H(\phi^A(0)) = f(0)\pi_x(\phi^A(0), 0) - u_{x\theta}(\phi^A(0), 0)$. It follows that when $f(0)\pi_x(\phi^A(0), 0) - u_{x\theta}(\phi^A(0), 0) > 0$ that $H(\phi^A(0)) > 0$ and the price schedule features a trial tier. $\square$

B.4 Proof of Lemma 1

Proof. Let $D : X \rightrightarrows X$ be a regular disposal correspondence and define $\delta(x) = \min D(x)$ for all $x \in X$. We next show that δ satisfies all the properties in the statement. First, note that δ is well-defined given that D is closed-valued. Also, note that $\delta(x) \leq x$ for all $x \in X$, given that D satisfies no over-utilization and δ is increasing given monotonicity of D and the properties of the strong set order. Next, observe that, for every $x \in X$, it holds $D(x) \subseteq [\delta(x), x]$ since $\delta(x) = \min D(x)$ and $x = \max D(x)$. Assume by contradiction that there exists $x \in X$ and $y \in [\delta(x), x]$ such that $y \in X \setminus D(x)$. In this case, we would have $y = \max\{y, \delta(x)\}$ and $\delta(x) = \min\{y, \delta(x)\}$ with $\delta(x) \in D(x)$ but $y \in D(y)$ and $y \in X \setminus D(x)$, which contradicts monotonicity of D in the strong set order. With this, we must have $D(x) = [\delta(x), x]$ for all $x \in X$. Next, we claim that, for every $x, y \in X$, if $y \in [\delta(x), x]$, then $\delta(y) = \delta(x)$. Fix $x, y \in X$ and assume that $y \in [\delta(x), x]$. We already know that $\delta(y) \leq \delta(x)$. Suppose by contradiction that $\delta(y) < \delta(x)$. But then, given the other properties of δ, for all $z \in (\delta(y), \delta(x))$ we would have that $z \in D(y)$ but $z \in X \setminus D(x)$ which would contradict transitivity as $y \in D(x) = [\delta(x), x]$. This establishes necessity. Conversely, fix a function $\delta : X \to X$ as in the statement. We want to show that $D : X \rightrightarrows X$ defined as $D(x) = [\delta(x), x]$ for all $x \in X$ is a regular disposal correspondence. Clearly, D satisfies reflexivity, no over-utilization, closed-valuedness, and monotonicity given that δ is increasing. Next, fix $x \in X$ and let $y \in D(x)$. Given the properties of δ we have $\delta(y) = \delta(x)$. But then we have $D(y) = [\delta(y), y] = [\delta(x), y] \subseteq [\delta(x), x]$. Given that x was arbitrarily chosen this shows that D satisfies transitivity and that is a regular disposal correspondence concluding the proof. $\square$

B.5 Proof of Lemma 2

Proof. We begin by eliminating the proposed allocation and transfers from the objective function of the seller. From the proof of Proposition 5, we have that transfers for any incentive compatible triple (ξ, ϕ, t) are given by:

$$t(\theta) = C + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \tag{88}$$

for some constant $C \in \mathbb{R}$. Thus, any ξ that supports ϕ leads to the same seller payoff and can therefore be made equal to ϕ without loss of optimality. Moreover, we know that ϕ being incentive compatible is equivalent to ϕ being monotone increasing and $\phi \leq \max\{\delta \circ \phi, \phi^A\}$. Finally, it is not optimal for the seller to exclude any agent from the mechanism as it is without loss to allocate any agent $x = 0$ rather than exclude them owing to the fact that

$\pi(0,\cdot) = 0$, $u(0,\cdot) = 0$, $u(x,\cdot)$ is monotone increasing over Θ, and $u(x,\theta)$ has single-crossing in (x,θ). In particular, for any incentive compatible allocation that excludes some type θ, it is without loss of optimality to instead set $\phi(\theta) = \xi(\theta) = t(\theta) = 0$. Each agent is indifferent between participation and not, and this does not change the principal's payoff.

Plugging in the expression (88), we can simplify the expression for the seller's total transfer revenue as the following:

$$\begin{aligned}\int_\Theta t(\theta)\,\mathrm{d}F(\theta) &= \int_\Theta \left(C + u(\phi(\theta),\theta) - \int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\right)\mathrm{d}F(\theta)\\ &= \int_\Theta (C + u(\phi(\theta),\theta))\,\mathrm{d}F(\theta) - \int_0^1\int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\,\mathrm{d}F(\theta)\end{aligned} \tag{89}$$

Using this expression for total transfer revenue, and the characterization of implementation from Proposition 5, we write the seller's problem as

$$\begin{aligned}&\max_{\phi,C}\ \int_\Theta \left(\pi(\phi(\theta),\theta) + C + u(\phi(\theta),\theta) - \int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\right)\mathrm{d}F(\theta)\\ &\text{s.t.}\quad \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \leq \max\{\delta(\phi(\theta)),\phi^A(\theta)\}\quad \forall\theta,\theta'\in\Theta:\theta'\geq\theta\\ &\qquad u(\phi(\theta),\theta) - \left(C + u(\phi(\theta),\theta) - \int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\right) \geq 0\quad \forall\theta\in\Theta\end{aligned} \tag{90}$$

We further simplify this by applying integration by parts on the double integral of $u_\theta(\phi(s),s)$ over θ and s:

$$\begin{aligned}\int_0^1\int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\,\mathrm{d}F(\theta) &= \left[F(\theta)\int_0^\theta u_\theta(\phi(s);s)\,\mathrm{d}s\right]_0^1 - \int_0^1 F(\theta)u_\theta(\phi(\theta),\theta)\,\mathrm{d}\theta\\ &= \int_0^1 (1-F(\theta))u_\theta(\phi(\theta),\theta)\,\mathrm{d}\theta\\ &= \int_0^1 \frac{(1-F(\theta))}{f(\theta)}u_\theta(\phi(\theta),\theta)\,\mathrm{d}F(\theta)\end{aligned} \tag{91}$$

Plugging into the seller's objective, we find that the principal solves:

$$\begin{aligned}&\max_{\phi,C}\ \int_\Theta (J(\phi(\theta)) + C)\,\mathrm{d}F(\theta)\\ &\text{s.t.}\quad \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \leq \max\{\delta(\phi(\theta)),\phi^A(\theta)\}\quad \forall\theta,\theta'\in\Theta:\theta'\geq\theta\\ &\qquad u(\phi(\theta),\theta) - \left(C + u(\phi(\theta),\theta) - \int_0^\theta u_\theta(\phi(s),s)\,\mathrm{d}s\right) \geq 0\quad \forall\theta\in\Theta\end{aligned} \tag{92}$$

It follows that it is optimal to set $C \in \mathbb{R}$ as large as possible such that:

$$V(\theta) = u(\phi(\theta), \theta) - \left(C + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\mathrm{d}s\right) \geq 0 \quad \forall \theta \in \Theta \tag{93}$$

We know that $V'(\theta) = u_\theta(\phi(\theta), \theta) \geq 0$ as $u(x, \cdot)$ is monotone over Θ. Thus, the tightest such constraint occurs when $\theta = 0$. Hence, the maximal C must satisfy:

$$V(0) = -C \geq 0 \tag{94}$$

Which implies that C is optimally 0 and ensures that the (IR) constraint holds for all types. Hence, the seller's program is:

$$\begin{aligned} \max_{\phi} \quad & \int_\Theta J(\phi(\theta))\,\mathrm{d}F(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \in [0, \phi^A(\theta)] \cup \{x \in X : \delta(x) = x\} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{95}$$

This completes the proof. □

B.6 Proof of Proposition 6

Proof. The upper-threshold disposal correspondence is given by:

$$D(x) = \begin{cases} [0, x] & \text{if } x < c, \\ \{x\} & \text{if } x \geq c. \end{cases} \tag{96}$$

