Materialistic Genius and Market Power:

Uncovering the best innovations*

E. Glen Weyl† Jean Tirole‡

August 4, 2010

Abstract

Most societies reward innovation with market power. Such an entrepreneurial system is optimal in as far as the materialistic genius celebrated, but not formalized, by many of its supporters accounts for most of the benefits from innovations. While market power distorts consumption, it also targets rewards for innovations towards those that generate the greatest consumer surplus per unit sold. Thus, optimal policy calls for some ex-post distortion but never full monopoly pricing. This mix can be calibrated empirically. The results address a number of classical problems and we develop tools for solving multidimensional screening models with endogenous information structures.

*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 gratefully for helpful comments provided by Daron Acemoglu, Bruce Caldwell, Sylvain Chassang, Michael Grubb, Sonia Jaffe, Nicolas Pistolesi, Ruth Raubitschek and especially Joshua Gans, Scott Duke Kominers, José Scheinkman and Scott Stern, seminar participants at Oxford and the Toulouse School of Economics (TSE) and most of all the excellent research assistance of Stephanie Lo, Vladimir Mukharlyamov, David Smalling, William Weingarten and especially Rui Wang and Andre Veiga. 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

> *When high roads, bridges, canals etc. are...supported by the commerce which is carried on by means of them, they can be made only where that commerce requires them, and consequently where it is proper to make them...A great bridge cannot be thrown over a river at a place where nobody passes...things which sometimes happen in countries where works of this kind are carried on by any other revenue than that which they themselves are capable of affording.*

–Adam Smith, *An Inquiry in the Nature and Causes of the Wealth of Nations*

> *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"

Knowledge diffuses most effectively[1] in the public domain. However, this efficient free flow of information affords no opportunity for innovators to recoup the opportunity cost of their creative work. Therefore, human societies have long debated the relative merits of prizes and exclusive use as means of rewarding innovation. The latter appears the overwhelming winner thus far, with intellectual property (IP) of various forms[2] (patents, copyrights, trade secrets) accounting for the lions share of innovator revenue. In fact, IP is one of the oldest tools of microeconomic policy, dating at least[3] to 7th century BCE Greece. Nonetheless, there has been substantial interest in recent years in supplementing or replacing IP with centrally-directed subsidies and prizes. In what follows we propose an approach, outlined in Section 1, to simplify the multidimensional screening problem inherent to jointly choosing prices to charge for and rewards to give to innovations. This allows us to derive empirically measurable quantities calibrating the optimal mix of IP and prizes.

[1] Recent persuasive empirical evidence is provided by Williams (2010).

[2] 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. This paper is agnostic as to when patents are "necessary" for appropriating market power (Boldrin and Levine, 2002, 2008) or even helpful in doing so (Kultti et al., 2007); instead our focus is on whether market power is appropriate, however implemented. This contrasts with the classic papers in the theoretical R&D literature (Gilbert and Shapiro, 1990; Green and Scotchmer, 1995; Klemperer, 1990; O'Donoghue et al., 2004), which assume rents may arise only from a pure patent system and study the least-cost provision of market power (e.g. breadth vs. length). See Gallini and Scotchmer (2002) for an overview.

[3] Athenaeus (c. 200-300), as translated by Jason Aftosmis, writes in Book XII verse 521 lines c8-d3 of IP in Sybaris dating at least to the 7th century BCE:

> *[Phylarchus, the 3rd century BCE historian, states that in Sybaris] if some cook or chef invented an extraordinary recipe of his own, no one but the inventor was entitled to use it for a year in order that during this time the inventor should have the profit and others might labor to excel in such endeavors.*

Our work addresses a broad debate in our field about how to foster and make the best use of innovations. Marshall (1890) famously advocated the limitation or even abolition of patents[4] and their replacement with prizes, subsidies to provide monopolies incentives to reduce their prices to cost[5] and the financing of public projects from public funds[6] to allow marginal cost pricing. These views were adopted by much of the profession (Pigou, 1920; Hotelling, 1938) and have become core tenets[7] of economics textbooks. Actual policies, of course, starkly diverge from these precepts: monopolies mostly do not receive specific subsidies, public infrastructure projects are often asked to cover their fixed costs, divisions may use full cost accounting and innovators invoke IP protection to exercise market power.

Such institutions are the heart of the idea of (intracorporate, social, political or commercial) entrepreneurship, whereby an inspired individual or small group undertakes a project for the prospect of a reward if it passes some form of market test. We seek to understand the circumstances under and the extent to which such entrepreneurial institutions, despite the distortions they cause, are nonetheless socially useful. In doing so, we draw on an idea commonly, but informally, expressed by the defenders of that system: the importance of materialistic genius.

In *Capitalism and Freedom*, Friedman (1962) defends[8] entrepreneurship by writing,

> *(T)he great advances of civilization...have never come from centralized government...Columbus did not set out...in response to a majority directive...Whitney, McCormick, Edison, and Ford...no one of these opened new frontiers...in technical possibilities...in response to governmental directives. Their achievements were the product of individual genius, of strongly held minority views, of a social climate permitting variety and diversity.*

The primary goal of our paper is thus to formalize these ideas and clarify their connection to the optimality of market power as a reward for innovation. To do this, we build a framework that simultaneously captures Smith's and Hotelling's competing arguments above, while ac-

[4]See Book IV, Chapter IX note 110, Chapter XI note 133 and Book VI, Chapter III more broadly.

[5]Book V, Chapter XIV notes 131 and 133.

[6]Book V, Chapter XIII, pp. 472-475.

[7]These views are well summed-up by Hotelling: public investment and the cost of innovations more generally "is best carried by the public treasury without attempting to assess it against the users of the particular commodity as such." Tirole (1988), pages 69-70, illustrates the tensions arising from the inconsistency of such standard prescriptions with the welfare standards used to study antitrust policy.

[8]Not only Friedman but other admirers of the entrepreneurial spirit (Schumpeter, 1942; Hayek, 1948; Rand, 1957; Kirzner, 1973) celebrate the importance of the Edison's, Ford's and Gates's of the world. For example, Rand (1970) argues that

> *(O)ne invention opens incalculable avenues to other inventions in other sciences... How can any person or group know what genius will be born where, and what ideas might occur to him. That's impossible by definition.*

Yet the connection between the existence of such genius and the entrepreneurial system is far from apparent. Many nineteenth century utopian socialists argued that capitalism would destroy genius by making all into petty bourgeois and even Marshall, one of capitalism's doughtiest defenders, argued in Book IV, Chapter VI that while in the "Middle Ages...genius...found vent in..work...the modern artisan is apt to be more occupied with management...or to collect a little store of capital."

commodating the quantity-dependent prizes advocated by Barder et al. (2006) and Berndt et al. (2007). We then develop technical tools to simplify the complex resulting multidimensional screening[9] with endogenous information structures.

These techniques, outlined in Section 1 and developed in detail in Sections 2-4, enable quantification of the apparently grandiose notion of the *value of materialistic genius*, based on intuitive measures of "value", "materialism" and "genius". Consider these each in turn:

1. *Value*: As in the literature on IP (Klemperer, 1990; Scotchmer, 1999) ,we assume the social value of innovations is proportional to the *monopoly* profits they could *potentially*[10] earn. Thus, it is appropriate to weight materialistic genius by (square[11] of) profits π.
2. *Materialism*: Materialistic individuals are highly responsive to incentives; materialism can thus be measured by the elasticity of supply of innovations, which we denote η.
3. *Genius*: In Friedman's view, a genius is an innovator whose "vision" cannot be understood by the rest of society; it is in the nature of genius to be indistinguishable from charlatanism by others. Thus, it is natural to formalize the degree of genius as residual tail uncertainty that society has about the innovator. A well-known[12] measure of such tail uncertainty is the variance of the logarithm. When prices are close to monopoly, society may approximately observe the monopoly profit a firm *could* make. Therefore, the relevant dimension of residual (to any information $\mathbf{I}$ available to the social planner) heterogeneity in innovations is that orthogonal to profits, the ratio of prices (or mark-ups) to quantities that we denote by x.

It is therefore natural to formalize the value of materialistic genius as

$$V \equiv E_{\pi^2}\left[\operatorname{Var}\left(\log(x)\middle|\,\mathbf{I},\pi\right)\eta\right]$$

[9]The growing literature on mechanism design for innovation incentives (Scotchmer, 1999; Cornelli and Schankerman, 1999; Hopenhayn and Mitchell, 2001; Hopenhayn et al., 2006; Chari et al., 2009) restricts instruments severely: the market price is either the monopoly price or the ex-post efficient one. Direct transfers from the government to the innovator are disallowed (Hopenhayn-Mitchell, Hopenhayn et. al., and Chari et. al. make a "no-money pump" assumption: $T \leq 0$, ruling out prizes) and no ex-post information (e.g. quality or price) is observable. Thus, the screening variable is patent length or breadth. These papers thus obtain non-responsiveness (Guesnerie and Laffont, 1984) results; oversimplifying, the social planner would particularly like major innovations to have limited protection. But producers of major innovations are also relatively more eager to get more protection as they enjoy a higher profit. So the"quality" of innovations is difficult to screen, often leading to a "one-size-fits-all" outcome. This provides an immediate solution to the screening problem in a way that is not possible when the broader instruments we study are allowed. Furthermore, all papers we know of in the literature consider only a single (effective) dimension of asymmetric information; while they may have heterogeneity in cost as well as benefit, only the latter has any interaction with preferences and thus can be screened.

[10]However, in sharp contrast to most of this literature, we *do not* assume it is proportional to the profits they actually earn, as we allow pricing above cost but below the monopoly optimum.

[11]The square is the relevant weighting, as we use the elasticity rather than the semi-elasticity and a one percent increase in innovation incentives obviously impacts innovations with larger (cost and benefit) scale more strongly.

[12]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 the notion of genius: the fine line separating it from insanity or sophistry.

We can then formalize Friedman's vague but intriguing conjecture: when V grows large, a system of complete monopoly power over innovations becomes optimal. However, we also prove a (partial) converse: when a closely related quantity grows small, it is optimal for innovations to be freely available and rewards for innovation to be provided by a centralized prize system. Our model therefore transforms broad philosophical arguments into empirically measurable quantities, allowing the adjudication of long-standing debates.

The key to our analysis is the notion that market power, despite its indisputable distortions, may provide market-based screen for low-surplus activities. This idea goes back at least to the work of Smith (1776) and has been echoed in the work of several of the great economists of the last two centuries (Dupuit, 1844; Mill, 1848; Coase, 1946; Vickrey, 1948). Yet as far as we know, the role of market power in avoiding "white elephants" has never been formalized. An important reason for this omission is the inherently multidimensional, and therefore technically challenging, nature of this screening problem. Smith's (surprisingly topical) fear of bridges to nowhere may be addressed simply by judging the merits of public projects based on usage. Only when innovations differ both[13] in the size of the market they create *and the value consumers take from using them* is market power necessary.

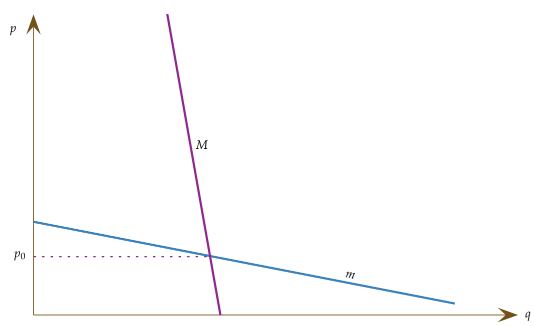

Figure 1: Distinguishing valuable from low-surplus projects

Figure 1 illustrates the basic idea: two equally costly innovations/products/projects: m ("minor", "me-too") and M ("Major"). Because willingnesses to pay are small, m creates little social value despite its large market size and does not vindicate the fixed cost of bringing it about. An example[14] might be one of the many expensive-to-develop but minor improvements on the widely marketed treatments for Type II diabetes that have come to market in recent years. In contrast, M, for example synthetic insulin that revolutionized the treatment of the rarer but deadlier Type I diabetes (Stern, 1995), adds substantial value.

[13] However, as we show in Subsection 5.5, if innovators may bribe consumers to purchase their products, even screening market size may require market power. This is a closely related, but distinct, rationale for market power.

[14] Another natural example pair would be ketchup, which is highly homogeneous but has undergone many minor widely marketed improvements, and mustard which has hundreds of highly prized but niche varieties (Gladwell, 2004). Finally, and more familiar to economists, an excellent example of a niche product that unpredictably created large consumer surplus is Honey Nut Cheerios (Hausman, 1997). We thank Scott Stern for guiding us to all of these examples.

A social planner who does not directly observe consumer surplus[15] and so, to tell m and M apart, needs to tease out information about the demand curve by using the property that the demand for a higher-quality product is less sensitive[16] 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 innovator depending only on demand at the marginal cost price charged – is unable to screen out m and screen in M.

Screening out costly, low-surplus projects is important to the extent that rare geniuses co-exist with them, as this makes it credible (and socially costly) for them to claim "strongly held minority beliefs" in the great social benefits of their projects. Furthermore, the more the supply of innovations responds[17] to incentives, the more important is Smith's concern with screening compared to Hotelling's fear about ex-post distortion. Thus, optimal level of market power is determined by the value of materialistic genius.

1 Road Map and Simplified Presentation

This section provides an independent, heuristic development and outline of the main arguments of the paper. Those interested in a more rigorous and complete treatment may therefore consider skipping it, while conversely those seeking a quick sense for the paper's thrust may view it as an alternative to reading the long paper that follows it.

To formalize the idea that authorities are not equipped to pick winners, we assume (beginning with Section 2) that the potential innovators are better informed not only about the expected cost c of their invention or a one dimension "benefit" as in Wright (1983), but also about the nature of demand for the product. Namely, individual innovations, each independent in production[18] and consumption[19] of one another, are characterized by two parameters:

[15] While it may seem feasible to measure consumer surplus without such great distortions by charging elevated prices for a short period of time, with a small probability or in a limited geographic area, such schemes may be open to easy manipulation, as we discuss in Subsection 5.6.

[16] This is dual to the classic identification result of Bresnahan (1982): just as demand twisters are needed to identify market power, market power is needed for the social planner to identify demand twisters. The relationship between quality and demand twisting is discussed more extensively in Subsection 6.5.

[17] Note, crucially, that because, in contrast to most of the literature, our model allows direct prizes, this is *not* simply an application of Nordhaus (1969)'s classic argument that, if market power is the only way to reward innovations, an increase in the elasticity of innovation supply calls for a (costly) increase in market power. In fact, in our baseline model, where the social planner maximizes social surplus, greater elasticity of innovation supply has no direct effect on the optimal level of transfers to innovators, just as an increase in the elasticity of demand has no effect on prices in a competitive market. Only if we introduce a distributive motive and thus a monopsony motive on the part of the social planner, as in 5.1, will greater elasticity raise optimal transfers. But even there, this channel is independent of that from greater elasticity to higher market power.

[18] A polar opposite of this assumption is that only one innovation may be created. In Subsection 5.3 we show that the need for market power, though not the other rich implications of our model, arises quite simply in this context as well.

[19] This may fail if innovations compete with or complement one another. In Subsection 5.2 we show that little about our baseline results change in this context.

the extent or size of the market, σ, and the quality of the innovation, m. The quantity, q, sold at price p is given by preferences representable by a natural[20] *stretch parametrization*,

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

that implies a perfect correlation between the monopoly price and average consumer surplus, if the same fraction of the monopoly price is charged for all goods. Thus, it is useful to normalize $Q(\cdot)$ so that m is the monopoly price and σ the demand at the ex-post efficient price 0.

The social planner faces a challenging three-dimensional screening problem, as the agent knows (σ, m, c) but she does not. To investigate the fundamental role market power plays, we allow the social planner to observe both price and quantity, and commit to a schedule of subsidies depending on these. By the revelation principle[21] (Gibbard, 1973), this is equivalent to the social planner asking the innovator to reveal $\left(\widehat{\sigma}, \widehat{m}\right)$ and committing to a fraction of the monopoly optimal price $p = m \cdot a\left(\widehat{\sigma}, \widehat{m}\right)$ that will be charged and a rewards $T\left(\widehat{\sigma}, \widehat{m}\right)$ that will be given to the innovator following each announcement.

She then monitors that the realized quantity[22] is consistent with the price charged to buyers and the innovator's announcement, providing the reward only if this occurs. Note that T is the total transfer and can take any form, including the patent system where $T(\sigma, m) = \sigma m Q(1)$ and $a(\sigma, m) = 1$ and the prize system where $T(\sigma, m)$ is a function of σ and $a(\sigma, m) = 0$.

The social planner would like to provide a greater reward to innovations that create high unobserved consumer surplus. However, doing so creates an incentive for the innovator to pretend he has developed a major innovation. As indicated in Figure 1, her ability to successfully fool the social planner depends on the prices charged. This leads us to the notion of an *imitation frontier*, which is the frontier of types $\left(\widehat{\sigma}, \widehat{m}\right)$ an innovator of type (σ, m) can mimic without the social planner's finding out, under a particular pricing policy $a(\cdot, \cdot)$. The innovator will aim at maximizing the transfer she receives over her imitation frontier, equating the marginal at which $\hat{\sigma}$ may be transformed into $\hat{m}$ along the frontier to the marginal rate at which substitution between these implied by the relative rewards given to each by the social planner. Therefore, the mechanism's "isoreward curves" must be tangent[23], as depicted in Figure 2 and detailed in Section 3. Interestingly, from both the economic and a technical perspectives, these imitation frontiers (and therefore the isoreward curves) depend on the (endogenous) price function, placing our problem in the class of multidimensional screening problems with endogenous information structures.

To bypass the technical issues such problems raise, the central exposition of the paper

[20] In Subsection 6.5 we discuss the wide range of standard demand functions consistent with this parameterization.

[21] A price theoretic approach that does not rely on the revelation principle is developed in Subsection 7.1.

[22] This strategy may fail either because the innovator has residual uncertainty about demand or may manipulate the quantity of the innovation sold to fool the social planner. While we show in Subsection 5.4 that little about our analysis changes in the first case, the second possibility imposes significant restrictions on the social planner and provides an alternative (but very closely related) rationale for market power, as we discuss in Subsection 5.5.

[23] This approach to multidimensional screening problems is grounded in an extension (Lemma 1) to multi-dimensional parameters and endogenous choice sets of Milgrom and Segal (2002)'s general envelope theorem.

