Market Power Screens Willingness-to-Pay*

E. Glen Weyl[†] Jean Tirole[‡]

June 10, 2011

Abstract

What is the best way to reward innovation? While prizes avoid deadweight loss, intellectual property screens out projects generating low consumer surplus per unit sold. We propose a stretch parameterization of demand under which innovations differ in both the size of the market they create and consumers' average willingness-to-pay for them. We solve the resulting multidimensional screening problem by decomposing the analysis into a separate choice of the level and structure of rewards for innovations. Optimal policy generally calls for some market power but never full monopoly pricing. The appropriate degree of market power is determined by a value-weighted average of the innovation supply elasticity multiplied by the log-variance of the ratio of the monopoly prices to quantities, opening our analysis to empirical calibration. Our results also shed light on the pricing of platforms, incentives within firms for product development and public infrastructure.

*Weyl acknowledges the support of the Milton Fund and Nokia. Tirole acknowledges the support of Microsoft and Nokia to the Institute for Industrial Economics (IDEI). The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 249429. We are grateful to many people for the helpful comments they provided on this paper, but we especially thank Roland Benabou, Joshua Gans, Scott Duke Kominers, José Scheinkman, Josh Schwartzstein, Yossi Spiegel, Scott Stern, Michael Whinston and Pai-Ling Yin. Most of all, we appreciate the excellent research assistance of Stephanie Lo, Vladimir Mukharlyamov, David Smalling, William Weingarten and especially Andre Veiga and Rui Wang. All errors are our own.

[†]Harvard Society of Fellows and TSE: 78 Mount Auburn Street, Cambridge, MA 02138, USA: weyl@fas.harvard.edu. The order of authors has been chosen to conform to the scientific convention of priority of contribution rather than the social convention of alphabetic precedence.

[‡]TSE, 21, allée de Brienne, Toulouse 31000, France: jean.tirole@tse-fr.eu.

For if the legislature should appoint a pecuniary reward for the inventors of new machines, etc., they would hardly ever be so precisely proportional to the merit of the invention as this is. For here, if the invention be good and such as is profitable to mankind, he will probably make a fortune by it; but if it be of no value he also will reap no benefit.

- Adam Smith, *Lectures on Jurisprudence*[1]

Nothing could be more absurd. Whether it was wise for the government to subsidize...Union Pacific Railroad...is an interesting historical question...but it would be better...to leave it unsolved than to ruin the country...by charging enormous freight rates and claiming that their sum constitutes a measure of the value...of the investment.

–Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and Of Railway and Utility Rates"

Can innovation be rewarded without the distortions created by intellectual property?[2] As reflected in Hotelling's quote above, at least since Marshall (1890) textbook economics has advocated replacing IP with a system of prizes and, more broadly, replacing monopoly by marginal cost pricing. Yet an even older tradition (Smith, 1762) has argued that without market-driven rewards, it is difficult to determine which projects merit the necessary development costs. In turn, Kremer (2000a,b) points out that both the ex-post pricing distortion and the waste of resources on useless innovations may simultaneously be avoided by basing prizes on consumption at the efficient price.[3] Yet this measurement of *market size* ignores a second dimension of heterogeneity among innovations, also necessary to determine their social value: *consumers' willingness to pay* for them. Our paper develops a model for trading off the

[1]We thank Joel Mokyr for this reference.

[2]Intellectual property, or IP as we will refer to it from here on, may take many forms: trade secrets, copyrights, patents, etc. We follow much recent literature in seeing the broad institution of market power as a reward for innovation (regardless of the exact form it takes) as separate from the specific institution of patents; our focus is on whether market power is appropriate, however implemented.

[3]Interest in centralized systems for stimulating innovation, including Kremer's "advance market commitments" (Barder et al., 2005), has risen in recent years with the resurgence of industrial policy (Economist, 2010). Note that quantity-dependent prizes, like our mechanisms, clearly require a standardized-within-market unit of quantity, such as individuals treated or smartphones having installed the application.

screening benefits of market power against the traditional distortion of consumption in this environment of multidimensional heterogeneity.

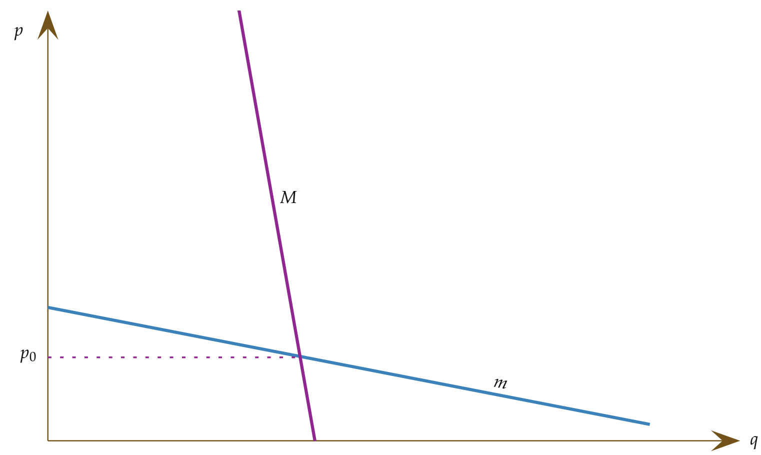

Figure 1: Distinguishing valuable from low-surplus projects

Figure 1 illustrates the basic idea. Consider two equally costly innovations: m ("me-too") and M ("Major"). Because willingnesses to pay are small, m creates little social value despite its large market size and does not justify the fixed cost of bringing it about. An example might be Netscape Navigator during the 1990's which, while widely adopted, sold at a low price because it offered little added value over its rivals. In contrast, M, such as America OnLine (AOL), brought substantial value to each consumer (and charged a correspondingly high price), though it had a smaller customer base.

A social planner who does not directly observe consumer surplus needs to separate m from M by using the property that the demand for a higher-quality product is (everywhere) less sensitive to price. This requires charging a price for M of at least p_0, the minimum price at which the demand for M actually exceeds that for m. At lower prices, m looks superior to M, being more widely adopted. Fundamentally, a prize system – a payment to the entrepreneur depending only on demand at the marginal cost (assumed to lie below p_0) – is unable to screen out m and screen in M.[4] This sorting role of market power arises not only in the choice

[4]Despite the extensive theoretical and policy interest in, and numerous successful historical examples (Kremer, 1998) of such a system, many consider it simply impractical. Yet it is hard to see what, other than informational asymmetries, could be the source of such "impracticality". Perhaps there is simply a fixed cost of government involvement; but then this could not explain the patent system (which clearly involves the government) or rationalize market power for large innovations as these would pass the fixed-cost threshold. Furthermore, market power is used in many situations, such as infrastructure procurement, where government involvement is inevitably highly intrusive;

between IP and prizes. As we discuss in Section 5, leading examples include bundling v. free pricing in online application and media markets, incentives for product development within firms and cost recovery rules in public infrastructure procurement.

Previous literature on the design of institutions for rewarding innovation has either ruled out prizes or considered environments with a single dimension of heterogeneity in which quantity-dependent prizes achieve optimal sorting. The first strand of literature assumes either explicitly that no transfers can be made from the planner to the entrepreneur (Gilbert and Shapiro, 1990; Klemperer, 1990; Green and Scotchmer, 1995; O'Donoghue et al., 1998; Scotchmer, 1999; Hopenhayn and Mitchell, 2001), ruling out prizes and R&D or output subsidies, or that private information is so severe that prizes can only drain the planner's resources (Cornelli and Schankerman, 1999; Gallini and Scotchmer, 2002; Hopenhayn et al., 2006; Chari et al., 2009). Thus this literature has focused on the allocation of market power across time or product space, rather than on its desirability per se. The second strand of literature allows for prizes, but only a single dimension of heterogeneity, in which case quantity-dependent prizes perfectly screen. This work thus either considers the comparison between quantity-independent prizes and IP (Shavell and van Ypersele, 2001) or concludes that quantity-dependent prizes achieve the first-best (Kremer, 2000a,b). By allowing *multidimensional heterogeneity*, our framework creates a *smooth trade-off* between screening and ex-post pricing distortion that determines which of a *range of institutions running between IP and quantity-dependent prizes* is optimal.

Our model, developed in Section 1, features an entrepreneur who is privately informed about both the size of the market for his innovation, σ, and consumers' willingness to pay for it, its quality m. Namely, we study a *stretch parametrization* of general demand functions Q,

$$q = \sigma Q\left(\frac{p}{m}\right).$$

The key assumption embodied in this parameterization is the proportionality of average consumer surplus and the monopoly price, m. A social planner only has a prior over (σ, m) and seeks to maximize total social welfare.[5] We illustrate the logic above in Subsection 1.2 with

see also our application to platform markets in Subsection 5.1.

[5]Section 4 extensively discusses our modeling approach, especially the structure of information, instruments and demand, and demonstrates the robustness of our conclusions to including a redistributive motive, externalities (such as competing or complementary/sequential innovations), moral hazard, residual uncertainty of the entrepreneur

an example showing that the tighter are the planner's priors, the more relatively attractive are quantity-dependent priors compared to IP.

In practice, however, we observe institutions (such as output subsidies, price controls and competition spurred by weak IP) that help limit prices without achieving full ex-post efficiency. We thus grant our social planner access to a range of policies *between* quantity-dependent prizes and IP. In particular, rewards t under a quantity-dependent prize system are a function of q, so we can write them as an increasing function $t\left(q^1 p^0\right)$, while under IP they are a function of profits $\pi = qp$, so we can write them as $t\left(q^{\frac{1}{2}} p^{\frac{1}{2}}\right)$. This suggests a natural intermediate class of policies with Cobb-Douglas *isoreward curves*, $t\left(q^{1-\alpha} p^{\alpha}\right)$. Raising α increases the reward given for a higher price and thereby the incentive for higher pricing. However, it also targets rewards to entrepreneurs with a high ratio of quality to market size ($x \equiv \frac{m}{\sigma}$). We show that this screening effect is beneficial so long as $\alpha < \frac{1}{2}$ because the stretch parameterization obeys Smith's hypothesis that the monopoly profit is "a pecuniary reward...precisely proportional to the merit of the invention."

Section 3 then quantifies this trade-off between ex-post distortion and screening. Raising α selects the best marginal innovations, which is important to the extent that these differ greatly from the worst innovations. However, it also increases deadweight loss on all innovations brought to market, not just those on the margin between being created or not. Thus the relative importance of screening also depends on the relative number of marginal and inframarginal innovations. Pure IP ($\alpha = \frac{1}{2}$) is never optimal as the benefits of sorting innovations dwindle as we reach perfect sorting, and pure quantity-dependent prizes ($\alpha = 0$) are (essentially) never optimal as marginal deadweight loss is 0 at the ex-post-efficient price.[6]

However, by the above logic, near-monopoly prices are optimal whenever the following quantity is large:

$$V_1 \equiv E_{\underbrace{\pi^2}_{\text{Value}}}\Big[\underbrace{\eta}_{\substack{\text{elasticity of}\\ \text{innovation supply}}} \cdot \underbrace{\operatorname{Var}\left(\log(x)|\,\mathbf{I},\pi\right)}_{\substack{\text{inequality of}\\ \text{innovation supply}}}\Big]$$

where π is the profit the innovation makes, $\mathbf{I}$ is any information available to the social planner ex ante and η is the elasticity of innovation supply with respect to rewards. In particular,

about demand conditions and the possibility of manipulation of sales by the entrepreneur.

[6] $\alpha > \frac{1}{2}$ is always worse than pure IP as it both creates greater deadweight loss and provides inferior screening as we show in Subsection 2.2.

when V_1 is large, we show that $\alpha^\star \approx \frac{1}{2} - \frac{\omega_1}{V_1}$, where ω_1 represents properties of the demand form assumed.[7] V_1 reflects the three key considerations described above:

1. The social *value* created by an innovation is proportional to the monopoly profit it could earn. Thus, weighting by the (square of) profit weighs the measure by value.[8]

2. The *elasticity of innovation supply* measures the relative weight of the marginal innovations subject to screening compared to the inframarginal innovations subject to deadweight loss.

3. The variance of the logarithm is a standard measure of tail uncertainty, and thus the *inequality of innovation supply* determines the importance of screening in high quality and out low quality innovations.[9]

When prices are close to the monopoly optimum, society may approximately observe the monopoly profit a firm *could* make. Thus $\mathrm{Var}\left(\log(x)|\,\mathbf{I},\pi\right)$, which is conditioned both on profits and any information available ex-ante to the social planner, represents the residual tail dispersion over the merits of innovations. Note that V_1 is an empirical quantity, consisting of an elasticity and a fairly easily-measurable variance. Thus, if V_1 is large, optimal policy may be approximated by relatively straightforward empirical measurements, as we illustrate in Section 6 through Weingarten (2011)'s calibration of our model for the iPhone App store. Conversely when V_1 is small, there is at least a strong case for moving somewhat away from common practices of market power as the primary means of rewarding innovation. Indeed, our techniques can also be applied in an industry where quantity-dependent prizes are initially employed.

Section 7 concludes the paper with a discussion of directions for future research. Detailed technical explanations, more general results and longer, less instructive proofs are collected into appendices following the main text. A more elaborate and technical extension of the results

[7] Note that because the logarithm and the elasticity are both unit-free, so is V_1.

[8] The square is the relevant weighting, as we use the elasticity rather than the semi-elasticity and a one percent increase in innovation incentives impacts larger innovations more strongly.

[9] In fact, the variance of the logarithm is one of the most commonly used measures of inequality (Creedy, 1977; Foster and Ok, 1999). It is even more intuitive in our setting because its well-known drawback of being hard to ground in utility theory (Dasgupta et al., 1973) is irrelevant, while its primary benefit of emphasizing the degree of extremely low and high outcomes (Sen, 1973) is central to capturing the elitist feature of innovation. We thank Nicolas Pistolesi for these references.

to isoreward curves outside the Cobb-Douglas class, along with a treatment of our problem using the revelation principle, appears in an online appendix at http://www.glenweyl.com.

1 Set-up

We first develop the basic model and provide a simple example showing when a market-power-based system can dominate an ex-post-efficient system.

1.1 Model

We assume no marginal cost of production, as ideas are typically free to distribute.[10] Potential innovations, each independent in production and consumption, are sponsored by entrepreneurs and characterized by three numbers:[11]

- c, or *cost*, an ex-ante cost of creating the innovation;
- σ, or *size of the market*, the demand at $p = 0$;
- and m, the monopoly price for the good, which we call *quality* for reasons that will become apparent below.

The entrepreneur's utility is $t - c$ if she innovates and 0 otherwise, where t is the reward the she receives. $\theta = (\sigma, m, c)$ is her private information. The social planner knows only that θ is distributed according to some smooth pdf f with full support $\mathbb{R}^3_{++}$ and all moments finite. He observes the price charged by the entrepreneur and the quantity she sells at this price and therefore announces a *reward* schedule $t(q, p)$ based on these. We restrict attention to t functions which weakly increase in each of p and q, to avoid incentives for disposing of quantity.

Demand for the innovation is characterized by a general function Q obeying standard assumptions discussed below. Thus, the quantity sold

$$q = \sigma Q\left(\frac{p}{m}\right)$$

[10]This is equivalent to *known* costs and an adjusted demand.
[11]See Subsections 4.2 and 4.5 for relaxations of independence in production and consumption respectively.

and has elasticity $\epsilon(a) \equiv -aQ'(a)/Q(a)$, where $a \equiv \frac{p}{m}$. We normalize $Q(0) = 1$, so ex-post-efficient demand is σ (thereby requiring finite demand at $p = 0$), and $\epsilon(1) = 1$ so that the monopoly optimal price is m. Q is assumed smooth, strictly decreasing wherever it is strictly positive, to have strictly declining marginal revenue and bounded $\frac{\epsilon''}{\epsilon}$, and to have vanishing marginal distortion at the ex-post-efficient price ($\lim_{a\to 0} aQ'(a) = 0$). A simple example is linear demand $q = \sigma(2m - p)/2m$; σ corresponds to the quantity-intercept of linear demand and m to half of the price-intercept as shown in Figure 2.

σ represents a horizontal stretching of inverse demand while m is a vertical stretching. We thus refer to this as the *stretch parametrization*. The crucial assumption inherent to the stretch parametrization is the proportionality between average social surplus and monopoly price when the same fraction of the monopoly price is charged for all innovations; the robustness of our results to relaxing this assumption is discussed in Subsection 4.1. In particular, under this parametrization, the social surplus created by an innovation is

$$\underbrace{p\sigma Q\left(\frac{p}{m}\right)}_{\text{profit}} + \underbrace{\sigma \int_p^\infty Q\left(\frac{\widetilde{p}}{m}\right) d\widetilde{p}}_{\text{net consumer surplus}} = \sigma m \left(aQ(a) + \int_a^\infty Q(\tilde{a})d\widetilde{a}\right) \equiv \sigma m S\left(a\right)$$

Thus if a is constant across types, so is the ratio of social surplus to profit $S(a)/aQ(a)$.