Consider the resulting pointwise maximization problem of the seller:

$$\max_{x \in X} J(x, \theta) \quad \text{s.t.} \quad \mathbb{I}[x < c]\left(\phi^A(\theta) - x\right) \geq 0 \tag{97}$$

where the constraint is the capping constraint derived in Proposition 5 in this context. The solution to this pointwise problem, by strict quasi-concavity of J in x for all $\theta \in \Theta$, is given by:

$$\phi^*(\theta) = \begin{cases} \phi^P(\theta) & \text{if } \phi^P(\theta) \geq c, \\ \phi^P(\theta) & \text{if } \phi^P(\theta) < c, \phi^P(\theta) \leq \phi^A(\theta), \\ \phi^A(\theta) & \text{if } \phi^P(\theta) < c, \phi^P(\theta) > \phi^A(\theta), J(\phi^A(\theta), \theta) \geq J(c, \theta), \\ c & \text{if } \phi^P(\theta) < c, \phi^P(\theta) > \phi^A(\theta), J(\phi^A(\theta), \theta) < J(c, \theta). \end{cases} \tag{98}$$

Recall that ϕ^* it is implementable when it is monotone by Proposition 5. Moreover, when it is implementable, it is optimal as it solves the program in Lemma 2. We now argue that the fact that $\Delta_c(\theta)$ crosses zero from below at most once is sufficient for ϕ^* to be monotone. Concretely, either (i) $\Delta_c(\theta)$ does not cross zero and only one of cases 3 and 4 above apply or (ii) $\Delta_c(\theta_1) = 0$ and case 3 applies for $\theta < \theta_1$ and case 4 applies for $\theta \geq \theta_1$. In cases 3 and 4, $\phi^A(\theta) < \phi^P(\theta) < c$, so going from $\phi^A(\theta)$ to c at θ_1 leads ϕ^* to increase. Moreover, we have that ϕ^P crosses ϕ^A from below at $\hat{\theta}$. Thus, case 2 obtains for $\theta \leq \hat{\theta}$. Then, if $\hat{\theta} < c$, case 3 obtains for $\theta \leq \theta_1$ and case 4 then holds from θ_1 until $\phi^P(\theta) = c$, which happens at θ_2. Thus, the contract is of the form claimed in the result. Moreover, it is monotone and therefore both implementable and optimal. This concludes the proof.

We can also provide conditions such that $\Delta_c(\theta)$ is single-crossing. A sufficient condition is that $\Delta_c'(\theta) \geq 0$. We can compute:

$$\Delta_c'(\theta) = J_\theta(c, \theta) - J_\theta(\phi^A(\theta), \theta) - J_x(\phi^A(\theta), \theta) \left(\phi^A\right)' (\theta) \tag{99}$$

This is greater than zero exactly when:

$$\int_{\phi^A(\theta)}^{c} J_{x\theta}(\tilde{x}, \theta) \, \mathrm{d}\tilde{x} \geq J_x(\phi^A(\theta), \theta) \left(\phi^A\right)' (\theta) \tag{100}$$

□

B.7 Proof of Lemma 4

Proof. First, we observe by monotonicity of ϕ and Lemma 1 that there are three cases to consider: (i) $\phi(s) = \delta_0$ for all $s \in [\theta, \theta']$ (ii) $\phi(s) > \delta_0$ for all $s \in [\theta, \theta']$ and (iii) there exists some $\hat{\theta} \in [\theta, \theta']$ such that $\phi(s) > \delta_0$ for all $s > \hat{\theta}$ and $\phi(s) = \delta$ for all $s \leq \hat{\theta}$. In the first case, observe that allocations are constant and so too are transfers. In the second case, fix any $s, s' \in [\theta, \theta']$ such that $s' \geq s$. By our construction of t, it is almost everywhere differentiable. Denote the countable set of discontinuity points of ϕ between any two types s, s' as $\mathcal{D}(s, s')$. It follows that we may write that:

$$t(s') - t(s) = \sum_{d \in \mathcal{D}(s,s')} \Delta t_d + \int_s^{s'} t'(z) \, \mathrm{d}z \tag{101}$$

where we define for any two monotone sequences $\{s_n\}_{n\in\mathbb{N}}$ such that $s_n \to^+ d$ and $\{s_n'\}_{n\in\mathbb{N}}$ such that $s_n' \to^- d$, for any discontinuity point, the value $\Delta t_d = \lim_{n\to\infty} t(s_n') - \lim_{n\to\infty} t(s_n)$.

It is simple to see that $t'(s) = u_x(\phi(s), s)\phi'(s) \geq 0$ as $\phi(s) \leq \phi^A(s)$ and $\phi' \geq 0$. We have that, $\phi(s_n') \to^- \overline{\phi}$ and $\phi(s_n) \to^+ \underline{\phi}$ for some $\underline{\phi} \leq \overline{\phi}$ as $\phi(s_n')$ and $\phi(s_n)$ are respectively monotone decreasing and increasing in $n \in \mathbb{N}$ and bounded. Hence, we can write:

$$\begin{aligned}
\Delta t_d &= \lim_{n\to\infty} t(s'_n) - \lim_{n\to\infty} t(s_n) \\
&= \lim_{n\to\infty} \left[u(\phi(s'_n), s'_n) - \int_0^{s'_n} u_\theta(\phi(z), z)\mathrm{d}z \right] \\
&\quad - \lim_{n\to\infty} \left[u(\phi(s_n), s_n) - \int_0^{s_n} u_\theta(\phi(z), z)\mathrm{d}z \right] \\
&= \lim_{n\to\infty} u(\phi(s'_n), s'_n) - u(\phi(s_n), s_n) \\
&= u(\overline{\phi}, d) - u(\underline{\phi}, d) \\
&\geq 0
\end{aligned} \tag{102}$$

as $\overline{\phi} \leq \phi^A(d)$ by the fact that $\phi(s) \leq \phi^A(s)$ for all $s, s' \in [\theta, \theta']$, and u is strictly quasiconcave. Thus, we have that $t(s') \geq t(s)$, as claimed.

For the third case, see that we can write:

$$t(s') - t(s) = t(s') - t(\hat{s}) + (t(\hat{s}) - t(s)) \geq 0 \tag{103}$$

The first term is positive by the argument made in case two, and the second term is zero by the argument made in case one. □

C Additional Results

C.1 Optimal Bunching and Free Disposal

As anticipated earlier in the paper (Section 3), this section extends our main analysis with $D = \overline{D}$ to cover cases in which the virtual surplus function J does not satisfy single-crossing. To achieve this relaxation and study scenarios where bunching is optimal, we apply techniques from Nöldeke and Samuelson (2007) to study the inverse problem of assigning types to consumption under the restriction that J is concave. Only to simplify exposition, we also assume that the agent's payoff function u admits a continuous third cross derivative $u_{xx\theta}$.

To this end, denote an inverse consumption function with $\psi : X \to \Theta$. This will be interpreted as a suitable inverse of the standard consumption function ϕ from our earlier analysis. In particular, ψ assigns to each outcome a type, whereas ϕ assigns to each type an outcome. For any monotone ψ, define further the correspondence:

$$\Psi(x) = \left[\lim_{y\to^- x} \psi(y), \lim_{y\to^+ x} \psi(y) \right] \tag{104}$$

which "fills in" discontinuities in the inverse consumption function.[31]

Our first result concerns implementation in this setting.

