The effect of consumer inferences on privacy regulation when information is costly

BY ABRAHAM L. WICKELGREN
University of Texas at Austin

Abstract

Consumers often value products whose purchase could also release sensitive data about themselves or otherwise intrude on their privacy. In many cases, consumers do not know exactly how great this privacy risk is , but they can engage in some research to learn more about it. Alternatively, the government can choose to prohibit products that pose sufficiently great privacy risks. Whether or not the government does so will affect both the consumers purchase and research decision. This paper analyses the effect of scuh bans when consumers can become informed at a cost. It also compares the regulatory policy of banning products because of their privacy risks to taxing them under the assumption that consumers can make better inferences from bans than taxes. Lastly, the paper analyzes how such regulations might affect ex ante investment incentives.

1 Introduction

Calls for privacy regulation typically give very little consideration that these regulations will limit voluntary transactions in the marketplace. As such, these regulations are typically justified in (implicitly) paternalistic grounds or on the idea that it is impossible for consumers to assess privacy harms. While it is likely true that assessing the privacy risks (if any) from using a particular service is difficult, the fact that so many people have concerns about privacy suggests that it is far from impossible. This suggests that any call for regulation should consider that consumers can, at some cost, learn about privacy risks. Thus, regulators should analyze how privacy regulation affects consumers incentive to acquire this information and the associated welfare consequences, both for static efficiency (incentives to use services and acquire information) and dynamic efficiency (incentives to create these services). This paper creates a model that addresses these issues as well as addressing the differing efficiency consequences of different forms of regulation (direct regulation versus taxes).

When deciding to purchase a good or use a service (possibly for free) that also provides consumer information to the supplying firm, a fully-informed consumer can simply trade-off their personal benefit from the service from any expected harm they might suffer (if any) from the release of their information. In this simple setting, there is no role for government regulation of privacy. Prohibiting firms from obtaining consumer information either causes them to exit the market or raise the price of their service (possibly from zero). Thus, either those consumers whose personal value from the service exceeds the expected harm from the loss of privacy lose the opportunity to make a trade that increases their utility or they are forced to pay for this service directly rather than through their data. At the same time, in the absence of regulation, those who value the service less than the harm they suffer from the privacy loss can simply refrain from purchasing it.

But, for many (possibly most) goods or services where firms collect consumer data, consumers will not have a good understanding of the full impact of consumer data collection. On the other hand, neither are consumers incapable of learning about those impacts, although such learning is costly. For example, when a consumer performs a search or browses a website, that consumer is unlikely to know exactly what data the website is collecting, who it is being shared with, or exactly how that sharing of data will impact them in the future. But, by doing some costly research on website's privacy policies and the uses of consumer data in general, this consumer can learn about how any data collection might affect them.

If the consumer does not invest in this learning, their cost-benefit calculation will depend on their prior distribution of the privacy harms. That prior distribution, in turn, will depend on the consumer's expectation of government regulations aimed at protecting consumer privacy or limiting the use of consumer data. If regulators have a relatively laissez-faire attitude, consumers would (rightly) believe that the possibility significant harm from consumer data collection is larger than if regulators scrutinized firm's data collection policies and regulated them accordingly.

How these beliefs affect a consumer's decision to learn the actual harm from data collection and their decision to use the service without doing such learning will vary with the consumer's value of the service or concern for privacy (we will focus on the case of heterogeneous weights on privacy harms). For consumers with very low vulnerability from privacy harms, there may be no reason to learn the true harm under any regulatory policy since the probability they would not want to use the service based on that information may be very low. Similarly, consumers with very high vulnerability would also have very low value of information since they are never going to use the service.

For consumers with fairly high, but somewhat less vulnerability, they might only spend the effort to learn the true harm if they believe the regulator would have prohibited firms from very intrusive data collection (or uses of that data). Otherwise, the probability of finding a low enough harm level to warrant use might be too small. On the other hand, for consumers with fairly low, but not too low vulnerability to privacy harm, they might use the service without learning the true harm if they believe the regulator would have prohibited very high harm levels, but they might be induced to learn the true privacy-harm if the regulator is unlikely to ban the service when intrusiveness is high.

We also consider the option to tax the intrusive service rather than ban it. Standard analysis would suggest that a tax must be weakly superior to a prohibition since one could think as a prohibition as akin to an infinite tax. In this model, however, this is not the case because we assume that consumers can more easily make inferences from prohibitions than they can from taxes. If regulators ban services, this easily communicates to consumers that sufficiently risky services are not available for sale. On the other hand, governments tax services for a wide variety of reasons, and many services are subject to many different taxes from different levels of government. Thus, it is much more difficult for a consumer to correctly infer the risk level from a tax (if this is not the case,then a tax is always weakly superior). In the model, therefore, we allow for consumers to make inferences when the government policy is to ban services that are too intrusive, but they can't make inferences from the level of tax.

In our model, the product or service has three possible levels of privacy harm (high, low, and zero). We find that it is taxes are superior to bans for the high harm product if the harm and cost of learning are sufficiently high that there is excess consumption of the high harm product without the tax. This leads to efficient allocation in this case. But, it also leads to greater costs from information acquisition. Banning the high harm product, by contrast, eliminates the gain from efficient purchases of this product, but it reduces information acquisition costs and leads to more efficient purchases when the product is harmless.

If consumers can become informed at very low cost, then the optimal tax is zero and banning the product is inefficient. As the cost of information increases, the optimal tax on the high-harm product becomes positive and banning the high-harm product can be optimal (at information costs either below or above the level at which a positive tax is optimal, depending on the values of other parameters).

The analysis of optimal regualtion is further complicated when one realizes that the distribution of product risk is likely affected by the expectation of regulation. If the potential seller expects that highly instrusive products are likely to be banned, it may have a greater incentive to invest in figuring out how to reduce privacy risks. On the other hand, it would also have less incentive to attempt to create the product at all. The model shows that it is possible that banning the most intrusive versions of products can potentially increase total profit and the incentives for ex ante investment.

This analysis has a few important implications. First, and most obviously, is the importance of clearly articulating and sticking to a consistent privacy policy so that consumers can determine how much research they need to do on their own about product risks. Second, when deciding on how to regulate privacy, it is important to consider how consumer perceptions will respond to that regulation. Third, while taxes can influence uninformed consumers' purchase decisions, if they do not send as clear sigals about a product's underlying risk, they may not be as effective at influencing consumers' decisions to learn about product risks. As a result, in some cases, an outright ban may be superior to taxes because it signals to consumers that the products that are not banned are safer in a way that taxes (or their absence) may not.

This analysis is based on a few assumptions. First, we assume that the government cannot eliminate (or greatly reduce) the consumers' cost of learning the true privacy risks through requiring warnings. While the government can require a simple product warning such as “this product may expose sensitive data,” for most consumers, this does not provide sufficient information about whether to purchase the product or not. The warning doesn't indicate what the risk of is, how severe the exposure may be, nor how severe the consequences may be. Moreover, the answers to all of these questions almost certainly depend on personal characteristics that the government does not know. While the government could provide the same data that the consumer can learn on their own, it cannot be done with a simple warning and interpreting that data will still require a substantial level of effort. Thus, in many cases, it is reasonable to assume that while the government ban a product for free, it cannot costlessly inform consumers so that they can make their own, individually-optimal decisions, without further cost to themselves. Of course, to the extent the goverment can inform consumers without limiting their options, that will always be superior with fully rational consumers (maybe because they do not believe consumers are fully rational). Nevertheless, governments do impose mandatory regulations, so it is important to understand how rational consumers respond to them. The fact that rational consumers can make inferences from these bans also suggests that the costs of imposing these bans on rational consumers (possible to protect irrational ones) may be smaller than previously appreciated.

Second, we assume that the consumers cannot learn the privacy risks by purchasing a small amount of the product (and, thereby, incurring a proportionately small amount of risk). For risks that are either low probability events or occur after a long period of time, which certainly describes the privacy risks of a great many products, this will be the case.