An entrepreneur (σ, m) chooses p so as to maximize her reward $t\left(\sigma Q\left(\frac{p}{m}\right), p\right)$. Equivalently chooses her optimal fraction of the monopoly price $a^\star\left(\sigma, m; t(\cdot,\cdot)\right)$ and creates the innovation if c is smaller than the resulting reward. The social planner seeks to maximize the *total social welfare* created by innovation:

$$\int_{\theta: c \leq t(\sigma Q(a^\star(\sigma,m;t(\cdot,\cdot))), ma^\star(\sigma,m;t(\cdot,\cdot)))} \left[\sigma m S\left(a^\star\left(\sigma, m; t(\cdot,\cdot)\right)\right) - c\right] f(\theta) d\theta. \tag{1}$$

1.2 An illustrative example

To fix ideas, let us compare two specific institutions, the prize and the IP systems, in the context of linear demand. Under the *prize system* (ex-post-efficient price and reward based only on demand at this price, $q = \sigma$), the expected welfare created by realized innovations

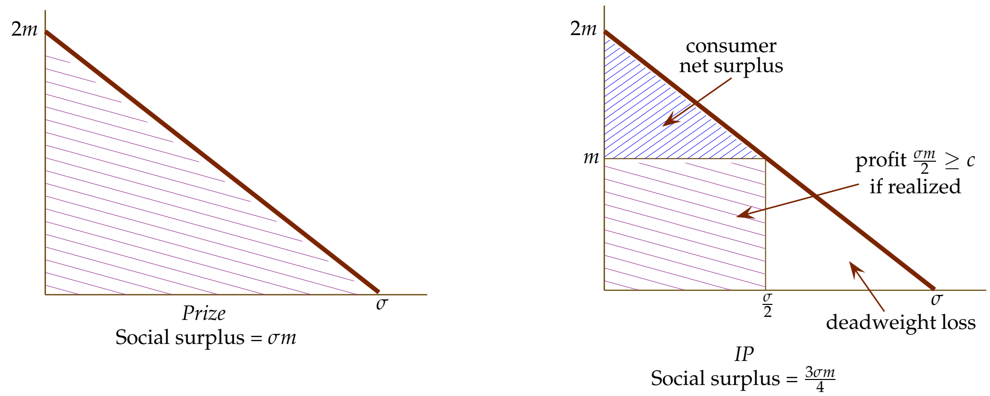

Figure 2: Linear demand under the stretch parametrization (left) and the division of potential gains from trade among deadweight loss, profits and consumer surplus at monopoly prices (right)

characterized by (c, σ) is

$$W_{\text{prize}} = \sigma E(m|\sigma, c) - c.$$

While the prize system realizes all potential gains from innovations that are created, it does nothing to screen out low m innovations. In fact, if $\sigma E(m|\sigma, c) < c$ is satisfied (for all σ and c), the average innovation that is created for any prize is not worth its cost. It is thus optimal to award no prizes at all, shutting down the market for innovations entirely even if many worthy innovations exist.[12]

Under the *IP system* (each innovation charges the monopoly price m and earns the monopoly profits $\frac{1}{2}\sigma m$), the innovation occurs if and only if $c < \frac{1}{2}\sigma m$. The social welfare created by the innovation is (only) $\frac{3}{4}\sigma m$ because of the deadweight loss associated with elevated prices as illustrated in the second panel of Figure 2. Thus, while the IP system destroys a quarter of the value created by each innovation, it robustly selects only innovations which are socially beneficial.[13]

To see the role of inequality of innovation supply in this tradeoff, consider the expected

[12]For example, if $m \sim \left[0, \frac{5}{2}\frac{c}{\sigma}\right]$ according to the decreasing triangular probability density function $f(m|\sigma, c) = \frac{4\sigma}{5c} - \frac{8\sigma^2}{25c^2}m$, despite all innovations with $m > \frac{c}{\sigma}$ being worthy, $\sigma E(m|\sigma, c) = \frac{5}{6}c < c$. Thus, it is optimal to shut down the market for innovations if one is constrained to prizes.

[13]In our "lemons" example from the previous footnote, the 4% of innovations with $m > 2\frac{c}{\sigma}$ are created, and all these are worth creating so the IP system is in this case superior.

welfare under the IP system when conditioning on (σ, c):

$$W_{\text{IP}} = \text{Prob}\left(m \geq \frac{2c}{\sigma}\right)\left[\sigma E\left(\frac{3m}{4}\middle|\sigma, c, m \geq \frac{2c}{\sigma}\right) - c\right]$$

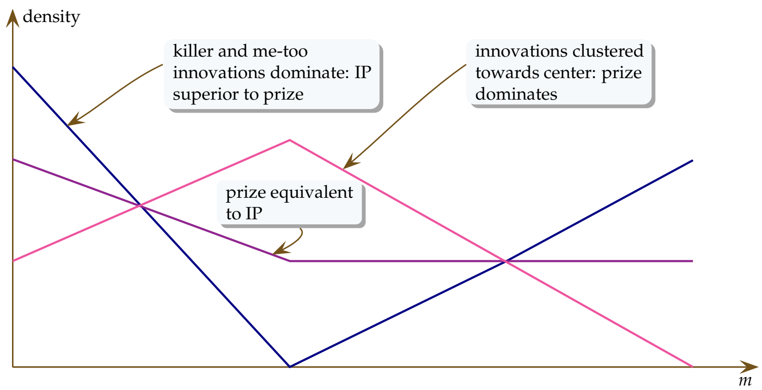

Figure 3: Bitriangular distributions with varying degrees of inequality of innovation supply

Because IP selects only innovations with relatively high m, the more important it is to select out these best innovations, the more valuable is the IP system relative to prizes. A simple example arises if m is distributed on $\left[0, 3\frac{c}{\sigma}\right]$ according to a bi-triangular pdf as shown in Figure 3.[14] When the distribution is peaked and thus innovations clustered towards the center, prizes are preferable. But when it reaches a sharp trough and thus the quality of innovations is highly unequal, IP performs much better.

However, in practice we observe a range of institutions, including specific subsidies on output, R&D subsidies, price controls, government bargaining over price (of medicines), limits on patent breadth or enforcement to encourage some competition and others. These restrain prices below the monopoly optimum without achieving full ex-post efficiency. As we will show, neither prizes or IP are optimal once we allow for a broader range of institutions, as an intermediate system outperforms both.

[14]In particular, if we let $\nu \equiv \frac{\sigma m}{c}$ and $\overline{f}$ represent the height of the peak/trough of the bi-triangular distribution

$$f(\nu|c,\sigma) = \begin{cases} \left(\frac{5}{3}\overline{f} + \frac{25}{36}\right)\nu + \frac{5}{6} - \overline{f} & \nu < \frac{6}{5} \\ \left(\frac{25}{81} - \frac{10}{9}\overline{f}\right)\nu - \frac{10}{27} + \frac{7}{3}\overline{f} & \nu \geq \frac{6}{5} \end{cases}$$

Then welfare from prizes is $.4 - .3\overline{f}$ while from IP is $\approx .47 - .57\overline{f}$.

2 The Isoreward Approach

In analyzing policies beyond the IP and quantity-dependent prizes discussed in the previous section, it is useful to consider a simple method decomposing policies into their *structure* and *level* of rewards. Each policy $t(q,p)$ is characterized by the *isoreward curves* along which it assigns constant rewards, and the actual level of rewards it assigns to each of these curves. Because entrepreneurs select their price to maximize reward along their demand curve, only the isoreward curves, and not the level of reward assigned to them, affect their pricing incentives. The level of the reward affects only their decision to enter. We refer to this decomposition as the *isoreward approach.*

Consider quantity-dependent prizes. These have rewards that depend only on quantity and thus the reward is $t(q)$.[15] By contrast, under classical IP, rewards depend on the product of price and quantity, qp. More generally any reward schedule $t(qp)$ has the same isoreward curves, leading the entrepreneur to maximize qp by charging the monopoly price.

These generalized IP rewards can be written as $t\left(q^{\frac{1}{2}}p^{\frac{1}{2}}\right)$ rather than $t(qp)$. Similarly, it is equivalent to write quantity-dependent prizes as $t\left(q^{1}p^{0}\right)$. Thus these two policies, one inducing monopoly pricing, the other inducing ex-post efficiency, have in common that their isoreward curves have a Cobb-Douglas form. A natural way to map between these extremes, therefore, is to consider the broader class of all policies with Cobb-Douglas isoreward curves, $t\left(q^{1-\alpha}p^{\alpha}\right)$, where the social planner's choice variable α determines the relative weight on price versus quantity.

As we show in our online appendix, the isoreward approach can be used to analyze a range of reward policies in which isoreward curves may be arbitrarily flexible smooth curves and thus far broader than the Cobb-Douglas class. This follows from the fact that *any* policy may be decomposed into its isoreward curves and the value assigned to these. In fact our online appendix establishes natural, if difficult to state, analogs of our results below that hold even allowing for this richer range of policies. However, we find it useful to restrict attention to Cobb-Douglas isoreward policies as they provide a parsimonious single-parameter class, allowing empirical calibration while still providing a continuous map between the extremes of

[15] We use t both as a function of p and q separately and as a function of the unidimensional isoreward curve.

generalized IP and quantity-dependent prizes.

2.1 Pricing

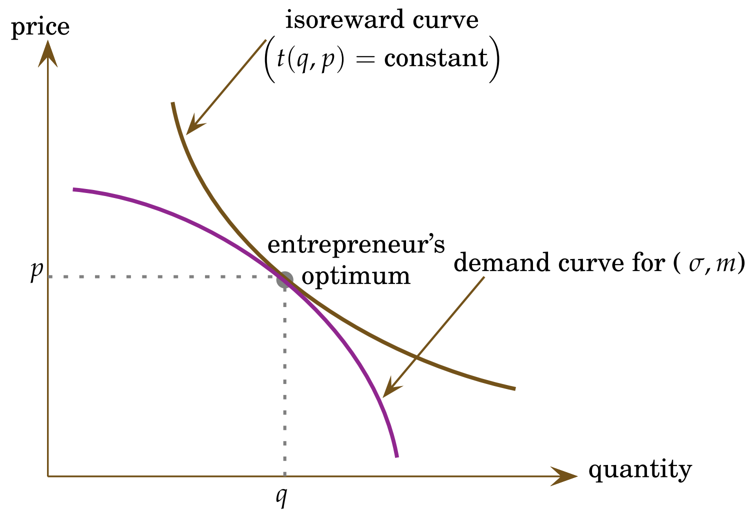

Figure 4: Isoreward curves must be tangent to demand at the entrepreneur's optimal price

Another useful property of the Cobb-Douglas isoreward class is, as we show in this subsection, that all innovators are incentivized to charge the same fraction a of the monopoly price. As pictured in Figure 4, to maximize rewards, the entrepreneur chooses the point on her demand curve at which it is tangent to the isoreward curve, given that the reward is (weakly) increasing as one moves out across isoreward curves. Cobb-Douglas isoreward curves are particular simple from this perspective as they have a constant elasticity $\frac{\alpha}{1-\alpha}$.[16] Thus every entrepreneur will choose a point that equates the elasticity of her demand curve, $\epsilon(a)$, to $\frac{\alpha}{1-\alpha}$.[17] Finite demand and zero marginal distortion at the ex-post-efficient price imply that $\epsilon(0) = 0$ and monopoly optimization that $\epsilon(1) = 1$. For marginal revenue to be decreasing, $\epsilon' > 0$ whenever $\epsilon \leq 1$.[18] Thus for any $\alpha \in \left[0, \frac{1}{2}\right]$, every entrepreneur will charge the same,

[16]Regardless of whether the isoreward curves are Cobb-Douglas, the entrepreneur will always choose a point on her demand curve whose elasticity equals the elasticity of the isoreward curve at that point.

[17]This can be seen because $q^{1-\alpha}p^{\alpha} = k^{1-\alpha} \implies q = kp^{\frac{\alpha}{1-\alpha}}$.

[18]$MR = p - \frac{p}{\epsilon} \propto a - \frac{a}{\epsilon(a)}$, so

$$Q'(MR)' > 0 \iff \left[a\left(1 - \frac{1}{\epsilon}\right)\right]' < 0 \iff \epsilon' > -\frac{\epsilon}{a}\left(1 - \frac{1}{\epsilon}\right) \iff \epsilon' > -\frac{\epsilon^2}{a}\left(1 - \frac{1}{\epsilon}\right).$$

This also implies that for all $\alpha > \frac{1}{2}$, $a > 1$ and thus $\alpha > \frac{1}{2}$ always leads to above-monopoly pricing.

unique $a(\alpha)$ solving

$$\epsilon\left(a(\alpha)\right) = \frac{\alpha}{1-\alpha}. \tag{2}$$

Thus we will alternate interchangeably in what follows between α and a.

2.2 The trade-off between ex-post distortion and screening

Because all entrepreneurs charge the same a, their prices and quantities are proportional to m and σ respectively as $p = a(\alpha)m$ and $q = \sigma Q\left(a(\alpha)\right)$. Thus an entrepreneur of type (σ, m) receives reward $\tau\left(\sigma^{1-\alpha}m^{\alpha};\alpha\right) \equiv t\left(Q\left(a(\alpha)\right)^{1-\alpha} a(\alpha)^{\alpha}\sigma^{1-\alpha}m^{\alpha}\right)$.[19] Additionally, the value of innovations, $\sigma m S\left(a(\alpha)\right)$ is proportional to σm. Thus Pigouvian payment in accordance to product would require the reward to be based on $\sigma^{\frac{1}{2}}m^{\frac{1}{2}}$ and thus is possible only with $\alpha = \frac{1}{2}$ (i.e. full monopoly pricing). The higher the value of α (below $\frac{1}{2}$), the closer we are to payment in accordance with product and thus perfect sorting among innovations. However, because $\epsilon' > 0, a' > 0$ and thus prices and ex-post distortion rise in α. We can thus refer to α as the degree of *market power* chosen by the planner. This establishes the basic tradeoff between sorting and ex-post efficiency that is the key to optimal policy.

This also immediately implies a useful and intuitive result: it is never optimal to choose $\alpha > \frac{1}{2}$ and thus induce prices above the monopoly optimum. This both worsens ex-post distortion and reduces the quality of sorting by over-rewarding quality relative to size

3 Optimal Rewards and Pricing

We can now decompose the planner's problem as

$$\max_{\alpha} W(\alpha) \text{ where } W(\alpha) = \max_{\tau(\cdot)} \int_{\theta: c \leq \tau(\sigma^{1-\alpha}m^{\alpha})} \left[\sigma m S\left(a(\alpha)\right) - c\right] f(\theta) d\theta,$$

which illustrates the three steps we follow in analyzing the problem. First, for a given market power α (or equivalently a) we perform the straightforward exercise of maximizing social welfare over choices of reward schedules τ. Then we establish the applicability of the envelope theorem to our context and thus the differentiability of the value function $W(\alpha)$. Finally, we

[19]From here on we use τ to represent rewards in the type space and t the same in price-quantity space.

use the envelope conditions derived in the first stage to determine the marginal benefits and costs of raising α, W'.

3.1 Optimal transfers

We first consider the optimal reward along each isoreward curve independently, a simple unidimensional problem. It is useful to re-parameterize the problem in terms of isoreward curves and the quality-market size ratio along a particular isoreward curve,

$$k = \sigma^{1-\alpha} m^{\alpha} \text{ and } x \equiv \frac{m}{\sigma},$$

respectively, rather than (σ, m). We will refer to the relevant (change-of-variables augmented) distribution function as $\tilde{f}(k, x, c)$ to distinguish it from $f(\sigma, m, c)$.[20] In this new notation, the social value created by an innovation, $\sigma m S(a(\alpha))$, becomes $k^2 x^{1-2\alpha} S(a(\alpha))$.

For a given reward $\tau(k)$, all innovations along the k isoreward curve with cost less than $\tau(k)$, those with $c \leq \tau\left(\sigma^{1-\alpha} m^{\alpha}\right)$ are created. If the "average marginal innovation" (i.e. the average innovation along isoreward curve k with cost $c = \tau(k)$) creates social value greater (lower) than $\tau(k)$, the social planner has an incentive to raise (lower) $\tau(k)$. Thus for a given α, the optimum is at a point at which these are exactly equated:

$$\tau^{\star}(k; \alpha) = k^2 S(a) E\left(x^{1-2\alpha} \middle| c = \tau^{*}(k; \alpha), k\right), \tag{3}$$

where the expectation is taken over x under $\tilde{f}$.[21] For implementation, the optimal transfer function in the (q, p) space is then derived simply by inverting the definition of τ from Subsection 2.2 above: $t^{\star}\left(\frac{k}{Q(a(\alpha))^{1-\alpha} a(\alpha)^{\alpha}}; \alpha\right) = \tau^{\star}(k; \alpha)$. The social value function $W(\alpha)$ associated with each pricing policy α can thus be computed.

[20] The formula for this transformation is provided in Appendix A.

[21] If c is not too affiliated with x under $\widetilde{f}$ given k, in a sense formalized in Proposition 3 in Appendix A, then there is a unique point at which this condition is satisfied and this constitutes the optimal (monotonicity-relaxed) transfer. Furthermore, if k is not too negatively affiliated under $\widetilde{f}$ with x given c (see again Appendix A) then $\tau^{\star}$ is monotone increasing and thus is the truly optimal transfer function. If either (but not both) of these conditions fail, standard ironing techniques can be used to determine optimal transfers as described in Appendix C. In any case the envelope conditions for $\tau^{\star}$ (the average marginal innovation's value equals the transfer) hold once innovations are pooled over the ironing region.

3.2 Optimal market power

Because $\tau^\star(k;\alpha)$ is chosen optimally, we can apply Milgrom and Segal (2002)'s envelope theorem for general choice sets and consider only the direct effect of an increase in a on social welfare, ignoring indirect effects through the optimal choice of $\tau^\star$.