Lemma 5. *A consumption allocation ϕ is incentive compatible if and only if there exists an inverse consumption $\psi : X \to \Theta$ that is monotone increasing, satisfies $\psi(x) \geq \left(\phi^A\right)^{-1}(x)$ for all $x \in X$, and is such that $\theta \in \Psi(\phi(\theta))$ for all $\theta \in \Theta$.*

Proof. This follows immediately from the proof of Proposition 5 and from Lemma 1 and Lemma 2 in Nöldeke and Samuelson (2007). □

Our previous result about the *cap* on optimal assignments of outcomes (Proposition 1) now translates into a *floor* in the inverse consumption function. The corresponding supporting price schedule can then be solved for in integral from:

$$T_\psi(x) = C + \int_0^x u_x(z, \psi(z))\,\mathrm{d}z \tag{106}$$

Most importantly, we now provide the solution to the principal's screening problem taking bunching into account. Toward simplifying the principal's problem, we define the following function:

$$\hat{J}(x,\theta) = u_x(x,\theta)(1 - F(\theta)) + \int_\theta^1 \pi_x(x,s)\,\mathrm{d}F(s) \tag{107}$$

Using this function and Lemma 5, we can re-express the principal's problem as:

$$\begin{aligned} &\max_\psi \int_0^{\overline{x}} \hat{J}(x,\psi(x))\,\mathrm{d}x \\ \text{s.t.}\quad &\psi(x') \geq \psi(x),\ \psi(x) \geq \left(\phi^A\right)^{-1}(x),\ \forall x', x \in X : x' \geq x \end{aligned} \tag{108}$$

This program can be solved pointwise under our assumptions:

Proposition 8. *The principal's Problem (1) is solved by the inverse consumption ψ^* defined for each $x \in X$ as:*

$$\psi^*(x) = \max\left\{\arg\max_{\theta \in [(\phi^A)^{-1}(x), 1]} \hat{J}(x,\theta)\right\} \tag{109}$$

Proof. We first show that $\hat{J}$ is supermodular. We follow Lemma 6 in Nöldeke and Samuelson (2007) and observe that the cross partial derivative of $\hat{J}$ is:

[31] Where we follow the convention from Nöldeke and Samuelson (2007) that:

$$\lim_{y \to^- 0} \psi(y) = 0, \quad \lim_{y \to^+ \overline{x}} \psi(y) = 1 \tag{105}$$

$$
\begin{aligned}
J_{x,\theta}(x,\theta) &= -\left[u_{xx}(x,\theta)+\pi_{xx}(x,\theta)\right]f(\theta)-\left[1-F(\theta)\right]u_{xx\theta}(x,\theta) && (110)\\
&= -J_{xx}(x,\theta)f(\theta) && (111)
\end{aligned}
$$

Given that J is concave, it follows that $\hat{J}$ is supermodular. Next, we argue that the correspondence $x \mapsto [(\phi^A)^{-1}(x), 1]$ is monotone in the SSO. Indeed, this immediately follows from the fact that $(\phi^A)^{-1}$ is strictly increasing. With this, we can apply Theorem 4' in Milgrom and Shannon (1994) to argue that ψ^* is monotone increasing. Given that by construction we have $\psi^* \geq (\phi^A)^{-1}$, the inverse consumption function ψ^* is implementable and a fortiori optimal. □

As in Nöldeke and Samuelson (2007), bunching manifests in the solution to this problem as discontinuity in the resulting inverse consumption function ψ. In particular, whenever ψ is discontinuous the outcome at the discontinuity is assigned to a positive measure of types.

C.2 Competition and Free Disposal

In this Appendix, we study the relationship between competition and optimal pricing under free disposal. We do this by comparing the monopoly screening benchmark with perfect competition. We show that our results are robust to this extension by demonstrating that zero marginal pricing is in fact more prevalent under perfect competition.

The nature of perfect competition we consider is that our monopolist faces a perfectly fringe of firms that can enter and displace them to serve the entire market. We make this simplifying assumption to avoid the complicated issue of endogenous market segmentation under competitive screening and study the cleanly competition's impact. In this case (as in e.g., Grubb, 2009), the equilibrium contract maximizes expected consumer surplus subject to our usual implementation constraints and a new constraint that the monopolist actually wishes to serve the market. That is, the screening problem becomes:

$$
\begin{aligned}
\sup_{\phi,\xi,T} \quad & \int_\Theta \left(u(\phi(\theta),\theta) - T(\xi(\theta))\right) \mathrm{d}F(\theta) \\
\text{s.t.} \quad & \phi(\theta) \in \arg\max_{x\in D(\xi(\theta))} u(x,\theta) \quad \text{for all} \quad \theta\in\Theta \\
& \xi(\theta) \in \arg\max_{y\in X} \left\{ \max_{x\in D(y)} u(x,\theta) - T(y) \right\} \quad \text{for all} \quad \theta\in\Theta \qquad (112)\\
& u(\phi(\theta),\theta) - T(\xi(\theta)) \geq 0 \quad \text{for all} \quad \theta\in\Theta \\
& \int_\Theta \left(\pi(\phi(\theta),\theta) + T(\xi(\theta))\right) \mathrm{d}F(\theta) \geq 0
\end{aligned}
$$

The last constraint encodes the effects of competition, and we refer to it as "monopolist's IR (individual rationality)." Toward characterizing the solution of this problem, define the total surplus function as:

$$S(x, \theta) = \pi(x, \theta) + u(x, \theta) \tag{113}$$

In analogy to our assumptions that J is strictly single-crossing and quasiconcave, we assume that S is strictly single crossing in (x, θ) and strictly quasiconcave in x. We further define the total surplus maximizing consumption level as:

$$\phi^O(\theta) = \arg\max_{x \in X} S(x, \theta) \tag{114}$$

Proposition 9. *The equilibrium consumption level under perfect competition is given by:*

$$\phi^{PC} = \min\{\phi^A, \phi^O\} \tag{115}$$

Proof. As in the proof of Lemma 2, we observe that agents' transfers under any locally incentive compatible menu are given by:

$$t(\theta) = C + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\mathrm{d}s \tag{116}$$

We can therefore rewrite the objective as:

$$-C + \int_\Theta \frac{1 - F(\theta)}{f(\theta)} u_\theta(\phi(\theta), \theta)\mathrm{d}F(\theta) \tag{117}$$

By integrating over types, we can then express the monopolist's IR constraint as:

$$\int_\Theta \left(\pi(\phi(\theta), \theta) + u(\phi(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(\phi(\theta), \theta)\right) \mathrm{d}F(\theta) + C \geq 0 \tag{118}$$

Thus, any optimal C sets:

$$-C = \int_\Theta \left(\pi(\phi(\theta), \theta) + u(\phi(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(\phi(\theta), \theta)\right) \mathrm{d}F(\theta) \tag{119}$$

Which implies that the objective function can be written as:

$$\int_\Theta \left(\pi(\phi(\theta), \theta) + u(\phi(\theta), \theta)\right) \mathrm{d}F(\theta) \tag{120}$$

Moreover, by the same arguments as in Proposition 5, the remaining implementation constraints are that $\phi(\theta) \leq \phi^A(\theta)$ for all $\theta \in \Theta$, ϕ is monotone increasing and $u(\phi(0), 0) - t(0) \geq 0$. Thus the problem under perfect competition solves:

$$\begin{aligned} \max_{\phi} \quad & \int_{\Theta} S(\phi(\theta), \theta)\, \mathrm{d}F(\theta) \\ \text{s.t} \quad & \phi(\theta) \leq \phi^A(\theta) \,\text{for all}\, \theta \in \Theta \\ & \phi \,\text{is monotone increasing} \\ & u(\phi(0), 0) - t(0) \geq 0 \end{aligned} \tag{121}$$

By identical arguments to Proposition 7 (as TS is strictly single-crossing and quasiconcave), it follows that the optimal consumption levels satisfy:

$$\phi^{PC}(\theta) = \min\{\phi^A(\theta), \phi^O(\theta)\} \tag{122}$$

which is monotone. Moreover, $t(0) = C + u(\phi^{PC}(0), 0) \leq 0$ as C is negative and $u(\phi^{PC}(0), 0) \geq 0$ as $\phi^{PC}(0) \in [0, \phi^A(0)]$. □

We now highlight the key difference between perfect competition and monopoly screening: zero marginal pricing is in fact more prevalent under perfect competition. To see this, observe that $\phi^O \geq \phi^P$. Thus, the constraint of free disposal is in fact more often binding under perfect competition than under monopoly. As a result, heightened competition in fact leads to a greater incentive for zero marginal pricing. The intuition behind this result is that there are no quantity distortions from information rents under the competitive solution. Thus, total surplus maximizing consumption is greater than virtual surplus maximizing consumption, and the inability of consumers to commit to consuming more leads to more frequent zero marginal pricing. This discussion is formalized in the following corollary:

Corollary 4. *The set of outcomes at which there is flat pricing under perfect competition includes the set of outcomes at which there is flat pricing under monopoly pricing.*