Third, we assume that consumers cannot make inferences from the price of the product. While in the model, cost of production and consumer demand are known, this is a modeling convenience. In the real world, almost no consumers know a firm's cost of production or the overall demand curve. As a result, consumers cannot observe a product's price and back out the underlying risks of the product. While one could imagine consumers getting a slightly more precise distribution of product risks if they had well-defined priors about demand and costs, the extra precision would be quite minor if these distribuitons were not very precise. An even more realistic model, would include the fact that making such an inference would require significant mental effort for most consumers, so that given the expected benefit, it would be very unlikely be worth it to make the effort.

This paper is related to the literature consumer information acquisition and inferences from disclosures. Chan and Leland (1982) first analyzed markets in which consumers can aquire information about quality at a cost. Katz (1990) shows that if costly quality is endogenous, then contracts cannot ensure efficient high-quality in the presence of reading costs.[1] Milgrom (2008) reviews the literature on persuasion games and what it suggests about mandated disclosure. Bar-Gill et al. (2019) show that consumers often draw the wrong inferences from mandated disclosures, which highlights another reason why warnings may not be effective. Armstrong (2015) discusses how differently informed consumers interact in the market, though his paper reviews models in which this difference in exogenous, not endogenously chosen.

Zhang (2014) considers the effect of consumer inference on government disclosure policy. In that paper, the consumer doesn't know whether a product has a risky ingredient in it or not. The government's policy choice is whether or not to require firms to disclose whether the product contains this ingerdient or not (at some cost). If the government knows the magnitude of the risk from this ingredient, then the consumer makes an inference about this risk from whether or not the government mandates disclosure. Because there is no option to ban the product, the decision to mandate disclosure signals that the ingredient is more harmful, while the decision not to signals the ingredient is less harmful.

While the Zhang paper shares the insight that consumers make inferences from government regulatory decisions, it differs from this paper in some important ways. First, as mentioned above, the most interventionist policy choice the government can make is to require disclosure of a given ingredient. This necessarily makes the inference from disclosure different than when the government also has the option to ban products entirely. That paper also does not consider taxes as an alternative policy instrument. Second, the consumer decision in binary, to buy or not to buy. They do not have the option to find out the harm at some cost. Thus, it does not examine how the regulation influences the decision to acquire this information.

1. Wickelgren (2011) shows that this result can be improved if there are standardized contracts and some competition. Klein and Leffler (1981) show that if consumers can observe quality costlessly and quickly enough, then prices above marginal cost can ensure high-quality even without contracts. But, if consumers observation of quality takes too long, the effective discount rate in their model will be too low to satisfy their condition enabling high quality.

The next section outlines the model and discusses the important assumptions that drive the results. Section 3 analyzes the laisssez faire regime. Section 4 analyzes the ban regime. Section 5 compares the two. Section 6 analyzes taxes. Section 7 discusses product development effects and Section 8 concludes.

2 Model

A monopoly firm produces a product at constant marginal cost c that causes a baseline privacy harm to consumers of either $\{h,m,l\}$, $1>h>m>l>0$. The full harm is the baseline harm times the idiosyncratic weight of $\theta\in[0,1]$; θ has probability density function g and associated distribution function G. We assume the government and the firm know both the support and distribution of g, but individual consumers do not know either. The firm knows the baseline harm level at the time of sale (and so does a government regulator). Consumers do not know the baseline harm level unless they spend k to learn the harm, but each consumer knows their θ. Consumers value for one unit of the product (gross of harm and price) is given by v . From here on we will often refer the the baseline harm as simply the harm when it creates no ambiguity.

In period -1, the government regulator commits to an observable regulatory policy. We will consider two possible policy choices. First, the regulator must decide whether or not to allow the product with harm h or m for sale in period 1. Second, the regulator can choose a tax t_h (t_m) to impose on the product that causes harm $h(m)$. We assume that the consumers can observe and make inferences from the presence or absence of regulations limiting sale but cannot observe (or, at least, make inferences from) taxes.

In period 0, the firm chooses an amount x to invest in product development and safety. Until the extension (section 6), we will suppress this stage and treat the probabilty of each level of harm as exogenous: the ex ante probability of harm $h(m)$ is $r_h(r_m)$, while the probability of l is $1-r_h-r_m$. In the extension section, we will no longer assume that product is developed for sure so that regulation can affect both the probability of product development and the type of product development.

In period 1, the government and firm observe the harm.[2] The description of the harm is too complex to explain to consumers in an easy to understand warning. (Because the precise harm for each consumer is different, and consumers would not be able to identify who the reference consumer is, the level of harm cannot be described by a simple

2. For a large population, the fact that the government can observe the harm without cost while the consumers incur a cost does not significantly affect the welfare calculation. On a per consumer basis, the government's cost of learning the harm is approximately zero whenever it would make sense for any consumer to learn the true harm.

number or index.) The government imposes the harm-contingent policy it set in period -1 (allowing, banning, or taxing the sale according to the harm). The firm then decides whether to offer the product for sale, and, if so, it chooses its price p. After observing p and (trivially) whether the product is available, consumers decide whether or not to spend k to learn the true harm and then make purchase decisions (fully informed, if they spent k, otherwise not).

2.1 Discussion of important assumptions

The model relies on a number of important assumptions. First, there is no product liability. If consumer harm and causation can be verified in court and firms have sufficient assets to pay for the harm, then liability can produce the first best–there is no need for regulation. This paper is relevant when those conditions don't hold, so that there is some role for ex ante regulation. In the context of privacy harms, proving causation is probably quite difficult. For some harms, it may not be obvious that they flow from an invasion of privacy (an employer, for example, may not explain that they declined to hire someone because of information they were able to access through the applicant's use of some digital platform). Even when the harm clearly stems from an invasion of privacy, it will often be difficult to determine what the source of the privacy loss was.

Second, the model assumes neither the government nor the firm can accurately convey the product's risks to the consumers at zero cost to those consumers. A product or service can come with privacy disclosures, but easy to understand disclosures often won't contain sufficient information for a consumer to assess the true risk. More detailed disclosures require significant consumer effort and learning to full learn the risk as the disclosure must detail the evidence that generates the expected risk estimate. Then each consumer must read and understand these risks and figure out their expected magnitude.[3] Because privacy risks likely carry different magnitudes for different people, assessing any disclosed risks will require non-trivial effort for each consumer.

Third, the model assumes that consumers do not make risk inferences from prices. As we will see below, this means the firm will often have a different optimal price depending on the level of harm. We disallow consumers to update their beliefs from prices because, in practice, consumers do not know enough about the overall demand curve or the marginal costs to make anything other than very noisy inferences. Moreover, making such an inference requires significant mental effort, which, given the likely increase in precision, is unlikely to be worth it (Kominers et al 2016).

3. As any economist who has been to the doctor knows, even experts do not present risks in a simple summary statistic that is sufficient for informed decision-making.

Fourth, for the same reasons, the model does not allow consumers to infer the precise harm from government taxes. Governments tax products for many reasons. Moreover, firms can include the tax in the price so that consumers are not directly aware of it. Of course, the government could announce it is taxing a service because of its privacy risks. This would be like a disclosure that might reduce the cost of the consumer learning the true expected harm, but would be unlikely to eliminate it for the same reasons that any other disclosure would not reveal the harm without cost to the receiver of that disclosure.

On the other hand, the model does allow consumers to make inferences from government regulations that ban dangerous products. Such regulations are easier for consumers to observe and interpret. While in reality, consumers are unlikely to know the precise threshold for banning products, they can easily observe whether a government is more interventionist or more lasseiz-faire and make inferences accordingly.

3 Laissez Faire Regime

In this section, we assume the government has not banned the product from sale nor imposed any taxes. Consumer decisions are simply based on the price p, the cost of information k, the value for the product v, and their beliefs about the probability distribution of harm ($r_h, r_m, 1-r_h-r_m$ for harm levels h, m, l), and their idiosyncratic weighting of the harm, θ.