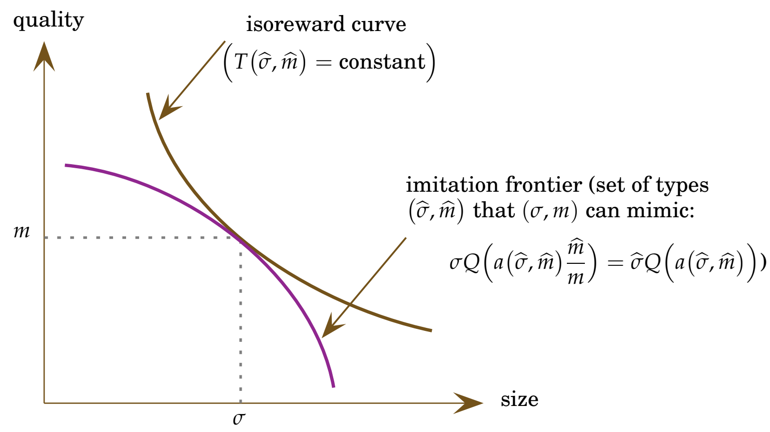

Figure 2: Imitation frontiers define the isoreward curves forming their upper envelope

initially restricts attention to a natural rule given our demand function, namely proportional pricing $P(\sigma, m) = am$. Despite its restrictiveness, it embeds both ex-post efficiency ($a = 0$) and monopoly pricing ($a = 1$) as special cases and allows a smooth transition between them with higher values of a corresponding to more ex-post distortion. Subsection 7.2 shows that all of our main results have natural analogs with more general pricing.

Our first theorem shows that Pigovian "payment in accordance with product" is feasible only for monopoly pricing and thus a value of a strictly between 0 and 1 is optimal. In the tradition of Baker (1992), we can compare the value that the agent creates for the principal, and what is actually measured by the latter. Due to the isoreward constraint, the social planner "observes" a performance index that puts too much weight on size relative to quality below monopoly pricing and too much weight on quality relative to size above monopoly pricing.

More technically, the total surplus created by the innovation is equal to

$$S(a)\sigma m$$

where $S(a)$ is a decreasing function of a, reflecting the increased distortion as price increases. Iso-reward curves also take a simple Cobb-Douglas form

$$\sigma^{\frac{1}{1+\varepsilon(a)}}\ m^{\frac{\varepsilon(a)}{1+\varepsilon(a)}} = \text{constant}$$

where $\varepsilon(a) \equiv -aQ'(a)/Q(a)$, the elasticity of demand, increases from 0 for ex-post efficient pricing to 1 for monopoly pricing. Ex-post efficient pricing, and more generally below-monopoly pricing, puts excessive weight on size while monopoly pricing achieves just the right balance between size and quality.

Choosing the degree of market power then boils down to a tradeoff between the ex-post distortion, minimized for $a = 0$, and the screening of socially optimal innovations, optimized for $a = 1$. Section 4 studies the design of this optimal mechanism, under which the optimal reward along an isoreward curve clears the market created by social demand for innovations and supply of innovations along that isoreward curve. The formula for optimal pricing applies Milgrom and Segal (2002)'s envelope theorem to illuminate the relevant tradeoffs. On the one hand, raising a allows greater rewards to be given to the best marginal[24] innovations, important to the extent that these (log-)differ from the worst innovations. However, it also distorts sales of all infra-marginal innovations, hence the role of elasticity of innovation supply. This establishes our formalization of Friedman's conjecture and its converse. We also derive empirical analogs of these quantities, showing how optimal market power can be calibrated and how the optimal mechanism can be implemented through a system of (non-linear) subsidies.

While we couch the analysis in a fairly specific context of incentives for innovation, it actually has a much broader scope. It may be extended, as we show in Section 5, to allow for distributional concerns between the innovators and the rest of society, externalities, mutually-exclusive innovations, residual uncertainty of the innovator about demand conditions, the manipulation of sales by the innovator and more sophisticated mechanisms by the social planner to measure demand. Given this, the analysis applies whenever a private or public institution wants to screen out activities that create little unobserved surplus and teasing out such information about quality requires allocative distortions. Section 6 shows how the framework applies to the bundling of applications by platforms, to incentives in conglomerates and to the incentives of multisided platforms to use participation by some users to select for the best users on other sides of the platform. We also discuss how the stretch parameterization is natural in many classical industrial organization problems.

Section 7 presents a price theoretic approach (analogous to the "demand profile" approach in non-linear pricing) to solving our model that relies less heavily on the revelation principle and uses this to generalize our results to broader pricing rules. Section 8 concludes by summarizing our modeling, technical and substantive contributions briefly and discussing directions future research might take. An appendix following the main text of the paper treats a number of technical issues that we deal with casually in the text. These include a supply-and-demand interpretation of optimal transfer policies that is valid under certain simplifying conditions, approaches to ironing optimal transfers when these fail, a version of the first-order condition which applies when optimal transfers are discontinuous, second-order conditions for optimal pricing and simulations of optimal policy under specific distributions. This acts as a short guide to a more extensive online appendix, available at http://www.glenweyl.com, which comprises the proofs of most results in the paper.

[24] By marginal we mean the technical notion of innovations just receiving enough reward to make their inventors indifferent between creating them and not and *not* the more common meaning of low-quality innovations.

2 Set-up

We begin by developing the basic model we use in the paper and providing a simple example that shows how a market power-based system can dominate an ex-post efficient system.

2.1 Model

Potential innovators (individuals or firms) are characterized by three numbers:

- c, or *cost*, an ex-ante cost of creating the innovation
- σ, or *size of the market*, the demand at the ex-post efficient price of 0
- m, the monopolist's optimal price for the good, which we call *quality* for reasons that will become apparent below.

Innovator's utility is $T - c$ if she innovates and 0 otherwise, where T is the reward the innovator receives; thus she chooses to innovate if and only if the reward she will receive exceeds her cost. We assume no marginal costs[25] of production, to focus on the case of IP.

$\theta = (\sigma, m, c)$ is private information of the innovator. The social planner knows[26] only that θ is distributed according to some smooth pdf f with full support $\mathbb{R}^3_{++}$ and all moments finite.

The social planner announces a *reward* to the innovator based on the price the innovator chooses to charge and the quantity of demand this leads to, a quantity observed by the social planner. By the revelation principle, we can instead think of the social planner as choosing a price to be charged for the good and a reward to be given conditional on a θ announced by the innovator, subject to constraints ensuring that the innovator will truthfully reveal her θ. It is easily demonstrated[27] that it is never incentive compatible for either price or reward to depend on c. Thus, we view the social planner as simply announcing two policies: a fraction of the monopoly optimal price conditional on market size and quality, $a(\sigma, m)$, and a reward conditional on these, $T(\sigma, m)$.

[25] But of course this is equivalent to *known* costs and an adjusted demand. This assumption is reasonable in many contexts, but is problematic in others; in fact firms' private information about cost is a focus of the extensive literature on regulation as pioneered by Baron and Myerson (1982) and treated fully by Laffont and Tirole (1993). Extending our analysis to this case is an exciting direction for future research.

[26] These priors are, of course, conditional on any information available to the social planner either directly or by incentive-compatible elicitation from other private agents. Kremer (1998) suggests eliciting such information while Chari et al. (2009) and others have argued such schemes are open to significant collusion. Our model is agnostic on this issue and valuable as long as, after all information is aggregated, there is still some residual asymmetric information between the innovator and social planner.

[27] Suppose two innovator types $\theta = (c, \sigma, m)$ and $\hat{\theta} = (\hat{c}, \sigma, m)$, differing only in cost, were assigned different rewards; without loss of generality assume $\hat{\theta}$ is given the higher reward. Then θ would always report $\hat{\theta}$ as $\hat{\theta}$ has exactly the same demand function as θ and could thus not be detected by the social planner. Finally, suppose the innovator breaks any indifference by choosing between rewards by choosing the lowest feasible price. Then, again, it is not feasible to assign different prices but the same reward to θ and $\widehat{\theta}$.

Demand for the innovation is characterized by a general function Q obeying standard assumptions[28] and the normalizations inherent[29] to our demand parameters. Thus, the quantity sold $q = \sigma Q\left(\frac{p}{m}\right)$ with corresponding elasticity $\epsilon(a) \equiv -aQ'(a)/Q(a)$, where $a \equiv \frac{p}{m}$. σ represents a horizontal stretching of inverse demand while m is a vertical stretching. We thus refer to this as the *stretch parametrization*; it is quite broad, as discussed in Subsection 6.5 below. 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 3. The crucial assumption inherent to the stretch parametrization is *a perfect correlation between average social surplus and monopoly prices* when a constant fraction of the monopoly price is charged: under this parametrization, the social surplus created by an innovation is

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

So if a is constant across types, so is the ratio of social surplus to profit.

2.2 An illustrative example

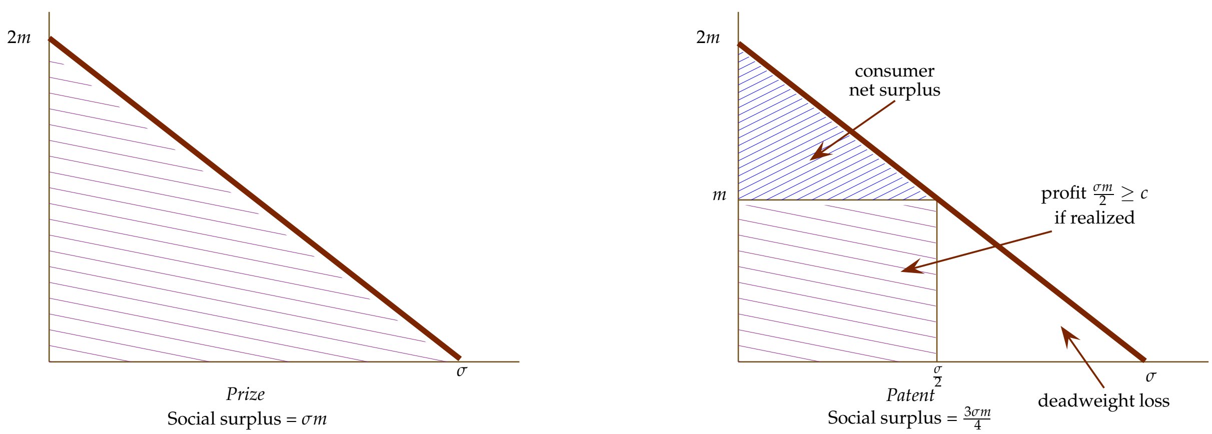

Figure 3: Linear demand under the stretch parametrization (left) and, under that demand, the division of potential gains from trade among deadweight loss, profits and consumer surplus at monopoly prices (right)

To build intuition, let us compare two specific institutions, the prize and the patent system, in the context of the linear demand curve illustrated in the first panel of Figure 3. Under the *prize system* (ex-post efficient prices and rewards based only on demand at these prices), the expected welfare created by innovations characterized by (c, σ) is

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

[28] Q is assumed smooth, strictly decreasing wherever it is strictly positive, to have strictly declining marginal revenue and bounded $\frac{\epsilon''}{\epsilon}$, and to obey $\lim_{a\to 0} aQ'(a) = 0$.

[29] $Q(0) = 1$, so ex-post efficient demand is σ, and $\epsilon(1) = 1$, so that the monopoly optimal price is m.

because every innovation with the same σ receives the same reward. Thus, 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 Akerlof (1970)'s condition that $\sigma E(m|\sigma,c) < c$ is satisfied (for all σ and c), the average innovation that is created for any prize given is not worth creating. It is thus optimal to award no prizes at all, shutting down the market for innovations entirely even if many worthy innovations exist. 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.

Under the *patent system* (each innovation charges the monopoly price m and earns the monopoly profits $\frac{\sigma m}{2}$), the innovation occurs if and only if $c < \frac{1}{2}\sigma m$, the monopoly profit, while 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 3. Thus, while the patent system destroys a quarter of the value created by each innovation, it robustly selects only innovations which are socially beneficial. In our "lemons" example, the one in every twenty five "genius" innovations with $m > 2\frac{c}{\sigma}$ are created, and all these are worth creating so clearly the patent system is in this case superior.

To see the role of genius in this tradeoff, consider the expected welfare under the patent system when conditional on (σ, c):

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

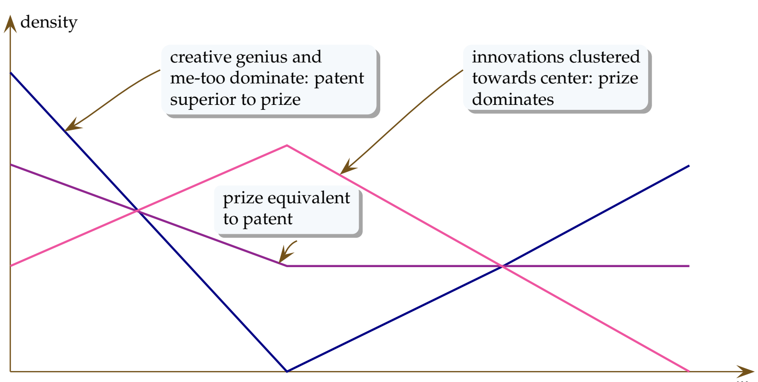

Figure 4: Bitriangular distributions with varying degrees of genius

Clearly the more important it is to select out the best, and only the best, innovations, the more valuable is the patent system relative to prizes. A simple example arises if m is again distributed on $\left[0, 3\frac{c}{\sigma}\right]$ according to a bi-triangular[30] pdf as shown in Figure 4. When the

[30]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

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

distribution is peaked and thus innovations clustered towards the center, prizes are preferable. But when it reaches a sharp trough, patents perform much better. The two systems perform equally well for an intermediate pdf that is close to uniform.

Of course, neither of these systems is optimal if we allow a broader range of institutions nesting both prizes and patents. Even if one were constrained to monopoly pricing, it would be better to pay the innovator $\frac{3}{2}$ her monopoly profit[31] to cause her to internalize the full value of social surplus. More importantly, full monopoly pricing is not necessary; a more moderate elevation of prices above cost may be sufficient and, as we show in the following section, is always optimal. The next two sections of this paper are devoted to solving for the optimal policy that trades off ex-post distortion against screening.

3 The Isoreward Approach

Ideally, the social planner would like every innovation yielding social value greater than its cost to be created. The natural solution to this problem, the principle of payment in accordance with product advocated by Pigou (1920), would be to give each innovation a reward equal to the social value it creates. However, given that he cannot observe σ and m is unable to perfectly implement payment in accordance with product. An innovator of type (σ, m) can pretend to be of another type $(\hat{\sigma}, \hat{m})$ if she cannot be distinguished by observing the demand that innovator generates. We assume *free disposal of demand*, that an innovator can freely reduce the demand for her product. Thus, she is able to imitate another innovator if, at the price that other is asked to charge, she would generate at least as much demand.

Formally, type (σ, m) can successfully imitate type $(\hat{\sigma}, \hat{m})$ if and only if

$$\sigma Q \left(a\left(\hat{\sigma}, \widehat{m}\right) \frac{\widehat{m}}{m} \right) \geq \hat{\sigma} Q\left(a\left(\hat{\sigma}, \widehat{m}\right)\right)$$

We will say that the points $(\hat{\sigma}, \hat{m})$ satisfying this with equality lie on (σ, m)'s *imitation frontier given pricing policy a*. This is a sort of "production possibilities frontier" for the innovator. To provide innovators with incentives to truthfully reveal their type, the social planner must provide at least as great a reward to each innovator type as that she could earn at any other point along her inside her imitation frontier.

Thus, the social planner's program is

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

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

[31] Of course, such Pigovian generosity is only optimal if the social planner is indifferent to transfers between the innovator and the rest of society, as we assume in our main analysis. Subsection 5.1 extends our analysis to the case when the social planner has a distributive motive.

subject to

$$\sigma Q\left(a\left(\hat{\sigma},\widehat{m}\right)\frac{\widehat{m}}{m}\right)\geq\hat{\sigma}Q\left(a\left(\hat{\sigma},\widehat{m}\right)\right)\implies T(\sigma,m)\geq T(\hat{\sigma},\hat{m}) \tag{2}$$

3.1 Technical challenges

A general solution to the above program is challenging for three reasons. First, Armstrong (1996) argues that it is often difficult to find or interpret solutions to the differential equations typically used to characterize multi-dimensional screening problems. We address this issue by developing, in this section, an intuitive "isoreward approach" to this class of multidimensional screening problems, anchored in a extension of the envelope theorem (Lemma 1) to multidimensional parameters and endogenous choice sets. This is approach is related to the method of characteristics[32] common in mathematics and physics, but allows for discontinuities and has a stronger intuitive economic interpretation in this context than is usually given to these techniques. Second, Rochet and Choné (1998) argue that the complex, global nature of incentive compatibility constraint like (2) makes even deriving such equations often infeasible. To avoid these complications, we propose a restricted, but intuitive, *proportional pricing* rule under which a is constant across innovations to ensure the validity of this approach (subject to the potential need for ironing). However we also demonstrate in Subsection 7.2 that our results are robust to relaxing this assumption. Finally, our setting poses an additional complication. Because the information structure is endogenous, the social planner's program, even once relaxed to a first-order partial differential equation, still constitutes a calculus of variations problem subject to a partial differential equations constraint, a frontier mathematics problem (Gregory and Pericak-Spector, 1999). Fortunately, the special nature of the objective function (that no derivatives directly enter it) allows us to analyze the effect of changing prices simply by drawing on Milgrom and Segal (2002)'s envelope theorem.

3.2 From imitation frontiers to isoreward curves

A standard approach to mechanism design problems is to reduce their often unwieldy global incentive compatibility constraints imposed by the necessity of global maximization by agents to local constraints at each point imposed by the necessary first-order conditions for those agents' maximization. This *first-order approach*, proposed in the single-dimensional context by Mirrlees (1971) and given rigorous foundations by Rogerson (1985), views then views the resulting envelope theorem linking the payoffs of different types of agents as the only, or at least the fundamental, constraint imposed by incentive compatibility. In this section we

[32]We thank Roland Fryer for this observation and are surprised that this approach has, to our knowledge, not been applied in the assumed-differentiable case to mulitidimensional screening. Basov (2005) provides some discussion of the method of characteristics in the context of multidimensional screening, but does not explicitly apply, or economically interpret, it.

develop an "isoreward" approach that acts as an intuitive multi-dimensional extension[33] of the Mirrlees-Rogerson first-order approach.

In doing so we draw an analogy[34] to a neoclassical production economy. We can think of the innovator as "producing" her report by choosing a point along her production possibilities frontier (imitation frontier) to maximize her reward. Thus, at the optimal report, the marginal rate at which the innovator can *transform* $\hat{\sigma}$ into $\hat{m}$ must be equated to the marginal rate at which the social planner rewards m relative to σ, as this is the innovator's marginal rate of substitution between m and σ. This local marginal rate of transformation is $-\frac{d\hat{\sigma}}{d\hat{m}}$, defined by implicit differentiation of the imitation frontier (her production possibilities frontier) at $(\hat{\sigma}, \hat{m}) = (\sigma, m)$, which after some algebraic manipulations yields the simple formula,

$$-\frac{d\hat{\sigma}}{d\hat{m}}\frac{\hat{m}}{\hat{\sigma}} = \epsilon\left(a(\sigma, m)\right) \tag{3}$$

Thus, local to the truth, a one percent increase in $\hat{m}$ requires a sacrifice of ϵ of a percent of $\hat{\sigma}$, as raising $\hat{m}$ by one percent forces the innovator to raise prices (locally) by one percent. Thus, crucially, an increase in $\hat{m}$ requires a sacrifice of $\hat{\sigma}$ only to the extent that a is large, as ϵ increases[35] from 0 to 1 as a does.