Proof. Define $H^{PC}(x) = S_x\left(x, \left(\phi^A\right)^{-1}(x)\right)$. We have that for all $x \in X$:

$$\begin{aligned} H^{PC}(x) &= S_x\left(x, \left(\phi^A\right)^{-1}(x)\right) = u_x\left(x, \left(\phi^A\right)^{-1}(x)\right) + \pi_x\left(x, \left(\phi^A\right)^{-1}(x)\right) \\ &\geq u_x\left(x, \left(\phi^A\right)^{-1}(x)\right) + \pi_x\left(x, \left(\phi^A\right)^{-1}(x)\right) - \frac{1 - F\left(\left(\phi^A\right)^{-1}(x)\right)}{f\left(\left(\phi^A\right)^{-1}(x)\right)} u_{x\theta}\left(x, \left(\phi^A\right)^{-1}(x)\right) \\ &= J_x\left(x, \left(\phi^A\right)^{-1}(x)\right) = H(x) \end{aligned} \tag{123}$$

Thus, $H(x) \geq 0 \implies H^{PC}(x) \geq 0$. Hence, by an identical argument to Proposition 2, whenever T^* is flat, so too is T^{PC}. $\square$

C.3 Optimal Bundling of Advertisements and Goods

For the main analysis, we took the principal's revenues π as given. In the context of digital goods pricing, as we have explained, it is common for a principal to derive revenue from advertising. Of course, the choice of how to integrate advertising into the product is often a choice variable, selected to balance consumer usage of the platform and revenue derived from advertising. In this section, we show how such an upstream bundling consideration can be studied as an extension of our main analysis, and show in an example how the scope for under-utilization reduces the optimal level of advertising that is bundled with the good.

The principal can choose some level of advertising $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ to bundle with the product. When they choose α, this leads to an agent utility function $u(\cdot;\alpha)$ and principal revenue function $\pi(\cdot;\alpha)$. Conditional on the choice of α, under our maintained assumptions, we have that the optimal contract is given by $\phi^*(\theta;\alpha) = \min\{\phi^P(\theta;\alpha), \phi^A(\theta;\alpha)\}$. Therefore the principal's payoff given a choice of advertising level is:

$$\Lambda(\alpha) = \int_0^1 J(\phi^*(\theta;\alpha), \theta; \alpha) \, \mathrm{d}F(\theta) \tag{124}$$

or the payoff from the optimal contract conditional on α. The principal's problem is then to choose $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ to maximize Λ. Denoting an optimal choice should one exist by α^*, one can interpret the model in the main text as taking agent utility as $u(\cdot;\alpha^*)$ and principal revenue as $\pi(\cdot;\alpha^*)$.

Of most interest given our analysis is understanding how under-utilization shapes the optimal level of advertising. To build intuition for this trade-off, we study a simple example in the remainder of this appendix. Agent type θ has payoffs:

$$u(x, \theta; \alpha) = (\theta - \kappa\alpha)x - \frac{x^2}{2} \tag{125}$$

for some $\kappa \in [0, 1]$ which indexes how much advertising reduces agents' enjoyment of the good. The types are uniformly distributed on $[0, 1]$. The principal's utility is given by:

$$\pi(x, \theta; \alpha) = \alpha x \tag{126}$$

and advertising is such that $\alpha \in [0, 1]$. In this model, advertising at intensity α creates $(1 - \kappa)a$ units of surplus per unit of consumption x, taking the difference of direct revenues and consumer annoyance in units of willingness to pay.

Plugging these payoff functions into our main analysis, we derive the virtual surplus function

$$J(x, \theta; a) = (2\theta + (1-\kappa)\alpha - 1)x - \frac{x^2}{2} \tag{127}$$

and the agent and principal optimal assignments

$$\begin{aligned} \phi^A(\theta; \alpha) &= \theta - \kappa\alpha \\ \phi^P(\theta; \alpha) &= 2\theta + (1-\kappa)\alpha - 1 \end{aligned} \tag{128}$$

where we have assumed that $X = [-M, M]$, for some sufficiently large M. Thus, the optimum for given α is given by:

$$\phi^*(\theta; \alpha) = \min\{\theta - \kappa\alpha, 2\theta + (1-\kappa)\alpha - 1\} \tag{129}$$

and the principal's objective (124) can be readily calculated in terms of primitives.

Let the optimal choice of advertising given free disposal be $\alpha^*(\overline{D})$ and the optimal choice of advertising given full contractibility of consumption be $\alpha^*(\underline{D})$. The result below establishes that advertising is weakly lower under free disposal, or $\alpha^*(\overline{D}) \leq \alpha^*(\underline{D})$. Intuitively, when agents can under-utilize the good, the impact of advertising reducing agents' bliss points and tightening the obedience constraint reduces the marginal benefit of additional advertising; or, more loosely, a world of feasible "pay to click" businesses would feature more advertising overall.

Corollary 5. *The optimal level of advertising is lower with free disposal than with perfect contractibility of consumption, $\alpha^*(\overline{D}) \leq \alpha^*(\underline{D})$.*

Proof. Define $\overline{\Lambda}(\alpha)$, and $\underline{\Lambda}(\alpha)$ respectively as $\Lambda(\alpha)$ under $\overline{D}$ (free disposal) and $\underline{D}$ (perfect contractibility). A sufficient condition for $\alpha^*(\overline{D}) \leq \alpha^*(\underline{D})$ is that $\overline{\Lambda}(\alpha) - \underline{\Lambda}(\alpha)$ is decreasing in α for all $\alpha \in [0, 1]$. Observe that we can write:

$$\overline{\Lambda}(\alpha) - \underline{\Lambda}(\alpha) = \int_{\{\theta: \phi^A(\theta;\alpha) \leq \phi^P(\theta;\alpha)\}} \left(J(\phi^A(\theta; \alpha), \theta; \alpha) - J(\phi^P(\theta; \alpha), \theta; \alpha) \right) \mathrm{d}\theta \tag{130}$$

In our parametric example, we have that $\phi^A(\theta; \alpha)$ crosses $\phi^P(\theta; \alpha)$ from above at some $\tilde{\theta}(\alpha) = 1 - \alpha \in [0, 1]$. Thus, we can simplify this expression to:

$$\overline{\Lambda}(\alpha) - \underline{\Lambda}(\alpha) = \int_{\tilde{\theta}(\alpha)}^{1} \left(J(\phi^A(\theta; \alpha), \theta; \alpha) - J(\phi^P(\theta; \alpha), \theta; \alpha) \right) \mathrm{d}\theta \tag{131}$$

We can then compute:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\overline{\Lambda}(\alpha)-\underline{\Lambda}(\alpha)\right] = -\tilde{\theta}'(\alpha)\Big(J(\phi^A(\tilde{\theta}(\alpha);\alpha),\tilde{\theta}(\alpha);\alpha) - J(\phi^P(\tilde{\theta}(\alpha);\alpha),\tilde{\theta}(\alpha);\alpha)\Big) \\ + \int_{\tilde{\theta}(\alpha)}^{1}\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[J(\phi^A(\theta;\alpha),\theta;\alpha) - J(\phi^P(\theta;\alpha),\theta;\alpha)\right]\mathrm{d}\theta\end{aligned} \tag{132}$$

Observe that the term on the first line is negative as $\tilde{\theta}'(\alpha)$ is negative and the integrand is negative by optimality of ϕ^P. To show that $\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\overline{\Lambda}(\alpha)-\underline{\Lambda}(\alpha)\right] \le 0$ for all $\alpha \in [0,1]$, it then suffices to show that for all $\theta \in [\tilde{\theta}(\alpha),1]$:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[J(\phi^A(\theta;\alpha),\theta;\alpha) - J(\phi^P(\theta;\alpha),\theta;\alpha)\right] \le 0 \tag{133}$$

Given our parametric assumptions, simple algebra reveals that:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[J(\phi^A(\theta;\alpha),\theta;\alpha) - J(\phi^P(\theta;\alpha),\theta;\alpha)\right] = \theta + \alpha - 1 \le 0 \tag{134}$$

whenever $\theta \ge \tilde{\theta}(\alpha) = 1-\alpha$, completing the proof. □

D Microfoundations of Revenue from Usage

In this appendix, we describe four more detailed microfoundations that describe how sellers may derive external revenue π in the manner described by our model from: Nash bargaining with external advertisers (D.1); cross-agent platform externalities (D.2); irrational addiction (D.3); and overconfidence (D.4), as in Grubb (2009).