Lemma 1. *$W(\alpha)$ is differentiable for all $\alpha \in \left(0, \frac{1}{2}\right)$ and its derivative may be evaluated by the envelope theorem, holding $\tau^\star$ fixed. Formally:*

$$W'(\hat{\alpha}) = \frac{\partial}{\partial \alpha}\left[\int_\sigma \int_m \int_{c=0}^{\tau^\star\left(\sigma^{1-\alpha}m^\alpha;\hat{\alpha}\right)} \left[\sigma m S\left(a(\alpha)\right) - c\right] f(\sigma, m, c) dc dm d\sigma\right]\Bigg|_{\alpha=\hat{\alpha}}$$

Proof. See Appendix B. □

In fact, this holds even if $\tau^\star$ is non-differentiable: the smoothness properties we have assumed on f, combined with the monotonicity of $\tau^\star$ are sufficient to establish the equidifferentiability and continuity conditions required by Milgrom and Segal. However, in most of what follows we will derive the formulae for the case when $\tau^\star$ is differentiable.[22]

Theorem 1: *An optimal value of α, α^*, exists. Either α^* is strictly between 0 and $\frac{1}{2}$, or $\alpha^\star = 0$ and $\tau^\star(k;\alpha)$ is constant in k.*

If transfers are everywhere constant when $\alpha = 0$, sorting need not have a local benefit as it is not used. However, such flat transfers can easily be ruled out by a significant weakening of our no-ironing condition (14) in Appendix C, which requires that σ not be too negatively affiliated with m. Thus when optimal transfers at $\alpha = 0$ are flat in k, this is likely to raise the *global* value of sorting as it implies sorting is very poor at $\alpha = 0$, as σ is a poor indicator of overall value as in our "lemons" example from Subsection 1.2. Thus, we conjecture that $\alpha = 0$ is never optimal.

Proof. See Appendix D. □

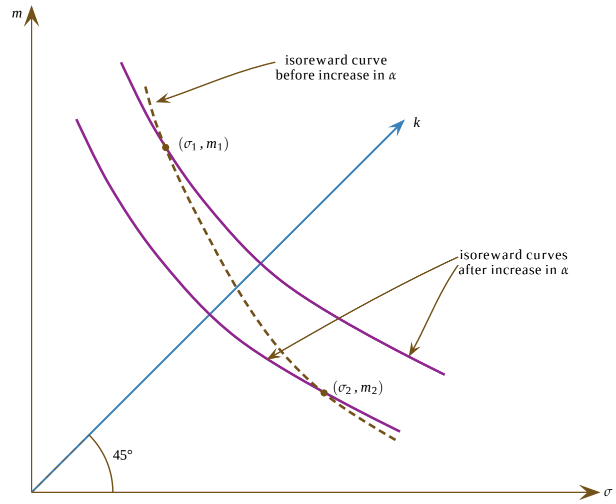

Figure 5: Increasing market power changes isoreward curves and therefore the reward given to different innovations, favoring high quality over large market size

3.3 Heuristics

Figure 5 first depicts an isoreward curve (the dashed curve) for a given α. Types (σ_1, m_1) and (σ_2, m_2) receive the same reward. An increase in α makes the isoreward curves flatter. This has the effect of shifting the high-quality type (σ_1, m_1) to a higher k (higher isoreward curve; note that $k = \sigma = m$ on the 45° line and is therefore independent of α) and the low-quality type (σ_2, m_2) to a lower k. This in turn implies that for locally fixed transfers $\tau^*(k;\alpha)$, a small increase in α at the margin will attract in marginal (cost equal to reward) high-quality projects and crowd out marginal low-quality projects.

The further points are from the $(\sigma = m = k)$ 45° line, the more quickly the k values corresponding to the point increases (if the point is above the line), or decreases (if it is below), in α. The rate of moving up (or down) isoreward curves $\frac{d\sigma^{1-\alpha}m^{\alpha}}{d\alpha}$ for an innovation of partial type x is proportional to $\log(x)$, just as the production of a Cobb-Douglas economy responds to a shift in shares at a rate proportional to the ratio of factors because isoquants are log-linear: $\log(k) = (1-\alpha)\log(\sigma) + \alpha\log(m)$.

The value of innovations along any particular isoreward curve is proportional to $x^{1-2\alpha}$

[22] The absence of ironing is sufficient to ensure this differentiability. Sufficient conditions for the absence of ironing are given in Appendix A. Appendix D presents analogous results when optimal transfers are non-differentiable. Additionally this approach may be applied, under appropriate conditions, when isoreward curves are characterized by (infinitely) many parameters as shown in our online appendix.

by Subsection 3.1. The social benefit associated with raising α is that high x innovations are created more frequently: to the extent that $x^{1-2\alpha}$ covaries with $\log(x)$ (which it must to some extent), more beneficial innovations will be selected by higher values of α.[23] Thus $\text{Cov}\left[x^{1-2\alpha}, \log(x)|k, c = \tau^\star(k;\alpha)\right]$, the covariance between the extra rewards given to (marginal) innovations and the value of these innovations, is a crucial quantity pushing towards greater market power. Furthermore, because $\lim_{\alpha\to\frac{1}{2}} \frac{x^{1-2\alpha}-1}{1-2\alpha} = \log(x)$, for α close to $\frac{1}{2}$,

$$\text{Cov}\left[x^{1-2\alpha}, \log(x)|k, c = \tau^\star(k;\alpha)\right] \approx (1-2\alpha)\,\text{Var}\left[\log(x)|\, k, c = \tau^\star(k;1)\right]$$

Thus, if one is considering near-monopoly pricing rules, the incentive for marginally higher α is nearly proportional to the (conditional) variance of the logarithm of x.

On the other hand, raising α reduces the value of all innovations that are created, not just the marginal ones (with $c = \tau^\star(k;\alpha)$). Thus, the relative weight on this distortion is the ratio of the mass of infra-marginal innovations to the density of marginal innovations; that is, the inverse hazard rate, or inverse semi-elasticity, of innovations with respect the reward/cost along a given isoreward curve. The harm per innovation is proportionate to the value of the innovations which is, by the optimality conditions for $\tau^\star$ closely tied to the reward given to innovations. Thus, we can normalize the semi-elasticity and obtain that it is t the (in)elasticity of innovations with respect to the rewards given them that determines the relative size of the disincentive to ex-post distortion.

3.4 Marginal costs and benefits of market power

The factors considered informally in the preceding subsection can be formally shown to be the key determinants of the net marginal social benefit of ex-post distortion.

Corollary 1: [24] *Assuming $\tau^\star(k;\alpha)$ is differentiable in k at α, $W'(\alpha) \propto$*

$$E\left[k^4\left[\underbrace{(1-2\alpha)\frac{\tau^{\star\prime}}{k}\eta Cov\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right)}_{sorting} - \underbrace{\frac{Q\alpha}{(1-\alpha)^3\epsilon'}E\left(x^{1-2\alpha}\right)E\left(x^{1-2\alpha}\middle|\, c<\tau^\star\right)}_{ex\text{-}post\ distortion}\right]\right], \tag{4}$$

[23] See Veiga and Weyl (2011) for a general analysis of this logic of smooth screening.

[24] This formula is a special case of the general first-order condition which applies even when $\tau^\star$ is not differentiable.

where the outer expectation is taken over k, the covariance and inner expectations over x, $\eta \equiv \frac{f_c(\tau^\star(k;\alpha)|k)\tau^\star(k;\alpha)}{F_c(\tau^\star(k;\alpha)|k)}$ is the elasticity of innovations with respect to reward, all quantities inside the expectation are evaluated conditional on k, $c = \tau^\star(k;\alpha)$ where not explicitly stated, f_c and F_c the (conditional) marginal pdf and cdf respectively of c and all expectations are taken under $\tilde{f}$. As usual, a necessary condition for the optimal choice of α is that this equal 0.

Proof. Assuming continuous differentiability of $\tau^\star$,

$$\frac{\partial}{\partial\alpha}\left[\int_\sigma\int_m\int_{c=0}^{\tau^\star\left(\sigma^{1-\alpha}m^\alpha;\hat{\alpha}\right)}\left[\sigma m S\left(a(\alpha)\right)-c\right]f(\sigma,m,c)dcdmd\sigma\right]\Bigg|_{\alpha=\hat{\alpha}}=$$

$$\int_\sigma\int_m\left(\sigma^{1-\hat{\alpha}}m^{\hat{\alpha}}\tau^{\star\prime}\log\left(\frac{m}{\sigma}\right)\left(\sigma m S-\tau^\star\right)f\left(\sigma,m,\tau^\star\right)+\int_{c=0}^{\tau^\star}\sigma m a' Q\epsilon f(\sigma,m,c)dc\right)dmd\sigma,$$

where all variables are evaluated at $\hat{\alpha}$. Recalling that $\epsilon\left(a(\alpha)\right)=\frac{\alpha}{1-\alpha}$, and thus, $a'=\frac{1}{(1-\alpha)^2\epsilon'}$, W' in the (k,x) space is

$$\int_k\int_x k\tau^{\star\prime}\log(x)\left(k^2x^{1-2\hat{\alpha}}S-\tau^\star\right)\tilde{f}\left(k,x,\tau^\star\right)-\int_{c=0}^{\tau^\star}k^2x^{1-2\hat{\alpha}}\frac{\hat{\alpha}Q}{(1-\hat{\alpha})^3\epsilon'}\tilde{f}(k,x,c)dcdxdk.$$

Letting $\tilde{f}_k(k)$ be the marginal distribution under $\tilde{f}$ of k (and similarly with other subscripts of densities), this becomes

$$\int_k E_{x,\tilde{f}}\left[\left[k\tau^{\star\prime}\log(x)\left(k^2x^{1-2\hat{\alpha}}S-\tau^\star\right)\Big|c=\tau^\star,k\right]\tilde{f}_k[\tau^\star;\hat{\alpha}]-k^2\frac{\hat{\alpha}Q}{(1-\hat{\alpha})^3\epsilon'}E_{x,\tilde{f}}\left(x^{1-2\hat{\alpha}}\Big|c<\tau^\star,k\right)\tilde{F}_k(\tau^\star;\hat{\alpha})\right]\tilde{f}(k)dk.$$

By the envelope conditions in equation (3), $\tau^\star=E_{x,\tilde{f}}\left(k^2x^{1-2\hat{\alpha}}S\big|\,c=\tau^\star,k\right)$, so that we can rewrite the derivative as

$$S\int_k k^4\left[\frac{\tau^{\star\prime}}{k}\mathrm{Cov}_{x,\tilde{f}}\left(\log(x),x^{1-2\hat{\alpha}}\big|\,c=\tau^\star,k\right)-\frac{\hat{\alpha}Q}{(1-\hat{\alpha})^3\epsilon'}\frac{E_{x,\tilde{f}}\left(x^{1-2\hat{\alpha}}\big|c<\tau^\star,k\right)E_{x,\tilde{f}}\left(x^{1-2\hat{\alpha}}\big|c=\tau^\star,k\right)}{\eta(\tau^\star|k)}\right]\tilde{f}_{kc}\left(k,\tau^\star\right)dk$$

where $\tilde{f}(k,c)$ is the marginal distribution of (k,c). If we now normalize by dividing by $S\int_k\tilde{f}\left(k,\tau^\star\right)dk$, remove remaining unnecessary arguments and subscripts and multiply and divide the first term in the appropriate places by $1-2\alpha$, we obtain the desired expression. □

For the first-order condition to actually characterize the optimum, the problem must be (quasi-)concave. While it is not typically tractable to determine simple conditions on primitives

to ensure this, we show in Appendix E that so long as the product of the inequality and elasticity of innovation supply does not increase too rapidly with α, especially for intermediate values of α, the problem is concave. All of our computational simulations thus far, some of which are described in the next subsection, exhibit such concavity.

3.5 Approximation theorems

Assuming quasi-concavity, we now consider the limit as the elasticity and the inequality of innovation supply grow large or small.

Theorem 2 (Optimal near-monopoly pricing)**:** *Let $\pi \equiv \sigma m Q(1)$ be the monopoly profit associated with an innovation and V_1 be defined by:*

$$V_1 \equiv \frac{E\left[\pi^2\, Var\left(\log(x)\middle|\, \pi, c=\frac{S(1)}{Q(1)}\pi\right)\eta\left(\frac{S(1)}{Q(1)}\pi\middle|\,\pi\right)\right]}{E\left[\pi^2\right]},$$

where the expectations are taken with respect to π under the distribution over (π, x, c) induced by $f(\sigma, m, c)$ and the variance is taken with respect to the conditional-on-π distribution of x.[25] *Then, if W is quasi-concave or if we consider only α values sufficiently near $\frac{1}{2}$, the optimal value of α is*

$$\alpha^\star = \frac{1}{2} - \frac{Q(1)}{S(1)\epsilon'(1)V_1} + O\left(\frac{1}{V_1^2}\right). \tag{5}$$

Thus for V_1 sufficiently large, $\alpha^\star$ approaches $\frac{1}{2}$; that is, as the value-weighted average of the product of the elasticity and inequality of innovation supply grow large, near-monopoly pricing becomes optimal.

Proof. By Corollary 1, rearranged slightly, the necessary condition for determining $\alpha^\star$ is

$$(1-2\alpha)\, SE\left[\frac{1}{S}\tau^{\star\prime}k^3\eta\mathrm{Cov}\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right)\right] = \frac{Q\alpha}{2(1-\alpha)^3\epsilon'}E\left[k^4 E\left(x^{1-2\alpha}\right)E\left(x^{1-2\alpha}\middle|\, c<\tau^\star\right)\right].$$

[25] Formally this is $\tilde{f}$ at $\alpha = \frac{1}{2}$.

Letting $\omega(\alpha) \equiv (1-2\alpha)\frac{S(1-\alpha)^3\epsilon'}{Q\alpha}$ and

$$\beta(\alpha) \equiv \frac{E\left[\frac{1}{S}\tau^{\star\prime} k^3 \eta \text{Cov}\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right)\right]}{V_1 E\left[k^4 E\left(x^{1-2\alpha}\right) E\left(x^{1-2\alpha} \middle| c < \tau^\star\right)\right]},$$

the first-order condition becomes $\omega(\alpha)\beta(\alpha) \equiv \kappa(\alpha) = \frac{1}{V_1}$. Proposition 5 in Appendix E shows that κ is locally decreasing near $\alpha = \frac{1}{2}$ and thus we may consider the local inversion of κ in this neighborhood $\kappa^{-1}_{\delta_{\frac{1}{2}}}$. Note that $\omega\left(\frac{1}{2}\right) = 0$. To determine $\beta\left(\frac{1}{2}\right)$ note first that $\tau^\star\left(k; \frac{1}{2}\right) = S(1)k^2$ by equation (3) and thus $\tau^{\star\prime}\left(k; \frac{1}{2}\right) = 2S(1)k$. Second, note that by L'Hôpital's rule, $\lim_{\alpha\to\frac{1}{2}} \frac{x^{1-2\alpha}-1}{1-2\alpha} = \log(x)$ so $\lim_{\alpha\to\frac{1}{2}} \text{Cov}\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right) = \text{Var}\left(\log(x)\right)$. Third, note that when $a = 1$, $k^2 Q(1) = \pi$ and that $E\left(x^{1-2\alpha}\right) = E\left(x^{1-2\alpha} \middle| c < \tau^\star\right) = 1$. Thus $\lim_{\alpha\to\frac{1}{2}} \beta(\alpha) = 2$. Thus Taylor expanding $\kappa^{-1}_{\delta_{\frac{1}{2}}}$ about 0 yields

$$\alpha^\star = \frac{1}{2} + \kappa^{-1\prime}_{\delta_{\frac{1}{2}}}(0)\frac{1}{V_1} + O\left(\frac{1}{V_1^2}\right). \tag{6}$$

By the inverse function theorem,

$$\kappa^{-1\prime}_{\delta_{\frac{1}{2}}}(0) = \frac{1}{\kappa'\left(\frac{1}{2}\right)} = \frac{1}{\omega'\left(\frac{1}{2}\right)\beta\left(\frac{1}{2}\right)} = \frac{1}{-2\frac{S(1)\left(1-\frac{1}{2}\right)^3\epsilon'(1)}{Q(1)\cdot\frac{1}{2}} \cdot 2} = -\frac{Q(1)}{S(1)\epsilon'(1)},$$

so that rearranging equation (6) yields the result. □

Intuitively, as the inequality and elasticity of innovation supply grow, the incentives for ex-post distortion grow until monopoly pricing becomes optimal in the limit. In this limit many of the complexities above disappear: isoreward curves become isoprofit sets, optimal transfers collapse to the social surplus of every innovation, namely $\frac{S(1)}{Q(1)}$ times its profit, and optimal pricing is near monopoly.