An alternative way to express the first-order conditions of a classical production economy is that, at the optimal bundle, the relevant production possibilities frontier curve is tangent to the indifference curve. Similarly, here, the curve along which rewards are constant passing through (σ, m) (the indifference curve) must be tangent to the (σ, m) imitation frontier at that point. In fact, this requirement is equivalent to the first-order conditions for optimization. Mathematically, the relevant Mirrlees-Rogerson first-order constraint is that the relative marginal rate of substitution, $\frac{T_m}{T_\sigma}\frac{m}{\sigma}$, be equated to the marginal rate of transformation derived above:

$$\frac{T_m}{T_\sigma} = \frac{\sigma}{m}\epsilon\left(a(\sigma, m)\right). \tag{4}$$

[33]Our extension is, like Milgrom and Segal (2002)'s extension of the Mirrlees-Rogerson approach, robust to the possibility of non-differentiable mechanisms

[34]The problem of incentive compatibility in our context can be seen as equivalent to that of market equilibrium in Rosen (1974)'s model of hedonic pricing in which every product exists (hence first-order conditions for its production by the most efficient producer, our incentive compatibility constraint, must be satisfied). However, our solution method via (in his context) isoprice curves, an alternative interpretation of his partial differential equations, has not, to our knowledge, been applied and might aid in the solution of such models.

[35]Given our normalizations and assumption of no marginal cost, this is equivalent to increasing elasticity for $0 < a < 1$:

$$MR = p - \frac{p}{\epsilon} \propto a - \frac{a}{\epsilon(a)}$$

so

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

Note that the right hand side of this last inequality must be strictly negative for $0 < a < 1$ as $\epsilon, a > 0$ and if $MR' < 0$ it must be positive for $a < 1$ so $\left(1 - \frac{1}{\epsilon}\right)$ must be positive. Thus, our class *cannot* accommodate constant elasticity demand, unsurprising given that this always has infinite demand at price 0.

Our (standard) assumptions imply that $\epsilon(0) = 0, \epsilon(1) = 1$, ϵ is differentiable and $\epsilon' > 0$. That is, the local reward given to m relative to σ is proportional to the elasticity of the demand curve for the innovation at prevailing prices, which increases as prices increase as long as[36] marginal revenue is declining.

Because the tangency conditions above must hold *for every* (σ, m) *pair*, and thus the marginal rate of substitution is everywhere proportional to $\frac{\sigma}{m}$, Condition (4) uniquely traces out a series of *isoreward curves* along which rewards must be constant, as formalized in the corollary below. These form the upper envelope of the imitation frontiers, a multi-dimensional extension of Harrod (1931) and Viner (1931)'s classic argument the long-run average cost curve is the lower envelope of short-run average cost curves for various levels of fixed investment.

Lemma 1. *Under an arbitrary differentiable pricing policy* $a(\cdot, \cdot)$*, incentive compatibility requires that* T *be weakly monotone in both its arguments and that rewards be constant along any curve* $\tilde{\sigma}(\tilde{m})$ *obeying for all* $\tilde{m}$

$$\tilde{m}'(\tilde{\sigma}; \sigma, m) = -\frac{\tilde{m}(\tilde{\sigma})}{\tilde{\sigma}\epsilon\left(a\left(\tilde{\sigma}, \tilde{m}\left(\tilde{\sigma}\right)\right)\right)}$$

and passing through a point (σ, m)*, except that, on a countable set of such curves, this may fail. However, changing* T *along such a set has no effect on the social planner's value function and thus* an *optimum for the social planner subject to incentive compatibility obeys this constraint. Every point* (σ, m) *is assigned to a unique such curve, whose value may be defined by the point at which it intersects the* $45°, \sigma = m$ *line.*

We have thus chosen the fairly intuitive convention of denoting isoreward curves by the point at which they intersect the $45°, \sigma = m$ line. We refer to this point k. Given free disposal, rewards clearly must be increasing in k.

Proof. See our Online Appendix Section 2. □

The proof of this result is by far the most subtle and challenging[37] of the paper. We extend a classic theorem of Young and Young (1924) to show that discontinuities of T lie along a countable set of curves. These curves are composed of at most a countable set of disconnected curves, or almost-curves from which a measure-zero set of points has been removed, of discontinuities. We then show that each of these (almost-)curves lies entirely along a single conjectured isoreward curve of the form stated in the lemma. This establishes that at most a countable number of conjectured isoreward curves contain points of discontinuity. Furthermore, we show that any conjectured isoreward curve along which T is continuous must in fact have constant rewards, that is, be an actual isoreward curve. Finally, we note that the

[36]We conjecture, but have not yet proven, that the assumption of declining marginal revenue can be dispensed with through ironing, because prices in the ironing range are never optimal.

[37]This proof was written jointly by Weyl and Michal Fabinger and will likely be spun off as Fabinger and Weyl (2011).

countable number of conjectured isoreward curves along which there may be discontinuities have no impact on maximized social welfare and thus can be ignored.

3.3 Proportional pricing

Under proportional pricing a constant fraction of the monopoly price is charged for each[38] innovation. This is a natural parameterization given our demand form. Because demand depends only on the fraction of the monopoly price charged, this parameterization implies that demand, $\sigma Q(a)$, is determined only by market size, and that the surplus created, $\sigma m S(a)$, depends only on the product of quality and market size. Thus, proportional pricing can also be seen as proportional production: it is equivalent to each innovation being asked to produce the same fraction of the socially optimal quantity.

In addition to its intuitive appeal, proportional pricing has a number of other benefits. First, it embeds both ex-post efficiency ($a = 0$) and monopoly pricing ($a = 1$) as special cases and allows a smooth transition between them by increasing a with the unambiguous interpretation that higher values of a correspond to more ex-post distortion. Second, it leads to isoreward curves taking a simple and familiar Cobb-Douglas form, as shown by the following corollary.

Corollary 1: *Under proportional pricing, incentive compatibility requires T being constant along (all but a countable set of) curves of the form $k = \sigma^{\frac{1}{1+\epsilon(a)}} m^{\frac{\epsilon(a)}{1+\epsilon(a)}}$.*

Proof. Lemma 1 implies rewards must be constant along any curve solving

$$\frac{d\tilde{\sigma}}{\tilde{\sigma}} = -\epsilon(a)\frac{d\tilde{m}}{\tilde{m}} \iff \epsilon(a)\log(\tilde{\sigma}) = -\log(\tilde{m}) + (1+\epsilon(a))\,k \iff \tilde{\sigma}^{\frac{1}{1+\epsilon(a)}} m^{\frac{\epsilon(a)}{1+\epsilon(a)}} = k$$

which yields the posited set of solutions. □

Finally, and most importantly, the validity of the isoreward approach is easy to show in this case. However, as we show in Subsection 7.1 a similar, if more technical, argument applies to any parametric or even non-parametric class of pricing rules obeying very weak (nearly necessary for incentive compatibility) assumptions. Furthermore, our results all have natural generalization in that broad context. Thus, proportional pricing is a natural class that simplifies our exposition, as well as the empirical burdens for identifying an optimum, without misleading an applied analyst about the more general features of the solution.

3.4 Validity of the isoreward approach

First-order conditions are, of course, not sufficient to ensure incentive compatibility in general. However our stretch parameterization and assumption of increasing elasticity make non-local

[38] Our solution may, of course, be applied in an industry specific manner so that a different a and T are assigned to each industry, or even sub-industry.

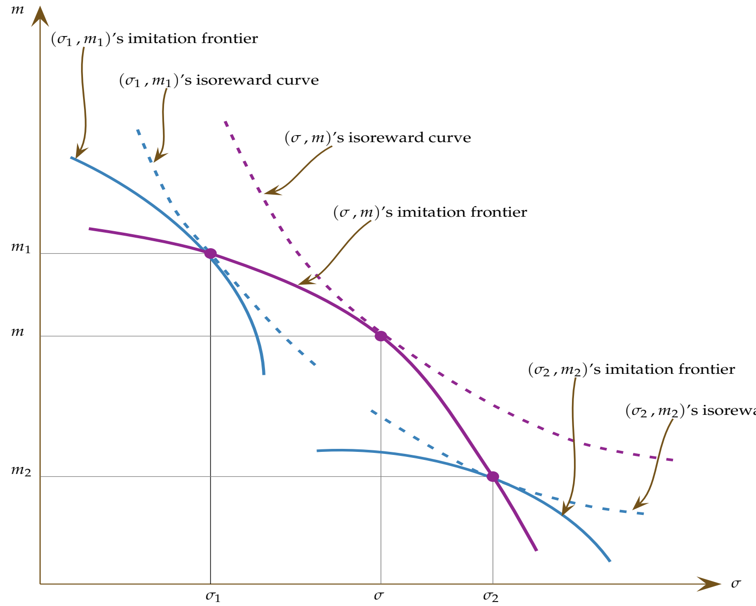

Figure 5: Justifying the isoreward approach based on the monotonicity of elasticity

deviations unattractive if local deviations are unattractive. As we show in Subsection 7.1, the isoreward approach is valid under any pricing policy for which $\epsilon\epsilon_{a_\sigma} < 1, \epsilon_{a_m} > -1$ and $\epsilon\epsilon_{a_\sigma} - \epsilon_{a_m} < 1$ hold for all (σ, m), where ϵ_{a_σ} and and ϵ_{a_m} respectively denote the elasticities of a with respect to σ and m respectively. This is particularly easy to illustrate in the case of proportional pricing, which clearly obeys these restrictions.

To see this, consider Figure 5 and suppose an innovator with type (σ, m) cheats by, say, over-reporting $m_1 > m$. She moves up along the demand curve $\sigma Q\left(a\frac{m_1}{m}\right)$ relative to the position of an innovator with a true value of $m = m_1$. This causes the elasticity of demand to be higher for the cheating innovator than for the innovator who truly is of type (σ_1, m_1) as shown in Figure 5, making further over-reporting $\hat{m} > m_1$ more costly than it would be to an innovator of true type (σ_1, m_1). Because the first-order conditions imply that type (σ_1, m_1) is (locally) indifferent to over- or under-reporting $\hat{m}$, (σ, m) will strictly prefer to reduce $\hat{m}$ back towards m as shown in Figure 5 by the fact that (σ, m) lies on a higher isoreward curve than does (σ_1, m_1). A reverse argument holds for under-reporting $m_2 < m$. Therefore, if no innovator has a local incentive to lie, any innovator imitating her will have a local incentive to move back towards the truth. Thus, so long as rewards increase across in k, the isoreward approach is valid. As a result, the isoreward and monotonicity constraints of Corollary 1 are *necessary and sufficient* for incentive compatibility under proportional pricing.

Corollary 2: *Under proportional pricing, the incentive compatibility constraint (2) is equivalent to the relaxed constraints of Corollary 1.*

Proof. See Subsection 7.1. □

So long as the optimal T along each isoreward curve monotonically increases in k, we can consider the optimal reward to assign to each isoreward curve in isolation. We focus below on the case when sufficient conditions for this are satisfied. When they are not, an ironing process (Guesnerie and Laffont, 1984) restores monotonicity: see Appendix B.

3.5 Sorting-optimality of monopoly pricing

Because social surplus is proportional to σm, payment in accordance to product would require that $\frac{T_m}{T_\sigma} = \frac{\sigma}{m}$ as social value created is proportional to σm. However, from equation (4) incentive compatibility restricts the relative rewards to quality compared to market size to $\frac{T_m}{T_\sigma} = \frac{\sigma}{m}\epsilon(a)$. Thus, Pigovian payment in accordance with product is feasible only when $a = 1$ (full monopoly pricing). Any degree of below-monopoly pricing implies that market size will be rewarded more than quality.

Conversely, any value of a above 0 involves ex-post distortion. The higher the value of a (below 1) the closer we are to payment in accordance with product and thus perfect sorting among innovations, but also the greater is ex-post distortion. This establishes the basic tradeoff between sorting and ex-post efficiency that is the key to optimal policy.

It also immediately implies a useful and intuitive result in a surprisingly clean fashion: it is never optimal to price (proportionately) above the monopoly optimum. Monopoly pricing gets sorting exactly right. Raising a above 1 both reduces the quality of sorting by over-rewarding quality relative to size and worsens ex-post distortion. As usual, given this, as we approach perfect sorting the marginal value of additional sorting diminishes smoothly to 0.

Theorem 1: *Either optimal rewards are constant everywhere and $a = 0$ or the optimal value of a is strictly between 0 and 1.*

If transfers are everywhere constant when $a = 0$, 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-monotonicity-ironing condition in Appendix B, which requires that σ not be too negatively affiliated with m in the sense of Milgrom and Weber (1982).

Proof. See Online Appendix Subsection 3.4. □

Subsection 7.2 derives a version of this result that hold under more general pricing. The local social screening benefit from higher prices outweigh the (zero) local social costs of ex-post distortion at every point beginning from ex-post efficiency and the local social ex-post efficiency benefits of lower prices always outweigh the local costs of poorer screening everywhere beginning from monopoly pricing.

4 Optimal Rewards and Pricing

Our solution now proceeds in three steps. First, we discuss briefly how the isoreward approach makes the derivation of optimal rewards straightforward. However, we do not discuss the mechanics of deriving these optimal transfers because they are not crucial to understanding the optimal choice of a. The reason, developed in our second step, is a variational calculus version of the envelope theorem, which obviates considering how optimal transfers are affected by a change in a. Finally, we use this envelope theorem to calculate the marginal benefits (from sorting) and costs (from ex-post distortion) of raising a.

4.1 Optimal transfers

Under the isoreward approach we consider the optimal reward along each isoreward curve independently, a simple unidimensional problem. It is therefore useful to re-parameterize the problem in terms of k, the isoreward curve, $x \equiv \frac{m}{\sigma}$, the quality-market size ratio denoting the position along a particular isoreward curve, and c, rather than (σ, m, c). We will refer to the relevant (change-of-variables augmented) distribution function[39] as $\tilde{f}$ to distinguish it from f. In this new notation, the social value created by an innovation, $S(a)\sigma m$, becomes $S(a)k^2 x^{\frac{1-\epsilon}{1+\epsilon}}$ because $k = \sigma^{\frac{1}{1+\epsilon}} m^{\frac{\epsilon}{1+\epsilon}}$.

If a reward $T(k)$ is given, all innovations along the k isoreward curve with cost less than $T(k)$ are created. If the average marginal innovation, the average innovation along isoreward curve k with cost $c = T(k)$, creates social value great than $T(k)$, the social planner has an incentive to raise $T(k)$; if it creates social value less than $T(k)$, the social planner has a local incentive to lower $T(k)$. Thus, the optimum is at a point where these are exactly equated: $T^\star(k;a) = k^2 S(a) E_{x,\tilde{f}}\left(x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} \middle| c = T^*(k;a), k \right)$.

If costs are not too correlated with x under $\widetilde{f}$ given k, in a sense formalized in Proposition 8 in Appendix A, then there is a unique point at which this condition is satisfied and this constitutes the optimal (monotonicity-relaxed) transfer $T^{\star\star}(k,a)$. Furthermore, if k is not too negative affiliated under $\widetilde{f}$ with x given c (see again Appendix A) then the $T^{\star\star}$ is monotone increasing and thus is the truly optimal transfer function. If either (but not both) of these conditions fail, standard ironing techniques can used to determine optimal transfers as described in Appendix B. The total social value $W(a)$ associated with each pricing policy a can thus be computed, reducing the complex program we began with to a standard single-variable calculus problem.

4.2 Optimal market power

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

Lemma 2. *$W(a)$ is differentiable for all $a \in (0,1)$ and its derivative may be evaluated by the*

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

envelope theorem, holding $T^\star$ fixed. Formally:

$$W'(\hat{a}) = \frac{\partial}{\partial a}\left[\int_\sigma \int_m \int_{c=0}^{T^\star\left(\sigma^{\frac{1}{1+\epsilon(a)}} m^{\frac{\epsilon(a)}{1+\epsilon(a)}};\hat{a}\right)} [\sigma m S(a) - c]\, f(c,\sigma,m) dc dm d\sigma\right]\Bigg|_{a=\hat{a}}$$

Proof. See Online Appendix Subsection 3.3. □

In fact, this holds even if $T^\star$ is non-differentiable: the smoothness properties we have assumed on f, combined with the monotonicity of $T^\star$ 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 $T^\star$ is differentiable (the absence of ironing[40] is sufficient to ensure this differentiability). Appendix C presents analogous results to what follows when optimal transfers are non-differentiable.

4.3 Heuristics

Recall that the relative rewards to m compared to σ are proportional to a. Thus, as a and therefore $\epsilon(a)$ grow, isoreward curves become less steep, allowing an increase in m to more easily move an innovator up the curves compared to the effect of an increase in σ. The derivative of $T^\star$ with respect to a is proportional to $\log(m/\sigma) \equiv \log(x)$ at that point.

To understand why, note that as a increases, innovations move across isoreward curves, *but the rewards given to these isoreward curves can be thought of as fixed* by the envelope theorem. Movement across isoreward curves is rapider for innovations that are far from the 45° line. These are innovations for which an increase most impacts relative rewards: they are most "un-diversified" across m and σ and thus most exposed to a change in the relative rewards given to one compared to the other. Innovations with high m relative to σ will quickly move up isoreward curves profiting from an increase in a while innovations with high σ relative to m will quickly move down, losing out.

Figure 6 first depicts an isoreward curve (the dashed curve) for a given pricing rule. Types (σ_1, m_1) and (σ_2, m_2) receive the same reward given a. An increase in a 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 a) and the low-quality type (σ_2, m_2) to a lower k. This in turn implies that for locally fixed transfers $T^*(k; a)$, a small increase in a at the margin will crowd in high-quality projects and crowd out equally costly $(c = T^*)$ low-quality projects.

The further points are from the $(\sigma = m)$ 45° line (the upwards-sloping line shown) which defines k values, 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 a. The rate of moving up (or down)

[40] In other words the obedience of conditions (11) and (12) in Appendix A.