3.1 Information acquisition decision

We first analyze the decision to become informed. Consumers with low idiosyncratic weight on the harm (less vulnerable consumers) are choosing between either buying the product without information or obtaining information. Let $\mathrm{EH} = r_h h + r_m m + (1-r_h-r_m)l$ be the expected harm level. If $\mathrm{EH} \geq m$, then any consumer who would buy the product without information, would also buy the product if they learned the harm was medium. Thus, for less vulnerable consumers, the value of information comes from not buying the product when learning it has high harms, making the value of information $r_h(p+\theta h - v)$. Thus, for these consumers, obtaining information is worth it if and only if $\theta \in \left(\frac{v-p+k/r_h}{h}, \frac{v-p}{\mathrm{EH}}\right)$. If θ is too low, the expected loss from buying the product when the harm is high is too small to justify cost of becoming informed. If θ is too large, then the consumer would not buy the product if uninformed, making value of information for the consumer comes from learning it is safe to purchase not learning that it isn't. If $\frac{v-p+k/r_h}{h} > \frac{v-p}{\mathrm{EH}}$, or $k > \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$, then there are no less vulnerable consumers who obtain information.

If EH $< m$, then some consumers who would buy the product if uninformed would not buy the product if they found out the harm was high or medium (consumers with $\theta > \frac{v-p}{m}$). These consumers have a higher value of information, $r_h(p+\theta h - v) + r_m(p+\theta m - v)$. If $k > \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$, then the marginal consumers who obtain information will not buy the product if they learn it has medium or high harm. In this case, the consumers who obtain information are those with $\theta \in \left(\frac{(r_h+r_m)(v-p)+k}{r_m m + r_h h}, \frac{v-p}{\mathrm{EH}}\right)$. No consumers will obtain information in this case if $\frac{(r_h+r_m)(v-p)+k}{r_m m + r_h h} > \frac{v-p}{\mathrm{EH}}$, or $k > \frac{(1-r_h-r_m)(\mathrm{EH}-l)(v-p)}{\mathrm{EH}}$.

Now consider the consumers who will not purchase the product if uninformed ($v < p + \theta\mathrm{EH}$), the value of information now is that if they learn the harm is low enough, they will purchase the product. If EH $< m$, then the these consumers will only purchase if they know the harm is low, so the value of information is $(1-r_h-r_m)(v-p-\theta l)$. These consumers only obtain information if $\theta \in \left(\frac{v-p}{\mathrm{EH}}, \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)$. If $\frac{v-p}{\mathrm{EH}} > \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}$, or $k > \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$, then there are no more vulnerable consumers who acquire information if EH $< m$.

If EH $\geq m$ and $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$, then while there will be some more vulnerable consumers that will purchase the product if the harm is medium, the marginal more vulnerable consumer (the one who is indifferent between obtaining information and not purchasing the product) will only buy if harm is low. Thus, this case is identical to the case above when EH $< m$.

If EH $\geq m$ and $k > \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$, the marginal more vulnerable consumers would purchase the product if harm were m, so the value of information is $(1-r_h-r_m)(v-p-\theta l) + r_m(v-p-\theta m)$. In that case, consumers obtain information if only if $\theta \in \left(\frac{v-p}{\mathrm{EH}}, \frac{(1-r_h)(v-p)-k}{(1-r_h)l + r_m(m-l)}\right)$. If EH $\geq m$, consumers will not obtain any information if $\frac{(1-r_h)(v-p)-k}{(1-r_h)l+r_m(m-l)} < \frac{v-p}{\mathrm{EH}}$, or $k > \frac{r_h(h-\mathrm{EH})(v-p)}{\mathrm{EH}}$.

Thus, we have the following lemma.

Lemma 1. *If the product is never banned from sale and (A)*EH $\geq m$

(i) If $k > \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$, *no consumers obtain information, and consumers purchase the product if and only if* $\theta < \frac{v-p}{\mathrm{EH}}$

(ii)If $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$*: consumers with* $\theta < \frac{v-p+k/r_h}{h}$ *purchase the product without information; consumers with* $\theta \in \left(\frac{v-p+k/r_h}{h}, \frac{(1-r_h)(v-p)-k}{(1-r_h)l+r_m(m-l)}\right)$ *obtain information, buying the product if and only if harm is low or medium; consumers with* $\theta \geq \frac{(1-r_h)(v-p)-k}{r_l l + r_m m}$, *do not acquire information and do not purchase the product.*

(iii) If $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$*: consumers with* $\theta < \frac{v-p+k/r_h}{h}$ *purchase the product without information; consumers with* $\theta \in \left(\frac{v-p+k/r_h}{h}, \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)$ *obtain information, buying the product if and only if harm is low or harm is medium and* $\theta < \frac{v-p}{m}$*; consumers with* $\theta \geq \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}$*, do not acquire information and do not purchase the product.*

(B) If $\mathrm{EH} < m$*:*

(i) If $k > \frac{(1-r_h-r_m)(\mathrm{EH}-l)(v-p)}{\mathrm{EH}}$*, no consumers obtain information, and consumers purchase the product if and only if* $\theta < \frac{v-p}{\mathrm{Eh}}$

(ii)If $k \in \left(\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}, \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}\right)$*: consumers with* $\theta < \frac{(r_h+r_m)(v-p)+k}{r_m m+r_h h}$ *purchase the product without information; consumers with* $\theta \in \left(\frac{(r_h+r_m)(v-p)+k}{r_m m+r_h h}, \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)$ *obtain information, buying the product if and only if the harm is low; consumers with* $\theta \geq \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}$*, do not acquire information and do not purchase the product.*

(iii) If $k \leq \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$*: consumers with* $\theta < \frac{v-p+k/r_h}{h}$ *purchase the product without information; consumers with* $\theta \in \left(\frac{v-p+k/r_h}{h}, \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)$ *obtain information, buying the product if and only if harm is low or harm is medium and* $\theta < \frac{v-p}{m}$*; consumers with* $\theta \geq \frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}$*, do not acquire information and do not purchase the product.*

Because the insights are substantially similar in the $\mathrm{EH} \geq m$ case and the $\mathrm{EH} < m$, from here on we will focus on the $\mathrm{EH} \geq m$ case. Furthermore, because this problem is interesting only if consumers sometimes acquire information, from here on, we will assume $k \leq \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$.

3.2 Pricing decision

We can use Lemma 1 to derive the demand curve for the product based on harm level. If harm is high, then the only consumers who purchase the product are those who purchase the product without information. Thus, demand is given by $G\left(\frac{v-p+k/r_h}{h}\right)$.

If harm is medium and $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$, then consumers purchase the product whenever $\theta < \frac{(1-r_h)(v-p)-k}{r_l l+r_m m}$ (all consumers who obtain information have a low enough θ to purchase when harm is medium), so demand is $G\left(\frac{(1-r_h)(v-p)-k}{r_l l+r_m m}\right)$. If $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$, then consumers with $\theta < \frac{v-p}{m}$ purchase the product (all of them obtain information), so demand is $G\left(\frac{v-p}{m}\right)$.

If harm is low, then all consumers who obtain information or purchase the product without information buy the product. If $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$, this means demand is $G\left(\frac{(1-r_h)(v-p)-k}{r_l l+r_m m}\right)$. If $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$, then demand is $G\left(\frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)$.

The following lemma gives profit-maximizing prices when consumer values are uniformly distributed.

Lemma 2. *If $g \sim U[0,1]$, then profit-maximing prices for each harm level are given by:* *(a)* $p_h = \frac{v+c}{2} + \frac{k}{2r_h}$; *(b)* $p_m = \frac{v+c}{2} - \frac{k}{2(1-r_h)}$ *if* $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$ *and* $p_m = \frac{v+c}{2}$ *if* $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$; *(c)* $p_l = \frac{v+c}{2} - \frac{k}{2(1-r_h)}$ *if* $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$ *and* $p_l = \frac{v+c}{2} - \frac{k}{2(1-r_h-r_m)}$ *if* $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$.