D.1 Advertisements

In the main text, we assumed that the principal receives payoff $\pi(x,\theta)$ when a type θ consumes x. In the advertising revenue interpretation of our model, we argued that this corresponds to physical revenue earned by the advertiser: when an agent of type θ spends x units of time on the platform, they will spend in expectation $\pi(x,\theta)$ on the good being advertised by the advertiser. Of course, at a deeper level, this interpretation can only arise as the outcome of some bargaining procedure between the platform (i.e., the "seller" in our main analysis) and the advertiser. In this appendix, we show that if the platform and advertiser engage in Nash bargaining, then the direct revenue is exactly correct, providing a deeper justification for the interpretation we provide.

By the taxation and revelation principles, as in our earlier analysis, it is without loss of generality to suppose that the platform and advertiser must, as the outcome of some bargaining process, design a tariff $T : X \to \mathbb{R}$. When the tariff is fixed, agents choose initial allocations:

$$\xi_T(\theta) \in \arg\max_{y \in X} \max_{x \in D(y)} u(x, \theta) - T(y) \tag{135}$$

Moreover, agents choose consumption:

$$\phi_T(\theta) = \arg \max_{x \in D(\xi_T(\theta))} u(x, \theta) \tag{136}$$

We suppose that payments between the platform and advertiser are possible and both have quasilinear preferences over money, the platform is a revenue maximizer, and as above that when an agent of type θ spends x units of time on the platform, they spend $\pi(x, \theta)$ on the good being advertised by the advertiser. Thus, total producer surplus under tariff T is given by:

$$\mathcal{S}(T) = \int_{\Theta} [\pi(\phi_T(\theta), \theta) + T(\xi_T(\theta))] \, \mathrm{d}F(\theta) \tag{137}$$

Note that this is the objective function we have considered throughout the paper.

The outside option of the advertiser is normalized to zero (they obtain no revenue in the absence of agreement) and the outside option of the principle is the total revenue under the revenue maximizing mechanism (*i.e.,* when $\pi \equiv 0$), which we denote by R. The Nash bargaining outcome (with bargaining weight ω for the advertiser) is then obtained as the solution of the following problem:

$$\begin{aligned} \max_{T} \max_{v_1, v_2} \quad & v_1^{\omega} (v_2 - R)^{1-\omega} \\ \text{s.t} \quad & v_1 + v_2 \leq \mathcal{S}(T), v_1 \geq 0, v_2 \geq R \end{aligned} \tag{138}$$

The ultimate choice of T in this program maximizes $\mathcal{S}(T)$. Thus, the mechanism selected by the Nash bargaining mechanism is the same as the optimal mechanism described in the main text. In other words, it is as if the principal treats the advertising revenues as its own.

Moreover, this conclusion would be true under *any* efficient bargaining procedure.

D.2 Platform Externalities

Our main analysis motivates, and solves for the optimal contract in, a model in which the principal gets direct external revenue from usage. But this model ignores another, complementary reason why providers prefer higher usage: they may make the platform intrinsically more valuable for other end users. Examples include matching services (e.g., dating apps like Tinder, Match.com, or OK Cupid), online games (e.g., Fortnite or World of Warcraft), and also the previously mentioned social media websites. We show how a variant model of that introduced in Section 2 with separable externalities born by the *agent* is almost identical to the model studied above.

To model this, we introduce a function $W : X \times \Theta \to \mathbb{R}_+$ which maps each agent's consumption to a positive externality. Agents' payoffs if they participate, given a consumption function ϕ, are:

$$v\left(x, \theta, (\phi(z))_{z\in[0,1]}\right) = u(x,\theta) + \int_0^1 W(\phi(z), z)\,\mathrm{d}F(z) \tag{139}$$

with the maintained assumption of a zero outside option otherwise. The rest of the model is as in Section 2. The externality of others' usage is obtained by an agent whenever they use the platform at the extensive margin. This makes the model amenable to settings where an agent may gain from participating, even if they do not regularly use the platform. One leading example among social media is the professional networking website LinkedIn. Having a LinkedIn profile generates the "passive" benefit of being findable by job recruiters, even if the user spends essentially zero time using the website.

The proof that this model is isomorphic to our general framework shows that the externality can be expressed as a payoff function for the principal through its appearance in the IR constraint. This summarizes two ideas. First, an agent of type θ contributes total surplus $W(x, \theta)$ via the externality when they consume x, provided that all types participate. Second, since the externality is *excludable* or not available to agents that do not participate in the mechanism, the principal can extract the full value of the externality as part of a "participation fee."

The formal argument is described and proven in the Lemma below:

Lemma 6. *Consider a variation of the baseline environment in which agents' preferences are given by the utility function in Equation 139. Suppose moreover that the modified virtual surplus function*

$$J^\dagger(x,\theta) = \pi(x,\theta) + u(x,\theta) + W(x,\theta) - \frac{1-F(\theta)}{f(\theta)} u_\theta(x,\theta) \tag{140}$$

is strictly quasiconcave in x and single-crossing in (x, θ). The optimal allocation is given by $\phi^*(\theta) = \min\{\phi^A(\theta), \phi^P(\theta)\}$, where $\phi^A(\theta) = \arg\max_x u(x,\theta)$ and $\phi^P(\theta) = \arg\max_x J^\dagger(x,\theta)$.

Proof. The non-standard feature is the externality. Observe first that the externality cannot affect the obedience constraint since it has no dependence on consumer choice. Observe next that, fixing an allocation ϕ, the reporting problem for type θ is

$$\max_{\theta'} u(\phi(\theta'), \theta) - t(\theta') + \int_0^1 W(\phi(z), z)\,\mathrm{d}F(z) \tag{141}$$

Because the externality is linearly separable and type-independent, it does not affect the reporting problem, so honesty is unaffected. The only relevant change is to the individual rationality constraint. The envelope theorem of Milgrom and Segal (2002) implies the same expression for transfers

$$t(\theta) = C + u(x, \theta) - \int_0^\theta u_\theta(\phi(z), z)\,\mathrm{d}F(z) \tag{142}$$

The same arguments from the proof of Lemma 2 imply that, without loss of optimality, we can restrict attention to allocations in which all agents participate (as $W \geq 0$) and such that $C = \int_0^1 W(\phi(z), z)\,\mathrm{d}F(z)$ and the objective of the monopolist is:

$$\int_0^1 \left(\pi(\phi(\theta), \theta) + W(\phi(\theta), \theta) + u(\phi(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(\phi(\theta), \theta) \right) \mathrm{d}F(\theta) \tag{143}$$

Observe that the externality is effectively an additional component of the principal's instrumental payoff. The claim of the result then follows by application of Proposition 1. □

A further implication, of course, is that appropriately amended versions of our results about pricing plans, welfare, and intermediate contractibility apply in this platform model. We omit formal versions of these results for brevity, as they all can be proved via the appropriate substitution of Lemma 6 for Proposition 1.

D.3 Irrational Addiction

As highlighted in the main text, addicted users are commonly cited as a major source of revenue in the context of apps and other digital goods (see, e.g., Allcott, Gentzkow, and Song, 2021). In this appendix, we describe a simple microfoundation of how external revenue could be derived from irrational addiction of consumers. That is, consumers' consuming more today is valuable because it increases their future willingness-to-pay and can lead to future revenue for the seller.