This particular approximation is also of interest for a pragmatic reason. Most advanced capitalist nations use primarily market-power-based schemes for rewarding innovation: thus current policy has α near $\frac{1}{2}$. Hence a local approximation to optimal policy near this point seems most useful for a gradualist interested in moving cautiously away from this point. If V_1

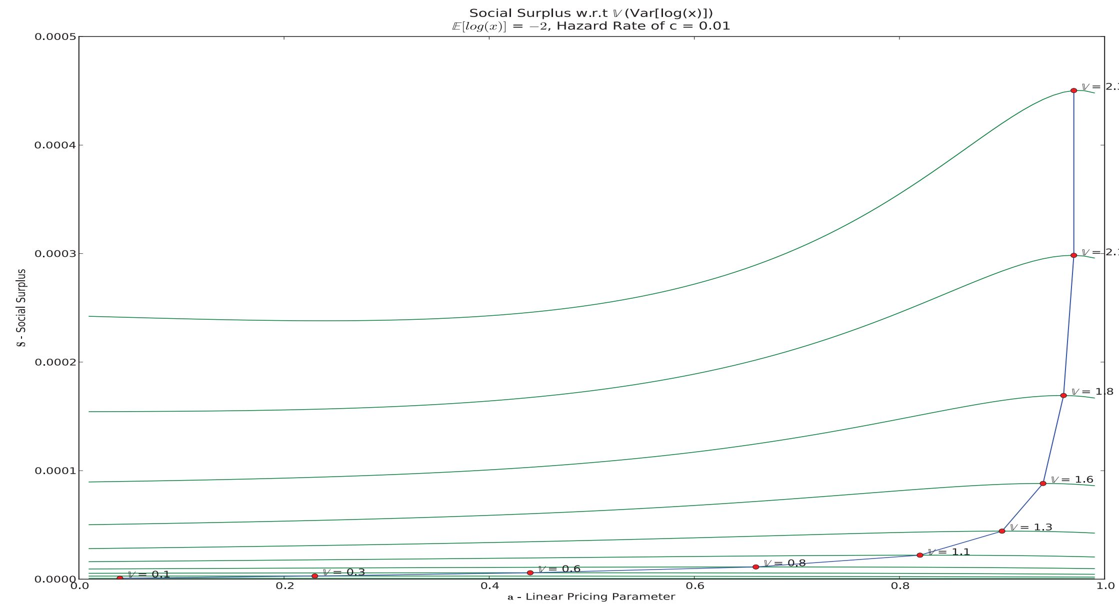

Figure 6: Increasing the inequality of innovation supply raises optimal market power

is not in fact measured to be large, one may only conclude it should move down significantly, but how far exactly is not clear as the approximation becomes poor. Empirical implementation of this approach is discussed further in Section 6. After policy has been changed marginally, the model may be recalibrated and further adjustments enacted, in the spirit of Chetty (2009). However, to illustrate the generality of the approach, we also consider an approximation about $a = 0$, which has the additional benefit of providing a near-converse toTheorem 2.

Theorem 3 (Partial converse)**:** *Let V_0 be defined by*

$$V_0 \equiv \frac{E\left[\sigma^4 \dfrac{\log\left(\tau_0\right)' \, Cov\left(\log(m), m|\sigma, c = \tau_0(\sigma)\right)\eta\left(\tau_0(\sigma)|\sigma\right)}{\sigma E\left(m|\sigma, c < \tau_0(\sigma)\right)}\right]}{E(\sigma^4)},$$

where τ_0, the optimal reward schedule at $\alpha = 0$, is defined implicitly by $\tau_0(\sigma) = \sigma S(0) E\left(m|c = \tau_0\left(\sigma\right), \sigma\right)$, the outer expectation is taken with respect to σ and the inner expectation and covariance are taken over m, all under the measure f. Then, if W is quasi-concave in α or if we consider only α values sufficiently near 0, the optimal value of α is

$$\alpha^\star = \epsilon'(0) V_0 + O\left(V_0^2\right). \tag{7}$$

Thus for V_0 sufficiently small, $\alpha^\star$ approaches 0.

Proof. Follows from analogous logic *mutatis mutandis* to the proof of Theorem 2 (Taylor expanding now about 0) as in this limit $k = \sigma$, $x^{1-2\alpha} = x$ when $\alpha = 0$ and $\left.\frac{Q}{(1-\alpha)^3\epsilon'}\right|_{\alpha=0} = \frac{1}{\epsilon'(0)}$. $\square$

Intuitively, as m becomes perfectly known or all innovations become infra-marginal, the screening benefit of ex-post distortion becomes small. Optimal policy becomes ex-post-efficient pricing coupled with demand-dependent prizes.

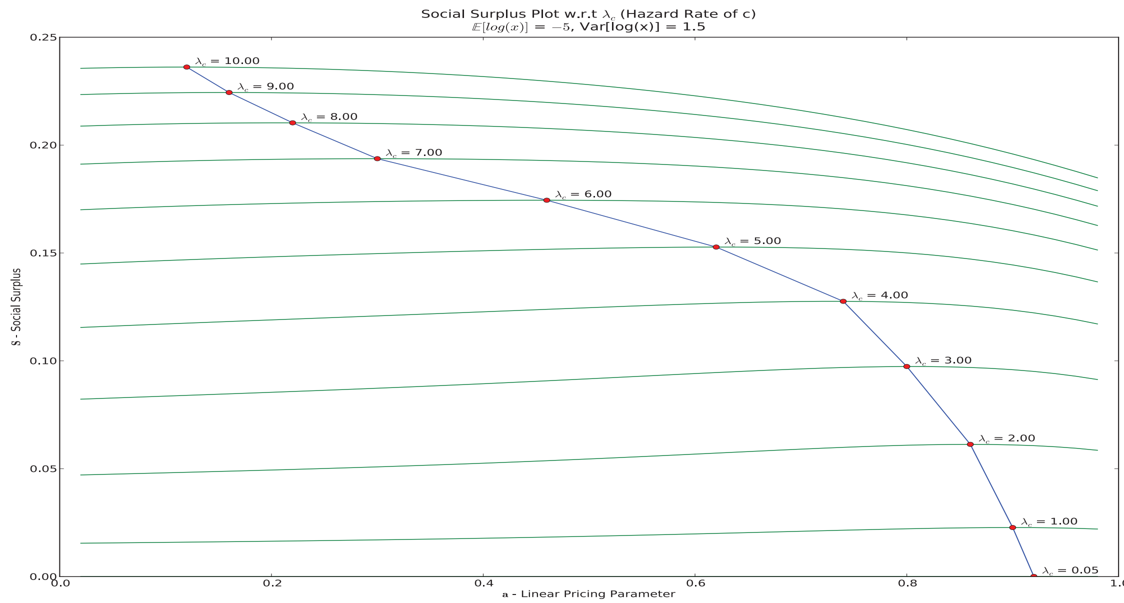

Figure 7: Increasing the elasticity of innovation supply raises optimal market power

While these two theorems are not quite converses, numerical simulations verify the validity of our intuitions. In particular we consider a simulation in which π is drawn from a uniform distribution, x is drawn log-normally and c is drawn exponentially (so that its elasticity can be unambiguously adjusted), all independent of one another. Figure 6, which pictures W for various values of the log-variance of x, shows that increasing the log-variance of x increases $a^\star \equiv a(\alpha^\star)$ through the full range of values. Figure 7 shows W for various values of the exponential parameter, showing that as the elasticity increases, so too does $a^\star$.[26]

[26]It should be noted that we have graphed the value $\tilde{W}$ that could be achieved if monotonicity constraints were not imposed, not the true W that would result from ironing. This is to avoid the difficulty of an ironing routine; however, it does not bias the results on $\alpha^\star$ as we have verified that, in every example, $\tau^{\star\star}$, the monotonicity relaxed optimal τ, is in fact monotone at $\alpha^\star$. It is only at sub-optimally low α that ironing may be necessary and there it only makes these a values further unattractive. Whether it is "typically" the case that ironing is unnecessary at or above optimal α is a question we hope to investigate in further simulations. Python code for our simulations is available at http://www.glenweyl.com. Appendices A and C discuss ironing in greater detail.

3.6 Discussion

A simple application of our results is to the relative attractiveness of prizes compared to IP in the development of new medicines compared to software and other high-technology. Kremer, his co-authors and many others have advocated prizes for the development of new medicines. More broadly, many developed countries have extensive price controls, centralized bargaining, subsidies and other non-market mechanisms for increasing access to medicines.[27] As in a prize system, these determine rewards per unit sold independently of the discipline of market prices. On the other hand, most advanced economies use, and most economists would advocate, market-oriented pricing of high technology products, from software to tablet computers.

These intuitions square with our reasoning because, at least in many cases, the per-consumer value of medicines can be judged somewhat objectively using clinical trials and standard calculations of the monetary value of added quality-adjusted life years. Thus the innovator has little private information on m ex post. Conversely high technology exhibits nearly the opposite pattern: it is typically far from apparent how much consumers value various new electronics and software until the market demonstrates this. Taking our example from the introduction, many engineers believed AOL would be worth little compared to Netscape because they failed to take into account the AOL's relative ease and social benefits for non-specialists.

One way to interpret the difference between these systems is the notion of "entrepreneurship". While this concept is taken to mean many things, one way to interpret an entrepreneur is as an individual who acts on the basis of strongly-held minority views about demand conditions. An "entrepreneurial" system is then one in which the set of innovations created or products brought to market is highly responsive to private information held by entrepreneurs about the value of the products, possibly at the expense of market power in the spirit of Schumpeter (1942). In this sense, our α can exactly be interpreted as the degree to which the reward structure is entrepreneurial as isoreward curves $\sigma^{1-\alpha}m^{\alpha}$ coincide with social isovalue curves $\sigma m S\left(a(\alpha)\right)$ to the extent that α is close to $\frac{1}{2}$. In fact, the covariance that determines the marginal screening benefits of α is exactly the increase in the covariance (along the set of

[27] These policies have traditionally been explained by the government discounting innovator surplus and/or low η at the country level, so that in the absence of cross-country coordination, each country internalizes too little of its effect on innovation supply. However, the difference between the medical and other sectors in this dimension suggests a role for private information about quality.

marginal innovations) between $\sigma^{1-\alpha}m^{\alpha}$ and σm.

Thus our results may be seen as establishing a link between belief in strong material motivation (elasticity of innovation supply) and heterogeneity of ability in entrepreneurship (inequality of innovation supply) on the one hand and belief in the desirability of an entrepreneurial system on the other. This may help explain formally the emphasis Hayek (1945) places on the private information of entrepreneurs and Friedman (1962) puts on materially-motivated "individual geniuses" in justifying such a system, as well as clarify which exact institutions "entrepreneurship" represents for these authors.

4 Modeling Choices and Robustness

This section discusses modeling choices and extends our basic framework in a number of directions, largely exhibiting the robustness of both its techniques and conclusions.

4.1 Implications of the stretch parameterization

We make extensive use of our stretch parameterization of demand, which significantly generalizes the linear specification of preferences typically used in multidimensional screening models (Rochet and Choné, 1998).[28] However, the parameterization still embodies a crucial assumption, namely that if a constant fraction of the monopoly price is charged for all innovations, then the average consumer surplus generated by an innovation is proportional to its monopoly optimal price. Effectively this rules out demand curves differing in their curvature. We suspect that in a more general model where curvature too could vary, our results would continue to hold if the correlation between curvature and (σ, m) were low. However, if curvature, and thus the ratio of average consumer surplus to monopoly price (Weyl and Fabinger, 2011), were highly positively correlated with m or negatively correlated with σ, the planner might want to have an incentive to set α above $\frac{1}{2}$ to further screen for curvature. On the other hand if curvature had a sufficiently opposite correlation, the role of market power in screening might be eliminated.

[28]Thus we suspect this parameterization for indifference curves may be useful more broadly in multidimensional screening.

An increase in σ corresponds to an increase in the traditional market size parameter in standard industrial organization (Bresnahan and Reiss, 1991) and international trade (Krugman, 1980). A shift in m, on the other hand, corresponds to a proportional increase in all consumers' willingness-to-pay, often represented (Berry et al., 1995) by a proportional increase in their income.

Among the classes of preferences representable by the stretch parametrization are all constant pass-through demand (Bulow and Pfleiderer, 1983) functions with common pass-through rate, the broader Apt demand class (Weyl and Fabinger, 2009) with common limiting pass-through and slope of pass-through that scales with m, any demand based on a statistical distribution with a constant location-to-scale parametrization and a single-product version of AIDS (Deaton and Muellbauer, 1980).[29]

Given its intuitiveness, this parameterization seems likely to be useful in other areas of industrial organization and beyond. As discussed in Subsection 4.2 it seems a natural parameterization for demand morphing in the spirit of Johnson and Myatt (2006). Similarly, increasing σ and m in proportion provides a simple manifestation of Bresnahan (1982)'s "demand shifters" and while changing their ratio while holding, say, their product constant manifests a demand "twister". It also offers a simple way to parameterize cases when, in the sense of Spence (1975), a monopolist has excessive or deficient incentives to supply quality, holding fixed quantity:

Proposition 1: *Holding fixed quantity, a monopolist will always have too little incentive to supply m. Furthermore, so long as demand is log-concave (linear-cost pass-through is less than 1), a monopolist will have excessive incentive to supply σ.*

Intuitively, most of the benefits of increased m go to infra-marginal consumers, but most of the benefits of increased σ go to marginal consumers.

Proof. Social value is $\sigma m S\left(Q^{-1}\left(\frac{q}{\sigma}\right)\right)$ while profits are $qmQ^{-1}\left(\frac{q}{\sigma}\right)$. Thus, the marginal social incentive to supply m is σS while the marginal private incentive is $qQ^{-1} = \sigma\frac{Qp}{m}$. But $\frac{Sm}{Qp}$ is exactly the ratio of average to marginal willingness-to-pay which is clearly above unity.

[29] See Weyl and Fabinger (2009) for an extensive list of statistical distributions used to generate demand functions. Two prominent examples are Gaussian or Type-I extreme value distribution with constant location to scale parameter ratios.

The social incentive to provide σ holding fixed q is

$$-\frac{S'}{Q'}\frac{q}{\sigma}m + mS = \frac{Q\epsilon Q}{Q'a}am + Sm = (S - aQ)m,$$

while the private incentive is

$$-\frac{q^2 m}{Q'\sigma^2} = -\frac{Q^2 m}{Q'} = \frac{Qam}{\epsilon}.$$

At monopoly optimal prices the first simplifies to $(S(1) - Q(1))\, m$ and the second to $Q(1)m$. $\frac{S(1)-Q(1)}{Q(1)}$ is the ratio of consumer to producer surplus at the monopoly optimal price, whose comparison to unity is dictated by the average pass-through rate at prices above the monopoly optimum (Weyl and Fabinger, 2011). □

4.2 Moral hazard

Society may face a trade-off between having particular entrepreneurs work on developing one or another innovation. An extreme example of such a trade-off is simple to model: the entrepreneur must choose which of many possible innovations to create. This leads to a natural moral hazard version of our model which, while too simple to formalize the notion of the inequality and elasticity of innovation supply, illustrates the robustness of our basic argument.

Suppose an entrepreneur can choose to create, for effort cost e, any innovation lying along the smooth curve $\sigma = h(m; e)$ where $h_m < 0 < h_e$. It is natural to assume that along an isoeffort curve $h(m; e)$, there is a unique point generating maximal potential surplus $mh(m; e)$ so that if the social planner were to select unconstrained, his problem would be quasi-concave.[30] A standard condition for this is increasing elasticity: $\frac{\partial \epsilon_h}{\partial m} > 0$ where $\epsilon_h \equiv -\frac{mh'}{h}$, where ϵ_g is the elasticity of an arbitrary function g.

The entrepreneur would like to minimize her effort cost of obtaining a reward t that the social planner offers her if she achieves a specified quantity-price target, which may be set in advance given the lack of ex-ante private information. From now on we will suppress the dependence of h on e, assuming the social planner has chosen (in some other stage of analysis) the optimal effort $e^\star$ to induce. Thus, we focus on the choice between quality and market size.

[30] Interestingly, Johnson and Myatt (2006) base their analysis on an assumption implying that the social planner's problem would be globally convex.

Incentive compatibility requires that the entrepreneur select m so to maximize $\sigma Q\left(\frac{p}{m}\right) = h(m)Q\left(\frac{p}{m}\right)$, or equivalently that the isoeffort curve be tangent to the demand curve at the requested price-quantity pair; otherwise the entrepreneur could achieve the desired price and quantity at a lower cost. Therefore, for (q, p) to be incentive compatible it must be that

$$\epsilon_h(m) = \epsilon\left(\frac{p}{m}\right). \tag{8}$$

Note that m is increasing in $a \equiv \frac{p}{m}$ because ϵ and ϵ_h are both increasing. Price must be higher to induce the entrepreneur to choose a higher quality product, as these are the products which fare better at higher prices.

Social welfare is $h(m)mS(a)$. If the social planner were unconstrained by incentives, he would choose $a = 0$ and $m = m^\star$ where $m^\star$ is the unique maximizer of $h(m)m$. However to achieve the surplus maximizing m would require $\epsilon_h = 1$, the social planner's problem is equivalent to the monopoly problem, which would, by incentive compatibility, require $a = 1$. This is exactly the same trade-off as in the adverse selection model and gives rise to a simple expression for the cost and benefit of ex-post distortion.

Proposition 2: *The first-order net benefit of increasing a in the moral-hazard model is proportional to*

$$\underbrace{(1-\epsilon)\,\epsilon_\epsilon}_{\textit{incentivizing high quality}} - \underbrace{\epsilon_{\epsilon_h}\epsilon_S}_{\textit{ex-post distortion}}.$$

Furthermore $0 < a^\star < 1$.

Proof. We can log-differentiate social welfare:

$$-\epsilon_S + (1-\epsilon_h)\frac{\partial \log(m)}{\partial \log(a)}$$

But by implicit differentiation of equation (8),

$$\frac{\partial \log(m)}{\partial \log(a)} = \frac{\epsilon_\epsilon}{\epsilon_{\epsilon_h}}$$

Combining this with the first expression, multiplying by ϵ_{ϵ_h} and noting that by incentive compatibility $\epsilon_h = \epsilon$ yields the result.