Proof. See Appendix. $\square$

The prices all have a term that is the typical monopoly price plus (or minus) a term related to the cost of information when the marginal consumer is one who is on the margin of acquiring information. So, the price of the high harm product is increasing in the cost of information because the marginal consumer is one who just does not acquire information. So, when the cost of information is larger, more consumers will buy without acquiring information. On the other hand, the price of the low harm product is decreasing in the cost of information because lower prices increase demand both directly and by inducing more information acquisition. The price of the medium harm product doesn't depend on the cost of information if that cost is small because the marginal consumer is already acquiring information, but once the information cost is high enough, the medium harm product price is decreasing in the cost of information for the same reason as the low harm product price.

Notice that prices do not directly depend on the actual harm. This is not because consumers are uninformed about the harm–when harm is low or medium large fractions of the consumers are informed. Rather, it stems from the multiplicative structure of the heterogeneity in consumer preferences. Greater harm shrinks demand, but does not change demand elasticity, so does not affect the optimal price when consumers are informed.[4]

4. An early version of the paper had heterogenous consumer values. In that case, the optimal monopoly price was decreasing in the harm.

4 High harm ban

Now let's examine what happens if consumers know the regulator will ban the product if the harm is high. The analysis will be similar except that now beliefs about the probability distribution of harm are $\{0, r_m^b, 1-r_m^b\}$ for harm levels $\{h, m, l\}$ if the product is for sale. For now, we will assume that the ex ante probabilities were fixed, so $r_m^b = r_m/(1-r_h)$. In section 6, we will relax this assumption.

4.1 Information acquisition decision

If consumers know the only possible harm levels are medium and low, then (assuming the product is for sale) both the upper and lower limit for obtaining information will fall. Consumers with $v-p \geq \theta(r_m^b m + (1-r_m^b)l)$ choose between buying the product without acquiring information and acquiring information. For these consumers, the value of information is $r_m^b(p+\theta m - v)$. Thus, these consumers acquire information if and only if $\theta > \frac{r_m^b(v-p)+k}{r_m^b m}$. Consumers with larger θ choose between acquiring information or not purchasing at all, thus their value of information is $(1-r_m^b)(v-p-\theta l)$. These consumers acquire information if and only if $\theta < \frac{(1-r_m^b)(v-p)-k}{(1-r_m^b)l}$. Thus, we have the following lemma.

Lemma 3. *If the regulator will ban the product when the harm is high:*

(a) Consumers with $\theta \leq \frac{r_m^b(v-p)k}{r_m^b m}$ *buy the product without acquiring information.*

(b) Consumers with $\theta \in \left(\frac{r_m^b(v-p)+k}{r_m^b m}, \frac{(1-r_m^b)(v-p)-k}{(1-r_m^b)l}\right)$ *spend k to learn the true harm and buy the product if and only if it is low harm.*

(c) Consumers with $\theta > \frac{(1-r_m^b)(v-p)-k}{(1-r_m^b)l}$ *do not acquire information or buy the product.*

Notice that consumers only ever acquire information if $k < \frac{r_m^b(1-r_m^b)(m-l)(v-p)}{\mathrm{EH}^b}$.

4.2 Pricing decision

If harm is medium, then demand only comes from consumers who do not acquire information, so is $G\left(\frac{r_m^b(v-p)+k}{r_m^b m}\right)$. If the product is harmless, then all consumers who acquire information also buy the product, so demand is $G\left(\frac{(1-r_m^b)(v-p)-k}{(1-r_m^b)l}\right)$.

The following lemma gives profit-maximizing prices when consumer values are uniformly distributed.

Lemma 4. *If* $g \sim U[0,1]$*, then profit-maximing prices if the regulator bans the high harm product are given by:*

$(a)\ p_m = \frac{v+c}{2} + \frac{k}{2r_m^b} \quad (b)\ p_l = \frac{v+c}{2} - \frac{k}{2(1-r_m^b)}$.

Proof. See Appendix. □

The medium harm product's price is now increasing in the cost of information while the low harm product's price is decreasing in the cost of information. This occurs because the medium harm seller sells only to those who do not obtain information, while the marginal buyer of the no harm product acquires information.

5 Welfare comparison

In this section, we compare welfare between the laissez faire regime and the ban regime. To analyze this formally, note that for small k expected welfare under the laissez faire regime is:

$$r_h\left\{\int_0^{\frac{v-p_h+k/r_h}{h}} (v-c-\theta h)d\theta - k\left[\left(\frac{(1-r_h-r_m)(v-p_h)-k}{(1-r_h-r_m)l}\right)-\left(\frac{v-p_h+k/r_h}{h}\right)\right]\right\}+ r_m\left\{\int_0^{\frac{v-p_m}{m}} (v-c-\theta m)d\theta - k\left[\left(\frac{(1-r_h-r_m)(v-p_m)-k}{(1-r_h-r_m)l}\right)-\left(\frac{v-p_m+k/r_h}{h}\right)\right]\right\}+(1-r_h-r_m)\left\{\int_0^{\frac{(1-r_h-rm)(v-p_l)-k}{(1-r_h-r_m)l}} (v-c-\theta l)d\theta - k\left[\left(\frac{(1-r_h-r_m)(v-p_l)-k}{(1-r_h-r_m)l}\right)-\left(\frac{v-p_l+k/r_h}{h}\right)\right]\right\} \quad (1)$$

The prices are given in Lemma 2 and the optimal tax in Lemma 3. The integral terms represent the welfare from consumption for each possible harm level. The terms in square brackets represent the probability of incurring information costs. The information cost terms differ for each harm level only because the prices differ.

Expected welfare under the policy of banning the product when harm is high is given by:

$$r_m^b(1-r_h^b)\left\{\int_0^{\frac{v-p_m^b+k/r_m^b}{m}} (v-c-\theta m)d\theta - k\left(\frac{(1-r_m^b)(v-p_m^b)-k}{(1-r_m^b)l} - \frac{v-p_m^b+k/r_m^b}{m}\right)\right\}+(1-r_m^b)(1-r_h^b)\left\{\int_0^{\frac{(1-r_m^b)(v-p_l^b)-k}{(1-r_m^b)l}} (v-c-\theta l)d\theta - k\left(\frac{(1-r_m^b)(v-p_l^b)-k}{(1-r_m^b)l} - \frac{v-p_l^b+k/r_m^b}{m}\right)\right\} \quad (2)$$

Notice that because we have defined r_m^b in the ban case as conditional on the product being sold, for welfare analysis, we have to convert them to unconditional probabilities for welfare analysis. Abusing notation somewhat, we define r_h^b as the unconditional probability of the high harm state when the high harm product is banned.

For the remaninder of this section, we will confine our analysis to the case of exogenous probabilities (i.e., we will will not allow for the regulation to affect the type of product developed), in which case we have $r_h^b = r_h$ and $r_l^b = r_l/(1-r_h)$. Prices are given by Lemma 4.

Because consumers are rational, banning the product can only reduce consumer welfare if harms are high–some consumers have low enough vulnerability to privacy harms that even very intrusive products are welfare-increasing for them. Those with higher vulnerability simply do not purchase the product. That said, the willingness to ban when harms are high creates a benefit from increasing consumption (which is efficient) for the medium and low harm product. When information costs are low, banning the high harm product increases consumption of the medium harm product by inducing purchases for $\theta \in \left(\frac{v-c}{2m}, \frac{v-c}{2m} + \frac{k}{2mr_m^b}\right)$ that would not occur (but are socially efficient) when the high harm product is not banned. While these purchases make the consumer worse off, because their value is less than the price, they increase total welfare because the value exceeds the social cost. These added purchases occur because consumers in this region do not obtain information, so they buy the product because it might be low harm. Consumption of the low harm product also increases because more consumers with high idiosyncratic weight obtain information and purchase the low harm product–this is efficient and increases consumer welfare.

Proposition 5. *Banning the high harm product reduces ex post welfare if the product harm is high, but increases ex post welfare from consumption if the product is medium or low harm.*

Proof. See Appendix □

While banning the product causes more of the more vulnerable consumers to obtain information, it also reduces the number of less vulnerable consumers who obtain information and, when the product is banned, no one obtains information. Thus, it is possible for total information costs to be higher or lower under a ban.