Suppose that agents live for two periods but are myopic. Let $x \in X$ be the agent's consumption today ($t = 0$) and $\tilde{x} \in X$ their consumption tomorrow ($t = 1$). An agent of type $\theta \in \Theta$ believes they have lifetime payoff from consumption x given by $u(x, \theta)$, where u satisfies our running assumptions. In reality, however, the agent also values consumption tomorrow. Moreover, the more (or less) that they consumed today the more (or less) they value consumption tomorrow. Thus, at $t = 1$, the agent has utility function $\tilde{u} : X^2 \times \Theta \to \mathbb{R}$, where $u(x, \tilde{x}, \theta)$ is their payoff. This *complete myopia* can be thought of as an extreme form of the inattention toward habit formation that Allcott, Gentzkow, and Song (2021) find is necessary to empirically rationalize total demand for six ubiquitous mobile apps (Facebook, Instagram, Twitter, Snapchat, web browsers, and YouTube).

Observe that given a full-revelation mechanism (or equivalently under observation of agent consumption under an implementable mechanism), the seller knows the agent's type tomorrow. Thus, when agents consume x today and their type is θ, tomorrow the monopolist sells them $\tilde{x}^*(x, \theta) \in \arg\max_{\tilde{x} \in X} \tilde{u}(x, \tilde{x}, \theta)$ and charges a transfer of:

$$\pi(x, \theta) = \tilde{u}(x, \tilde{x}^*(x, \theta), \theta) \tag{144}$$

to extract full surplus. Thus, from the perspective of today, the monopolist faces the non-linear pricing problem we study in the main text, with an external revenue function given by Equation 144, which can capture the gains from addicting a user through contemporaneous consumption and extracting this surplus from them in the future.

To interpret our conditions for the presence and shape of multi-part tariffs, it is useful to calculate marginal usage-based revenues

$$\begin{aligned} \pi_x(x, \theta) &= \tilde{u}_x(x, \tilde{x}^*(x, \theta), \theta) + \tilde{u}_{\tilde{x}}(x, \tilde{x}^*(x, \theta), \theta)\tilde{x}^*_x(x, \theta) \\ &= \tilde{u}_x(x, \tilde{x}^*(x, \theta), \theta) \end{aligned} \tag{145}$$

where the second equality applies the envelope theorem. Thus, the marginal revenue of consumption today comes from its marginal payoff consequence tomorrow. The concavity of usage-based revenues is similarly

$$\begin{aligned} \pi_{xx}(x, \theta) &= \tilde{u}_{xx}(x, \tilde{x}^*(x, \theta), \theta) + \tilde{u}_{\tilde{x}\tilde{x}}(x, \tilde{x}^*(x, \theta), \theta)(\tilde{x}^*_x(x, \theta))^2 + \tilde{u}_{\tilde{x}}(x, \tilde{x}^*(x, \theta), \theta)\tilde{x}^*_{xx}(x, \theta) \\ &= \tilde{u}_{xx}(x, \tilde{x}^*(x, \theta), \theta) + \tilde{u}_{\tilde{x}\tilde{x}}(x, \tilde{x}^*(x, \theta), \theta)(\tilde{x}^*_x(x, \theta))^2 \end{aligned} \tag{146}$$

where we again apply the envelope theorem in the second step. Concavity arises if payoffs are directly concave in previous consumption, or if payoffs are concave in contemporaneous consumption *and* contemporaneous consumption increases in previous consumption. The latter combination of assumption is standard in the theoretical literature on addiction as habit (Becker and Murphy, 1988) and empirically verified in the structural estimation of Allcott, Gentzkow, and Song (2021).

D.4 Overconfidence

Here, we study over-confidence at the participation stage, as in Grubb (2009). We show explicitly, building on results in Grubb's original analysis, how to recast agents' over-confidence about being a high-demand type as a source of revenue for the principal, which depends on agents' forecasted *consumption* rather than forecasted allocation.

A natural reason why a principal may allocate more of a good than an agent wants *ex post* is that the agent actually wanted even more *ex ante*, via different and inconsistent preferences. Such a story is at the heart of Grubb (2009)'s analysis of selling to overconfident consumers and his leading example of pricing cell phone plans, a context in which individuals regularly (based on anecdotes and empirical exploration) over-estimate their demand in advance. We now illustrate how over-confidence at the participation stage nests naturally in our framework as an example of a positive payoff externality.

This model is nested in the canonical monopoly pricing model, with zero payoff externalities and zero production costs, with one twist: agents decide whether to participate *ex ante* without knowing their type θ, but with a prior belief $\theta \sim \check{F}(\theta)$ which may differ from the objective truth $\theta \sim F(\theta)$. The common individual rationality constraint for all consumers is that the expected payoff at the allocation $(\phi(\theta), \xi(\theta), t(\theta))_{\theta\in\Theta}$ exceeds the outside option 0, or

$$\int_0^1 (u(\phi(\theta), \theta) - t(\theta))\,\mathrm{d}\check{F}(\theta) \geq 0 \tag{147}$$

As demonstrated in Grubb (2009), the above after substituting in the envelope expression for transfers (local incentive compatibility) translates to the following upper bound for the transfer for the lowest type agent:

$$t(0) \leq u(\phi(0), 0) + \int_0^1 (1 - \check{F}(\theta))u_\theta(\phi(\theta), \theta)\,\mathrm{d}\theta \tag{148}$$

Observe first that, in a classical model with correctly specified expectations $\check{F}(\theta) = F(\theta)$, (148) reveals that consumers of every type, lacking any private information, must surrender all expected surplus (information rents) to the planner *ex ante*. With mis-specified $\check{F}(\theta) \neq F(\theta)$, it means that consumers of every type pay for the surplus they *expect* to receive, which may be incorrect.

Using this expression, and the principal's desire for individual rationality to bind at equality, leads to the following simplification of the problem which has equivalent content to part 1 of Proposition 1 in Grubb (2009):

Lemma 7. *The optimal assignment in the monopoly problem studied by Grubb (2009) is* *identical to the assignment from the problem introduced in 2 with external revenue*

$$\pi(x, \theta) = \frac{1 - \check{F}(\theta)}{f(\theta)} u_\theta(x, \theta) \tag{149}$$

Proof. This follows immediately from comparing our Lemma 2 and Proposition 1 in Grubb (2009). $\square$

Intuitively, this says that the marginal cost of "producing" x for type θ is negative, in proportion to how much it increases agents' expectations of surplus (from information rents) *ex ante*. It results in an equivalent allocation to the original problem but a different transfer, off by the integral term in (148) which is now rebated to the producer as negative production cost.

The main takeaway from this analysis is that the combination of time-inconsistency and free disposal in Grubb (2009) generates multi-part tariffs for the same reason highlighted in Corollary 1. Our comparative statics results can be applied directly to this context. Our normative analysis applies given the appropriate re-interpretations of consumer surplus being based on *ex post* preferences or the objective type distribution.

E Additional Applications

E.1 Regulation With Quality Grades

In this appendix, we study the problem of a government regulating a monopolist with unknown costs, as formalized by Baron and Myerson (1982). The original analysis shows how the regulator can optimally design such a contract under the assumption that the regulator can actually monitor and/or contract upon the choice variable, which we may interpret as the "quality" of the provided service. But, in practice, an attribute like "quality" (e.g., for an infrastructure project like a bridge or airport) may not be easily described *ex ante* under the payoff-relevant, continuous measure, but instead only in coarser bins. This maps to our *grades* model of partial contractibility from Section 4.