When $a = 0$, $\epsilon_S = 0$ as there is no first-order distortion from raising a but $(1-\epsilon)\epsilon_\epsilon > 0$ (as $\epsilon = 0$ and thus, in fact, $\epsilon_\epsilon = \infty$) and thus there is a first-order benefit from raising a. When $a = 1$ there is no first-order benefit from raising a as $\epsilon = 1$, but there is a first-order loss from the distortion thus caused as $\epsilon_{\epsilon_h}, \epsilon_S > 0$. Therefore, it is always optimal to choose an $a^\star \in (0,1)$. □

More detailed comparative statics may easily be derived. If demand is very elastic, even for low values of a, it will be optimal to choose a low. If the isoeffort curve is very elastic even for low values of a, then it is optimal to chose high values of a. In a broad sense, our core model can be seen as providing structure to these relative elasticities and tying them thereby to statistical properties of the distribution of innovations.

4.3 Sales manipulation

In line with the literature on advance market commitments and output subsidy policies, we have assumed so far that sales are verifiable by the government. At the very least, such verifiability requires the existence of either exclusive resale outlets with trustworthy record keeping or an encryption device preventing inflated sale claims. Yet, even if actual sales are verifiable, the entrepreneur may still want to manipulate sales figures by asking friends and affiliates to purchase on her behalf. Such manipulation may provide a separate, but closely related, rationale for above-cost pricing.

Allowing for such manipulation, the scheme $t(q,p)$ is non-manipulable if such purchases are not profitable. That is, for any (q,p) in the equilibrium support, $t(q+\Delta,p) - p\Delta$ must be maximized in the range $[0,\infty)$ at $\Delta = 0$. If t is differentiable, this adds the following non-manipulability constraint:

$$t_q(q,p) \leq p$$

The non-manipulability constraint is inconsistent with low mark-ups. For instance, a quantity-responsive prize system (i.e. $p = 0$, $t(q)$ increasing) is no longer feasible, let alone approximately optimal. Specializing to the Cobb-Douglas isoreward curves we use above, this non-

manipulability constraint can be rewritten as

$$\frac{p}{q} \geq \left(\frac{t'\left(q^{1-\alpha}p^{\alpha}\right)(1-\alpha)}{q^{1-\alpha}p^{\alpha}} \right)^{\frac{1}{1-2\alpha}}.$$

A low price policy leads to this being violated for a wide range of x values, unless the transfer policy is highly unresponsive. In this setting, screening even the size of the market requires market power. Simple IP ($t = qp$) satisfies the non-manipulability constraint everywhere with equality. Indeed simple IP is optimal in the class of Cobb-Douglas isoreward policies.

The optimal scheme under the non-manipulability constraint lies outside the scope of this paper, but we conjecture that the optimal α is higher than in its absence and that simple IP is optimal with sales manipulation under weaker conditions than $V_1 \to \infty$.

4.4 Multiple price observations

For analytical convenience, we have presumed that the social planner does not require the entrepreneur to randomize over prices. Randomization could facilitate sorting at a lower distortion cost (Brynjolfsson and Zhang, 2006) by improving the social planner's information about the demand curve without forcing all consumers to pay higher prices.

There are several ways in which multiple prices might emerge: pure stochastic pricing (the entrepreneur may randomize prices and rewards may be contingent on the realized outcome, as well as on the randomization chosen), geographically differentiated prices and intertemporally differentiated prices. The latter two forms of price variability require the absence of consumer arbitrage: geographical price dispersion is not sustainable if resale across territories is feasible.[31] Similarly, for a durable good, Coasian arbitrage limits the forms of intertemporal discrimination that can be achieved.

But even in the absence of consumer arbitrage, price variability in general would not obviate the need for pricing in the upper part of the demand curve sufficiently long, with sufficient probability. A general treatment of this point lies outside the scope of the paper and we content ourselves with a simple illustration.

[31] Of course, the incentives for arbitrage are greater for higher prices and smaller geographic area (or period of time) over which they are applied, further emphasizing the difficulty of obtaining "free" information about demand.

An example. Consumers have willingness to pay for the innovation equal to v_H (fraction m) or v_L (fraction $1-m$), where $v_H > v_L$. Let $\overline{v} \equiv m v_H + (1-m) v_L$ denote the social surplus at ex-post-efficient pricing ($p \leq v_L$). There are two types of innovations: minor (c_1, σ_1, m_1) in proportion f_1, and major (c_2, σ_2, m_2) in proportion f_2, with $f_1 + f_2 = 1$. One has $\sigma_1 \geq \sigma_2$, $m_1 < m_2$ and, with obvious notation, $\overline{v}_2 - c_2 > 0 > \overline{v}_1 - c_1$ so the social planner would like to screen in major innovations and out minor ones.

Consider first *intertemporal price variations.* One might intuit that the social planner should mandate high prices for a short while in order to learn about the demand curve and then, conditional on the observed demand having passed the market test, award a prize to the inventor and have the innovation turned to the public domain. This reasoning, however, ignores two related features: First, for exogenous reasons, demand may develop faster or slower than thought and so early demand observations will be noisy. Second, the entrepreneur may use marketing and more generally non-price effort to frontload realizations of demand. We formalize the latter possibility in a straightforward way: let time run from 0 to 1 and there be no discounting for notational simplicity. Let $z_k^t = 1$ if type $k \in \{1, 2\}$ charges $v_L < p_t \leq v_H$ and $z_k^t = 0$ if $p_t \leq v_L$. To formalize the feasibility of moving demand across time, we allow the entrepreneur to choose demand size path $\{\sigma_k^t\}_{t \in [0,1]}$ subject to the constraint $\int_0^1 \sigma_k^t dt \leq \sigma_k$ (total demand across time is constant).

To mimic the major innovation, the minor innovation must choose $\sigma_1^t = \sigma_2^t m_2 / m_1$ for all t such that $z_2^t = 1$ and $\sigma_1^t = \sigma_2^t$ otherwise. Letting $z \equiv \left(\int_0^1 \sigma_2^t z_2^t dt \right) / \sigma_2$ denote the fraction of the demand for time at which the major innovation leads to a distortion, screening out the minor innovation requires that the latter be forced to frontload sufficient demand that the planner later finds out:

$$\sigma_1 < \sigma_2 \left(1 + \frac{m_2 - m_1}{m_1} z \right)$$

This condition puts a lower bound on the fraction of time z over which a high price must be charged. Similarly, rather than moving demand across time, the entrepreneur may *manipulate sales*, as studied in the previous subsection. Let $X \in [0, 1]$ denote the fraction of time for which $p_t = v_H$ and $1 - X$ the fraction of time for which $p_t = v_L$. To make up for the sale shortage $(m_2 - m_1)$ when $p_t = v_H$, the producer of the minor innovation must spend $(m_2 - m_1) v_H$,

and so incentive compatibility requires that

$$c_2 - c_1 \leq (m_2 - m_1)v_H X$$

Again, this sets a lower bound on the fraction of time for which the monopoly price v_H must be charged if sorting is to occur.

Finally, let us use this example to illustrate that *random schemes* similarly require a sufficient probability of a high price. Under risk neutrality, this would not be the case: it would suffice to implement a high price with a small probability, and then give a very high transfer in that state of nature if demand turns out to be sufficient. There are however two limits to this argument. The first is that the random scheme is highly manipulable in the sense of Subsection 4.3: a small probability of teasing out the demand curve requires a very high transfer in order to make up for the R&D cost c_2. And so the cost of manipulation $(m_2 - m_1)v_H$ is lower than this transfer. Second, entrepreneur risk aversion also constrains the use of random schemes.[32]

The bottom line is that the social planner must trade off the reduced cost of sorting and the cost of randomization when contemplating the use of multiple price observations (temporal, geographic, randomized). Furthermore, these mechanisms for attaining multiple price observations would likely be fragile to the introduction of additional dimensions of heterogeneity (in the duration or geographical distribution of demand for the product or in the risk aversion of the entrepreneur). Finally, the possibilities for manipulation will likely grow with more sophisticated mechanisms, potentially making consumer surplus unverifiable in the absence of market power even if it is somewhat observable. Thus, the basic insights of this paper seem robust. However, the appropriate role of more sophisticated schemes to partially mitigate the costs of market power, especially in the presence of richer heterogeneity, is an exciting direction for future research.

[32] Suppose for example that the entrepreneur's utility is $u(T) = T$ for $0 \leq T \leq \overline{T}$ and $u(T) = \overline{T}$ for $t \geq \overline{T}$, where $\overline{T} > c_2$. Let X denote the probability that $p = v_H$ and $(1 - X)$ the probability that $p = v_L$. It is optimal to set $t = \overline{T}$ when $p = v_H$ and demand is m_2. Let $\underline{T}$ denote the reward when $p = v_L$. Then $X\overline{T} + (1 - X)\underline{T} \geq c_2$ for the major innovation to be created. On the other hand sorting requires that $(1 - X)\underline{T} \leq c_1$. And so

$$X \geq \frac{c_2 - c_1}{\overline{T}}.$$

Again the probability of monopoly pricing cannot be too small.

4.5 Distribution, externalities and residual uncertainty

Finally, let us discuss informally some extensions; formal results can be found in the Appendix.

a) *Distributional concerns*

In many applications, transfers to entrepreneurs should not be viewed as socially neutral. A simple way to incorporate this into our model is to assume the social planner puts a weight of only $\lambda \in [0, 1)$ on the welfare of the entrepreneurs compared to that of the government and consumers. Then her program is exactly as in expression (1), but with $\lambda c + (1 - \lambda)T$ replacing c, and is subject to the same incentive compatibility conditions.

Analyzing $a^\star$ is also straightforward. The two primary effects we emphasized above, sorting and ex-post distortion, persist. The additional element added is the effect that an increase in a has on the rewards given to marginal compared to infra-marginal entrepreneurs (Spence, 1975).

b) *Externalities*

Innovations often "stand on the shoulders" of previous innovations. On the other hand, many innovative products compete with existing products. If the value of such spillovers is not a direct function of the demand parameters, we need to consider further heterogeneity; the analysis is then beyond the scope of direct extension of our model. However in the, not unreasonable, simple cases when they are proportional either to the potential or actual net surplus created by an innovation (σm and $\sigma m S(a)$ respectively), a fairly straightforward analysis is possible. All of our theorems can be extended to this context. More explicit and detailed modeling of complementary and substitutable innovations in our framework remains an important direction for future research, however, for the light it might shed on competition policy.

c) *Residual uncertainty*

We assumed above that the entrepreneur knows, at the point of undertaking the innovation, the exact demand her product will face. This seems unrealistic for two reasons. First, the entrepreneur likely learns a significant amount about the demand from the time of undertaking the innovation to when she brings it to market (e.g. through market research). Before sinking c, the entrepreneur knows only a signal of her eventual demand. Such uncertainty does not

significantly change our approach. By the time the product comes to market, the entrepreneur will know the parameters and thus still charge $a(\alpha)$. Additionally her expected rewards will be $E\left[\tau\left(\sigma^{1-\alpha}m^{\alpha}\right)\right]$ and thus nearly the same analysis will follow. However, it will be the *residual* private information about demand that will be relevant, as heterogeneity in x that the entrepreneur cannot predict will not affect her incentives to develop products.

Second, even at the point of bringing the innovation to market the entrepreneur may be uncertain about the demand that will be realized once she chooses a price. For example, suppose that given (σ, m) realized demand is given by a distribution $H\left(q|\sigma Q(p/m)\right)$. Again this has little impact: changing σ in a multiplicative manner has no effect on the elasticity of demand as a function of a and thus the entrepreneur will charge the same $a(\alpha)$. Similarly, entrepreneurs with a higher expected $\frac{q}{p}$ (a higher x) will generate more social value and expect higher rewards when α is higher. Again, the appropriate expectations must be taken, but otherwise the analysis is unchanged.

5 Other Applications

This section demonstrates how the techniques developed above apply to a range of problems more general than IP or even the optimal distortion of prices that we focus on above.

5.1 Platforms

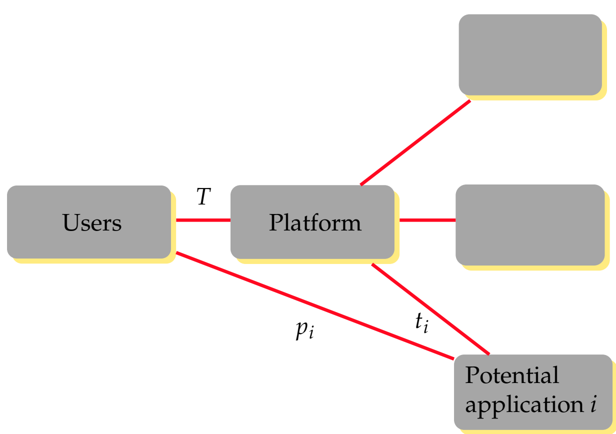

Figure 6: Application development incentives in a two-sided market

A two-sided platform (as shown in Figure 6 above), such as an operating system, must attract multiple sides of a market, say end-users and application developers.[33] A key decision such platforms face is how to "regulate" the relationship between application developers and end-users. Should it let application developers charge a monopoly price ($p_i = m_i$) for their application, effectively giving them IP? Or should it bundle the applications, free to the consumers, with the platform, paying the developer a prize-like up-front fee?

The analogy between a platform and a social planner can be made more formal and is particularly precise when end-users differ in an idiosyncratic parameter of taste for the platform and there is a large number of applications.[34] It can then be shown that the platform aims at maximizing total surplus (end-user gross surplus) minus the rewards given to developers. Therefore, the extension of the paper's analysis to the case where the social planner cares nothing for developers ($\lambda = 0$ in Subsection 4.5 above) carries through without any modification.

This analysis qualifies the classic result of Armstrong (1999), Bakos and Brynjolfsson (1999) and Geng et al. (2005), who showed that a platform with full knowledge about the quality of a large number of independent applications optimally bundles them with access to the platform. Bundling is no longer optimal when the platform is unsure which applications bring value to the end-user; thus the laissez-faire (monopoly pricing) policy for applications is nearly optimal if a few "killer apps" make most of the platform's value and the elasticity of innovation supply is high.

[33] See Rysman (2009) for a recent survey.

[34] Let $i \in [0,1]$ denote the set of potential innovations, $\delta_i = 1$ if the innovation is developed and $\delta_i = 0$ otherwise. Consumers are indexed by k, uniform in $[0,1]$. Consumer k's net payoff is $\int_0^1 V_i(p_i\,;\sigma_i\,,m_i)\delta_i di - T - \xi_k$, where ξ_k is distributed according to $H(\cdot)$. Application developer i introduces the innovation if and only if $t_i \geq c_i$. The platform's profit is then $\left[T + \int_0^1 \left[\pi_i(p_i\,;\sigma_i\,,m_i) - t_i\right]\delta_i di\right]\left[1 - H\left(t - \int_0^1 V_i(p_i\,;\sigma_i\,,m_i)\delta_i di\right)\right]$, or after a change in variables and using $S_i \equiv V_i + \pi_i$, $\left[\widehat{T} + \int_0^1 \left[S_i(p_i\,;\sigma_i\,,m_i) - t_i\right]\delta_i di\right]\left[1 - H\left(\widehat{T}\right)\right]$. Thus the platform first maximizes $\int_0^1 \left[S_i(p_i\,;\sigma_i\,,m_i) - t_i\right]\delta_i di$ and then chooses the surplus-adjusted price $\widehat{T}$. Put differently, the planner behaves exactly as a social planner with strong redistributive concerns ($\lambda = 0$).

Of course this assumption is likely rarely satisfied exactly. Yet the fact that most first-party applications are bundled with most platforms (e.g. Apple's basic apps) suggests that it is close to correct; at least, an incentive problem towards application developers likely accounts for a large part of the asymmetry between the pricing of internal and external applications. Furthermore, despite the fact that little that is known about it, it seems likely that even optimal price discrimination (constrained to a single tariff, say) would involve charging only a small fraction of the monopoly optimal price for an individual product as the platform would seek only to extract that part of surplus associated with income.

In practice we observe online music stores, such as Apple's iTunes store and Rhapsody, using price caps, fixed or proportional licensing fees and low (or purely bundled) prices to consumer. In our model, this is rationalized by the fact that only a small fraction of the total revenue for a song comes from any of these individual stores and thus the elasticity of innovation supply with respect to a change in their revenue from one of these stores is small. Contrast this with the Apple App store, which allows total pricing freedom to App developers and gives them a share of revenue. Given that applications are developed almost exclusively for the iPhone or another individual platform, the elasticity of innovation supply is likely much higher and thus there is more burden on Apple to sort out which applications deserve greater rewards by using market power. Section 6 discusses a more detailed application to this market.

5.2 Intrapreneurship

A similar tradeoff arises when the "application developer" is an internal division and the platform wants to provide it with incentives to develop useful innovations. A division manager is endowed with a project for a new product. If authorized, the manager will enjoy a private benefit or cost, and headquarters will observe price and sales, but not the resulting spillovers. Spillovers can be traced to the existence of either repeat purchases (e.g., due to lock-in) or the sale of complementary products; the spillovers will benefit the conglomerate, but not (at least not fully) the division manager.

The unobserved profit from spillovers is the counterpart of the unobserved net consumer surplus in our main analysis. Assume (reasonably) that spillovers are larger when consumer's willingnesses to pay for the division's product are higher. Spillovers are then larger when the demand curve is less price sensitive. This setting is perfectly analogous to the platform setting discussed in the previous subsection, except that the relevant incentives are internal to the firm. Our paper thus provides a rationale for allowing intrapreneurs to receive rewards proportional to the profits generated *on their specific product*, even though this causes multi-marginalization problems for the firm, helping to address recent debates about such incentive schemes (Hunt and Lerner, 1995).