We can also compare overall welfare from consumption between banning and not banning the product. If information costs are very small, then there will always be greater welfare from not banning the product because almost all consumers will be informed and make efficient purchase decisions. On the other hand, if information costs are so high that no consumers acquire information, then the banning the high harm product raises total welfare if and only if the high harm product is sufficiently harmful, as indicated in the following result.

Lemma 6. *If k is sufficiently large that consumers never obtain information whether the high harm product is banned or not, then expected welfare from consumption of the high harm product is positive if and only if* $h < 4\mathrm{EH}$*, but it is efficient to ban the high harm product for any* $h > \bar{h}$ *where* $\bar{h} < 2\mathrm{EH}$.

Proof. *See Appendix* □

While the result that the ban is optimal whenever the high harm product is sufficiently harmful is fairly obvious, it is more interesting to note that banning the high harm product is still optimal for a substantial range of harm where expected welfare from consumption of the high harm product is positive. This follows because when the product has lower harms, the fact that the product would have been banned had the harms been high leads to greater consumption when the harm is lower.

When information costs fall just below the level at which no one obtains information, there is a discrete jump in welfare because prices for the low and medium harm product fall discontinuously to induce more information acquistion while prices for the high harm product just increase marginally to stop consumers from acquiring information. Thus, the initial positive effect on consumption when the harm is medium or low outweighs the negative effect when harm is high.

It is possible for the minimum information cost at which consumers just start to obtain information to be either larger or smaller when the high harm product is banned. Intuitively, this is because banning the product can either increase or decrease uncertainty. For example, if high harm is very likely, then there is not much uncertainty under laisssez faire, but there coul be under a ban if medium and low harm are equally likely and very different from each other. On the other hand, if medium harm is very unlikely, then banning the high harm product means the low harm product is very likely, so there isn't much point to acquiring information. In turns out, however, that if information costs are large enough that consumers only learn in one case (which ever case it is) that welfare is larger in the ban case as long as the high harm is sufficiently high.

Proposition 7. *There exists a* $\tilde{k} < \mathrm{Max}\left\{\frac{r_h(v-c)(h-\mathrm{EH})}{2\mathrm{EH}}, \frac{r_m^b(1-r_m^b)(m-l)(v-c)}{2\mathrm{EH}^b}\right\}$ *and a* $\tilde{h} < 2\mathrm{EH}$ *such that some consumers obtain information in at least one of the two regimes in which banning the high harm product is superior to laissez faire for* $k > \tilde{k}$ *and* $h > \tilde{h}$ *whenever* $\frac{r_h(v-c)(h-\mathrm{EH})}{2\mathrm{EH}} < \frac{r_m^b(1-r_m^b)(m-l)(v-c)}{2\mathrm{EH}^b}$ *(consumers obtain information only with a ban) or* $r_h > 4/15$.

Proof. See Appendix. $\square$

The advantage of the laisssez faire regime is enabling efficient allocation when harm is high. The advantage of banning the high-harm product is that then if the product is for sale, consumers know the harm either medium or low. This improves both their information acquisition decision in this case and their purchase decision. If consumer information is relatively easy to obtain, then it is never optimal to ban the high harm product. But, this proposition shows that if information is costly, then even when some consumers obtain information, it can be better to ban the high harm product rather than allow consumers to purchase it so as to induce more consumption of the low and medium harm products.

6 Taxes

In addition to banning products with high privacy harms, regulators could also tax them so as to discourage consumption. If taxes convey the same information as product bans, then taxes will always be (weakly) superior because a large enough tax on a product will effectively eliminate consumption of that product. It is possible, however, that taxes are less salient and/or communicate less information than outright bans. As discussed in the introduction, firms can choose to include the tax in the price to make it less obvious to consumers that the government is taxing the product. In addition, because governments tax many products for a variety of reasons, it may be harder for consumers to infer the government's information from the tax. To examine the relative effect of taxes if this is the case, we now examine regulation through taxes if consumers do not make inferences from taxes.

6.1 Taxes without product bans

Because we have assumed that taxes convey no information to consumers, taxes only affect welfare through the effect on prices. They affect prices by raising the marginal cost for the firm from c to $c+t$. Thus, if the government imposes a tax of t_i on a product with harm level i, the prices are adjusted from those in Lemma 2 by replacing c with $c+t_i$. Using this, we can analyze ex post welfare effect of taxes. We will explicitly analyze the low information costs casse, $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$. The case of higher information costs is similar, though the exact size of the optimal tax is somewhat different.

If harm is high and information costs are low ($k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$), ex post welfare is given by:

$$\int_0^{\frac{v-p+k/r_h}{h}} (v-c-\theta h)d\theta - k\left[G\left(\frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)-G\left(\frac{v-p+k/r_h}{h}\right)\right] \tag{3}$$

Taxes operate through prices, and increasing the price shifts the range of values that acquire information but also reduces the size of the range because the lower limit decreases by $1/h$, which is less than $1/l$, the size of the decrease in the upper limit.[5] Thus, if G is uniform, then taxes decrease affect information acquisition costs in addition to affecting welfare through the consumption decision.

In the absence of taxes, there are two distortions that affect consumption of the high harm product that work in opposite directions. First, because consumer's do not know the harm, those who do not acquire information believe the expected net value of the product is higher than it is, leading to excess consumption. Second, the firm's market power leads to prices above the full information marginal cost, and their desire to reduce consumer information acquisition further induces them to charge higher prices. Both of these lead to insufficient consumption

If information costs are small, the net effect is that there is insufficient consumption of the high harm product, so there is no benefit to taxing the product (we do not allow subsidies). But, if information costs are sufficiently large, there would be excess consumption of the high harm product without taxes. In this cases, taxes can induce efficient consumption decisions. That said, because taxes also affect information costs, the optimal tax is slightly larger and stil linduces somewhat less consumption than optimal.

We summarize these in the following lemma.

Proposition 8. *If the product can be sold at any harm level and* $k \leq \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$*, then it is never optimal to tax the medium harm product or the low harm product. It is optimal to tax the high harm product if and only if* $k > \frac{r_h(v-c)l}{2r_h(h-l)+l}$*, in which case the optimal tax is* $t_h = k\left(\frac{1}{r_h}+\frac{2(h-l)}{l}\right)-(v-c)$*, and consumers with* $\theta < \frac{v-c}{h} - k\frac{h-l}{lh}$ *purchase the high harm product.*

Proof. See Appendix. ☐

5. The the medium information costs case differs in that the effect of price on the upper limit for when consumers acquire information is different. It is still the case that this effect on the upper limit exceeds (in magnitude) the effect on the lower limit, but this difference is less. As a result, optimal taxes are slightly smaller.

If the expected surplus from purchasing the product is sufficiently large (for at least some consumers) relative to the cost of information, then it is socially optimal to tax the high harm product at least a little. These taxes are set so that somewhat fewer consumers than optimal purchase the product (none of these consumers obtain information) in order to reduce total information acquistion costs. Everyone with a lower value obtains information and does not purchase the product. For the medium and low harm product, however, there is never excessive consumption because the marignal consumers always obtain information and so know the true harm level when the high harm product is not banned. Thus, the monopoly distortion leads to insufficient consumption without any counterveiling distortion in the other direction. Given the maximum information cost, the benefit of the tax in reducing information acquisition costs never outweighs the cost of reducing efficient consumption.

6.2 High harm product banned

When the product harm is medium, welfare is given by:

$$\int_0^{\frac{r_m^b(v-p)+k}{r_m^b m}} (v-c-\theta m)g(v)d\theta - k\left[G\left(\frac{(1-r_m^b)(v-p)-k}{(1-r_m^b)l}\right) - G\left(\frac{r_m^b(v-p)+k}{r_m^b m}\right)\right] \tag{4}$$

The analysis for the medium harm product parallels the analysis of the high harm product in the last section (as it is now the highest possible harm product). If k is low, there is insufficient consumption due to monopoly pricing. If k is large enough, however, there can be excessive consumption as the number of consumers who buy without obtaining information increases. The following lemma provides the analogous result on optimal taxation in the ban case.