We now use our framework to illustrate the economic consequence of limited, grade-based contractibility in the simplest possible case of Baron and Myerson (1982) with constant marginal cost and linear demand.[32] Firms can produce quality levels $x \in [0, \overline{x}]$ and customers have an inverse demand curve $P(x) = \overline{x} - x$, where $\overline{x} > 2$, which reflects decreasing *marginal*

[32] This is the simplest illustrative case under which we have the required concavity of the monopolist's revenue function and the virtual surplus function. It would be trivial to extend our results to a larger class of demand and cost functions under the appropriate assumptions.

willingness to pay for quality. Firms are differentiated by their marginal cost $1-\theta$, where $\theta \sim U[0,1]$. Firms' payoffs, as a function of their type θ and quantity x, are their profits:

$$u(x,\theta) = (\overline{x} - x)\cdot x - (1-\theta)\cdot x \tag{150}$$

The government maximizes the sum of consumer surplus, the transfers, and a fraction $\tau \in [0,1)$ of the producer surplus. It is simple to derive that consumer surplus is given by $\frac{x^2}{2}$. With some minor algebraic manipulation, the regulator's payoff from a producer of type θ producing x is then:

$$\pi(x,\theta) + t(\theta) = \frac{x^2}{2} + \tau \cdot u(x,\theta) + (1-\tau)t(\theta) \tag{151}$$

We capture the aforementioned scope for non-contractibility and lack of monitoring with the disposal correspondence D^g introduced in Section 4.1. There is a "high grade" defined by quality $x \in [c, \overline{x}]$ and a "low grade" defined by quality $x \in [0, c)$. A producer called upon to produce a good of quality y can actually provide any lower quality within the grade in which y lies, or faces the disposal correspondence

$$D^g(y) = \begin{cases} [0,y] & \text{if } y < c, \\ [c,y] & \text{if } y \geq c. \end{cases} \tag{152}$$

To describe the optimal contract, we first define the following function which, as we will soon see, defines the government's preferred level of quality under full contractibility:

$$\phi^P(\theta) = \min\left\{\overline{x} - (2-\tau)(1-\theta), \overline{x}\right\} \tag{153}$$

We next define the monopolist's profit maximizing choice absent any regulation as

$$\phi^A(\theta) = \frac{\overline{x}}{2} - \frac{1-\theta}{2} \tag{154}$$

Using these expressions, we now state the following result which describes the optimal contract:

Corollary 6. *In the optimal allocation, the contracted quality is*

$$\phi^*(\theta) = \begin{cases} \phi^P(\theta) & \textit{if } \theta < \theta_0, \\ \phi^A(\theta) & \textit{if } \theta \in [\theta_0, \theta_1), \\ \max\{c, \phi^A\} & \textit{if } \theta \geq \theta_1 \end{cases} \tag{155}$$

where

$$\begin{aligned}\theta_0 &= 1 - \frac{\overline{x}}{3 - 2\tau} \\ \theta_1 &= \theta_0 + \frac{c}{3 - 2\tau}\end{aligned} \tag{156}$$

Proof. We first derive the expressions given in the main text for ϕ^A and ϕ^P. See that firms' profits are given by

$$u(x, \theta) = (\overline{x} - x) \cdot x - (1 - \theta) \cdot x \tag{157}$$

The principal maximizes the expected value of expression (151). It is equivalent if the principal maximizes the expected value of this objective divided by $(1-\tau)$. The transformed objective has the form $\tilde{\pi}(x, \theta) + t(\theta)$ with

$$\tilde{\pi}(x, \theta) = \frac{1}{1 - \tau} \frac{x^2}{2} + \frac{\tau}{1 - \tau} \cdot u(x, \theta) \tag{158}$$

The associated virtual surplus is

$$\begin{aligned}J(x, \theta) &= \frac{1}{1 - \tau} \frac{x^2}{2} + \frac{\tau}{1 - \tau} \cdot u(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta) \\ &= \frac{1}{1 - \tau} \left(\overline{x} \cdot x - \frac{x^2}{2} - (2 - \tau)(1 - \theta) x \right)\end{aligned} \tag{159}$$

where the second line uses the definition of $u(x, \theta)$, the hazard rate expression $(1-F(\theta))/f(\theta) = 1 - \theta$ for the uniform distribution, and the substitution $u_\theta(x, \theta) = x$. This virtual surplus function is strictly concave in x, strictly single-crossing in (x, θ) and maximized at

$$\phi^P(\theta) = \overline{x} - (2 - \tau)(1 - \theta) \tag{160}$$

We now derive consumption in the optimal contract. Combining Lemma 2 with the characterization of grades contractibility via its minimal selection

$$\delta^g(y) = \begin{cases} 0 & \text{if } y < c \\ c & \text{if } y \geq c \end{cases} \tag{161}$$

we can re-write the point-wise program

$$\begin{aligned}\phi^*_{pw}(\theta) \in \arg\max_x \; & J(x, \theta) \\ \text{s.t.} \; & \begin{cases} x \leq \phi^A(\theta) & \text{if } x < c \\ x \leq \max\{\phi^A(\theta), c\} & \text{if } x \geq c \end{cases}\end{aligned} \tag{162}$$

Moreover, if ϕ^*_{pw} is monotone increasing, then $\phi^* = \phi^*_{pw}$.

Since $J(\cdot, \theta)$ is quasiconcave for any θ, we can show that ϕ^*_{pw} has the following form:

$$\phi^*_{pw}(\theta) = \begin{cases} \phi^P(\theta) & \text{if } \phi^P(\theta) \leq \phi^A(\theta) \\ \phi^A(\theta) & \text{if } \phi^P(\theta) > \phi^A(\theta), J(\phi^A(\theta), \theta) \geq J(c, \theta) \\ c & \text{if } \phi^P(\theta) > \phi^A(\theta), J(\phi^A(\theta), \theta) < J(c, \theta) \end{cases} \tag{163}$$

Moreover, since $J(\cdot, \theta)$ is quadratic for any θ, strictly concave, and maximized at $\phi^P(\theta)$, we observe that $J(\phi^A(\theta), \theta) \geq J(c, \theta)$ if and only if $|\phi^A(\theta) - \phi^P(\theta)| \leq |c - \phi^P(\theta)|$. Using this observation, and direct calculation, we can derive Equation 155 as a proposed point-wise solution to the problem, and then immediately verify that it is monotone. This completes the proof. □

The contract of Corollary 6 can take on a number of forms depending on the value of primitives. One particularly interesting case is obtained when $\phi^P > \phi^A$ for all $\theta \in \Theta$, or the principal would strictly prefer higher quality than the market would provide in autarky regardless of the firm's costs. In this case, it is straightforward to note that $\theta_0 < 0$ and hence the optimal allocation takes the form

$$\phi^*(\theta) = \begin{cases} \phi^A(\theta) & \text{if } \theta < \theta_1 \\ \max\{c, \phi^A\} & \text{if } \theta \geq \theta_1 \end{cases} \tag{164}$$

By inspection, given that all firms are producing either at their bliss points or at the cut-off c, it is evident that the government cannot subsidize production of quality on the margin. They can, however, provide a lump sum payment to firms that produce quality in excess of the grade threshold c. We can in fact calculate explicitly the tariff as

$$T(x) = \begin{cases} u(\phi^A(0), 0) & \text{if } x < \phi^A(\theta_1) \\ u(c, \theta_1) - u(\phi^A(\theta_1), \theta_1) + u(\phi^A(0), 0) & \text{if } x \geq c \end{cases} \tag{165}$$

For low-type (high-cost) firms, the government tolerates monopoly and charges an entry fee that extracts fully the willingness to pay of the highest-cost firm. For high-type (low-cost) firms, the government pays a subsidy $u(\phi^A(\theta_1), \theta_1) - u(c, \theta_1)$ that compensates the marginal high-grade firm for producing at c. In practice, this contract looks like a two-tiered lump-sum payment to high-cost and low-cost firms.

E.2 Optimal Income Taxation with Over-reporting

Here, we study optimal income taxation envisioned as a screening problem in the tradition of Mirrlees (1971), and augmented with externalities. A common claim by politicians and lobbying groups is that high income taxes stifle innovation and job creation. In more formal economic language, productive effort may have positive productivity spillovers, as in many leading models of economic growth, or demand externalities, as in standard Keynesian macroeconomic models.[33] We introduce a stylized model of taxation of such a scenario to study the implications for taxation with and without the ability to monitor agents *over-reporting* their income.