5.3 Public infrastructure

The traditional approach to building a new highway or new train tracks is to enter a procurement contract with an infrastructure builder, and then to turn to a separate infrastructure operator to manage it; the infrastructure may then be accessed at a relatively low price. By contrast, under a public-private partnership (PPP), the builder of the new infrastructure derives substantial revenue from its later operations. Such a long-term approach links builder compensation to actual revenue derived from the end-user and is often vaunted as a way to screen out white elephants.

Purely public projects may be seen in a similar light if we consider the limiting case of observed costs and a political entrepreneur seeking funding from the public purse for a project, motivated to have a successful project by the prestige, career enhancement or other non-pecuniary benefit it brings. In Appendix D we show that in this case an optimal scheme (*the* optimal scheme if transfers are – at least slightly – socially costly) is to reimburse the known cost provided that the innovation satisfy a minimum scoring rule: $q^{1-\alpha}p^{\alpha} \geq k$, and nothing if this score is not reached. When α is close to $\frac{1}{2}$, this minimum score is equivalent to a minimum profit level. Thus we may interpret policy regimes under which politicians or bureaucrats are under more pressure to recover costs by charging higher prices as high α regimes and ones under which the emphasis is on consumer surplus and subsidies are required as corresponding to low α. The role of the inequality and elasticity (now in the threshold level rather than reward space) of innovation supply is unchanged.

6 Empirical Calibration

Weingarten (2011) calibrates our model in the context of the smartphone applications markets discussed in Subsection 5.1. While the validity of his results are not the focus of this paper, his approach illustrates a method for empirically calibrating our model.

In particular he studies Apple's iPhone (and more recently iPad) Apps market, in which developers receive 70% of the revenues derived from a free choice of their own price. Price data and a wide range of other application characteristics are scraped directly from online. Because quantity data (the number of phones on which the application is running) is proprietary,

Weingarten uses the number of ratings an application has received as a proxy (only application owners may give ratings), which he shows is fairly accurate in various ways.

Then, following Acemoglu and Linn (2004), he measures the elasticity of innovation supply by considering an event which (exogenously) raised the demand for, and thus revenue of, applications (namely the introduction of the iPhone on Verizon) and observing the increase in the number of applications created. His point estimate of elasticity of innovation supply is .95.[35]

To measure the inequality of innovation supply, Weingarten assumes the set of marginal applications has the same residual log-variance as the whole set of applications, to increase his data set. Letting x_i be the ratio of an individual product price to the quantity proxy he uses and π_i be their product while $\mathbf{I}_i$ is a vector containing all other covariates he collects (such as application category, size, average rating, etc. which are available to the platform independent of market prices), he runs various specifications of the regression

$$\log(x_i) = \gamma\left(\pi_i, \mathbf{I}_i\right) + \epsilon_i, \tag{9}$$

weighted by π_i^2 to recover the residuals ϵ_i. In his most robust (to cross-validation) and thus preferred specification he measures the inequality of innovation supply to be 5.5. Together, these yields an estimate of V_1, $\hat{V}_1 = 5.23$, assuming no correlation across isoprofit curves between the elasticity and inequality of innovation supply.

He plugs this value into equation (7), assuming a Bulow and Pfleiderer (1983) constant pass-through demand, yielding $\alpha^\star \approx \frac{1}{2} - \frac{\rho}{(\rho+1)\hat{V}_1}$ or $a\left(\alpha^\star\right) \approx 1 - \frac{4\rho^2}{(\rho+1)\hat{V}_1}$, where ρ is the pass-through rate. While he experiments with a variety of pass-through rates, we focus on $\rho = .5$ (linear demand) as this is most conservative in deviating from current practice (leads to highest α), as seems sensible. This yields $\hat{\alpha}^\star = .44$ or $\hat{a}^\star = .87$.

Weingarten then goes on to use this estimate to calculate isoreward curves and applies

[35] A striking result based on this comes from using the "distributional concerns" version of our model described in Subsection 4.5 in which the platform has no concern for entrepreneur's profits. Solving for optimal rewards under linear demand and a constant elasticity of .95 for innovation supply yields that, constrained to $\alpha = \frac{1}{2}$, the optimal reward to entrepreneurs is 73% of profits, strikingly close to what Apple (and other application stores such as the Android Market) provide.

equation (3) to compute optimal rewards to each isoreward curve by running the regression

$$\pi_i = \gamma\left(p_i^{.44}\hat{q}_i^{.56}, \mathbf{I}_i\right) + \epsilon_i \tag{10}$$

and using the resulting model to assign rewards to innovations. He then shows how this system would make an extremely popular but relatively low-priced game, Angry Birds, earn about 10% higher revenue while an expensive and less well-selling game, Infinity Blade, would receive only 64% of the revenue it was previously receiving. Consumer prices of each would be expected to fall by 13%.

Of course, the resulting rule should be interpreted very cautiously: Weingarten's data suffers from a lack of access to proprietary information, and his analysis could be improved by incorporating factors such as externalities across applications, residual uncertainty, etc. However, his exercise provides a method for applying our results to the structural estimation of optimal innovation policy.

7 Conclusion

This paper aspires to make three contributions. First, in terms of modeling, we propose the intuitive stretch parameterization of demand to allow a smooth trade-off between quantity-dependent prizes and IP. Second, on a technical level, we develop a simple isoreward approach to analyzing multidimensional screening problems: parameterize policies based on the shape of the isoreward curves they create and then solve separately for the structure and level of rewards using the envelope theorem. Finally, and substantively, we show how the inequality and elasticity of innovation supply are tightly connected to the optimality of market power as a reward for innovation, making precise and empirically testable the conjectures of classical thinkers.

Needless to say, our framework requires further elaboration in order to help fashion policy. Furthermore, given the foundational role that many of the issues we address in this paper play in several areas of price theory, we believe that our work opens a number of promising directions for future research. First, our general formula ought to be calibrated in specific industries

beyond Weingarten's smartphone applications example. Second, several extensions would test the robustness of our insights: the demand function could be generalized beyond the stretch parameterization and the optimal structure, not just level, of market power could be more fully analyzed; although we have presented arguments that make us hopeful that our insights will carry over, only a rigorous analysis can vindicate such a claim. Along the same lines, a general analysis of demand uncertainty (under entrepreneur limited liability or risk aversion), as well as richer private information (duration, marketing, price discrimination, marginal costs) and richer instruments for screening demand would be welcome. In particular, the latter would be crucial in allowing our analysis to shed light on the design of patent length and breadth. Finally, the extension of our techniques to accommodate R&D races, licensing competition and cumulative innovation stands high on the research agenda. For instance, our techniques are likely to be helpful in analyzing the validity of the notion (central to antitrust doctrine) that acquired market power may be maintained but should not be extended. The acceptability of vertical foreclosure practices is often felt to depend on the extent of innovation/investment; while market power gained through horizontal mergers and predation are frowned upon in the absence of substantial efficiency gains. Formal analyses would be useful to help guide policy in these matters.

Appendices

A Supply, demand and optimal transfers

This appendix solves for the monotonicity-relaxed optimal $\tau^{\star\star}(k;\alpha)$ function and conditions under which this function is, in fact, monotone. The change of variables from (σ, m) to (k, x) requires transforming the distribution of values according to

$$\tilde{f}(k,x,c;\alpha) \equiv f\left(kx^{-\alpha}, kx^{1-\alpha}, c\right) kx^{-2\alpha}.$$

We can then rewrite the social planner's problem, with the substitution and by switching the order of integration, as

$$\max_{\tau(\cdot)} \int_k \int_{c=0}^{\tau(k)} \int_x \left(k^2 x^{1-2\alpha} S\left(a(\alpha)\right) - c\right) \tilde{f}(k,x,c;\alpha) dxdcdk \text{ s.t. } \tau \text{ increasing.} \tag{11}$$

Consider the marginal cumulative distribution of innovations in terms of their cost of creation c, integrating out over x, lying along a particular isoreward curve, $F(\tau;k,a) \equiv \frac{\int_{c=0}^{\tau}\int_x \tilde{f}(k,x,c;\tilde{\alpha})dxdc}{\int_{c=0}^{\infty}\int_x \tilde{f}(k,x,c;\tilde{\alpha})dxdc}$. This is the fraction of innovations that will be created if a reward τ is offered along this curve.

$\Sigma(r;k,\alpha) \equiv F^{-1}(r;k,\alpha)$ is then the (clearly increasing) inverse *supply of innovations* lying along isoreward curve k, namely the reward necessary to induce a fraction r of innovations lying along that curve to be created.

We can similarly define the social inverse *demand for innovations*. First, let us define the average value of an innovation lying on isoreward curve k with cost c by

$$\overline{S}(c;k,\alpha) \equiv k^2 S\left(a(\alpha)\right) E_{x,\tilde{f}}\left[x^{1-2\alpha}|k,c\right] \equiv k^2 S\left(a(\alpha)\right) \cdot \frac{\int_x x^{1-2\alpha}\tilde{f}(k,x,c;\alpha)dx}{\int_x \tilde{f}(k,x,c;\alpha)dx} \tag{12}$$

Then $D(r;k,\alpha) \equiv \overline{S}\left(\Sigma(r;k,\alpha);k,\alpha\right)$ is the average value of a marginal innovation lying along isoreward curve k, given that a fraction r of innovations lying on that isoreward curve have been created.

The optimal reward along the isoreward curve is the intersection of the supply and demand curves for innovations, assuming these intersect only once. To the extent that $\alpha < \frac{1}{2}$, D will slope downwards if x varies negatively with c given k and upwards in the reverse case, in a sense made more rigorous below. Both effects are dampened for large α. So long as D does not increase too quickly, supply and demand will have a unique intersection corresponding to the optimal quantity of innovations and reward along k given α. These optimal rewards ignore the monotonicity constraint that higher-k isoreward curves must receive higher rewards. When α is relatively low, if k is sufficiently negatively affiliated with x given c, the optimal reward unconstrained by monotonicity may be decreasing in k; ruling out such strong negative affiliation ensures that the relaxed solution is in fact optimal.

Proposition 3: *Suppose that for all k, c and a fixed α,*

$$Cov_{x,\tilde{f}}\left[x^{1-2\alpha}, \frac{\partial \log(f)}{\partial c}\middle| k,c\right] \leq \frac{1}{k^2 S(a)} \tag{13}$$

and

$$2E_{x,\tilde{f}}\left[x^{1-2\alpha}|k,c\right] \geq -k\,Cov_{x,\tilde{f}}\left[x^{1-2\alpha}, \frac{\partial \log(\tilde{f})}{\partial k}\middle| k,c\right] \tag{14}$$

Then the optimal reward function $\tau^\star(k;\alpha)$, given α, is defined for each k by the unique value at which $D(\cdot;k,\alpha)$ and $S(\cdot;k,\alpha)$ intersect if $D(0;k,\alpha) > S(0;k,\alpha)$, $D(1;k,\alpha) < S(1;k,\alpha)$ for all k, a or by the appropriate boundary solutions otherwise.

These conditions are intuitive extensions of the classic Mirrlees (1979)-Rogerson (1985) monotone likelihood ratio property that ensures validity of first-order approaches in classical, single-dimensional screening problems. If as x increases $\frac{\partial \log(f)}{\partial c}$ also increases, this exactly represents x having a strong monotone likelihood ratio relationship with c. Thus, condition (13) can be viewed as stating that c and x are not "too" affiliated (Milgrom and Weber, 1982), while condition (14) can be seen as stating that x is not too negatively affiliated with k. Note that ironing-free approach is always valid for sufficiently high values of α. This is illustrated by Figure 7, which shows $\tau^{\star\star}(k;\alpha)$ for the simulation we describe in the text. For high values of α it is monotone increasing, but must be ironed for low values of α.

Proof. The first-order condition for the monotonicity-unconstrained optimum is exactly that $\overline{S}(\tau;k,\alpha) = \tau$. This condition is sufficient for the monotonicity-unconstrained optimum if the first-order derivative with respect to t, $\overline{S} - \tau$, is monotone decreasing in τ (Guesnerie and Laffont, 1984). Because these expressions are clearly differentiable, given the smoothness

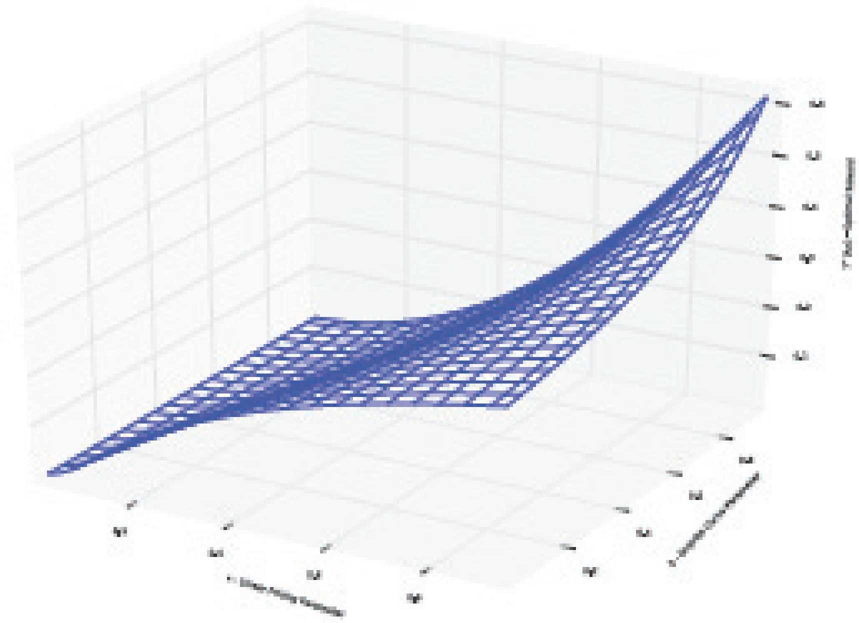

Figure 7: $\tau^{\star\star}\left(k; a^{-1}(a)\right)$ in our simulation as described in Subsection 3.5

assumed, this is equivalent to

$$\frac{\partial\overline{S}}{\partial c} \leq 1 \iff k^2 S(a)\frac{\partial E_{x,\tilde{f}}\left[x^{1-2\alpha}\middle| k, c\right]}{\partial c} \leq 1 \iff \mathrm{Cov}_{x,\tilde{f}}\left[x^{1-2\alpha}, \frac{\partial \log\left(\tilde{f}\right)}{\partial c}\right] \leq \frac{1}{k^2 S(a)}.$$

Because $\tilde{f}$ differs from f as a function of c only by a multiplicative factor, we can replace $\tilde{f}$ with f and obtain inequality (13) in the proposition. Furthermore an upward shift in $\overline{V}$ must then increase this optimal transfer (Milgrom and Shannon, 1994). Thus we are guaranteed that $\tau^{\star\star}$ will be monotone so long as $\overline{V}$ increases in k. Again by differentiability this is equivalent to (taking logs)

$$\frac{\partial\overline{S}}{\partial k} \geq 0 \iff \frac{2}{k} + \frac{\mathrm{Cov}_{x,\tilde{f}}\left[x^{1-2\alpha}, \frac{\partial \log(\tilde{f})}{\partial k}\middle| k, c\right]}{E_{x,\tilde{f}}\left[x^{1-2\alpha}\middle| k, c\right]} \geq 0$$

which simplifies to condition (14). Thus if this condition also holds the $\tau^{\star\star}$ derived from supply and demand is in fact $\tau^{\star}$. $\square$

As discussed in Subsection 2.2, a (grossly) sufficient condition to ensure that when $\alpha = 0$, $\tau^{\star}$ is not flat and therefore that $\alpha^{\star} \in \left(0, \frac{1}{2}\right)$ is obedience of inequality (14) when $\alpha = 0$. At $\alpha = 0$ this condition simplifies to (at each (σ, c)) $2E_{m,f}\left[m|\sigma, c\right] \geq -\sigma \mathrm{Cov}_{m,f}\left[m, \frac{\partial \log(f)}{\partial \sigma}\middle| \sigma, c\right]$. That is, σ is not too negatively affiliated with m.

B Envelope Theorem

The application of the general Milgrom and Segal (2002) envelope theorem at $\hat{\alpha}$ requires equidifferentiability *across feasible reward functions* of social welfare with respect to α, holding fixed the reward function, and continuity of the derivative *given the (locally) optimal choice of the reward function* , both at $\hat{\alpha}$. This appendix therefore establishes two sublemmata prior to the proof of Lemma 1:

1. First, we establish equidifferentiability in Sublemma 1 by deriving a general expression for the first-order derivative of social welfare holding fixed the reward function that is valid even when $\tau^\star$ is not differentiable. This expression, for any reward function, can be broken into two pieces, one along the boundary and one on the interior. The second always converges "quickly" because the derivative is proportional to S, which is constant across τ. The first does as well, because the mass along the boundary is bounded above by monotonicity and the rate of movement across the boundary by uniform bounds we establish on the covariance of $x^{1-2\alpha}$ and $\log(x)$.
2. Second, we establish continuity of the derivative in Sublemma 2 from the continuity of the optimal reward boundary; this does not require that the optimal rewards, as a function of k, be continuous but rather that the curve dividing innovations that are created and those that are not continuously deform as α changes. Combined with the smoothness of $\tilde{f}$ both in its arguments and in α suffices to establish continuity of the derivative.