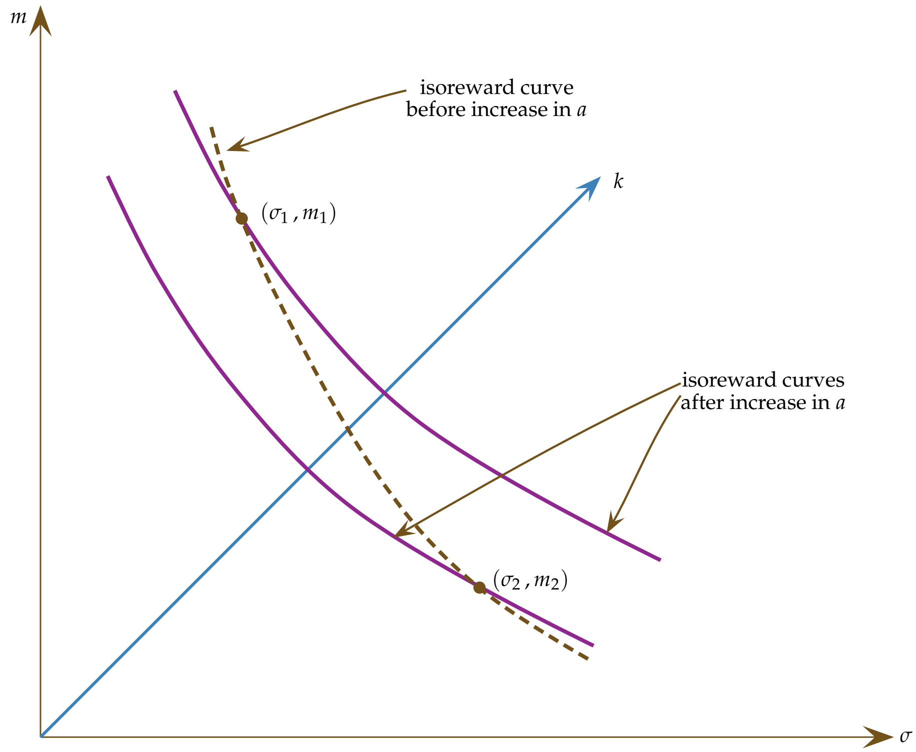

Figure 6: Increasing market power changes isoreward curves and therefore the reward given to different innovations, favoring high quality over large market size

isoreward curves $\left(\frac{d\left(\sigma^{\frac{1}{1+\epsilon(a)}} m^{\frac{\epsilon(a)}{1+\epsilon(a)}}\right)}{da}\right)$ for an innovation of partial type x is proportional to $\log(x)$, just as the production of a Cobb-Douglas economy[41] responds to a shift in shares at a rate proportional to the ratio of factors because isoquants are log-linear.

The value of innovations, along any particular isoreward curve, is proportional to $x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}$. The social benefit associated with raising a is that high x innovations are created more frequently; that is, to the extent that $x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}$ covaries with $\log(x)$ (which it must to some extent), more beneficial innovations will be selected by higher values of a. Thus, it should be clear that $\text{Cov}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}, \log(x)|k, c = T^{\star}(k)\right]$, the covariance between the extra rewards given to (marginal) innovations and the value of these innovations, is a crucial quantity pushing towards greater ex-post distortion. Furthermore, because $\lim_{a\to 1} \frac{x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}-1}{1-\epsilon(a)} = \frac{\log(x)}{2}$, for a close to 1

$$\text{Cov}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}, \log(x)|k, c = T^{\star}(k;a)\right] \approx \frac{1-\epsilon(a)}{2} Var\left[\log(x)|\, k, c = T^{\star}(k;1)\right]$$

Thus, near monopoly pricing, the incentive for ex-post distortion (higher a) is closely connected to the (conditional) variance of the logarithm of x.

[41] If q is a Cobb-Douglas isoquant with inputs K and L then $\log(q) = \alpha \log(K) + (1-\alpha)\log(L)$ so if we write $r \equiv \frac{K}{L}$ then the derivative of the isoquant shifts at a rate proportional to $\log(r)$. In fact, a number of the economists who have proposed this form for isoquants have derived it from this property (Lloyd, 2001).

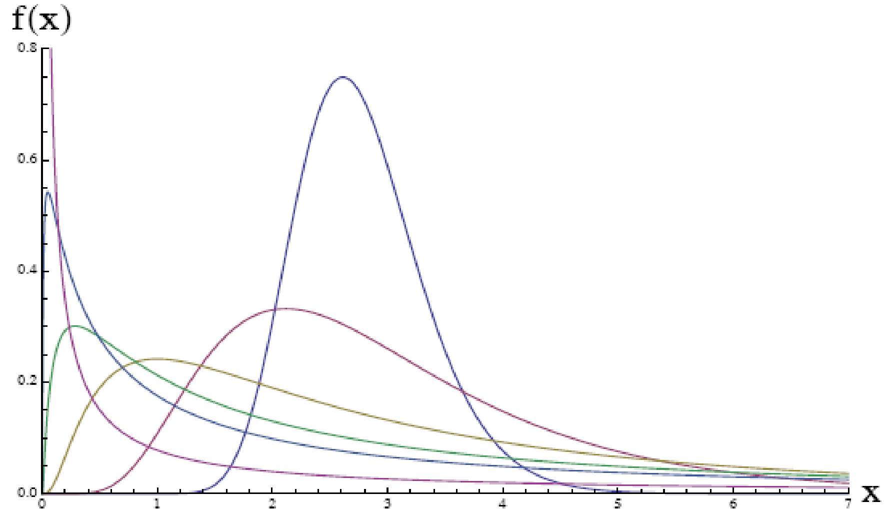

Figure 7: Log normal probability density functions with log-mean 1 and various values of log-variance. As log-variance grows tails (both near 0 and large values) grow while the middle range shrinks.

On the other hand, raising a reduces the value of innovations that are created. Unlike the sorting benefits of higher a described above, these harms apply to *all* innovations, not just those on the margin (with $c = T^{\star}(k; a)$). Thus, the relative weight on this distortion is closely related to the ratio of the number of infra-marginal innovations to the number of marginal innovations; that is, the inverse hazard rate, or inverse semi-elasticity, of innovations with respect to c conditional on k. The harm per innovation from increasing the price is proportionate to the value of the innovations which, given the logic above, is closely related to the reward given to innovations. Thus, it is in fact the (in)elasticity of innovations with respect to the rewards given them that determines the relative size of the disincentive to ex-post distortion.

Thus, we can finally return to discuss the notions of "genius" and "materialism" that we defined in the introduction. Consider the probability density function of a log-normal distribution with fixed (log-)mean 1 and various levels of variance as shown in Figure 7 above. As the variance increases, weight in the tails (both near zero and large values) grows, more than the traditional variance per se. Similarly, the elasticity of innovation supply, the percent increase in innovations for a percent increase in rewards, is a natural measure of the degree of "materialism" of innovators in precisely the same way that the labor supply elasticity can be seen as a measure of the materialism of workers: the extent of responsiveness to incentives. This combination of materialism with genius is particularly important; geniuses, such as the stereotypic brilliant scientist[42], are not always strongly responsive to incentives and thus

[42]However, such a scientist's ideas, even if not directly materially motivated, may require financial support from a backer sharing his vision or only succeed with such small probability even when they are brilliant that to justify the opportunity cost of work these must be handsomely rewarded in the case of success.

Furthermore, reinterpreted, our analysis may apply to academics who are more often motivated (at least directly and primarily) by their colleagues' esteem. For example, consider the, perhaps superficial, impression an outsider "screening" them for rewards is not important.

4.4 Full first-order condition

These forces can be formally shown to be the key determinants of the net marginal social benefit of ex-post distortion.

Corollary 3: *Assuming $T^\star$ is differentiable in k at a, $W'(a) \propto$*

$$E_{k,\tilde{f}}\left[k^4\left[\underbrace{(1-\epsilon)\frac{T^{\star\prime}}{k}\frac{\epsilon'}{(1+\epsilon)^2}\frac{\eta Cov_{x,\tilde{f}}\left(\log(x), x^{\frac{1-\epsilon}{1+\epsilon}}\right)}{1-\epsilon}}_{sorting} - \underbrace{\epsilon Q E_{x,\tilde{f}}\left(x^{\frac{1-\epsilon}{1+\epsilon}}\right) E_{x,\tilde{f}}\left(x^{\frac{1-\epsilon}{1+\epsilon}}\middle| c < T^\star\right)}_{ex\text{-}post\ distortion}\right]\right] \quad (5)$$

where η is the elasticity of innovations with respect to reward and all quantities inside the expectation are evaluated conditional on a, k and $c = T^\star(k;a)$ where not explicitly stated. As usual, a necessary condition for the optimal choice of a is that this equal 0*.*

This formula is a special case of the general first-order condition which applies even when $T^\star$ is not differentiable, as stated in Proposition 13 in Appendix C. Intuitively, by Leibnitz's rule, there are two effects of an increase in a on social welfare. First, increasing a changes boundary of the integrating region: the rewards given to different innovations and thus which marginal innovations are created shift, as described above. We refer to this as the *sorting effect*. Second, increasing a changes the interior of the integrals, reducing the surplus of all innovations that are created, which we refer to as the *ex-post distortion effect*. The first effect is always positive, the second always negative. The optimal level for a balances these two incentives.

Proof. See Online Appendix Subsection 3.4. □

Of course, for the equation of these marginal benefits and costs to actually characterize the optimum, the problem must be (quasi-)concave. While, as in the Mirrlees problem, it is not typically tractable to determine simple conditions on primitives to ensure this, we show in Appendix D that so long as the product of materialism and genius does not increase too rapidly, especially for intermediate values of a, the problem is concave. All of our computational simulations thus far, some of which are described in Subsection E below, exhibit such concavity.

4.5 Limit theorems and Friedman's conjecture

Our analytical results therefore depend on assuming quasi-concavity. These results consider the limit as genius and materialism grow large or small and can be seen as formal proofs of Friedman's conjecture and its converse.

would have of the process of acquiring academic prestige in mathematics compared to economics. In mathematics the importance of a result is typically judged at the time the result is produced. This reduces the need for the costly jostling for citations so important to provide a "market test" for ideas in economics. Of course, much more detailed data on the role of citations, the predictability of the importance of various results and the responsiveness to prestige incentives in the two fields would be needed to draw such conclusions more firmly.

Theorem 2 (Friedman's Conjecture)**:** *Let* π *be the monopoly profit associated with an innovation and* V_1 *the* value of materialistic genius near monopoly *be*

$$\frac{E_{\pi,\tilde{f}}\left[\pi^2 Var_{x,\tilde{f}}\left(\log(x)|\, k=\sqrt{\frac{\pi}{Q(1)}}, c=\frac{S(1)}{Q(1)}\pi\right)\eta\left(\frac{S(1)}{Q(1)}\pi\middle|\, k=\sqrt{\frac{\pi}{Q(1)}}, a=1\right)\right]}{E_{\pi,\tilde{f}}\left[\pi^2\right]}.$$

Then, within a class of distributions for which W *is quasi-concave, those with sufficiently high values of* V_1 *have* $a^\star$ *arbitrarily close to* 1*. That is, as the value of materialistic genius near monopoly grows large, monopoly pricing becomes optimal.*

Proof. See Online Appendix Subsection 3.6. □

Intuitively, as the extent of genius and materialism grows, 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, $\frac{S(1)}{Q(1)}$ of its profit, and pricing is near monopoly-optimal.

Theorem 3 (Partial converse of Friedman's Conjecture)**:** *Let* V_0*,* the value of materialistic genius near ex-post efficiency *be*

$$\frac{E_{\sigma,f}\left[\sigma^4\frac{\log(T^\star(\sigma))' Cov_{m,f}(\log(m),m|\sigma,c=T^\star(\sigma))\eta(T^\star(\sigma)|\sigma,a=0)}{\sigma E(m|\sigma,c<T^\star)}\right]}{E_{\sigma,f}(\sigma^4)}$$

Then, within a class of distributions for which W *is quasi-concave, those with sufficiently low values of* V_1 *have* $a^\star$ *arbitrarily close to* 0*. That is, as the value of materialistic genius near ex-post efficiency grows small, ex-post efficiency becomes optimal.*

Proof. See Online Appendix Subsection 3.6. □

Intuitively, as m becomes perfectly known or all innovations become infra-marginal, the incentives for ex-post distortion become small and are overwhelmed by even the small distortion it causes. Optimal policy becomes ex-post efficient pricing coupled with pure (ex-post demand-dependent) rewards.

These two theorems are not quite converses of one another. They consider only limit cases and use related but not identical measures of genius and materialism. Nonetheless, we believe they establish a tight connection between the extent of materialistic genius and the justification of market power.

4.6 Examples

The comparative statics arising from our model match common intuitions about the relative merits of entrepreneurial and more centralized or bureaucratic systems of procurement. Consider the standard comparison between product and process innovations. It is typically difficult to predict the success of a new product and thus a large degree of asymmetric innovation is likely to exist between the innovator and the firm or government "procuring" the innovation. Thus, it is natural, in the context of our model, that a fairly entrepreneurial approach should be taken to such innovations as we see in most societies. Process innovations, on the other hand, are for the most part bureaucratized and procured in a much more hierarchical manner within firms. This accords with our results, because the potential benefits of such innovations are far more predictable and measurable before the product goes to market. Similarly, one might compare the recent enthusiasm for prizes and patent buyouts in the medical sphere to the almost complete absence of interest in such schemes for high technology. While the benefits of a new medicine may be measured to at least some extent independent of the willingness of consumers to pay for the medicine, such is nearly impossible for new consumer technologies which may, unpredictably, being of great or no value at all.

Similar comparisons appear in settings which differ in the elasticity of innovation supply. Consider the comparison between Apple's iTunes Music Store, or other online music sites, and their iPhone App Store. As discussed in Subsection 6.1 below, the optimal strategy[43] for such a site selling media strongly complementary to a platform, such as the iPod or iPhone, closely resembles that of a social planner. In practice we observe the online music store 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. Compare this to the 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.

4.7 Empirical counterparts

Beyond these broad associations, Theorem 2 above has natural and more precise empirical analogs. This is analogous to Saez (2001)'s argument that the Mirrlees model of optimal taxation could be calibrated using labor supply elasticities, distributions of income and social welfare weight. The two crucial quantities are materialism and genius. While many caveats[44] must be applied in interpreting any attempt to calibrate these, a relatively simple procedure[45]

[43]William Weingarten is currently working on calibrating our model to apply to this context.

[44]The caveats with such a calibration should be interpreted are numerous, as with most empirical calibration of theoretical models; nonetheless, such calibrations, such as Saez (2001), can be quite informative by setting a quantitative benchmark. Many of the assumptions we have made throughout may be relaxed in empirical estimation, exploiting the extensions we develop in Section 5. Specific cautions are discussed in footnotes throughout this subsection.

[45]The exercise depends on assuming current rewards to innovation are optimal conditional on the existing pricing, which is clearly a stretch. If one were to ask what the optimal (approximate) value of a, holding fixed any (known) T

seems to give a reasonable first pass.

Measuring materialism is essentially similar to measuring the elasticity of labor supply: feasible and a focus of much research but carrying many familiar difficulties. Measurement of genius is, at least on the first pass, somewhat more straightforward. Essentially, it consists of taking a series of innovations[46] in an industry to which a uniform policy would be thought to apply and collecting data on all products innovated in that industry over that period. For each product one would measure a mark-up and quantity (likely some average over time), as well as any information available to individuals other than the innovator at the time of patent but before the product came to market. Letting x_i be the ratio of an individual product of absolute mark-up (here, the price) to quantity (which is a valid approximation for a close to 1) and π_i be their product while $\mathbf{I}_i$ is a vector containing all other covariates available prior to marketing, one would run the regression

$$\log(x_i) = \gamma\left(\pi_i, \mathbf{I}_i\right) + \epsilon_i \tag{6}$$

to recover the residuals ϵ_i. The variance of these residuals[47] (in a regression weighted by π_i^2) would then represent the degree of genius.

With these two estimates, call them M and G respectively, in hand one can calibrate our model in a fairly straightforward manner. In the proof of Theorem 2 in Online Appendix Subsection 3.6 we show that when a is near 1, as our society's widespread use of market power indicates we are, the first-order condition in (5) simplifies to

$$a^{\star} \approx 1 - \frac{4Q(1)}{S(1)\left(\epsilon'(1)\right)^2 MG} \tag{7}$$

Note that, in addition to materialism and genius, two other unknowns appear: the ratio of (fraction of total potential) social surplus to the fraction of total potential quantity at monopoly prices and the slope of demand elasticity. These are properties of demand curvature[48] about which empirical work typically makes assumptions rather than measuring; thus, in what follows

policy, not necessarily the optimal one, an analogous formula with slightly altered numbers would apply but would include terms related to a social desire to increase or decrease the level of rewards. Assuming away such a social motive is equivalent to assuming the optimality of T. Thus, as long as one is primarily interested in adjustments to the shape (induced prices) of innovation incentives rather than their level, our approach should apply reasonably well.

[46]Genius should be measured for marginal innovations only. However, to provide a reasonable data set, this would have, in practice, to be based on nearly all innovations. It would thus rely on the assumption that the degree of marginal genius does not differ significantly from infra-marginal genius. This might potentially be improved upon using the same instrument that identifies the elasticity of innovation supply to pick out marginal innovations.

[47]The measurement of G above is really an upper bound on the degree of genius, as not all of the residual variance in $\log(x_i)$ represents asymmetric information; much of it is unknown to the innovator as well. See Subsection 5.4 for more details on how to adjust the formula to allow such residual uncertainty.

[48]It may be feasible to measure, or at least gain some intuition, about the second of these as $\epsilon'(1)$ is simply the inverse of the rate at which the monopolist finds it optimal to increase prices in response to a tax at the monopoly optimal price. Weyl and Fabinger (2009) discuss ways to calibrate intuitions about such *pass-through rates.* $\frac{S(1)}{Q(1)}$ can also be expressed as an average over a wide range potentially far from the monopoly optimal price.

we make a standard assumption that ties down these quantities and use this as an example.

Along those lines, suppose that demand is linear so pass-through is constant and equal to $\frac{1}{2}$ (i.e. $\epsilon'(1) = 2$) and $\frac{S(1)}{Q(1)} = \frac{3}{2}$. Then the formula simplifies to $1 - \frac{2}{3CM}$. Suppose that the elasticity of innovation supply were .14 as estimated by Dubois et al. (2010) in the pharmaceutical industry and two typical (squared-profit-weighted) innovations with the same ex-post profit differed by about half order of magnitude in their ratio of mark-ups to profits. Then optimal value of a would be approximated negative, clearly implying that the approximation is invalid and that something quite far from monopoly pricing is likely optimal. On the other hand, suppose the elasticity of innovation supply were 4, as estimated by Acemoglu et al. (2006) (also in pharmaceuticals), and those two typical innovations differed by three orders of magnitude. Then the optimal value of a would be near .98 and this would likely be a fairly accurate estimate. Thus, it should be clear that our model does not immediately bias the result of an analysis either in favor of market power or against it: it leaves optimal policy as very much an open empirical question.