Proposition 9. *If the high harm product is banned, then it is never optimal to tax the no harm product. It is optimal to tax the medium harm product if and only if* $k > \frac{r_m^b(v-c)l}{l+2r_m^b(m-l)}$, *in which case the optimal tax is* $t_m = k\left(\frac{1}{r_h} + \frac{2(m-l)}{l}\right) - (v-c)$, *and consumers with* $\theta < \frac{v-c}{m} - k\frac{m-l}{lm}$ *purchase the medium harm product.*

Proof. See Appendix. $\square$

It isn't necessarily the case that optimal taxation occurs at higher or lower information costs when there is a ban. The fact that the top harm is smaller in this case makes the right hand side larger, so this effect makes some tax less likely to be optimal. If $r_l^b > r_h$, then this conclusion is reinforced. But, if $r_l^b < r_h$, then it's possible that it is optimal to tax the low harm product when the high harm product is banned even if taxing the high harm product is not optimal.

7 Extension: Product development investment

We now consider the period 0 investement decision. To distinguish between the ex post probabilities of each type of product when we know there is a product and the ex ante probabilities as a function of investment, we change our notation slightly. We now use $q_i(x)$ to reflect the probability of developing a product of type i, and allow for the possibility that $q_h(x)+q_m(x)+q_l(x)<1$, so product development may not occur.[6] For simplicity, we assume a uni-dimensional investment choice that affects both the overall probability of development and the expected harm. We assume that $q_l'(x)>0$ and $q_m'(x)>0$, more investment increases the probability of the no harm and low product. We also assume that $q_h'(x)+q_m'(x)+q_l'(x)>0$, more investment increases the probability of product development. Lastly, we assume that $q_i''(x)<0$.

We denote the firm's expected profit from developing product i as π_i in the absence of a ban on the high harm product. If the high harm product is banned, then denote the expected profit from product $i\in\{h,l\}$ as π_i^b. Then, if the high harm product will not be banned, the firm chooses x to maximize:

$$q_h(x)\pi_h+q_m(x)\pi_m+q_l(x)\pi_l-x$$

If the high harm product will be banned, the firm's objective function is:

$$q_l(x)\pi_l^b+q_m(x)\pi_m^b-x$$

Thus, the first order conditions for profit-maximizing investment for the no ban and ban regimes are:

$$\begin{aligned} q_h'(x)\pi_h+q_m'(x)\pi_m+q_l'(x)\pi_l&=1\\ q_l'(x)\pi_l^b+q_m'(x)\pi_m^b&=1 \end{aligned}$$

Note that in our model, profit for the low and no harm products is larger if the high harm product is banned under a wide range of conditions, as the following lemma shows.

Lemma 10. *Whenever at least some consumers acquire information, then $\pi_m^b>\pi_m$ and $\pi_l^b>\pi_l$.*

Proof. *See Appendix.* □

6. Thus, the probabilities in the prior sections denoted by r_i can be thought of as $r_i=q_i/(q_h+q_l+q_0)$.

This result suggests that in many situations, while banning the high harm product obviously reduces the firm's profit if the product causes high harm, it can increase the profits for the low and no harm products. Thus, the effect of the ban on the incentive to invest in the product ex ante could be positive or negative. The following result establishes a sufficient condition for a ban on the high harm product to increase investment.

Proposition 11. *If $\pi_l^b - \pi_l > 0$ and $\pi_0^b - \pi_0 > 0$ and either (a) $q_h'(x^{\mathrm{LF}}) < 0$ (investment decreases the probability of the high harm product being developed at the optimal level of investment under laissez faire) or (b) $q_l(x)/q_h(x)$ and $q_0(x)/q_l(x)$ are increasing in x (the monotone likelihood ratio property holds) and total expected profits are greater when high harm is banned, then the firm invests more under a ban.*

Proof. See Appendix. □

This result indicates that while it is certainly possible that banning high harm products will decrease investment in creating new products, that is not necessarily the case. If this investment increases the probability of lower harm products relative to higher harm products or if without the ban greater investment decreases the probability of the high harm product, and then such regulation can increase investment if it increases total profits. While that is unlikely if high harm products are much more likely to be developed, it can be the case if (after optimal investment) the probability of developing a high harm product is sufficiently small. In paticular, if the optimal level of investment is already fairly high without a ban, then it is more likely that further investment makes high harm less likely, which would mean that a ban stimulates investment. By making consumers less wary of purchasing the product, bans make these these lower harm products more profitable.

8 Conclusion

As consumers, we often purchase products without precise knowledge of any possible adverse privacy effects they may have on us. We typically do not know the risks of sharing our data when we use various technology products and services. While in principle, product liability might could make us indifferent to our lack of knowledge, in practice that is rarely the case. Suits are expensive, causation is difficult to prove, and for harms that come regardless of a design defect, recovery may not be available at all, especially if there was some type of warning. If product liability were perfect, of course, there would be no reason to regulate any dangerous or hazardous products.

Because liability and warnings are imperfect, consumers are often left with a choice not only whether to purchase a product or not but also whether to gather more information about the risk. Gathering such information is costly. In this paper, we show that this cost can affect both the optimal tax for products with potentially large privacy harms and whether or not it is optimal to ban them outright. When government can commit to banning these products, this provides consumers with valuable information about the hazards of products that are still on the market. Because making similar inferences from taxes is much harder, we show that sometimes committing to ban products with only small amounts of possible social surplus can be optimal ex ante, even though it foregoes possible ex post social surplus associated with an optimal tax regime.

Another possible benefit to committing to banning products with large privacy risks is that it can possibly increase the marginal private benefit to investing in making products safer. On the other hand, it also might make the expected private benefit of developing a new product less than cost, further complicating the optimal regulatory policy. In the current model, we show that under some plausible conditions, if total profits increase from banning the most risky products, then such regulations can increase investment.

Lastly, this analysis has some implications for changing regulatory policy. When regulatory preferences or information changes, optimal regulatory policy might remain sticky. For example, say new regulators believe the fraction of very high-valued consumers (who buy without research) is somewhat lower than previous regulators did. This might mean that more restrictive regulation would have been optimal initially, but it might not warrant banning exisitng products if many consumers have already incurred the research costs of learning those risks.

Because optimal regulation depends both on the risks of the product and the distribution of benefits, changing regulation because of new information about one of these factors also runs the risk of sending the wrong message to consumers. For example, if new regulators place more weight on the importance of enabling purchases by high-valued consumers, and thereby allow the sale of previously banned products, some consumers might wrongly infer that the change was due to new information about the harmfulness of the product. While it is possible to mitigate this by accurately conveying the reasons for the change, it is unlikely to fully resolve the problem. This suggests another reason why optimal regulation might be sticky, at least with respect to changes in other paramters aside from harmfulness.

9 Appendix

Proof. (Lemma 2) These follow directly from maximizing $(p-c)\alpha(p)$, where $\alpha(p)$ is the maximum θ given p which purchases the product, derived above for each harm level and cost of information. $\square$

Proof. (Lemma 4)These follow directly from maximizing $(p-c)\alpha(p)$, where $\alpha(p)$ is the maximum idiosyncratic weight which purchases the product, derived above for each harm level and cost of information. $\square$

Proof. (Proposition 5) If the high harm product is not banned, then the under the optimal tax, $p_h = \frac{v+c+t_h}{2} + \frac{k}{2r_h}$. The most vulnerable consumer who consumers the product is $\theta_{h,\max} = \frac{v-c}{h} - k\frac{h-l}{lh} < \frac{v-c}{h}$. Thus, consumption of the high harm product always creates positive social welfare. Total welfare, net of information costs, when the product is high harm is:

$$\int_0^{\frac{v-p_h+k/r_h}{h}} (v-c-\theta h)d\theta - k\left[\left(\frac{(1-r_h-r_m)(v-p_h)-k}{(1-r_h-r_m)l}\right)-\left(\frac{v-p_h+k/r_h}{h}\right)\right]$$

With the prices and optimal tax, this becomes:

$$\left\{\frac{(hk-l(v-c))^2}{hl^2}+\frac{2k^2(r_h^2+(1-r_h)(1-r_m))}{r_h(1-r_h-r_m)l}\right\}>0$$

For the medium harm product, for low information costs, the most vulnerable consumers who consumer the medium harm product without the ban have $\theta = \frac{v-c}{2m}$. If the high harm product is banned, then the most vulnerable consumer who purchases the medium harm product is $\theta = \frac{v-c}{2m} + \frac{k}{2mr_m^b}$. Whenever k is such that consumers acquire information in the ban case, this is always less than $\frac{v-c}{m}$. So purchases of the medium harm product always increase social welfare under the ban and occur for more values of θ.