In what follows we will use the (slightly abusive of our earlier) notation $W(\alpha;\tau(\cdot))$ to denote the social welfare when pricing policy is α and the transfers assigned to each k (given α) are $\tau(k)$:

$$W(\alpha;\tau(\cdot)) \equiv \int_\sigma \int_m \int_{c=0}^{\tau\left(\sigma^{1-\alpha} m^\alpha\right)} \left(\sigma m S\left(a(\alpha)\right) - c\right) f(\sigma,m,c) dc dm d\sigma$$

Sublemma 1: *The class of functions $W(\alpha;\tau(\cdot))$ are equidifferentiable in α across all monotone increasing $\tau(\cdot)$ at each $\alpha \in \left[0,\frac{1}{2}\right]$.*

Proof.

$$W'(\alpha;\tau(\cdot)) = \lim_{\delta\to 0} \frac{\int_\sigma \int_m \left(\int_{c=0}^{\tau\left(\sigma^{1-\alpha-\delta} m^{\alpha+\delta}\right)} (\sigma m S(a(\alpha+\delta)) - c) f(\sigma,m,c) dc - \int_{c=0}^{\tau\left(\sigma^{1-\alpha} m^{\alpha}\right)} (\sigma m S(a(\alpha)) - c) f(\sigma,m,c) dc \right) dm d\sigma}{\delta} =,$$

letting $S_\delta \equiv S\left(a(\alpha+\delta)\right)$,

$$\lim_{\delta\to 0} \frac{\int_x \int_k \left(\int_{c=0}^{\tau\left(kx^\delta\right)} \left(k^2 x^{1-2\alpha} S_\delta - c\right) \tilde{f}(k,x,c) dc - \int_{c=0}^{\tilde{t}(k)} \left(k^2 x^{1-2\alpha} S - c\right) \tilde{f}(k,x,c) dc \right) dk dx}{\delta} =,$$

dropping the arguments where possible,

$$\lim_{\delta\to 0} \frac{\int_x \int_k \left(\int_{c=\tau(k)}^{\tau\left(kx^\delta\right)} \left(k^2 x^{1-2\alpha} S_\delta - c\right) \tilde{f} dc + \int_{c=0}^{\tau(k)} k^2 x^{1-2\alpha} \left(S_\delta - S\right) \tilde{f} dc \right) dk dx}{\delta} =$$

$$\lim_{\delta\to 0}\frac{\int_x\int_k\int_{c=\tau(k)}^{\tau\left(kx^{\delta}\right)}\left(k^2x^{1-2\alpha}S_\delta-c\right)\tilde{f}dcdkdx}{\delta}-\frac{Q\alpha}{2(1-\alpha)^3\epsilon'}\int_k k^2\int_x\int_{c=0}^{\tau(k)}x^{1-2\alpha}\tilde{f}dcdkdx \tag{15}$$

by differentiability of S and ϵ and the argument given in the proof of Corollary 1. The second term is as we want it, so we focus on the first. We can break it into the sum of three terms: an integral over the region where $x>1$, the point where $x=1$ and the integral over the region where $x<1$. The second of these is identically zero as the integral over c always runs over a degenerate region. The third can be transformed in a manner parallel to that of the first, so we focus on the first in most of the remaining proof; thus the quantity we will analyze is

$$\lim_{\delta\to 0}\frac{\int_{x=1_+}^{\infty}\int_k\int_{c=\tau(k)}^{\tau\left(kx^{\delta}\right)}\left(k^2x^{1-2\alpha}S_\delta-c\right)\tilde{f}dcdkdx}{\delta} \tag{16}$$

The integral over k, which runs from 0 to ∞ is just the limit of an integral running from bounds $\underline{k}(\delta,z)$ to $\overline{k}(\delta,z)$ as $z\to\infty$ so long as $\lim_{z\to\infty}\underline{k}(\delta,z)=0$ and $\lim_{z\to\infty}\overline{k}(\delta,z)=\infty$. In particular let $\underline{k}(\delta,z)\equiv\frac{1}{\overline{k}(\delta,z)}$ and $\overline{k}(\delta,z)\equiv x^{\delta\left\lceil\frac{\log(z)}{\delta\log(x)}\right\rceil}$. Clearly $\lim_{z\to\infty}\overline{k}(\delta,z)=\infty$ and thus $\lim_{z\to\infty}\underline{k}(\delta,z)=0$. Then we can re-write expression (16) as

$$\lim_{\delta\to 0}\lim_{z\to\infty}\frac{\int_{x=1_+}^{\infty}\int_{k=\underline{k}(\delta,z)}^{k=\overline{k}(\delta,z)}\int_{c=t(k)}^{t\left(kx^{\delta}\right)}\left(k^2x^{1-2\alpha}S_\delta-c\right)\tilde{f}dcdkdx}{\delta}. \tag{17}$$

Furthermore, we may approximate the integral over k by a Riemann sum, the intervals of which are given by

$$\left(\underline{k}(\delta,z),\underline{k}(\delta,z)x_\delta\right),\left(\underline{k}(\delta,z)x_\delta,\underline{k}(\delta,z)x_\delta^2\right),\ldots,\left(\overline{k}(\delta,z)x_\delta^{-1},\overline{k}(\delta,z)\right)$$

Let us number these intervals starting from 1 as both down and up: "up" interval 1 is $\left(1,x^\delta\right)$ and "down" interval 1 is $\left(x^{-\delta},1\right)$. There are then $\left\lceil\frac{\log(z)}{\log(x_\delta)}\right\rceil$ of each up and down intervals and the length of the ith up interval is $x^{\delta(i-1)}\left(x^\delta-1\right)$ while of the ith down interval is $x^{-i\delta}\left(x_\delta-1\right)$. Thus note that the upper bound on the length of any interval, $\overline{k}(\delta,z)\left(x^\delta-1\right)\to 0$ as $\delta\to 0$, so that any choice of a point within the interval at which to evaluate the Riemann sum will lead to a sum converging to the integral as $\delta\to 0$. Thus, if we evaluate the Riemann sum at the bottom of the interval, the expression (17) becomes the limit as δ becomes small and z becomes large of

$$\frac{\left(x^\delta-1\right)\int_{x=1_+}^{\infty}\left(\sum_{N=1}^{\left\lceil\frac{\log(z)}{\delta\log(x)}\right\rceil}x^{-\delta N}\int_{c=\tau\left(x^{-\delta N}\right)}^{\tau\left(x^{-\delta(N+1)}\right)}\left(x^{1-2(\alpha+\delta N)}S_\delta-c\right)\tilde{f}dc+x^{\delta N}\int_{c=\tau\left(x^{\delta(N-1)}\right)}^{\tau\left(x^{\delta N}\right)}\left(x^{1-2[\alpha-\delta(N-1)]}S_\delta-c\right)\tilde{f}dc\right)dx}{\delta}=$$

$$\lim_{\delta\to 0}\lim_{z\to\infty}\int_{x=1_+}^{\infty}\frac{x^\delta-1}{\delta}\int_{c=\tau(\underline{k}(\delta,z))}^{\tau\left(\overline{k}(\delta,z)\right)}\tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)\left(\tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)^2x^{1-2\alpha}S_\delta-c\right)\tilde{f}dcdx=$$

where, in the cases where τ^{-1} is not single valued, it is defined to pick out the single value to make this equality correct, and where c is not in the range of t because of a discontinuity, $\tau^{-1}(c)$ is taken to be the unique value of k such that $\tau(k_-)\le c\le\tau(k_+)$. Thus if $\lim_{c\to 0}\tau(c)=\underline{c}>0$

then for all $c < \underline{c}$, $\tau(c) = 0$ as clearly $\tau(0) = 0$.

$$\lim_{\delta\to 0} \int_{x=1_+}^{\infty} \frac{x^{\delta}-1}{\delta} \int_{c=0}^{\overline{c}} \tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)\left(\tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)^2 x^{1-2\alpha} S_{\delta} - c\right)\tilde{f}dcdx \tag{18}$$

The equidifferentiability of the second term of (15) follows from the finiteness of the moments of f and $\tilde{f}$; by the same argument we can, for the purposes of establishing equidifferentiability reduce equation (18) to

$$\lim_{\delta\to 0} \int_{c=0}^{\overline{c}} \tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)\left(\tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)^2 x^{1-2\alpha} - c\right)\tilde{f}dc.$$

We must show this converges uniformly across feasible τ. Clearly $\left|x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor} - c\right| \leq x^{\delta}c$ and thus

$$\left|\int_{c=0}^{\overline{c}} \tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)\left(\tau^{-1}\left(x^{\delta\left\lfloor\frac{\log(c)}{\delta\log(x)}\right\rfloor}\right)^2 x^{1-2\alpha} - c\right)\tilde{f}dc - \int_{c=0}^{\overline{c}} \tau^{-1}(c)\left(\tau^{-1}(c)^2 x^{1-2\alpha} - c\right)\tilde{f}dc\right| <$$

$$\max\left\{\left|x^{1-2\alpha}\int_{c=0}^{\overline{c}}\left[\tau^{-1}\left(cx^{\delta}\right)^3 - \tau^{-1}(c)^3\right]\tilde{f}dc\right|, \left|\int_{c=0}^{\overline{c}} c\left[\tau^{-1}\left(cx^{\delta}\right) - \tau^{-1}(c)\right]\tilde{f}dc\right|\right\}$$

For brevity's sake, we only show that the second of these must converge uniformly, as the argument that the first does follows by the same logic.

$$\left|\int_{c=0}^{\overline{c}} c\left[\tau^{-1}\left(cx^{\delta}\right) - \tau^{-1}(c)\right]\tilde{f}dc\right| = \left|\int_{c=0}^{\overline{c}} c\tau^{-1}\left(cx^{\delta}\right)\tilde{f}dc - \int_{c=0}^{\overline{c}} c\tau^{-1}(c)\tilde{f}dc\right| =,$$

by the change of variables $\tilde{c} = cx_{\delta}$ on the first integral,

$$\left|\int_{\tilde{c}=0}^{x^{\delta}\overline{c}} \tilde{c}\tau^{-1}(\tilde{c})\tilde{f}\left(\cdot,\cdot,\tfrac{\tilde{c}}{x^{\delta}}\right)d\tilde{c} - \int_{c=0}^{\overline{c}} c\tau^{-1}(c)\tilde{f}(\cdot,\cdot,c)dc\right| = \int_{c=\frac{\overline{c}}{x^{\delta}}}^{\overline{c}} c\tau^{-1}(c)\tilde{f}dc + \int_{c=0}^{\frac{\overline{c}}{x^{\delta}}} c\tau^{-1}(c)\left(\tilde{f}\left(\cdot,\cdot,\tfrac{c}{x^{\delta}}\right) - \tilde{f}(\cdot,\cdot,c)\right)dc$$

But both of these clearly converge uniformly across τ as thefirst term is just some upper tail of the kc moment, which is finite, and the second term is, in the limit, just the (bounded by smoothness) partial slope of $\log(\tilde{f})$-weighted value of the kc moment along the τ^{-1} curve given that finite moments imply finite moments along any one-dimensional curve. $\square$

Sublemma 2: $W_1(\alpha;\tau^{\star}(\cdot;\alpha)) \equiv \frac{W(\alpha;\tau^{\star}(\cdot;\alpha))}{\partial\alpha}$ *is continuous on* $\left[0, \frac{1}{2}\right]$ *as a function of* α.

Proof. We wish to show that

$$\lim_{\delta\to 0} W_1(\alpha+\delta, \tau^{\star}(\cdot, \alpha+\delta)) = \lim_{\delta\to 0} W_1(\alpha-\delta, \tau^{\star}(\cdot, \alpha-\delta))$$

It is clearly from the reasoning in the proof of Sublemma 1 that $W_1(\alpha, \tau(\cdot))$ is continuous in α for any $\tau(\cdot)$ so it suffices to show that

$$\lim_{\delta\to 0} W_1(\alpha, \tau^{\star}(\cdot, \alpha+\delta)) = \lim_{\delta\to 0} W_1(a, \tau^{\star}(\cdot, \alpha-\delta))$$

Now, abbreviating $W\left(x, \tau^\star(\cdot, y)\right)$ to $W(x, y)$, this is equivalent to:

$$\lim_{\delta\to 0}\left|\lim_{\eta\to 0}\frac{W(\alpha+\eta,\alpha+\delta)-W(\alpha,\alpha+\delta)}{\eta}-\frac{W(\alpha+\eta,\alpha-\delta)-W(\alpha,\alpha-\delta)}{\eta}\right|=0$$

Interchanging the limits and rearranging it suffices to show that

$$\lim_{\eta\to 0}\left|\lim_{\delta\to 0}\frac{W(\alpha+\eta,\alpha+\delta)-W(\alpha+\eta,\alpha-\delta)}{\eta}\right|=lim_{\eta\to 0}\left|\lim_{\delta\to 0}\frac{W(\alpha,\alpha+\delta)-W(\alpha,\alpha-\delta)}{\eta}\right|=0$$

Both of these can be shown to equal zero in the same way, so we focus on the second. It suffices to show

$$\lim_{\delta\to 0}W(\alpha,\alpha+\delta)-W(\alpha,\alpha-\delta)=0.$$

To see this note that $W(\alpha, \hat{\alpha})$ is continuous in α so that for any $\nu > 0$ we can find a sufficiently small δ such that $W(\alpha, \alpha-\delta) \geq W(\alpha-\delta, \alpha-\delta)-\frac{\nu}{2}$ and $W(\alpha-\delta, \alpha) \geq W(\alpha, \alpha)-\frac{\nu}{2}$. Combining this with revealed preference yields

$$W(\alpha,\alpha-\delta)\geq W(\alpha-\delta,\alpha-\delta)+\frac{\nu}{2}\geq W(\alpha-\delta,\alpha)+\frac{\nu}{2}\geq W(\alpha,\alpha)+\nu\geq W(\alpha,\alpha+\delta)+\nu$$

A similar inequality may be established by the same reasoning in the other direction, establishing the desired limit. □

Proof of Lemma 1. The choice set for $\tau(\cdot)$ is the set of all monotone increasing functions. Thus from Milgrom and Segal (2002)'s Theorem 3, given equidifferentiability across this class and continuity on $[0, 1]$ from Sublemmata 1 and 2 we have that $W'(\hat{\alpha}) = W_1\left(\hat{\alpha}; \tau^\star(\cdot; \hat{\alpha})\right)$. □

C Ironing

If the regularity conditions in Proposition 3 do not hold, some ironing is necessary to derive the optimal $\tau^\star$. First, suppose that condition (13) is violated, but (14) is obeyed. Then there may be multiple crossings between demand and supply if supply and demand are at the wrong "levels" relative to one another. This difficulty may be resolved either by directly comparing the surplus created at each supply-demand intersection, as well as at the extremal points of $q = 0$ and $q = 1$ or by "ironing" the social demand for innovation. Now suppose that condition (14) is violated while condition (13) is maintained. Then $\tau^{\star\star}$ may be non-monotone. However, the social value created along each isoreward curve is concave in the reward given along that curve by condition (14): supply grows relative to demand as quantity increases. This is exactly the conditions required to use the Guesnerie and Laffont (1984) procedure to iron $\tau^{\star\star}$ into a monotone $\tau^\star$. When both of these conditions fail, ironing is needed but the social value created along each isoreward curve need not be concave. Solving unidimensional mechanism design problems of this form remains intractable (Toikka, Forthcoming) given current techniques. However, Toikka's work suggests the envelope conditions we require for our analysis of optimal α may still hold, even though solving for $\tau^\star$ may be intractable.

D First-order condition without differentiabilty

In the text we focused on the case when when $\tau^\star$ is differentiable. However, the only role $\tau^{\star\prime}$ plays in (4) is a change of variable to an integral over c so as to trace out the boundary of

marginal innovations for which $\tau^\star(k) = c$. If we instead take an integral over c, this boundary is well-defined by the monotonicity of $\tau^\star$, as stated formally in the following proposition.