4.8 Implementation

Policy makers are unlikely to literally run a direct revelation mechanism. Instead, they would like to provide monetary transfers or taxes to firms that are granted a patent, dependent on their prices, thus inducing them to behave as described by the mechanism. The implementation of the optimal scheme is quite simple to describe in these terms, because, under proportional pricing, there is a direct relationship between prices and m on the one hand and quantity and σ on the other. In particular, implementing $a^\star$ requires providing innovators a net payoff of

$$\hat{T}^\star\left(p^{\frac{\epsilon(a)}{1+\epsilon(a)}} q^{\frac{1}{1+\epsilon(a)}}\right) \equiv T^\star\left(\left(\frac{p}{a}\right)^{\frac{\epsilon(a)}{1+\epsilon(a)}}\left(\frac{q}{Q(a)}\right)^{\frac{1}{1+\epsilon(a)}}\right)$$

or, in other words, a gross subsidy (after the firm has collected its profits) of

$$\hat{T}^\star\left(p^{\frac{\epsilon(a)}{1+\epsilon(a)}} q^{\frac{1}{1+\epsilon(a)}}\right) - pq$$

Thus, the net rewards to innovations have Cobb-Douglas level sets in p and q, but need not be homogeneous of degree 1. While simple to state, it might be challenging to implement[49] a non-linear scheme of this form, just as it has often been argued that it is difficult to implement optimal non-linear tax schemes. However, just as with any standard mechanism, fairly close approximations to a policy of this sort may be based on piecewise linear[50] consumption

[49]However, the empirical calibration of the appropriate level of the elasticity of demand, as opposed to a, can actually be performed under weaker pass-through rate assumptions. Given that this is all that is necessary for the optimal implementation of policy, determining optimal policy in practice may be easier than determining the optimal value of a.

[50]See, for example, Babayev (1997) for state-of-the-art algorithms for constructing such approximations.

subsidies.

Example (verifiable cost). Suppose that the social planner observes the cost c, as might be the case for public infrastructure projects as discussed in Subsection **??** below. An optimal scheme (*the* optimal scheme if transfers are – at least slightly – socially costly) is to reimburse the cost provided that the innovation satisfy a minimum scoring rule: $p^{\frac{\varepsilon(a)}{1+\varepsilon(a)}} q^{\frac{1}{1+\varepsilon(a)}} \geq k$, and nothing if this score is not reached. When a is close to 1, this minimum score is equivalent to a minimum profit level.

5 Extensions

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

5.1 Distributional concerns

In many applications, including straight IP in a society with distributional motives, transfers to innovators 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 innovators compared to that of the government and consumers. Then her program is exactly as in (1), but with $\lambda c + (1-\lambda)T$ replacing c, and is subject to the same incentive compatibility conditions. Note that while we no longer have a guarantee that $a^\star \in (0,1)$ by our simple argument[51] in Theorem 1, all of the rest of Section 3 goes through exactly the same.

Lerner (1934)'s formula[52] gives us that at the optimal reward, the innovator receives a fraction $\frac{\eta}{1-\lambda+\eta}$ of the marginal value of her innovation given pricing rule a. Suppose we were interested in conditions under which not only monopoly pricing, but the exact monopoly rent was optimal. This would require

$$\frac{\eta}{1-\lambda+\eta} = \frac{\pi}{S}$$

or $\lambda = 1 - \frac{S-\pi}{\pi}\eta$. Note that this can only be satisfied if the elasticity of innovation supply is less than $\frac{\pi}{S-\pi}$ (the ratio of profit to net consumer surplus, e.g. for linear demand, the elasticity of supply is less than two). Thus, there is a tension between monopoly pricing and monopoly rents: to the extent innovators are materialistic, it is unwise to try to cheat them of their full social rents. But this is exactly the setting where monopoly pricing is beneficial (holding constant G).

Analyzing $a^\star$ is similarly straightforward. The two primary effects we emphasized above, sorting and ex-post distortion, persist. The additional element added is the effect that an

[51] In particular, it is now not obvious that only sorting operates locally when $a = 0$ or that only ex-post distortion operates when $a = 1$, as the Spencian effects discussed below arise in both cases.

[52] To derive optimal transfers we can follow nearly the same procedure, except that the supply curve for innovations (as described in Subsection 7.2 below) must be replaced with a $\lambda/1-\lambda$ mix of the supply curve and the monopsonist's (Bulow and Roberts, 1989) *marginal cost*.

increase in a has on the rewards given to marginal compared to infra-marginal innovators. To the extent that these rewards, proportional as we discussed in Subsection 4.3 to $\log(x)$, are equal, they can be fully internalized by adjusting rewards optimally. When raising a gives more less to infra-marginal innovators, it will be more[53] attractive. This is exactly the logic of Spence (1975)'s model of quality-choosing monopoly, expressed precisely in the following proposition.

Proposition 1: *The derivative of social welfare W^λ maximized over transfers (with maximizer T^λ) with respect to a is proportional to expression (5) if*

$$(1-\lambda)\frac{\epsilon'}{(1+\epsilon)^2}E_{k,\tilde{f}}\left[k^4\frac{T^{\lambda\prime}}{k}E\left(x^{\frac{1-\epsilon}{1+\epsilon}}|c=T^\lambda\right)\left[E\left(\log(x)|c=T^\lambda\right)-E\left(\log(x)|c<T^\lambda\right)\right]\right]$$

is also added to it and that η is replaced by $1-\lambda+\eta$.

Proof. See Online Appendix Subsection 4.1. □

5.2 Externalities

Many innovations generate externalities. Some (e.g. green technology) are orthogonal to the nature of products as innovations, but many arise directly from the innovative nature of the products. In particular, many innovations "stand on the shoulders" of previous innovations (Scotchmer, 1991; Green and Scotchmer, 1995), implying that innovations generate positive spillovers to later innovations that build on them. On the other hand, many innovative products compete with existing products, generating negative spillovers (Spence, 1976; Dixit and Stiglitz, 1977; Loury, 1979). To the extent that the value of such spillovers is not a direct function of the demand parameters we consider, the analysis of these effects requires introducing further heterogeneity which is 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, a fairly straightforward analysis is possible, as we develop below.

Let us begin with the case when the externality is proportional to the potential (σm), rather than the actual ($S(a)\sigma m$), surplus generated by the innovation. In this case, nothing changes in our analysis in Section 4 other than the optimal transfers $T^\star$.

Proposition 2: *Let the social surplus generated by an innovation be $[\gamma+S(a)]\,\sigma m$. Then Corollary 3 applies except that $T^\star(k)=[\gamma+S(a)]\,k^2E_{x,\tilde{f}}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}\middle|k,c=T^\star(k)\right]$ rather than $S(a)k^2E_{x,\tilde{f}}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}\middle|k,c=T^\star(k)\right]$.*

[53]Suppose that, along an isoreward curve, c is affiliated with x: higher quality (compared to size) innovations are more costly. Then marginal innovations will receive a larger increase in rewards from higher a compared to infra-marginal innovations and thus a distributional motive will (ignoring effects through $T^\star$ and the increased elasticity) tend to increase the incentive for market power relative to the case of social surplus maximization. We conjecture this effect will dominate the others, but have not found conditions under which we can prove this.

Proof. See Online Appendix Subsection 4.2. □

Raising $T^\star$ by raising γ may increase or decrease optimal market power in this case. On the one hand, to the extent that the elasticity of innovation supply is declining over the relevant range, as we might expected to be, higher transfers will encourage lower market power. On the other hand higher transfers will likely make ${T^\star}'(k)$ greater, largely offsetting the first effect, though of course either could dominate in practice.

If instead the spillover is proportional to realized surplus, there is, relative to the first case, more motive to reduce market power when the spillover is more positive, as the reduced consumption induced by higher prices also reduces the spillover.

Proposition 3: *Let the social surplus generated by an innovation be $(1+\gamma)S(a)\sigma m$. Then Corollary 3 applies except that $T^\star(k) = (1+\gamma)S(a)k^2 E_{x,\tilde{f}}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}\middle| k, c = T^\star(k)\right]$ rather than $S(a)k^2 E_{x,\tilde{f}}\left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}\middle| k, c = T^\star(k)\right]$ and the second, ex-post distortion term is scaled up (or down) by $1+\gamma$.*

Proof. See Online Appendix Subsection 4.2. □

If ex-post distortion also reduces the amount of spillovers, there is more of an incentive to hold down (or up, for negative spillovers) the price of innovations. This is consistent with the argument of Bessen and Maskin (2009) that innovations likely to have large spillovers may be best priced nearer to cost to encourage follow-on innovation, so long as the profits thereby lost can be (more than) made up through subsidies (including that directly coming from lower licensing fees). For the most part, however, the basic conclusions of our analysis are robust to the existence of consumption spillovers. In fact, all of our theorems could be easily 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.

5.3 Mutually-exclusive innovations

Another way that project may fail our assumption of complete independence is on the production side. As Aghion et al. (2008) and Murray et al. (2009) argue, there are often trade-offs between different research agendas. An extreme example of such trade-offs is simple to model: there is a single agent who 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 genius or materialism, illustrates the robustness of our basic argument.

Suppose an innovator can choose to create, for effort cost e, any innovation lying along the smooth curve $\sigma = h(m; e)$ where $h_m < 0 < h_e$. The innovator would like to minimize her effort cost of obtaining a prize T 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 asymmetric 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, the choices of interest to us are those between quality and quantity, as in Lewis and Sappington (1994), Ottaviani and Prat (2001) and Johnson and Myatt (2006). It is natural to assume that along an isoeffort curve $h(m)$, there is a unique point generating maximal potential surplus $mh(m)$ so that if the social planner were to select unconstrained, his problem[54] would be quasi-concave. A standard condition for this is increasing elasticity: $\frac{\partial \epsilon_h}{\partial m} > 0$ where $\epsilon_h \equiv -\frac{mh'}{h}$.

The social planner can announce a price-quantity pair and penalize the innovator if she fails to reach that price-quantity pair. Incentive compatibility requires that the isoeffort curve is tangent to the demand curve at the requested price-quantity pair; otherwise the innovator 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. Intuitively prices must be higher to induce the innovator to choose a product that fairs better when prices are higher.

Social welfare is $S(a)h(m)m$. 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 $mh(m)$. 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 $a = 1$. But to maximize surplus given the innovation that is created requires $a = 0$. This is exactly the same trade-off in the case of independent projects and gives rise to a simple expression for the costs and benefits of ex-post distortion.

Proposition 4: *The first-order net benefits of increasing a in the mutually-exclusive innovations model are proportional to*

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

where ϵ_f is the elasticity of the function f.

When $a = 0$, $\epsilon_S = 0$ as there is no first-order distortion from raising a but $\frac{\epsilon_\epsilon}{\epsilon_{\epsilon_h}} > 0$ and thus there is a first-order benefit from raising a. When $a = 1$ there is no first-order benefit from raising a as $\epsilon_h = 1$, but there is a first-order loss from the distortion thus caused as $\epsilon_{\epsilon_h}, \epsilon_S > 0$. Therefore, it will always be 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 will be

[54]Interestingly, Johnson and Myatt base their analysis on an assumption implying that the social planner's problem would be globally convex. This implication of their assumption does not appear to be widely appreciated.

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.

Proof. Denoting the logarithm of each variable with $\sim$, we can log-differentiate social welfare:

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

But by implicit differentiation of equation (8),

$$\frac{\partial \tilde{m}}{\partial \tilde{a}} = \frac{\epsilon_\epsilon}{\epsilon_{\epsilon_h}}$$

Combining this with the first expression and multiplying by ϵ_{ϵ_h} yields the expression in the proposition. □

5.4 Residual uncertainty

We assumed above that the innovator knows, at the point of undertaking the innovation, the exact demand her product will face. This seems unrealistic for two reasons. First, the innovator 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). Second, even at the point of bringing the innovation to market the innovator may be uncertain about the demand that will be realized once she chooses a price, as in Baron (1971), Leland (1972) and Holthausen (1976). In this section we therefore show that our basic results are robust to such residual uncertainty and in particular we illustrate how the empirical formula of Subsection 4.7 may be extended[55] to these settings.

5.4.1 Ex-ante uncertainty

Suppose that, at the point of going to market the innovator knows (σ, m) but at the point of deciding whether to make the investment or not she knows her cost c of innovating and some[56] signal, s, of her eventual demand. The key feature of this setting is that ex post, when innovations go to market, nothing differs from our standard model above. Thus, our approach to incentive compatibility applies exactly as before. Then a similar reinterpretation of the arguments used to derive the first-order derivative with respect to a in the case of no residual uncertainty yields an analogous first-order derivative in this case.

[55] Similar elaborations are possible for the other extensions in this section, but we omit these for brevity, using this subsection as an illustrative example of how to undertake such an extension.

[56] s is drawn, jointly with the eventual realizations of (σ, m) and the contemporaneous c, from $\mathbb{R}^N$ for some integer N according to the (smooth and finite moment) joint distribution $g(\sigma, m, s, c)$, which can be transformed as above into $\tilde{g}(k, x, s, c)$.

Proposition 5: *In the model with ex-ante uncertain demand, suppose $T^\star$ is differentiable at a and, as described more formally*[57] *in our online appendix, that conditional on the fact that an innovation ends up on isoreward curve k and is marginal, the average distribution of $\hat{k}$ and $\hat{x}$ that the social planner infers the innovator anticipated are independent and the distribution of x does not depend on k. Assume also a similar independence condition on each s. Then* $W'(a) \propto$

$$E_k\left[k^4\left[(1-\epsilon)\frac{E_s\left[E_{\hat{k}}\left[\hat{k}^2\middle|s\right]\middle|k\right]}{k^2}\frac{2\epsilon'}{(1+\epsilon)^2}\frac{\eta\left(T^\star|k\right)Cov_s\left(E_x[\log(x)|s],E_x\left[x^{\frac{1-\epsilon}{1+\epsilon}}|s\right]\middle|k,c=T^\star(k)\right)}{(1-\epsilon)}-\epsilon QE_x\left[x^{\frac{1-\epsilon}{1+\epsilon}}|k,c\leq T^\star(k)\right]x^{\widetilde{\frac{1-\epsilon}{1+\epsilon}}}\right]\right]$$

Proof. See Online Appendix Subsection 4.3. □

Expression (5) differs from expression (5) in two ways. First, and crucially, the expectations and covariance over x are, as in Holmström (1979), taken only including the uncertainty *over* s, not that given s. Thus, it is only the *asymmetric* information between the innovator and society and not the social planner's total uncertainty over x that is relevant. Similarly, in some places k, or functions of it, must be replaced with the average value they would have been expected to take on for an innovation that ends up with such a k value.

Our empirical formula therefore generalizes in a simple manner, described in more detail[58] in subsection 4.3 of our online appendix: only the variance in $\log(x)$ (given profits and the social planner's prior information) *over* the private information of the innovator, not given it, is relevant. Rather than simply taking the variance of the residuals from the regression described in Subsection 4.7 above, one must multiply this variance by R^2 of a regression of these residuals on the information of the innovator.

5.4.2 Ex-post uncertainty

Alternatively, we may consider a case where the innovator does not know the demand even ex-post. Suppose that given (σ, m) realized demand is given by a distribution

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

Again, the imitation frontiers are unchanged: first, the social planner cannot distinguish between (σ, m) and $(\widehat{\sigma}, \widehat{m})$ at price $\widehat{p}$ as long as

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

[57] It is possible to derive extensions versions of our empirical formula that apply to this model without such stringent assumptions. The required adjustments are no more burdensome empirically than those based on this formula, but they are more cumbersome to describe and thus we omit them here.

[58] In particular, the relative incentive for ex-post distortion is now scaled up by a ratio of the average anticipated to actual value of k^2.

since the distributions of quantities are exactly the same in both cases. Second, and conversely, if the latter equality is not satisfied, the social planner can costlessly tell (σ, m) and $(\widehat{\sigma}, \widehat{m})$ apart by asking the innovator to predict demand: see Osband (1989) for an illustration of how "scoring rules" can be employed to that effect.

The analysis can then be generalized. A particularly straightforward case in point is that of multiplicative (ξ_m, with $E(\xi_m) = 1$) shocks that are independent of (σ, m, c). That is,

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

so that

$$S(p;\, \sigma, m) = \xi_m \left[p\left(\sigma Q\left(\frac{p}{m}\right)\right) + \int_p^\infty \sigma Q\left(\frac{\widetilde{p}}{m}\right) d\widetilde{p}\right]$$

which is proportional to the welfare earlier. As shown in Online Appendix Subsection 4.3, our empirical formula must be altered in similar, but not identical, manner[59] to the case of ex-ante uncertainty. Now in calculating x we take not the ratio of realized prices to quantities but the ratio of prices to the best projection of the innovator of the quantity.

5.5 Sales manipulation

In line with the literature on advance market commitments and output subsidies 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 innovator 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.

With affiliated purchases, 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[60], 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 prize system ($p = 0$, T is an increasing function of q) is no longer feasible, let alone approximately optimal.

Recalling that the optimal transfer in the absence of such manipulations writes:

$$T = \hat{T}\left(\hat{k}\right) \quad \text{where} \quad \hat{k} \equiv p^{\frac{\varepsilon(a)}{1+\varepsilon(a)}}\, q^{\frac{1}{1+\varepsilon(a)}},$$

[59] Of course, both of these filters might be applied to reduce the degree of asymmetric information implied by the formula in Subsection 4.7 when appropriate.

[60] Otherwise its right Dini derivatives must both obey this bound.

this non-manipulability constraint can be rewritten as

$$x \geq \left(\frac{\hat{T}'}{(1+\epsilon(a))\,k} \right)^{\frac{1+\epsilon(a)}{1-\epsilon(a)}}$$

. A low mark-up leads to this being violated for a wide range of x values, unless the policy is highly unresponsive. Thus, the only feasible means of using market signals is allowing ex-post distortion, providing an closely related alternative to our rationale for market power: in this setting, screening even the size of the market requires market power. The monopoly solution ($a = 1$, and so $k = \sqrt{\pi}$, yielding $T(k) = k^2$) satisfies the non-manipulability constraint everywhere with equality.

The characterization of the optimal scheme under the non-manipulability constraint lies outside the scope of this paper. However the tools we develop here will likely be helpful in addressing it; we naturally conjecture that the optimal a is higher under the constraint than in its absence.

5.6 Multiple price observations

For analytical convenience, we have presumed that the social planner does not require the innovator to randomize over prices. This could facilitate sorting at lower distortion cost 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 (each announcement of (σ, m) generates a price distribution), 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. Similarly, for a durable good, Coasian arbitrage limits the forms of intertemporal discrimination that can be achieved (Coase, 1972; Bulow, 1982).

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.

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 innovator 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 innovator to choose demand path $\{\sigma_k^t\}_{t \in [0,1]}$ subject to the constraint $\int_0^1 \sigma_k^t dt = \sigma_k$ (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 and so the planner later finds out:

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

Rather than moving demand across time, the innovator 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 5.5: 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, innovator risk aversion also constrains the use of random schemes.[61]

The bottom line of this section is that the social planner must trade off the reduced cost

[61] Suppose for example that the innovator'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

of sorting and the cost of randomization when contemplating the use of multiple price observations (temporal, geographic, randomized). The basic insights of this paper seems highly robust. However the appropriate role of more sophisticated schemes to partially mitigate the costs of market power is an exciting direction for future research.