For the low harm product, the most vulnerable consumer who consumes without the ban has $\theta = \frac{v-c}{2l} - \frac{k}{2l(1-r_h-r_m)}$. If the high harm product is banned, then the most vulnerable consumer who purchases the low harm product is $\theta = \frac{v-c}{2l} - \frac{k}{2l(1-r_m^b)}$. Notice that both of these cutoffs are less than the efficient cutoff of $\frac{v-c}{l}$. So purchases of the low harm product always increase social welfare in both cases. The upper limit on θ is greater under the ban because $1-r_m^b$ is the probability of the product being low harm under the ban while $1-r_h-r_m$ is the probability of the product being low harm without the ban, which must be smaller. $\square$

Proof. (Lemma 6) If consumers do not obtain information, then welfare from the high harm product without taxes is:

$$\int_0^{\frac{v-p_h}{\mathrm{EH}}} (v-c-\theta h)d\theta$$

With no consumer information gathering, the profit-maximizing price is just $(v+c)/2$. This generates welfare of $\frac{(v-c)^2(4\mathrm{EH}-h)}{8\mathrm{EH}^2}$, which is positive if and only if $h<4\mathrm{EH}$. Taxes will be optimal if and only if $\frac{v-c}{2\mathrm{EH}}>\frac{v-c}{h}$ (there is too much consumption of the high harm product). In that case, taxes are set so that $\frac{v-p_h}{\mathrm{EH}}=\frac{v-c}{h}$ and ensure optimal consumption when harm is high, so there is always social gain from consuming the high harm product.

Expected total welfare in the no ban regime is given by

$$r_h\int_0^{\frac{v-p_h}{\mathrm{EH}}}(v-c-\theta h)d\theta+r_m\int_0^{\frac{v-p_m}{\mathrm{EH}}}(v-c-\theta m)d\theta+(1-r_h-r_m)\int_0^{\frac{v-p_l}{\mathrm{EH}}}(v-c-\theta l)d\theta$$

With no consumer information gathering, the optimal price in all cases is $(v+c)/2$. Exptected total welfare is $\frac{3(v-c)^2}{8\mathrm{EH}}$.

If the high harm product is banned and consumers do not obtain information, then exptected total welfare is:

$$r_m\int_0^{\frac{v-p_m}{\mathrm{EH}^b}}(v-c-\theta m)d\theta+(1-r_h-r_m)\int_0^{\frac{v-p_l}{\mathrm{EH}^b}}(v-c-\theta l)d\theta$$

Again, the optimal price is $(v+c)/2$. Expected total welfare is $\frac{3(1-r_h)(v-c)^2}{8\mathrm{EH}^b}$. Expected total welfare is greater in the ban case if and only if $(1-r_h)\mathrm{EH}>\mathrm{EH}^b$ or $(1-r_h)^2h>(2-r_h)(r_m m+(1-r_h-r_m)l)$. If $h\geq 2\mathrm{EH}$, this is always satisfied whenever $\mathrm{EH}\geq m$.

$\square$

Proof. (Proposition 7) In the no ban case, at least some consumers obtain information for $k<\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$. In the ban case, at least some consumers obtain information for $k<\frac{r_m^b(1-r_m^b)(v-p)(m-l)}{\mathrm{EH}^b}$. Because if no consumers obtain information, $p=(v+c)/2$ whether there is a ban or not, we use this price in both expressions to find the minimum information cost for no consumers to acquire information. Once consumers have some incentive to obtain information, the price of the low harm product drops discontinuously, leading to even more consumers obtaining information if the harm is low. The price for the high harm product (or medium, if the high harm product is banned) increases. This increase, however, is only marginal, however, because once the price has increased enough that consumers don't obtain information, there is no reason to increase it further since it is above the monopoly price. Thus, the effect on welfare in the ban case is the following:

$$(1-r_h-r_m)\left\{\int_{\frac{v-c}{2\mathrm{EH}^b}}^{\frac{(1-r_m^b)(v-p_l^b)-k}{(1-r_m^b)l}}(v-c-\theta l)d\theta-k\left(\frac{(1-r_m^b)(v-p_l^b)-k}{(1-r_m^b)l}-\frac{v-p_l^b+k/r_m^b}{m}\right)\right\}$$

This follows because when no consumers obtain information, consumers purchase the product if and only if $\theta\leq\frac{v-p}{\mathrm{EH}^b}=\frac{v-c}{2\mathrm{EH}^b}$. Once information costs drop below $k<\frac{r_m^b(1-r_m^b)(v-p)(m-l)}{\mathrm{EH}^b}$, so that some consumers obtain information, then consumers purchase the product if and only if $\theta\leq\frac{(1-r_m^b)(v-p_l^b)-k}{(1-r_m^b)l}\ll\frac{v-c}{2\mathrm{EH}^b}$ because the price falls from $p=(v+c)/2$ to $p_l^b=\frac{v+c-k/(1-r_m^b)}{2}$.

Using $k=\frac{r_m^b(1-r_m^b)(v-p)(m-l)}{\mathrm{EH}^b}$ with $p=(v+c)/2$ (to determine k) and a price of $p_l^b=\frac{v+c-k/(1-r_m^b)}{2}$ for purchases, this is:

$$\frac{m(1-r_h)(r_m(m-l)+2(1-r_h)l)(7r_m(m-l)+6(1-r_h)l)-4r_m^2(1-r_h-r_m)(m-l)^3}{32(1-r_h)ml(r_mm+(1-r_h-r_m)l)}(1-r_h-r_m)(v-c)^2>0$$

Thus, if $\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}<\frac{r_m^b(1-r_m^b)(v-p)(m-l)}{\mathrm{EH}^b}$, then there exists a k so that consumers only obtain information in the ban case and welfare is greater with a ban.

If $\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\geq\frac{r_m^b(1-r_m^b)(v-p)(m-l)}{\mathrm{EH}^b}$, then for k just smaller than $\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$ consumers will obtain information in the laissez faire case but not in the ban case. In this region, welfare in the no ban case is:

$$r_h\int_0^{\frac{v-p_h}{\mathrm{EH}}}(v-c-\theta h)d\theta+r_m\left\{\int_0^{\frac{(1-r_h)(v-p_m)-k}{(1-r_h)l+r_m(m-l)}}(v-c-\theta m)d\theta-k\left(\frac{(1-r_h)(v-p_m)-k}{(1-r_h)l+r_m(m-l)}-\frac{v-p_m+k/r_h}{h}\right)\right\}+(1-r_h-r_m)\left\{\int_0^{\frac{(1-r_h)(v-p_l)-k}{(1-r_h)l+r_m(m-l)}}(v-c-\theta l)d\theta-k\left(\frac{(1-r_h)(v-p_l)-k}{(1-r_h)l+r_m(m-l)}-\frac{v-p_l+k/r_h}{h}\right)\right\}$$

This follows because consumers only obtain information in the low or medium harm case where the price drops to induce more information acquisition. Using $k=\frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}$ with $p=(v+c)/2$ and $p_m=p_l=\frac{v+c}{2}-\frac{k}{2(1-r_h)}$ as given by Lemma 2 in the case of medium information costs, this is:

$$\frac{(v-c)^2}{32h\mathrm{EH}^2(r_mm+(1-r_h-r_m)l)}\{r_h^2(1-r_h)^2(3+4r_h)h^3+2r_h(1+r_h)(8-11r_h+6r_h^2)h^2(r_mm+(1-r_h-r_m)l)+(12-4r_h-5r_h^2+12r_h^3)h(r_mm+(1-r_h-r_m)l)^2+4r_h^2(r_mm+(1-r_h-r_m)l)^2\}$$