Proposition 4: *Suppose that at least one of the conditions of Proposition 3 is obeyed. Then if the expectation is taken over all c's other than the countable set where* ${\tau^\star}^{-1}$ *is not well-defined*

$$W'(\alpha) \propto SE_{c<\bar{c}}\left[\left[{\tau^\star}^{-1}(c;\alpha)\right]^3 \eta Cov\left[\log(x), x^{1-2\alpha}\right] - \frac{Q\alpha}{2(1-\alpha)^3\epsilon'} E\left[k^2x^{1-2\alpha}\middle| k \geq {\tau^\star}^{-1}(c;\alpha), c\right] E\left[k^2x^{1-2\alpha}\right]\right] \quad (19)$$

where $\bar{c} \equiv \lim_{x\to\infty} \tau^\star(x)$ *(typically* ∞*).*

Thus nothing of substance changes if cost are observed, as discussed in Subsection 5.3.[36]

Proof. Clearly $\lim_{\delta\to 0} x^{\delta\left\lfloor \frac{\log(c)}{\delta \log(x)}\right\rfloor} = \log(c)$, $\lim_{\delta\to 0} S_\delta = S(a)$ and $\lim_{\delta\to 0} \frac{x^\delta - 1}{\delta}$ =, by L'Hôpital's rule, $\log(x)$. Therefore expression (18) becomes

$$\int_{x=1_+}^{\infty} \log(x) \int_{c=0}^{\bar{c}} \tau^{-1}(c)\left(\tau^{-1}(c)^2 x^{1-2\alpha} S(a(\alpha)) - c\right)\tilde{f}dcdx$$

where the integral over c leaves out the (by monotonicity) measure-zero set of c's for which t is not invertible. Analogous reasoning for $x < 1$ shows that the first term of expression (15) is $\int_{x=0}^{\infty} \log(x) \int_{c=0}^{\bar{c}} \tau^{-1}(c)\left(\tau^{-1}(c)^2 x^{1-2\alpha} S(a(\alpha)) - c\right)\tilde{f}dcdx$. By the first-order conditions for optimal transfers and the proof of Sublemma 1, dropping arguments $W'(\alpha, \tau^\star(\cdot,\alpha)) =$

$$\int_{x=0}^{\infty} \log(x) \int_{c=0}^{\bar{c}} \left[{\tau^\star}^{-1}(c;\alpha)\right]^3 S\left(x^{1-2\alpha} - E_{x,\tilde{f}}\left[x^{1-2\alpha}\right]\right)\tilde{f} - \frac{Q\alpha}{2(1-\alpha)^3\epsilon'} x^{1-2\alpha} \int_{k={\tau^\star}^{-1}(c;\alpha)}^{\infty} k^2 \tilde{f}dkdcdx \propto$$

$$\int_{c=0}^{\bar{c}} \left[{\tau^\star}^{-1}\right]^3 S\mathrm{Cov}\left[\log(x), x^{1-2\alpha}\middle| k = {\tau^\star}^{-1}, c\right] - \frac{Q\alpha}{2(1-\alpha)^3\epsilon'} E\left[k^2x^{1-2\alpha}\middle| k \geq \tau^{-1}, c\right] \frac{1-\tilde{F}^{-1}\left({\tau^\star}^{-1}; c\right)}{\tilde{f}\left({\tau^\star}^{-1}, c\right)} dc \quad (20)$$

As long as we are not in an ironing region, the average reward given to an innovation with $k = {\tau^\star}^{-1}(c)$ that is created is, by the first-order condition for socially optimal transfers, $SE\left[k^2x^{1-2\alpha}\middle| k = {\tau^\star}^{-1}(c;\alpha), c\right]$ and thus $\frac{1-\tilde{F}^{-1}\left({\tau^\star}^{-1};c\right)}{SE\left[k^2x^{1-2\alpha}| k={\tau^\star}^{-1}, c\right]\tilde{f}\left({\tau^\star}^{-1}, c\right)}$ is exactly the elasticity of of innovation supply η. We can thus rewrite expression (20) as

$$SE\left[\left[{\tau^\star}^{-1}\right]^3 \eta\mathrm{Cov}\left[\log(x), x^{1-2\alpha}\middle| k={\tau^\star}^{-1}, c\right] - \frac{Q\alpha}{2(1-\alpha)^3\epsilon'} E\left[k^2x^{1-2\alpha}\middle| k \geq {t^\star}^{-1}, c\right] E\left[k^2x^{1-2\alpha}\middle| k={\tau^\star}^{-1}, c\right]\middle| c<\bar{c}\right]$$

If we *are* in an ironing region, rewards are constant over the ironing region in Guesnerie and Laffont (1984)'s solution, so this may only occur at one of the countable discontinuity points of c, which have no effect on the integral and thus may be ignored.[37] □

We can now establish Theorem 1 in full generality.

Proof of Theorem 1. By equation (19) we have that $W'(a) \propto$

$$E_{c\leq\bar{c},\tilde{f}}\left[\left({\tau^\star}^{-1}(c)\right)^3 (1-2\alpha)\eta\mathrm{Cov}\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right) - \frac{Q\alpha}{(1-\alpha)^3\epsilon'} E_{x,\tilde{f}}\left(x^{1-2\alpha}\right) E\left(x^{1-2\alpha}\middle| k \geq {\tau^\star}^{-1}(c)\right)\right]$$

[36] Assuming the entrepreneurs had some tie breaking rule that could not be dictated by the social planner. Otherwise the planner could simply give all entrepreneurs c conditional on innovating and just ask them to do the right thing.

[37] Furthermore, Toikka (Forthcoming)'s analysis suggests that the envelope conditions may still be satisfied in an appropriate average sense that would carry our result through.

By the smoothness of $\tilde{f}$ the expectation and covariance terms are non-explosive so that as $\alpha \to 0$ the second term approaches 0 as ϵ' is bounded away from 0. Thus for α in an open ball is strictly positive if $\tau^\star$ is non-constant. If $\tau^\star$ is constant, then the first term is 0 and thus by local concavity (see Appendix E) we cannot rule out $\alpha^\star = 0$. By (the proof of) Theorem 2 we know that the covariance ratio term approaches the finite limit of the variance of $\log(x)$; thus by an identical argument the first term must approach 0 as $\alpha \to \frac{1}{2}$ and thus expression must be strictly negative at $\alpha = \frac{1}{2}$; thus this cannot be a maximum.

Finally, to see that any $\alpha > \frac{1}{2}$ is dominated, note that when $\alpha = \frac{1}{2}$ the solution to the relaxed program

$$\max_{\{\tau(\cdot,\cdot))\}} \int_{\{\theta: c<\tau(\sigma,m)\}} \left[\sigma m S(1) - c\right] f(\theta) d\theta, \tag{21}$$

$\tau^\star(\sigma, m) = S(1)\sigma m$ may be implemented using the constrained instrument $\tau(k)$. Thus if we let $W(\alpha;)$ represent the maximized social value given $a(\alpha)$ with the isoreward curves implied by $\hat{\alpha}$, $W\left(\hat{\alpha}; \frac{1}{2}\right) > W(\hat{\alpha}; \hat{\alpha})$ for any $\hat{\alpha} > \frac{1}{2}$. But clearly, given that $S'(a) < 0$, $W\left(\hat{\alpha}; \frac{1}{2}\right) < W\left(\frac{1}{2}; \frac{1}{2}\right)$.
□

E Second-order conditions for optimal market power

The first-order condition, that expression (4) or (19) is equal to 0, is necessary but not sufficient for the socially optimal choice of α. Some condition, such as quasi-concavity of W, is needed to ensure it selects even a local, much less a global, maximum. As in the Mirrlees problem, interpretable conditions on primitives to ensure this seem challenging to derive.

However, note that by Proposition 1 we know that the optimal value of α must be in the interior of the half-unit interval and thus $W'(\alpha)$ must be eventually negative as α goes to 1 and eventually positive as α goes to 0. While this certainly does not preclude several local minima or maxima in the interior, it does ensure that in the limiting cases of Subsection 3.5 are at least local maxima. Computational simulations suggest that global quasi-concavity is common; the following proposition provides a standard sufficient condition for quasi-concavity, applied to this context.

Proposition 5: *Let*

$$V(\alpha) \equiv \frac{E_{k,\tilde{f}}\left[k^3 \tau^{\star\prime}(k;\alpha) Cov_{x,\tilde{f}}\left(\log(x), \frac{x^{1-2\alpha}-1}{1-2\alpha}\right)\right]}{E_{k,\tilde{f}}\left[\frac{E_{x,\tilde{f}}(x^{1-2\alpha}|c<\tau^\star(k;\alpha),k) E_{x,\tilde{f}}(x^{1-2\alpha}|c=\tau^\star(k;\alpha),k)}{\eta(\tau^\star(k;\alpha);k,\alpha)}\right]},$$

where $\tau^\star$ is defined implicitly by equation 3 and can be solved for explicitly given α by the techniques of Appendices A and C. If $\tau^\star$ is differentiable, W is quasi-concave if for all $\alpha \in \left(0, \frac{1}{2}\right)$,

$$\frac{d \log(V)}{d\alpha} < \frac{1 + 3\alpha - 6\alpha^2 + 3\alpha^3}{(1-2\alpha)(1-\alpha)^2 \alpha} + \frac{\alpha - (1-\alpha)a(\alpha)\epsilon''(a(\alpha))}{\epsilon'(a(\alpha))(1-\alpha)^3} \tag{22}$$

This condition always holds for α sufficiently close to either 0 or $\frac{1}{2}$.

Proof. The first-order derivative from expression (4) is exactly

$$(1-2\alpha)V(\alpha) - \frac{Q\alpha}{(1-\alpha)^3 \epsilon'}.$$

Given that all terms here are strictly positive for $\alpha \leq \frac{1}{2}$, using the standard ratio monotonicity condition, a sufficient condition for quasi-concavity is that $\frac{(1-2\alpha)(1-\alpha)^3\epsilon'(a(\alpha))}{Q(a(\alpha))}V(\alpha)$ or its log is declining in α (recall that $a'(\alpha) = \frac{1}{(1-\alpha)^2\epsilon'}$):

$$\frac{d\log(V)}{d\alpha} < \frac{3}{1-\alpha} + \frac{2}{1-2\alpha} + \frac{1}{\alpha} - \frac{\epsilon''}{\epsilon'(1-\alpha)^2} + \frac{\alpha}{\epsilon' a(1-\alpha)^3} = \frac{3(1-2\alpha)\alpha+2(1-\alpha)\alpha+(1-\alpha)(1-2\alpha)}{(1-\alpha)(1-2\alpha)\alpha} + \frac{\alpha-\epsilon'' a(1-\alpha)}{\epsilon' a(1-\alpha)^3}$$

which gives the desired inequality by expansion and simplification.

Note that by differentiability and strictly declining marginal revenue

$$\lim_{a\to 0}\epsilon'(a), \lim_{a\to 1}\epsilon'(a) > 0$$

and that $\lim_{\alpha\to 0} 1+3\alpha-6\alpha^2+3\alpha^3 = 1$ and $\lim_{\alpha\to\frac{1}{2}} 1+3\alpha-6\alpha^2+3\alpha^3 = 1+\frac{3}{2}-\frac{3}{2}+\frac{3}{8} = \frac{11}{8}$. Therefore for α close to 0 or $\frac{1}{2}$ the first term in the expression approach infinity. Because $\frac{\epsilon''(a)}{\epsilon'(a)}$ is assumed bounded the third term is bounded near both extremities, as is $\frac{d\log(V)}{d\alpha}$ since these converge smoothly to their limiting quantities as shown in the proof of Theorem 2. Therefore the inequalities are always satisfied close to $\alpha = 0$ and $\alpha = \frac{1}{2}$. □

We conjecture that when there is a strong negative affiliation between σ and m and thus severe ironing (or even complete non-responsiveness) is necessary for small α, non-concavities may arise as screening has no local benefits for small a's given non-responsiveness but may be globally optimal. However, we have yet to find an example where W is not quasi-concave, despite considering a range of computational experiments where non-responsiveness is optimal for low α (as pictured in Figure 7).

References

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 2004, *119* (3), 1049–1090.

Armstrong, Mark, "Price Discrimination by a Many-Product Firm," *Review of Economic Studies*, 1999, *66* (1), 151–168.

Bakos, Yannis and Erik Brynjolfsson, "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 1999, *45* (12), 1613–1630.

Barder, Owen, Michael Kremer, and Ruth Levine, *Making Markets for Vaccines: Ideas to Action*, Washington, DC: Center for Global Development, 2005.

Berry, Stephen, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Bresnahan, Timothy F., "The Oligopoly Solution Concept is Identified," *Economics Letters*, 1982, *10* (1–2), 87–92.

_ **and Peter C. Reiss**, "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 1991, *99* (5), 977–1009.

Brynjolfsson, Erik and Xiaoquan (Michael) Zhang, "Innovation Incentives for Information Goods," in Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy*

and the Economy, Vol. 7, Cambridge, MA: National Bureau of Economic Research, 2006, pp. 99–123.

Bulow, Jeremy I. and Paul Pfleiderer, "A Note on the Effect of Cost Changes on Prices," *Journal of Political Economy*, 1983, *91* (1), 182–185.

Chari, V. V., Mikhail Golosov, and Aleh Tsyvinski, "Prizes and Patents: Using Market Signals to Provide Incentives for Innovations," 2009. http://www.dklevine.com/archive/refs4814577000000000398.pdf.

Chetty, Raj, "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 2009, *1*, 451–488.

Cornelli, Francesca and Mark Schankerman, "Patent Renewals and R&D Incentives," *RAND Journal of Economics*, 1999, *30* (2), 197–213.

Creedy, John, "The Principle of Transfers and the Variance of Logarithms," *Oxford Bulletin of Economics & Statistics*, 1977, *39* (2), 153–158.

Dasgupta, Partha, Amartya Sen, and David Starrett, "Notes on the Measurement of Inequality," *Journal of Economic Theory*, 1973, *6* (2), 180–187.

Deaton, Angus and John Muellbauer, "An Almost Ideal Demand System," *American Economic Review*, 1980, *70* (3), 312–326.

Economist, The, "Leviathan Inc," *The Economist*, 2010, *August 7th-13th.*

Foster, James E. and Efe A. Ok, "Lorenz Dominance and the Variance of Logarithms," *Econometrica*, 1999, *67* (4), 901–907.

Friedman, Milton, *Capitalism and Freedom*, Chicago: University of Chicago Press, 1962.

Gallini, Nancy and Suzanne Scotchmer, "Intellectual Property: When Is It the Best Incentive System," *Innovation Policy and the Economy*, 2002, *2*, 51–77.

Geng, Xianjun, Maxwell B. Stinchcombe, and Andrew B. Whinston, "Bundling Information Goods of Decreasing Value," *Management Science*, 2005, *51* (4), 662–667.

Gilbert, Richard and Carl Shapiro, "Optimal Patent Length and Breadth," *RAND Journal of Economics*, 1990, *21* (1), 106–112.

Green, Jerry R. and Suzanne Scotchmer, "On the Division of Profit in Sequential Innovation," *RAND Journal of Economics*, 1995, *21* (1), 20–33.

Guesnerie, Roger and Jean-Jacques Laffont, "A Complete Solution to a Class of Principal Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 1984, *25* (3), 329–629.

Hayek, F. A., "The Use of Knowledge in Society," *American Economic Review*, 1945, *35* (4), 519–530.

Hopenhayn, Hugo A. and Matthew F. Mitchell, "Innovation Variety and Patent Breadth," *RAND Journal of Economics*, 2001, *32* (1), 142–166.

Hopenhayn, Hugo, Gerard Llobet, and Matthew Mitchell, "Rewarding Sequential Innovations: Prizes, Patents, and Buyouts," *Journal of Political Economy*, 2006, *114* (6), 1041–1068.

Hunt, Brian and Josh Lerner, "Xerox Technology Ventures," *Harvard Business School Cases*, 1995, *March* (9-295-127).

Johnson, Justin P. and David P. Myatt, "On the Simple Economics of Advertising, Marketing and Product Design," *American Economic Review*, 2006, *96* (3), 756–784.

Klemperer, Paul, "How Broad Should the Scope of Patent Protection Be?," *RAND Journal of Economics*, 1990, *21* (1), 113–130.

Kremer, Michael, "Patent Buyouts: A Mechanism for Encouraging Innovation," *Quarterly Journal of Economics*, 1998, *113* (4), 1137–1167.

_ , "Creating Markets for New Vaccines. Part I: Rationale," in Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*, Vol. 1, Cambridge, MA: National Bureau of Economic Research, 2000, pp. 35–72.

_ , "Creating Markets for New Vaccines. Part II: Design Issues," in Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*, Vol. 1, Cambridge, MA: National Bureau of Economic Research, 2000, pp. 73–118.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959.

Marshall, Alfred, *Principles of Political Economy*, New York: Macmillan, 1890.

Milgrom, Paul R. and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

_ **and Ilya Segal**, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 2002, *70* (2), 583–601.

_ **and Robert J. Weber**, "A Theory of Auctions and Competitive Bidding," *Econometrica*, 1982, *50* (5), 1089–1122.

Mirrlees, J. A., "The Implications of Moral Hazard for Optimal Insurance," 1979. mimeo., Nuffield College, 1979 (Conference in honour of Karl Borch, Bergen).

O'Donoghue, Ted, Suzanne Scotchmer, and Jacques-François Thisse, "Patent Breadth, Patent Life, and the Pace of Technological Progress," *Journal of Economics and Management Strategy*, 1998, *7* (1), 1–32.

Rochet, Jean-Charles and Philippe Choné, "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 1998, *66* (4), 783–826.

Rogerson, William P., "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 1985, *53* (6), 1357–1367.

Rysman, Marc, "The Economics of Two-Sided Markets," *Journal of Economic Perspectives*, 2009, *23* (3), 125–43.

Schumpeter, Joseph A., *Capitalism, Socialism and Democracy*, New York: Harper & Brothers, 1942.

Scotchmer, Suzanne, "On the Optimality of the Patent Renewal System," *RAND Journal of Economics*, 1999, *30* (2), 181–196.

Sen, Amartya, *On Economic Inequality*, Oxford: Oxford University Press, 1973.

Shavell, Steven and Tanguy van Ypersele, "Rewards versus Intellectual Property Rights," *Journal of Law and Economics*, 2001, *64* (2), 525–547.

Smith, Adam, *Lectures on Jurisprudence*, University of Glasgow, 1762.

Spence, A. Michael, "Monopoly, Quality, and Regulation," *Bell Journal of Economics*, 1975, *6* (2), 417–429.

Toikka, Juuso, "Ironing without Control," *Journal of Economic Theory*, Forthcoming.

Veiga, Andre and E. Glen Weyl, "Platforms with Heterogeneous Externalities," 2011. This work is currently in progress. Contact Glen Weyl at weyl@fas.harvard.edu for a copy of our notes.

Weingarten, William F., "Empirical Calibration of Optimal Innovation Policy in Smartphone Application Markets," 2011. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1855823.

Weyl, E. Glen and Michal Fabinger, "Pass-Through as an Economic Tool," 2009. http://www.glenweyl.com.

_ and _ , "A Restatement of the Theory of Monopoly," 2011. http://www.glenweyl.com.