6 Other Applications

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

6.1 Platforms

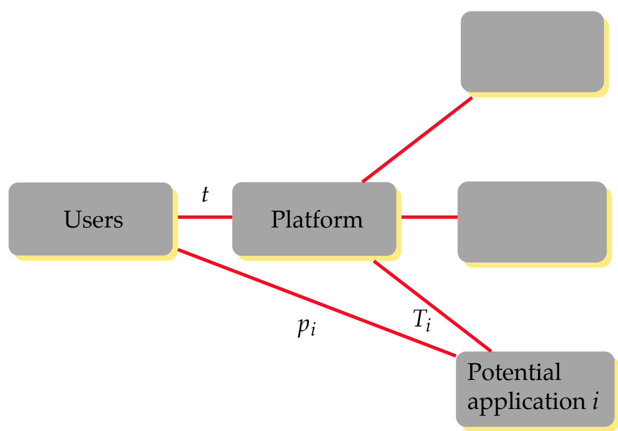

Figure 8: Application development incentives in a two-sided market

A two-sided platform[62] (as shown in Figure 8 above), such as an operating system, must attract multiple sides of a market, say end-users and application developers. As discussed in Subsection 4.6 above, one of the key decisions faced by such platforms 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, as described formally in Online Appendix Subsection 4.4, end-users differ in an idiosyncratic parameter of taste for the platform and there is a large number of

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

Again the probability of monopoly pricing cannot be too small.

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

applications. 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 5.1 above) carries through without any modification.

This analysis qualifies the classic result of Armstrong (1999) and Bakos and Brynjolfsson (1999), 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 iPhone App Store's policy may be nearly optimal if a few killer apps make most of the platform's value and the elasticity of innovation supply is high.

6.2 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 applications. 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 *by the profits on their product along*, even though this causes multi-marginalization problems for the firm, helping to address recent debates about such incentive schemes (Hunt and Lerner, 1995).

6.3 Infrastructure procurement

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, discussed in the example of Subsection 4.8, of known costs and consider the innovator as a political entrepreneur seeking funding from the public purse for a project on which she stakes her reputation. Our analysis above showed that, in this case, an optimal rule is to provide the funding if the entrepreneur agrees to accept public condemnation (and thus political ruin) if the project fails to achieve a threshold score and public praise (and perhaps reelection) if it passes that threshold. This score is based on a mix of revenue recovery and quantity consumed (and thus consumer surplus). The optimal degree of market power then simply represents the optimal elasticity of the threshold score with respect to each of these, which will determine the political entrepreneur's pricing incentives.

6.4 Platforms with heterogeneous externalities

The basic idea implicit in endogenous information structures has applications ranging even further away from the structure of our model. One application, being considered in work in progress by Veiga and Weyl (2010), is to the theory of multi-sided platforms, which typically makes the restrictive assumption that the externalities delivered by any participant on a particular side of the market are either identical or can be fully third-degree price discriminated. This clearly fails in newspapers, for example, where wealthy readers are more valuable to advertisers than are poor readers, but cannot be charged a different price. The difficulty to this point in solving a model with heterogeneous externalities has been analyzing the value that, say, adding an additional advertiser to a paper brings in terms of *sorting* rich from poor consumers. If rich marginal consumers dislike advertising more than poor marginal consumer do, for example, then advertisements will be less attractive than if these preferences were reversed. This effect can be quantified exactly through the sort of envelope theorem and covariance approaches we took above.

6.5 Applications of the stretch parameterization

We make extensive use of our stretch parameterization of demand, which significantly generalizes the linear specification of preferences typically used even in the most general multidimensional screening models (Rochet and Choné, 1998). In this subsection we briefly describe the breadth of this parameterization and some other potential interpretations and applications of it.

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) if the slope-of-pass-through parameter scales appropriately to m and limiting pass-through is common, any demand based on a statistical distribution with a constant location-to-scale parametrization[63] and a single-product version of AIDS (Deaton and Muellbauer, 1980). An increase in σ corresponds to an

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

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 (proportional decrease in their "price coefficient"), often represented (Berry et al., 1995) by a proportional increase in their income.

Given its intuitiveness, this parameterization seems likely to be useful in other areas of industrial organization and beyond; we provide a few examples. As discussed in Subsection 5.3 it seems a natural parameterization for demand morphing in the spirit of Johnson and Myatt (2006). Similarly, σ compared to m provides a simple manifestation of Bresnahan (1982)'s distinction between demand shifters and twisters. It 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. A monopolist will always have too little incentive to supply m holding fixed quantity[64] as most of the benefits of higher m accrue to infra-marginal consumers but, so long as demand is log-concave (linear-cost pass-through is less than 1), a monopolist will have excessive[65] incentive to supply σ. More broadly, in multi-dimensional screening stretch parameterizations of indifference curves may extend many of the useful properties traditionally attributed to linear indifference curves.

7 General Pricing

This section considers the generalization of our analysis to broader pricing rules.

7.1 Incentive compatibility with general pricing

Virtually all of our analysis above adopts the standard contract-theoretic approach of invoking the revelation principle and solving for optimal assignments of transfers and prices to types of innovators subject to incentive compatibility. The alternative (and equivalent) price theoretic approach (Bulow and Roberts, 1989; Wilson, 1993; Milgrom, 2004), often called the "demand

[64] Social value is $S\left(Q^{-1}\left(\frac{q}{\sigma}\right)\right)\sigma m$ 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.

[65] The social incentive is now

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

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 $\left(S(1) - Q(1)\right)m$ and the second to $Q(1)m$. $\frac{S(1)-Q(1)}{Q(1)}$ is the ratio of consumer to producer surplus at monopoly optimal prices, whose comparison to unity is dictated by the average pass-through rate at prices above the monopoly optimum (Weyl and Fabinger, 2009). At prices other than monopoly optimal a similar result may be shown, but we omit it in the interests of brevity.

profile approach in the context of non-linear pricing, is, for some problems, more fruitful. Among these is the analysis of incentive compatibility with more general pricing that we consider in this subsection. We therefore begin this subsection by developing such an approach, using it to derive more general sufficient conditions for incentive compatibility.

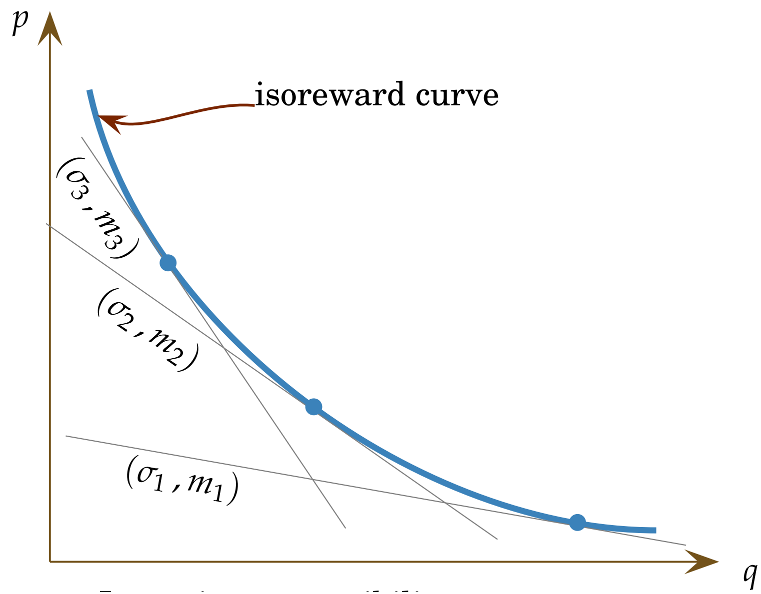

Figure 9: Isoreward curves in the (q, p) space under the price theoretic approach

Let us consider the analysis of $\hat{T}(q, p)$ directly. Suppose that for every (q, p) pair there is *some*[66] innovator whom we would like to produce q and charge p. By free disposal, $\hat{T}$ must be increasing in (q, p) for this to be feasible. If innovator (σ, m) is to choose a quantity price pair $\left(\sigma Q\left(\frac{p}{m}\right), p\right)$ then the isoquant (isoreward curve) of $T(q, p)$ must be tangent to (σ, m)'s demand curve at $\left(\sigma Q\left(\frac{p}{m}\right), p\right)$ (up to a set of such curves of measure zero; see Subsection 7.1 below). Thus, the simple demand curves play the role of imitation frontiers in the price theory approach. This is pictured in Figure 9, where isoreward curves form the upper envelope of demand curves at the points along these curves assigned to the respective innovators.

Proportional pricing would impose that we wish to implement $p = am$ for some constant $a \in [0, 1]$. However sufficient incentive compatibility conditions are straightforward to interpret for general incentive compatible pricing rules based on the price theory approach. We thus move immediately to these.

Suppose we wish to implement some continuous mapping $(q, p) : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ which prescribes the quantity $q(\sigma, m)$ and price $p(\sigma, m)$, which any innovator is instructed to produce and charge, respectively. Clearly to be implementable the price-quantity pair must lie along the appropriate demand curve:

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

[66] Note that even if this fails it is irrelevant and we can always restore monotonicity, as any time there *is* an innovator at (q, p) the reward there must be greater than at any point dominated by this.

What else is required of (q, p) to be implementable? Suppose that q were to be decreasing in σ or p decreasing in m. This would effectively ask types with a comparative advantage in producing demand or prices to produce less of these than another type with a comparative disadvantage. The classic logic of Spence (1973) and Mirrlees shows this may not be incentive compatible. Moreover, in two variables individual (weak) monotonicity in each argument is not sufficient (McAfee and McMillan, 1988); if (q, p) is differentiable the Jacobian of the transformation must be positive semidefinite (it cannot flip points across quadrants in coordinate systems based at any point). Such a differentiable mapping is known as a monotone orientation-preserving weak self-diffeomorphism (weak MOPSD) of $\mathbb{R}^2_+$; a strict MOPSD has a positive definite Jacobian.

While this condition is quite imposing as a formal statement, it is really just the most natural two-dimensional generalization of the standard monotonicity condition for one-dimensional implementation. In algebraic terms it simplify states that $\epsilon\epsilon_{a_\sigma} < 1$ (where ϵ_{a_σ} is the elasticity of a with respect to σ), $\epsilon_{a_m} > -1$ and that $\epsilon\epsilon_{a_\sigma} - \epsilon_{a_m} < 1$. The first condition is that increasing σ should not increase prices so far as to offset the direct increase in quantity this causes, so that q remains monotone in σ. The second condition posits that a should not fall so rapidly in m that p is actually declining in m. The final condition is equivalent to the condition that moving along the local demand function towards higher prices also moves towards higher m and lower σ. This last is the requirement, in addition to monotonicity, that ensures the preservation of orientation.

Lemma 3. *Suppose that (q, p) is a strict MOPSD of $\mathbb{R}^2_+$ with $q(\sigma, m) = \sigma Q\left(\frac{p(\sigma,m)}{m}\right)$. Then the conditions in Lemma 1 are necessary and sufficient for incentive compatibility. Almost conversely if (q, p) is differentiable then any incentive compatible $\hat{T}$ implementing (q, p) is constant over any neighborhood where (q, p) fails to be MOPSD.*

Thus, if we are willing[67] to assume (q, p) is differentiable, we may use the isoreward approach except when there is bunching (a weak MOPSD or no MOPSD at all with flat $\hat{T}$). However, any weak MOPSD is the limit of a series of strict MOPSD and thus little is lost by restricting attention to the latter.

Proof. See Online Appendix Subsection 5.1. □

7.2 Substantive results with general pricing

Building on the preceding section we may now investigate the properties of more general, incentive compatible pricing rules. Our analysis is less complete than under proportional pricing, but we provide four results. First, we present a general first-order derivative with respect to adjusting prices at any point and use it to provide a generalization of Theorem 1. Second, we discuss a version of our analysis that may apply even absent proportional pricing: a

[67]Fabinger and Weyl (2011) are working to relax this and other technical assumptions.

first-order derivative with respect to the overall level of market power. Third, we use a similar strategy to prove a generalized version of Friedman's Conjecture and its converse. Finally, we discuss which direction, starting from proportional pricing, may be optimal to move the price schedule. This provides some limited justification for proportional pricing. For simplicity we restrict attention to the case when $T^\star$ is differentiable, but, as shown in Appendix C below, little changes if this fails.

Just as in Section 4, under general pricing we can change variables from (σ, m) to (k, x). Here k again represents the isoreward curve measured by the point at which it intersects the $45°, \sigma = m$ line and x is again $\frac{m}{\sigma}$. However, now that pricing is not proportional, demand elasticity is not the same at all points. It is thus necessary to consider, for any point (k, x), the elasticity of demand at that point $\epsilon(k, x)$; we also use this convention for other quantities such as S and Q. However, not only the elasticity at a point is relevant; the value of k corresponding to any (σ, m) is determined by the elasticity *at every point along the isoreward curve passing through* (σ, m) between x and the 45° line. Thus, one must consider the *average*[68] *elasticity of demand*:

$$\tilde{\epsilon}(k, x) \equiv \frac{1}{\overline{\frac{1}{1+\epsilon}}(k, x)} - 1$$

where

$$\overline{\frac{1}{1+\epsilon}}(k, x) \equiv \int_{z=1}^{x} \frac{1}{z \log(x) \left[1 + \epsilon(k, z)\right]} dz$$

An innovation's social value is, after some algebraic manipulations, $S\left(k, x\right) k^2 x^{\frac{1-\tilde{\epsilon}(k,x)}{1+\tilde{\epsilon}(k,x)}}$. The benefits of raising a again arise from sorting and those of lowering it from reducing ex-post inefficiency, both locally holding fixed rewards given to each k by the envelope theorem; the optimum balances these two incentives. However for general pricing we must consider this trade-off at each (σ, m), or (k, x), pair.

Proposition 6: *If $T^\star$ is differentiable in k, the first-order net benefit of increasing a at $\left(\hat{k}, \hat{x}\right)$ beginning from a strict MOPSD pricing policy $a(\cdot, \cdot)$ is, if $x \geq 1$, proportional to*

$$\underbrace{\frac{T^{\star\prime}}{\hat{k}} \frac{\epsilon'\left(\hat{k},\hat{x}\right)}{\hat{x}\left[1+\epsilon\left(\hat{k},\hat{x}\right)\right]^2} \left[1+\log(\hat{x}) \overline{\frac{\epsilon'\left(\epsilon_{a_m}-\epsilon_{a_\sigma}\right)}{(1+\epsilon)^2}}\left(\hat{k},\hat{x}\right)\right] \left(\frac{E_{\tilde{f},x>\hat{x}}\left[Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]}{E_{\tilde{f},x}\left[Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]} - 1\right) \eta\left(T^\star|\hat{k}\right)}_{sorting} - \underbrace{Q\epsilon\hat{x}^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}} H\left(\hat{x}|\hat{k},T^\star\right) \frac{E\left[\tilde{f}\left(\hat{x}|\hat{k},c\right)|c<T^\star,\hat{k}\right]}{\tilde{f}\left(x|c=T^\star,\hat{k}\right)}}_{ex\text{-}post\ distortion}$$

where H is the (conditional) hazard rate of x under $\tilde{f}$ and if $x < 1$, proportional to

$$\frac{T^{\star\prime}}{\hat{k}} \frac{\epsilon'\left(\hat{k},\hat{x}\right)}{\hat{x}\left[1+\epsilon\left(\hat{k},\hat{x}\right)\right]^2} \left[1+\log(\hat{x}) \overline{\frac{\epsilon'\left(\epsilon_{a_m}-\epsilon_{a_\sigma}\right)}{(1+\epsilon)^2}}\left(\hat{k},\hat{x}\right)\right] \left(\frac{E_{\tilde{f},x>\hat{x}}\left[Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]}{E_{\tilde{f},x}\left[Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]} - 1\right) \eta\left(T^\star|\hat{k}\right) - Q\epsilon\hat{x}^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}} H\left(\hat{x}|\hat{k},T^\star\right) \frac{E\left[\tilde{f}\left(\hat{x}|\hat{k},c\right)|c<T^\star,\hat{k}\right]}{\tilde{f}\left(x|c=T^\star,\hat{k}\right)}$$

where R is the reversed hazard rate. A necessary condition for a strict MOPSD (no-bunching)

[68]As described in Online Appendix Subsection 5.2, this is a (transformed) log-average elasticity along the isoreward curve from point (k, x) to point $(k, 1)$.

solution is that these equal zero at every point in the plane.

This is just the standard calculus-of-variations first-order derivative at a point in this context, using the envelope theorem.

Proof. See Online Appendix Subsection 5.2. □

Three things may be gleaned from these relatively dense expressions. First, despite their complexity, the same basic forces are at work as with proportional pricing. The first term is the product of materialism and the degree to which raising a is able to sort for the best innovations. This is quite naturally measured by the ratio of the social value of average innovations with a higher value of x than $\hat{x}$, along isoreward $\hat{k}$, to the social value of an overall-average innovation along that isoreward curve. Thus, the basic logic of our analysis carries through more generally. Second, there seems be a strong indication, discussed more extensively later, that along an isoreward curve a will optimally decline in x (decline in m and/or increase in σ), at least over a significant range. We will discuss the reasoning behind this more extensively below.

Finally, supposing this is the case, it is worth noting that this creates both greater, and lesser, selective pressure in favor of high x innovations. On the one hand, it raises selective pressure as their S values are higher as well. On the other hand in this case ϵ_{a_m} is likely smaller than ϵ_{a_σ} (so that a declines in x along k) so that the rather odd term

$$\log(\hat{x})\overline{\frac{\epsilon'\left(\epsilon_{a_m}-\epsilon_{a_\sigma}\right)}{(1+\epsilon)^2}}\left(\hat{k},\hat{x}\right)$$

becomes negative and thus depresses the incentives for market power. The basic source of this term, explained extensively in Online Appendix Subsection 5.2, is that changing pricing alters the set of prices through which (σ, m)'s isoreward curve passes as it approaches the 45° line and thus indirectly affects the rewards given to (σ, m) to the extent that a is not constant.

This analysis may be used to generalize Theorem 1.

Theorem 4: *At global ex-post efficient pricing there is a local incentive at all points to raise prices at any (σ, m) for which $T^\star(\sigma)$ is not constant in the neighborhood of σ. At global monopoly pricing there is a local incentive at all points to lower prices.*

Proof. See Online Appendix Subsection 5.2. □

Another way of recovering some broader results of our analysis is to focus on its primary goal: determining the optimal "level", rather than structure, of market power. One of the necessary conditions for optimality is, of course, choosing this level correctly. In particular, we can consider the first-order costs and benefits of uniformly lowering $\frac{1}{1+\epsilon}$ at every point by a small amount; this particular direction is chosen for the analytic simplifications it allows. This gives a similar expression to our baseline proportional pricing first-order condition, as shown in the following proposition. However the additional forces identified above still play a role.