Agents have a payoff over consumption c and human capital investment $x \in [0, \overline{x}]$ that is quasilinear in consumption and quadratic in human capital investment $U(c, x, \theta) = c - (1-\theta)\frac{x^2}{2}$. A higher type θ corresponds with lower costs to obtaining human capital. Types are distributed $\theta \sim U[0,1]$. Each agent has a budget constraint $c \leq y - t$ which bounds consumption by income y net of the tax t. Income depends on both one's own human capital and an average human capital externality, or $y = x + \int e(\phi(z))\,\mathrm{d}F(z)$ defined up to some increasing, differentiable, and concave $e : \mathbb{R}_+ \to \mathbb{R}_+$. Using the constraint and definition of income, the payoff function for agents of type θ when they have human capital x and and all other agents have human capital given by ϕ is given by:

$$u\left(x, \theta; (\phi(z))_{z\in\Theta}\right) = x - (1-\theta)\frac{x^2}{2} + \int e(\phi(z))\,\mathrm{d}F(z) \tag{166}$$

The government's payoff is the sum of government revenues plus utilitarian welfare multiplied by some $\tau \in [0, 1)$. This objective can be written as the following:

$$\pi\left(x, \theta; (\phi(z))_{z\in\Theta}\right) + t(\theta) = \tau u\left(x, \theta; (\phi(z))_{z\in\Theta}\right) + (1-\tau)t(\theta) \tag{167}$$

As highlighted above, the existence of externalities might prompt the government to subsidize agents' incomes via a negative marginal tax. However, the government may not have the ability to observe an agent over-reporting their income to take advantage of such a policy. Formally, this corresponds to an agent with allocated human capital investment of $\xi(\theta)$ actually making a human capital investment of $x \in D(\xi(\theta))$. The optimal mechanisms, and some salient properties thereof, are summarized below:

Corollary 7 (Taxation and Monitoring)**.** *The optimal mechanism with full monitoring $D =$*

[33] This basic idea animates a recent literature on optimal income taxation with externalities (see, e.g., Rothschild and Scheuer, 2016; Lockwood, Nathanson, and Weyl, 2017; Badel, Huggett, and Luo, 2020).

$\underline{D}$ *allocates*

$$\phi^P(\theta) \in \arg\max_{x \in [0,\overline{x}]} \left\{ x - (2-\tau)(1-\theta)\frac{x^2}{2} + \tau e(x) \right\} \tag{168}$$

for all $\theta \in [0,1]$. *The optimal mechanism without monitoring* $D = \overline{D}$ *allocates*

$$\phi^*(\theta) = \begin{cases} \phi^P(\theta) & \text{if } \phi^P(\theta) < (1-\theta)^{-1} \\ \min\{\overline{x}, (1-\theta)^{-1}\} & \text{if } \phi^P(\theta) \geq (1-\theta)^{-1} \end{cases} \tag{169}$$

to all agents $\theta \in [0,1]$. *Moreover, the following properties hold:*

1. *Both with and without monitoring, worker welfare surplus increases in* τ*;*
2. *Workers benefit from from full monitoring relative to the free-disposal benchmark.*

Proof. First, see that the consumer utility function

$$u(x,\theta; (\phi(z))_{z\in\Theta}) = x - (1-\theta)\frac{x^2}{2} + \int e(\phi(z))\,\mathrm{d}F(z) \tag{170}$$

satisfies, for every fixed value of the externality term, single-crossing with respect to (x,θ), is monotone increasing in θ, and is strictly quasiconcave in x. Thus it satisfies the conditions for our implementation result in Proposition 5.

Next, we apply the same technique from the proof of Corollary 6 to re-scale the principal's objective to put weight 1 on transfers. In particular, we define

$$\tilde{\pi}(x,\theta; (\phi(z))_{z\in\Theta}) = \frac{\tau}{1-\tau}\left(x - (1-\theta)\frac{x^2}{2} + \int e(\phi(z))\,\mathrm{d}F(z) \right) \tag{171}$$

Since $1/(1-\tau) > 0$, maximizing the expected value of (167) is the same as maximizing the expected value of $\tilde{\pi}(x,\theta) + t(\theta)$. Moreover, the principal's total payoff simplifies to

$$\int_\Theta J(\phi(\theta),\theta)\,\mathrm{d}F(\theta) = \frac{1}{1-\tau}\int_\Theta \left(\phi(\theta) + \tau e(\phi(\theta)) - (2-\tau)(1-\theta)\frac{\phi(\theta)^2}{2} \right) \mathrm{d}F(\theta) \tag{172}$$

See that the new form of the objective, up to scale, implies an outcome-equivalent model in which the planner's objective function is $\hat{\pi}(x,\theta) = \tau e(x)$; the agent's utility function is $\hat{u}(x,\theta) = x - (1-\theta)\frac{x^2}{2}$; and the outside option is 0. We can use Lemma 6 to derive the optimal contract with and without free disposal, with the caveat that type-specific utility $\hat{V}(\theta)$ in this model corresponds with the actual model's utility minus the externality, or

$$V(\theta) = \hat{V}(\theta) + \int e(\phi(z))\,\mathrm{d}F(z) \tag{173}$$

We now prove the comparative statics results. Observe first that increasing τ weakly increases $\phi^P(\theta)$. Since the program defining $\phi^P(\theta)$ is continuous, twice-differentiable, and globally concave, it is sufficient to verify the following:

$$\frac{\partial^2}{\partial x \partial \tau}\left(x-(2-\tau)(1-\theta)\frac{x^2}{2}+\tau e(x)\right)=(1-\theta)\frac{\mathrm{d}e(x)}{\mathrm{d}x}\geq 0 \tag{174}$$

where the inequality follows as $e(\cdot)$ is an increasing function. Welfare for type θ is given in the optimal mechanism by

$$V(\theta)=\int_0^\theta u_\theta(\phi^*(s),s)\,\mathrm{d}s+\int_0^1 e(\phi^*(s))\,\mathrm{d}s=\int_0^\theta\frac{(\phi^*(s))^2}{2}\mathrm{d}s+\int_0^1 e(\phi(s))\,\mathrm{d}s \tag{175}$$

Both integrands weakly increases point-wise, so $V(\theta)$ also increases.

For the second comparative statics result, we first apply Proposition 3 to show that $\hat{\overline{V}}(\theta;a^*)\leq\hat{\underline{V}}(\theta;a^{*'})$, where $a^*,a^{*\prime}$ are respectively the optimal allocations under free disposal and no disposal. See next that

$$\overline{V}(\theta;a^*)-\underline{V}(\theta;a^{*'})=\left(\hat{\overline{V}}(\theta;a^*)-\hat{\underline{V}}(\theta;a^{*'})\right)+\left(\int\left(e(\phi^*(s))-e(\phi^{*\prime}(s))\right)\mathrm{d}s\right) \tag{176}$$

or the increase in welfare can be decomposed into the component characterized above and the increase in the externality between the two optimal mechanisms. But $\phi^{*\prime}(\theta)\geq\phi^*(\theta)$ is established in the proof of Proposition 3; this implies that $e(\phi^*(s))\leq e(\phi^{*\prime}(s))$. Putting the two ingredients together implies $\overline{V}(\theta;a^*)-\underline{V}(\theta;a^{*'})\leq 0$, which proves the claim.

□

The government wants to boost human capital investment beyond the bliss point when the externality is sufficiently high. With a concave external benefits function, marginal externalities are highest for the lowest type workers. The policymaker, without monitoring, implements a zero marginal tax for these workers; with monitoring, they implement a negative marginal tax.

The negative marginal tax for low incomes resembles the Earned Income Tax Credit (EITC) in the United States. Our model micro-founds why such negative marginal taxes may exist; why monitoring lower incomes is essential for supporting this system; and why, despite the implication of standard Mirleesian models that the only relevant deviations of agents are to *under*-report income, the US Tax Code uniformly treats under- or over-reporting income as equivalent forms of tax fraud.[34] Our welfare results moreover show that, through the

[34]See Title 26, Section 7206, of the US Tax Code ("Fraud and false statements") available at: https://www.law.cornell.edu/uscode/text/26/7206.

lens of the current model, the monitoring underlying successful implementation of the EITC benefits *all* agents in the economy. Every dollar given as subsidy to an EITC recipient is quite literally a dollar saved by agents earning any higher income in our model, because higher retention (lower taxes) is required to incentivize truthful reporting. Thus our model implies very broad benefits, across incomes, of having the legal capacity to implement negative marginal tax rates for low incomes.