Subtracting the welfare from the no ban case with no information acquisition gives:

$$\frac{r_h(v-c)^2}{32h\mathrm{EH}^2(r_mm+(1-r_h-r_m)l)}\{-r_h(1-r_h)^2(9-4r_h)h^3-2(4-21r_h+17r_h^2-6r_h^3)h^2(r_mm+(1-r_h-r_m)l)+(20-17r_h+12r_h^2)h(r_mm+(1-r_h-r_m)l)^2+4r_h(r_mm+(1-r_h-r_m)l)^2\}$$

This expression is decreasing in h for $h\geq 2\mathrm{EH}$. At $h=2\mathrm{EH}$, it is:

$$\frac{2r_h(4-15r_h)(v-c)^2(r_mm+(1-r_h-r_m)l)^2}{32h\mathrm{EH}^2(1-2r_h)^3}$$

Thus, for $r_h > 4/15$ and $h \geq 2\text{EH}$, it is optimal to ban the high harm product when k is at the maximum level for some consumers to obtain information if there is no ban even if no consumers obtain information in the no ban case. □

Proof. (Propostion 8). First, consider the optimal tax on the high harm product. If $k \leq \frac{(\text{EH}-l)(1-r_h-r_m)(v-p)}{\text{EH}}$, then for uniform G, the first ordere condtion for the social welfare expression in (1) for the high harm product is:

$$\frac{k(l+2r_h(h-l))-r_h l(v-c+t_h)}{4r_h hl}=0 \tag{5}$$

This can only be satisfied for positive t_h if and only if $k > \frac{r_h(v-c)l}{2r_h(h-l)+l}$, in which case the solution is $t_h = k\left(\frac{1}{r_h}+\frac{2(h-l)}{l}\right)-(v-c)$. Using this t_h and the price from Lemma 2, the upper bound on consumers who purchase without obtaining information is $\theta < \frac{v-c}{h} - k\frac{h-l}{lh}$.

If harm is medium and $k \leq \frac{(\text{EH}-l)(1-r_h-r_m)(v-p)}{\text{EH}}$, then welfare is:

$$\int_0^{(v-p)/m} (v-c-\theta m)d\theta - k\left[G\left(\frac{(1-r_h-r_m)(v-p)-k}{(1-r_h-r_m)l}\right)-G\left(\frac{v-p+k/r_h}{h}\right)\right] \tag{6}$$

If G is uniform, then the derivative with respect to t_m evaluated at $t_m = 0$ and the maximum k is:

$$-\frac{(v-c)\}h(hr_h+mr_m)+(1-r_h-r_m)[hl-(h-l)m((r_h(h-l)+r_m(m-l)))]\}}{4h\,\text{mEH}} \tag{7}$$

This is always negative (the numerator is decreasing in l and always positive at $l = 0$), so it is never optimal to tax the product if harm is low if the cost of information is low.

If it's not optimal to tax the medium harm product, then it also won't be optimal to tax the low harm product. □

Proof. (Proposition 9) This proof is exactly analogous to the proof of Lemma 3 with the adjusted welfare function–using (2) instead of (1). □

Proof. (Lemma 10) If $k \leq \frac{(\text{EH}-l)(1-r_h-r_m)(v-p)}{\text{EH}}$, then under laissez faire profits from the medium harm product are $(v-c)^2/4m$ while profits from the low harm product are:

$$\int_0^{\frac{(1-r_h-rm)(v-p_l)-k}{(1-r_m-r_h)l}} (p_l-c)d\theta = \frac{((1-r_h-r_m)(v-c)-k)^2}{4(1-r_h-r_m)^2 l}$$

With a ban, profits from the medium harm product are:

$$\int_0^{\frac{(v-p_m)+k/r_m^b}{m}} (p_m-c)d\theta = \frac{(r_m(v-c)+(1-r_h)k)^2}{4r_m^2 m} > (v-c)^2/4m$$

and profits from the low harm product are:

$$\int_0^{\frac{(v-p_l)-k/(1-r_m^b)}{l}} (p_l-c)d\theta = \frac{((1-r_h-r_m)(v-c)-(1-r_h)k)^2}{4(1-r_h-r_m)^2 l} > \frac{((1-r_h-r_m)(v-c)-k)^2}{4(1-r_h-r_m)^2 l}$$

If $k \in \left(\frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}, \frac{r_h(v-p)(h-\mathrm{EH})}{\mathrm{EH}}\right)$, then in the laissez faire regime, demand for the medium and low harm product is $\frac{(1-r_h)(v-p)-k}{r_l l + r_m m}$. Recall that for smaller k, demand for the medium harm product is $\frac{(v-p)}{m}$. These are equal if $k = \frac{(m-l)(1-r_h-r_m)(v-p)}{m} < \frac{(\mathrm{EH}-l)(1-r_h-r_m)(v-p)}{\mathrm{EH}}$. Thus, medium harm profits are lower in the medium k region than in the low k, ensuring that they are also lower than in the ban regime (in which the profit function doesn't change as k moves into this region).

When the product is low harm, demand is determined by those who acquire information but wouldn't purchase if uninformed. Because the probability of learning the product is low harm is greater with the ban, the incentive for those high θ types to acquire information is greater with the ban, meaning that demand must be larger. So low harm profits must be larger with the ban. □

Proof. (Proposition 11) The difference in the marginal benefit from investment between a ban and no ban at the same investment level is:

$$q_l'(x)(\pi_l^b - \pi_l) + q_0'(x)(\pi_0^b - \pi_0) - q_h'(x)\pi_h$$

If $q_h'(x) < 0$, then this is positive because $\pi_l^b - \pi_l > 0$ and $\pi_0^b - \pi_0 > 0$. If $q_h'(x) > 0$, then this is positive if and only if the following is positive:

$$\frac{q_l'(x)}{q_h'(x)}(\pi_l^b - \pi_l) + \frac{q_0'(x)}{q_h'(x)}(\pi_0^b - \pi_0) - \pi_h > \frac{q_l(x)}{q_h(x)}(\pi_l^b - \pi_l) + \frac{q_0(x)}{q_h(x)}(\pi_0^b - \pi_0) - \pi_h$$

The last inequality follows from MLRP and is positive if and only if total expected profits are larger with the ban. □

Bibliography

Armstrong, Mark, “Search and Rip-off Externalities,” 47 Review of Industrial Organization 273-302 (2015).

Bar-Gill, Oren, David Schkade, and Cass R. Sunstein, “Drawing false inferences from mandated disclosures,” 3 Behavioral Public Policy 209-227 (2019).

Chan, Yuk-Shee and Hayne Leland, “Prices and Qualities in Markets with Costly Information,” 49 Review of Economic Studies 499-516 (1982).

Katz, Avery, “Your Terms or Mine? The duty to read the fine print in contracts,” 21 RAND Journal of Economics 518-37 (199).

Klein, Benjamin and Keith B. Leffler, “The Role of Market Forces in Assuring Contractual Performance,” 89 Journal of Political Economy 615-41 (1981).

Kominers, Scott Duke, Xiaosheng Mu, and Alexander Peysakhovich, “Paying (for) Attention: The Impact of Information Processing Costs on Bayesian Inference,” Working paper (2016).

Milgrom, Paul, “What the Seller Won't Tell You: Persuasion and Disclosure in Markets,” 22 Journal of Economic Perspectives 115-131 (2008).

Wickelgren, Abraham L. “Standardization as a Solution to the Reading Costs of Form Contracts,” 167 Journal of Institutional and Theoretical Economics 30-44 (2011).

Zhang, Juanjuan, “Policy and Inference: The case of product labeling,” Working paper available at https://www8.gsb.columbia.edu/programs/sites/programs/files/finance/COI.pdf (2014).