Proposition 7: *Starting from any strict MOPSD policy a with $a < 1$ everywhere and assuming $T^\star$ is differentiable at this policy, the first variation of W in the direction $a + \frac{(1+\epsilon)^2}{\epsilon'}$ (uniform decrease in $\frac{1}{1+\epsilon}$) is proportional to*

$$E_{k,\tilde{f}}\left[k^4\left(\frac{\eta T^{\star\prime}}{k}\,Cov_{x,\tilde{f}}\left[\frac{\epsilon'}{(1+\epsilon)^2}\left(1+\overline{\log\left(\frac{x}{z}\right)(\epsilon_{a_m}-\epsilon_{a_\sigma})}\right)\log(x), Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]\right)-E_{\tilde{f},x}\left[Q\epsilon x^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\middle| k, c<T^\star(k)\right]E_{\tilde{f},x}\left[Sx^{\frac{1-\tilde{\epsilon}}{1+\tilde{\epsilon}}}\right]\right]$$

Proof. See Online Appendix Subsection 4.3. □

Similarly, we may consider the benefits of moving towards monopoly pricing from *any pricing policy* arbitrarily close to it, or towards ex-post efficiency from any pricing policy arbitrarily close to it, to derive a general version of Friedman's Conjecture, or its converse.

Theorem 5: *Beginning from any a sufficiently, uniformly close to uniform monopoly pricing ($a = 1$ everywhere) but with $a < 1$ everywhere, if V_1 (of Theorem 2) is sufficiently large there are first-order benefits from moving a small amount (uniformly) towards uniform monopoly pricing. Beginning from any strict MOPSD a sufficiently, log-uniformly close to ex-post efficiency ($a = 0$ everywhere) but with $a > 0$ everywhere, if V_0 (of Theorem 3) is sufficiently small there are first-order benefits from moving a small amount (uniformly) towards ex-post efficiency.*

The proof is effectively identical to that of Friedman's Conjecture and its converse, with only slight complexities in simplifying the more general first-order condition local to the proportional policies of ex-post efficiency and monopoly pricing.

Proof. See Online Appendix Subsection 4.3. □

Finally, to provide at least some notion of what the optimal structure of market power may look like, we can consider evaluating the first-order condition for optimal pricing at each point, beginning from optimal proportional pricing. This provides at least some indication of the optimal structure of market power.

Corollary 4: *If $T^\star$ is differentiable in k, the first-order net benefit of increasing a at $\left(\hat{k}, \hat{x}\right)$ given beginning from a proportional pricing policy a is, if $x \geq 1$, proportional to*

$$\frac{\epsilon' T^{\star\prime}\left(\hat{k}\right)}{(1+\epsilon)^2\,\hat{k}\hat{x}^{\frac{2}{1+\epsilon}}}\left(\frac{E_{\tilde{f},x>\hat{x}}\left[x^{\frac{1-\epsilon}{1+\epsilon}}\middle|\hat{k},T^\star\right]}{E_{\tilde{f},x}\left[x^{\frac{1-\epsilon}{1+\epsilon}}\middle|\hat{k},T^\star\right]}-1\right)\eta\left(T^\star|\hat{k}\right)-Q\epsilon H\left(\hat{x}|\hat{k},T^\star\right)\frac{E\left[\tilde{f}\left(\hat{x}|\,\hat{k},c\right)\middle| c<T^\star,\hat{k}\right]}{\tilde{f}\left(x|\,T^\star,\hat{k}\right)}$$

and if $x < 1$

$$\frac{\epsilon' T^{\star\prime}\left(\hat{k}\right)}{(1+\epsilon)^2\,\hat{k}\hat{x}^{\frac{2}{1+\epsilon}}}\left(1-\frac{E_{\tilde{f},x<\hat{x}}\left[x^{\frac{1-\epsilon}{1+\epsilon}}\middle|\hat{k},T^\star\right]}{E_{\tilde{f},x}\left[x^{\frac{1-\epsilon}{1+\epsilon}}\middle|\hat{k},T^\star\right]}\right)\eta\left(T^\star|\hat{k}\right)-Q\epsilon R\left(\hat{x}|\hat{k},T^\star\right)\frac{E\left[\tilde{f}\left(\hat{x}|\,\hat{k},c\right)\middle| c<T^\star,\hat{k}\right]}{\tilde{f}\left(x|\,T^\star,\hat{k}\right)}$$

With some additional simplifying assumptions these formula may give some insight. Assume that k, x and c are all independent, which implies that $T^\star$ is quadratic in k. Furthermore, suppose that $\frac{E_{\tilde{f},\hat{x}}\left[\hat{x}^{\frac{1-\epsilon}{1+\epsilon}}\middle|\hat{x}>x,k,c=T^\star(k)\right]}{E_{\tilde{f},\hat{x}}\left[\hat{x}^{\frac{1-\epsilon}{1+\epsilon}}\middle|k,c=T^\star(k)\right]}-1=\gamma x^{\frac{1-\epsilon}{1+\epsilon}}$ as in a generalized Pareto distribution. Then the first formula simplifies to

$$\frac{\gamma\epsilon'\eta}{(1+\epsilon)^2 x}-Q\epsilon H(x)$$

This setting implies that the local incentives for distortion are independent of k. However it also gives a sense that, after some point at least, the local incentive for raising prices is likely to decline in x along an isoreward curve. This occurs for two reasons. First, if we have the standard increasing hazard rate condition, H grows in x. This is just the logic behind the classic "no distortion at the top" result of Mirrlees (1971), though based on sorting rather than rent extraction. The sorting benefits of higher prices only affect innovations with higher x than that at which prices are distorted; if the mass of such upper-tail innovations shrinks relative to that of those being distorted as hazard rates are increasing, there will be little incentive to create such distortion. However there is also an additional motive for lower distortions for higher x here: the multiplicative, log-linear nature of the stretch parameterization[69] implies that the upper tail is less affected by shifts in elasticity at high x values than those at low values (above unity). In some sense the closer x is to 1 the more dramatically a change in the elasticity at that point shifts the isoreward curve and therefore the greater its sorting value. While for $x<1$ things are a bit subtler, there seems to be some weak, but general, indication[70] that optimal policy calls for a values declining along isoreward curves.

8 Conclusion

This paper aspires to make three contributions. First, in terms of modeling, we develop a multidimensional screening framework and introduce the intuitive stretch parameterization of demand to formalize Smith's argument that market power helps screen for high surplus innovations. Second, on a technical level, we develop techniques (the isoreward approach and application of the envelope theorem) to derive a solution for the broad but unstudied class of multidimensional screening problems with endogenous information structures. Finally, and

[69]It would be interesting to know if this is a more general property of multidimensional screening problems.

[70]This extremely tentative conclusion merits two comments. First, if this is the case it might point towards bunching (a weak MOPSD) being optimal as a declining with rising x pushes up against order-preservation and/or monotonicity. This would be consistent with the arguments of Armstrong (1996) and Rochet and Choné (1998) that bunching is quite common in multidimensional screening and is an interesting observation from an economic perspective, as it would imply the easy-to-implement institution of price controls potentially forming part of optimal policy. Second, it provides at least some reassurance that proportional pricing is not too wildly off as a policy prescription. There are limits to how rapidly a may increase with x and still be a MOPSD and therefore incentive compatible; had it been optimal for a to decrease in x there would have been no limit to how far off proportional pricing might have been. Obviously all of this analysis is exceptionally preliminary and the more complete analysis of our model with general pricing remains an important topic for future research.

substantively, we quantify a notion of "the value of materialistic genius", which we show is tightly connected to the optimality of market power as a reward for innovation, making precise 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 our work opens a number of promising directions for future research. First, our general formula ought to be calibrated empirically in specific industries. Second, several extensions would test the robustness of our insights: the demand function could be generalized beyond the stretch representation 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 inventor limited liability or risk aversion), as well as richer asymmetric information (duration, marketing, price discrimination, marginal costs) and richer instruments for screening demand to accompany would be welcome. In particular, the latter would be crucial in allowing our analysis to shed light on the design of patent length and breath. Finally, the extension of our techniques to accommodate R&D races, licensing competition and cumulative innovation, and thus tradeoffs between regulation and competition policy in these settings, 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 (Rey and Tirole, 2007); 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.

References

Acemoglu, Daron, David Cutler, Amy Finkelstein, and Joshua Linn, "Did Medicare Induce Pharmaceutical Innovation," *American Economic Association Papers & Proceedings*, 2006, *96* (2), 103–107.

Aghion, Philippe, Mathias Dewatripont, and Jeremy Stein, "Academic Freedom, Private-Sector Focus, and the Process of Innovation," *RAND Journal of Economics*, 2008, *39* (3), 617–635.

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Armstrong, Mark, "Multiproduct Nonlinear Pricing," *Econometrica*, 1996, *64* (1), 51–75.

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

Athenaeus, *Deipnosophistae (The Scholars at Dinner)* c. 200-300.

Babayev, Djangir A., "Piece-Wise Linear Approximation of Functions of Two Variables," *Journal of Heuristics*, 1997, *2* (4), 313–320.

Baker, George P., "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 1992, *100* (3), 598–614.

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

Barder, Owen, Michael Kremer, and Heidi Williams, "Advance Market Commitments: A Policy to Stimulate Investment in Vaccines for Neglected Diseases," *The Economists' Voice*, 2006, *3* (3).

Baron, David P., "Demand Uncertainty in Imperfect Competition," *International Economic Review*, 1971, *12* (2), 196–208.

_ **and Roger B. Myerson**, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 1982, *50* (4), 911–930.

Basov, Suren, *Multi-dimensional Screening*, Berlin and Heidelberg, Germany: Springer-Verlag, 2005.

Berndt, Ernst, Rachel Glennerster, Michael Kremer, Jean Lee, Ruth Levine, Georg Weizsäcker, and Heidi Williams, "Advance Market Commitments for Vaccines Against Neglected Diseases: Estimating Costs and Effectiveness," *Health Economics*, 2007, *16* (5), 491–511.

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

Bessen, James and Eric Maskin, "Sequential Innovation, Patents and Imitation," *RAND Journal of Economics*, 2009, *40* (4), 61–635.

Boldrin, Michele and David K. Levine, *Against Intellectual Monopoly*, Cambridge, UK: Cabmridge University Press, 2008.

_ **and David Levine**, "The Case against Intellectual Property," *American Economic Review*, 2002, *92* (2), 209–212.

Bresnahan, Timothy F., "The Oligopoly Solution Concept is Identified," *Econoimcs 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.

Bulow, Jeremy I., "Durable-Goods Monopolists," *Journal of Political Economy*, 1982, *90* (2), 314–332.

_ **and John Roberts**, "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, 1989, *97* (5), 1060–1090.

_ **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.

Coase, Roland H., "Durability and Monopoly," *Journal of Law and Economics*, 1972, *15* (1), 143–149.

Coase, Ronald H., "The Marginal Cost Controversy," *Economica*, 1946, *13* (51), 169–182.

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.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 1977, *67* (3), 297–308.

Dubois, Pierre, Olivier de Mouzon, Fiona Scott-Morton, and Paul Seabright, "Market Size and Pharmaceutical Innoation," 2010. This paper is in preparation and a draft is available on request from the authors.

Dupuit, Arsène Jules Étienne Juvénal, *De la Mesure de L'utilité des Travaux Publics*, Paris, 1844.

Fabinger, Michal and E. Glen Weyl, "A Multidimensional Envelope Theorem with Endogenous Choice Sets," 2011. This paper is in preparation.

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.

Gibbard, Alan, "Manipulation of Voting Schemes: A General Result," *Econometrica*, 1973, *41* (4), 587–602.

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

Gladwell, Malcolm, "The Ketchup Conundrum," *New Yorker*, 2004, *September 6.*

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

Gregory, John and Kathleen Pericak-Spector, "A Bliss-Type Multiplier Rule for Constrained PDE Variational Problems," *Utilitas Mathematica*, 1999, *56*, 143–151.

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.

Harrod, R. F., "The Law of Decreasing Costs," *The Economic Journal*, 1931, *41* (164), 566–576.

Hausman, Jerry A., "Valuation of New Goods under Perfect and Imperfect Competition," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, Vol. 58 of National Bureau of Economic Research*, Studies in Income and Wealth*, Chicago and London: University of Chicago Press, 1997, pp. 209–37.

Hayek, F. A., *Individualism and Economic Order*, Chicago: University of Chicago Press, 1948.

Holmström, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, 1979, *10* (1), 74–91.

Holthausen, Duncan M., "Input Choices and Uncertain Demand," *American Economic Review*, 1976, *66* (1), 94–103.

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.

Hotelling, Harold, "The Economics of Exhaustible Resources," *Journal of Political Economy*, 1931, *39* (2), 137–175.

_ , "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, 1938, *6* (3), 242–269.

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.

Kirzner, Israel M., *Competition and Entrepreneurship*, Chicago: University of Chicago Press, 1973.

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.

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

Kultti, Klaus, Tuomas Takalo, and Juuso Toikka, "Secrecy versus Patenting," *RAND Journal of Economics*, 2007, *38* (1), 22–42.

Laffont, Jean-Jacques and Jean Tirole, *A Theory of Incentives in Procurment and Regulation*, Cambridge, MA: MIT Press, 1993.

Leland, Hayne E., "Theory of the Firm Facing Uncertain Demand," *American Economic Review*, 1972, *62* (3), 278–291.

Lerner, A. P., "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, 1934, *1* (3), 157–175.

Lewis, Tracy R. and David E. M. Sappington, "Supplying Information to Facilitate Price Discrimination," *International Economic Review*, 1994, *35* (2), 309–327.

Lloyd, P. J., "The Origins of the von Thüen-Mill-Pareto-Wicksell-Cobb-Douglas Function," *History of Political Economy*, 2001, *33* (1), 1–19.

Loury, Glenn C., "Market Structure and Innovation," *Quarterly Journal of Economics*, 1979, *93* (3), 395–410.

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

McAfee, R. Preston and John McMillan, "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 1988, *46* (2), 335–354.

Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 1981, *12* (2), 380–391.

_ , *Putting Auction Theory to Work*, Cambridge, UK: Cambridge University Press, 2004.

_ **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.

Mill, John Stuart, *Principles of Political Economy with some of their Applications to Social Philosophy*, London: John W. Parker, West Strand, 1848.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

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

Murray, Fiona, Philippe Aghion, Mathias Dewatripont, Julian Kolev, and Scott Stern, "Of Mice and Academics: Examining the Effect of Openness on Innovation," 2009. http://www.economics.harvard.edu/faculty/aghion/files/Ofemics.pdf.

Nordhaus, William D., *Invention, Growth, and Welfare*, Cambridge, MA: MIT Press, 1969.

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*, 2004, *7* (1), 1–32.

Osband, Kent, "Optimal Forecasting Incentives," *Journal of Political Economy*, 1989, *97* (5), 1091–1112.

Ottaviani, Marco and Andrea Prat, "The Value of Public Information in Monopoly," *Econometrica*, 2001, *69* (6), 1673–1683.

Pigou, Arthur C., *The Economics of Welfare*, London: MacMillan, 1920.

Rand, Ayn, *Atlas Shrugged*, New York: Penguin, 1957.

_ , "Question and Answer session following lecture "The Anti-Industrial Revolution"," in Robert Mayhew, ed., *Ayn Rand Answers: The Best of Her Q&A*, New York: New American Library, 1970.

Rey, Patrick and Jean Tirole, "A Primer on Foreclosure," in Mark Armstrong and Robert H. Porter, eds., *Handbook of Industrial Organization*, Vol. 3, Amsterdam: North-Holland, 2007, pp. 2145–2220.

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.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1974, *82* (1), 34–55.

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

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *The Review of Economic Studies*, 2001, *68* (1), 205–229.

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

Scotchmer, Suzanne, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 1991, *5* (1), 29–41.

_ , "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.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations*, London, 1776.

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

Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

_ , "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 1976, *43* (2), 217–235.

Stern, Scott, "Incentives and Focus in University and Industrial Research: The Case of Synthetic Insulin," in Nathan Rosenberg, Annetine C. Gelijns, and Holly Dawkins, eds., *Sources of Medical Technologies: Universities and Industry*, Washington, DC: Natoinal Academy Press, 1995, chapter 7.

Tirole, Jean, *Theory of Industrial Organization*, Cambridge, MA: MIT Press, 1988.

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

Veiga, Andre and E. Glen Weyl, "Multi-Sided Platforms with Heterogeneous Externalities," 2010. This work is currently in progress. Contact Glen Weyl at weyl@fas.harvard.edu for a copy of our notes.

Vickrey, William, "Some Objections to Marginal-Cost Pricing," *Journal of Political Economy*, 1948, *56* (3), 218–238.

Viner, Jacob, "Cost Curves and Supply Curves," *Zietschrift für Nationalökonomie*, 1931, *3*, 23–46.

Weyl, E. Glen and Michal Fabinger, "Pass-Through as an Economic Tool," 2009. http://www.fas.harvard.edu/~weyl/research.htm.

Williams, Heidi, "Intellectual Property Rights and Innovation: Evidence from the Human Genome," 2010. http://www.nber.org/~heidiw/papers/5_12_10a_hlw.pdf.

Wilson, Robert B., *Nonlinear Pricing*, Oxford: Oxford University Press, 1993.

Wright, Brian D., "The Economics of Invention Incentives: Patents, Prizes, and Research Contracts," *American Economic Review*, 1983, *73* (4), 691–707.

Young, W. H. and Grace Chisolm Young, "On the Discontinuities of Monotone Functions of Several Variables," *Proceedings of the London Mathematical Society*, 1924, *s2-22* (124–142).

Appendix

A Supply, demand and optimal transfers

In this subsection we consider, given a fixed value of a, the problem of solving for the monotonicity-relaxed optimal $T^{\star\star}(k;a)$ function and then the conditions under which this is monotone and thus is the monotonicity-constrained optimal $T^{\star}(k;a)$.

We begin with a change of variables from (σ, m) to (k, x); this requires transforming the distribution of values according $\tilde{f}(k,x,c;a) \equiv f\left(kx^{-\frac{\epsilon(a)}{1+\epsilon(a)}}, kx^{\frac{1}{1+\epsilon(a)}}, c\right) kx^{-\frac{2\epsilon(a)}{1+\epsilon(a)}}$. We can then rewrite the social planner's problem, with the substitution and by switching the order of integration, as

$$\max_{\tilde{T}(\cdot)} \int_k \int_{c=0}^{\tilde{T}(k)} \int_x \left(k^2 x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} S(a)a - c\right) \tilde{f}(k,x,c;a)\,dx\,dc\,dk \tag{9}$$

subject to the constraint that $\tilde{T}(\cdot)$ is monotonically increasing, which we ignore for the remainder of this subsection and return to in the following subsection.

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:

$$F(T; k, a) \equiv \frac{\int_{c=0}^{T} \int_x \tilde{f}(k, x, c; a) dx dc}{\int_{c=0}^{\infty} \int_x \tilde{f}(k, x, c; a) dx dc}$$

This is the fraction of innovations that will be created if a reward T is offered along this curve. $\Sigma(r; k, a) \equiv F^{-1}(r; k, a)$ 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

$$\overline{S}(c; k, a) \equiv k^2 S(a) E_{x, \tilde{f}} \left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} | k, c \right] \equiv k^2 S(a) \cdot \frac{\int_x x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} \tilde{f}(k, x, c; a) dx}{\int_x \tilde{f}(k, x, c; a) dx} \tag{10}$$

Then $D(r; k, a) \equiv \overline{S}\left(\Sigma(r; k, a); k, a\right)$ is the average value of a marginal innovation lying along isoreward curve k, given that a fraction q of innovations lying on that isoreward curve have been created. We refer to this as the social inverse *demand*[71] *for innovations*.

The optimal reward along the isoreward curve can then simply be found as the intersection of the supply and demand curves for innovations, assuming these intersect only once. In fact, if c and x are independent given k, the demand for innovations is flat so the optimal reward is simply its value for all r. It is clear from equation (10) that $D(r; k, a)$ is constant in r if either c is independent of x given k, if a is close to 1 or both. More generally for $a < 1$, 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 a. 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 a.

These optimal rewards, however, ignore the monotonicity constraint that higher k isoreward curves must receive higher rewards. Because higher k isoreward curves move out toward higher values of σ and m, it is natural that they should have higher average social value for any given cost and thus higher optimal rewards. However, when a is relatively low, if k is sufficiently negatively affiliated with x given c, optimal rewards unconstrained by monotonicity may be decreasing in k; ruling out such strong negative affiliation ensures that the relaxed solution is in fact optimal.

Proposition 8: *Suppose that for all k, c and a fixed a,*

$$Cov_{x, \tilde{f}} \left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}, \frac{\partial \log(f)}{\partial c} \middle| k, c \right] \leq \frac{1}{k^2 S(a)} \tag{11}$$

and

$$2E_{x, \tilde{f}} \left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} | k, c \right] \geq -k \, Cov_{x, \tilde{f}} \left[x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}}, \frac{\partial \log(\tilde{f})}{\partial k} \middle| k, c \right] \tag{12}$$

Then the optimal reward function $T^{\star}(k; a)$, given a, is defined for each k by the unique value at which $D(\cdot; k, a)$ and $S(\cdot; k, a)$ intersect if $D(0; k, a) > S(0; k, a), D(1; k, a) < S(1; k, a)$ for

[71] Note that this approach directly incorporates one of the key incentive constraints, namely that rewards cannot depend on costs and thus that the first innovations created along any isoreward curve will be the cheapest. This is analogous to the standard "integration by parts" approach to incentive constraints in screening problems.

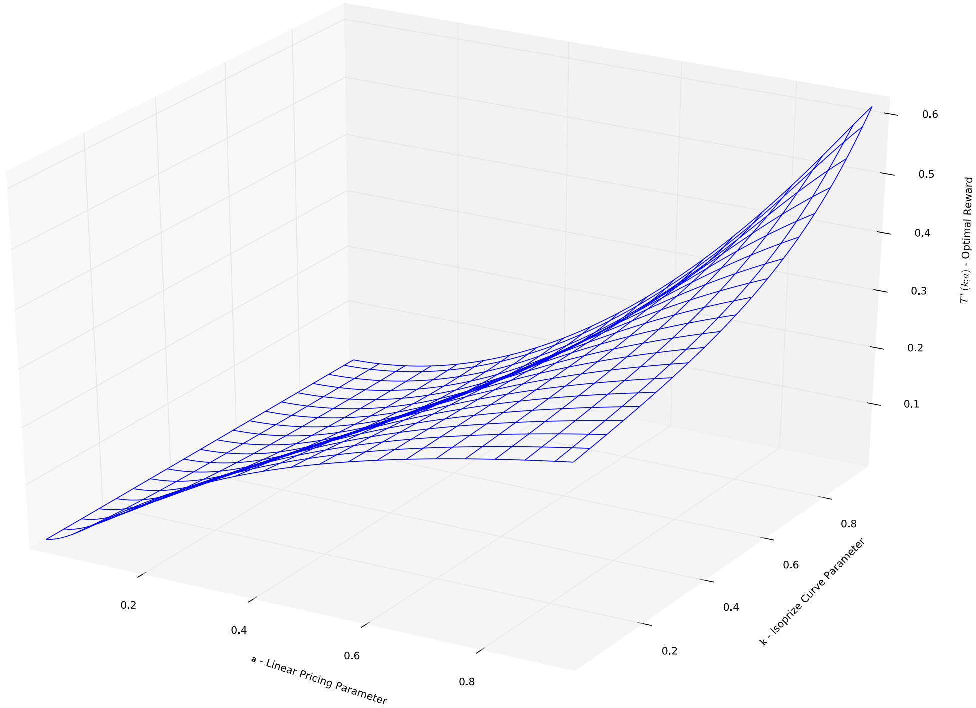

Figure 10: $T^{\star\star}(k;a)$ in our simulation as described in Subsection E

all k, a. If $D(0; k, a) \leq S(0; k, a)$ then the optimal reward is 0 and if $D(1; k, a) \geq S(1; k, a)$ then the optimal reward is anything exceeding the maximal possible cost $\overline{c}$ given a and k, or infinite if no such cost exists.

These conditions are intuitive extensions[72] 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 (or some monotone function of it) increases $\frac{\partial \log(f)}{\partial c}$ also increases, this exactly represents x having a strong monotone likelihood ratio relationship (Milgrom, 1981) with c. Thus, condition (11) can be viewed as stating that c and x are not "too" affiliated (Milgrom and Weber, 1982), while condition (12) can be seen as stating that x is not too negatively affiliated with k. Note that the simple approach is always valid for sufficiently high values of a. This is illustrated by Figure 10, which shows $T^{\star\star}(k; a)$ for the simulation we describe in Subsection 6 below. For high values of a it is monotone increasing, but must be ironed for low values of a.

Proof. See Online Appendix Subsection 3.1. $\square$

As discussed in Subsection 3.5, a (grossly) sufficient condition to ensure that when $a = 0$, $T^\star$ is not flat and therefore that $a^\star \in (0, 1)$ is obedience of inequality 12 when $a = 0$. This implies that at this pricing policy, reward monotonicity ironing is unnecessary. At $a = 0$ this condition simplifies to (at each (σ, c)):

$$2E_{mf}\left[m|\sigma, c\right] \geq -\sigma \mathrm{Cov}_{m,f}\left[m, \frac{\partial \log(f)}{\partial \sigma}\middle| \sigma, c\right]$$

[72]At some we hope to reduce these conditions to applications of the Rogerson result.

That is, σ is not too negatively affiliated with m. Again this condition is grossly sufficient; all that needs to be avoided is ironing *over the entire range of the curve*; this condition rules out ironing *anywhere along the curve*.

B Ironing

If these regularity conditions do not hold, some ironing is necessary to derive the optimal $T^\star$. Here, we briefly discuss the classical ironing techniques that can be applied if either of these conditions is violated individually.

First, suppose that condition (11) is violated, but (12) 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 in the spirit of Hotelling (1931) relative to the supply curve.

Proposition 9: *If condition (11) is violated but condition (12) obeyed, then $T^\star$ is determined, as in Proposition 8, but with the "Hotelling ironed" demand curve that never increases relative to the supply curve replacing the demand curve.*

Proof. See Online Appendix Subsection 3.2. □

Now suppose that condition (12) is violated while condition (11) is maintained. Then $T^{\star\star}$ may be non-monotone. However, the social value created along each isoreward curve is concave in the reward given along that curve by condition (12): supply grows relative to demand as quantity increases. This is exactly the conditions required to use the Guesnerie and Laffont (1984) procedure to iron $T^{\star\star}$ into a monotone $T^\star$.

Proposition 10: *If condition (11) is obeyed but condition (12) violated, then $T^\star$ is the Guesnerie and Laffont (1984) ironing of $T^{\star\star}$.*

Proof. See Online Appendix Subsection 3.2. □

When both of these conditions fail, ironing is needed but the social value created along each isoreward curve need not be concave. Little is known (Toikka, Forthcoming) about how to solve mechanism design problems of this form and making an attempt to do so is outside the scope of our paper.

C First-order condition with non-differentiable rewards

We focused above on the first-order condition for $a^\star$ when $T^\star$ is differentiable. However, the case of non-differentiable $T^\star$ does not pose substantial difficulties. While $T^{\star\prime}$ appears in (5), the role it plays there is just that of a change of variable to an integral over c so as to trace out the boundary of marginal innovations for which $T^\star(k) = c$. If we instead take an integral over c, this boundary is well-defined by the monotonicity of $T^\star$, regardless of whether it is differentiable, as stated formally in the following proposition.

Proposition 11: *Suppose that at least one of the conditions of Proposition 8 is obeyed. Then if the expectation is taken over all c's other than the (at most) countable set where ${T^{\star}}^{-1}$ is not well-defined $W'(a) \propto$*

$$\frac{S\epsilon'}{[1+\epsilon]^2} E_{c<\overline{c},\tilde{f}} \left[\left[{T^{\star}}^{-1}(c;a) \right]^3 \eta Cov_{\tilde{f},x} \left[\log(x), x^{\frac{1-\epsilon}{1+\epsilon}} \right] - Q\epsilon E_{x,k,\tilde{f}} \left[k^2 x^{\frac{1-\epsilon}{1+\epsilon}} \middle| k \geq {T^{\star}}^{-1}(c;a), c \right] E_{x,k,\tilde{f}} \left[k^2 x^{\frac{1-\epsilon}{1+\epsilon}} \right] \right] \tag{13}$$

where, again, if not otherwise stated, expectations are taken over the marginal set for which $T^{\star}(k;a) = c$, η is the elasticity of innovation supply and $\overline{c} \equiv \lim_{x\to\infty} T^{\star}(x)$ (typically ∞).

Note that this demonstrates that essentially nothing in our formulae would change if cost were observed[73], as discussed in Subsection 4.8. The proof of this formula uses a Riemann sum representation of the relevant integrals to exactly represent the set of marginal types as an integral over costs with k viewed only as term entering the quantity to be integrated. This formulation is crucial in establishing the smoothness conditions required for Lemma 2.

Proof. See Online Appendix Subsection 3.4. □

In Online Appendix Subsection 3.4 we show how this collapses to the formula in Proposition 3 when $T^{\star}$ is differentiable.

D Second-order conditions for optimal market power

Of course, the first-order condition, that expression (5) or (13) is equal to 0, is necessary but not sufficient for the socially optimal choice of a. 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 directly on primitives to ensure this seem challenging to derive.

However, note that by Proposition 1 we know that the optimal value of a must be in the interior of the unit interval and thus $W'(a)$ must be eventually negative as a goes to 1 and eventually positive as a goes to 0. While this certainly does not preclude several local minima or maxima in the interior, these basic forces point towards W "typically" being quasi-concave. Some of these simulations are discussed in the following appendix.

Proposition 12: *Let*

$$M(a) \equiv \frac{1}{E_{k,\tilde{f}} \left[\frac{E_{x,\tilde{f}}\left(x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} | c<T^{\star}(k;a),k \right) E_{x,\tilde{f}}\left(x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} | c=T^{\star}(k;a),k \right)}{\eta(T^{\star}(k;a);k,a)} \right]}.$$

and

$$G(a) \equiv E_{k,\tilde{f}} \left[k^3 {T^{\star}}'(k;a) \frac{Cov_{x,\tilde{f}}\left(\log(x), x^{\frac{1-\epsilon(a)}{1+\epsilon(a)}} \right)}{1-\epsilon(a)} \right]$$

If $T^{\star}$ is differentiable, W is quasi-concave if for all $a \in (0,1)$,

$$\frac{d\log(GM)}{da} < \frac{\epsilon(a)}{a} + \epsilon'(a) \left(\frac{1+3\epsilon(a)-2\epsilon^2(a)}{\epsilon(a)\left[1-\epsilon^2(a)\right]} \right) - \frac{\epsilon''(a)}{\epsilon'(a)}$$

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

This condition always holds for a sufficiently close to either 0 *or* 1.

Proof. See Online Appendix Subsection 3.5. □

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 a, 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 a (as pictured in Figure 10).

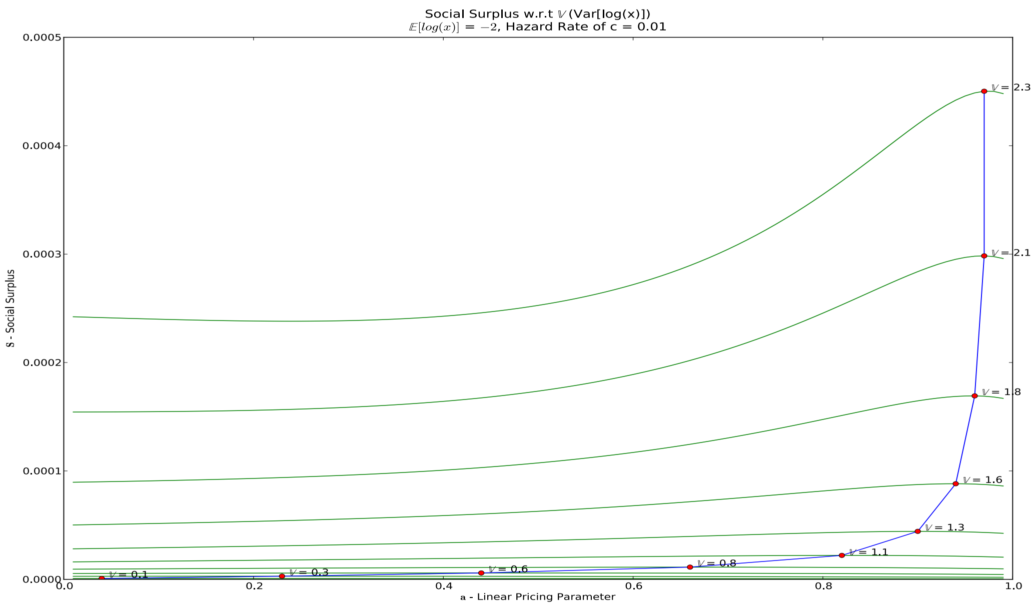

Figure 11: $\tilde{W}(a)$ and $a^\star$ for various values of the variance of the logarithm of x given log-mean of -2 and a .01 hazard rate of c

E Simulations

Another illustration of our results may be obtained by deriving exact optimal policies for a particular distribution f. This helps to check the validity of our theory and requires the development of techniques that allow experimentation with various distributions. It also provides the foundations for Figure 10.

A natural class of distributions to try are ones with parameters corresponding clearly to the notions of genius and materialism, as these will help determine both the validity of our theoretical predictions as well as its robustness across a values. One simple way to do this is to have $x = \frac{m}{\sigma}$ be distributed log-normally, conditional on σm, so that its variance can be manipulated and have c be distributed exponentially so that its (reversed) hazard rate and thus elasticity can be manipulated. It is easy to show that at every point the elasticity of innovation supply is negatively related to the hazard rate of the exponential distribution of costs.

A simple way to implement this basic idea is to allow σm to have some arbitrary, simple distribution, such as uniform over some interval, and then have c and x distributed exponentially and log-normally respectively, independently of one another. This independence both

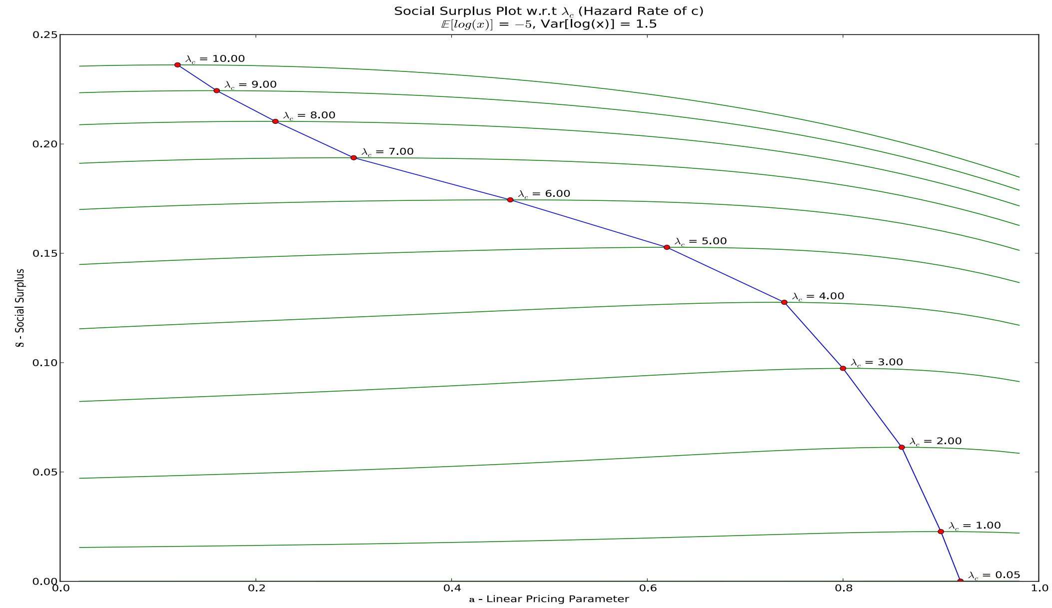

Figure 12: $\tilde{W}(a)$ and $a^\star$ for various values of the variance of the hazard rate of c (elasticity increases as it declines) given log-mean of -5 and log-variance of 1.5 for x.

simplifies calculations, as we will see, and obviates the need for demand ironing as it implies that demand curves are flat[74] in cost.

An intuitive result is shown in Figure 11. As the variance of the log of x increases, $a^\star$ steadily rises to 1, while as it declines, $a^\star$ gradually falls to 0. This result is robust across all specifications we have tried and confirms[75] the basic intuition of Theorem 2. Similarly, Figure 12 shows that as the hazard rate of c declines, and thus the elasticity of innovation supply (materialism) increases, $a^\star$ gradually rises towards 1. This confirms our prediction about the effect of materialism. Python code for our simulations is available at http://www.glenweyl.com.

[74]Because demand is flat under independence, the optimal reward is always the value of demand at any point and is therefore just an expectation. The log-normal form for the distribution of x and uniform for σm allows this to be computed analytically. Social surplus along each isoreward curve can then also be calculated analytically requiring computation only to determine W. Graphing W and/or applying Newton's method easily yields $a^\star$ as we find in every case that W is quasi-concave. Comparative statics can then easily be performed by perturbing parameter values.

[75]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 $a^\star$ as we have verified that, in every example, $T^{\star\star}$ is in fact monotone at $a^\star$. It is only at sub-optimally low a 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 a is a question we hope to investigate in further simulations.