

# THE STRANGE CASE OF PRIVACY IN EQUILIBRIUM MODELS

RACHEL CUMMINGS\*    KATRINA LIGETT†    MALLESH M. PAI‡    AARON ROTH§

August 14, 2015

## ABSTRACT

We study how privacy technologies affect user and advertiser behavior in a simple economic model of targeted advertising. In our model, a consumer first decides whether or not to buy a good, and then an advertiser chooses an advertisement to show the consumer. The consumer’s value for the good is correlated with her type, which determines which ad the advertiser would prefer to show to her—and hence, the advertiser would like to use information about the consumer’s purchase decision to target the ad that he shows.

In our model, the advertiser is given only a differentially private signal about the consumer’s behavior—which can range from no signal at all to a perfect signal, as we vary the differential privacy parameter. This allows us to study equilibrium behavior *as a function of the level of privacy provided to the consumer*. We show that this behavior can be highly counter-intuitive, and that the effect of adding privacy in *equilibrium* can be completely different from what we would expect if we ignored equilibrium incentives. Specifically, we show that increasing the level of privacy can actually *increase* the amount of information about the consumer’s type contained in the signal the advertiser receives, lead to decreased utility for the consumer, and increased profit for the advertiser, and that generally these quantities can be non-monotonic and even discontinuous in the privacy level of the signal.

---

\*Computing and Mathematical Sciences, California Institute of Technology, Email: [rachelc@caltech.edu](mailto:rachelc@caltech.edu). Supported in part by a Simons Award for Graduate Students in Theoretical Computer Science, NSF grant CNS-1254169, US-Israel Binational Science Foundation grant 2012348, and a Google Faculty Research Award.

†Computing and Mathematical Sciences, California Institute of Technology, Email: [katrina@caltech.edu](mailto:katrina@caltech.edu). Supported in part by NSF grant CNS-1254169, US-Israel Binational Science Foundation grant 2012348, the Charles Lee Powell Foundation, a Google Faculty Research Award, an Okawa Foundation Research Grant, and a Microsoft Faculty Fellowship.

‡Department of Economics, University of Pennsylvania. Email: [malleshe@econ.upenn.edu](mailto:malleshe@econ.upenn.edu). Supported in part by NSF Grant CCF-1101389.

§Department of Computer and Information Science, University of Pennsylvania. Email: [aaroth@cis.upenn.edu](mailto:aaroth@cis.upenn.edu). Supported in part by NSF Grant CCF-1101389, an NSF CAREER award, and an Alfred P. Sloan Foundation Fellowship.

# 1 INTRODUCTION

As advertising becomes increasingly targeted and data driven, there is growing concern that the algorithms driving these personalization decisions can be *discriminatory*. For example, [Datta et al. \[2015\]](#) highlighted potential gender bias in Google’s advertising targeting algorithms, giving evidence that male job seekers were more likely to be shown ads for high paying jobs than female job seekers. Similarly, the FTC has expressed concern that data mining of user online behavior, which data brokers use to categorize users into categories such as “ethnic second-city struggler” and “urban scrambler”, is used to selectively target users for high interest loans [[Ryan, 2014](#)].

One tempting response to such concerns is regulation: for example, we could mandate the use of privacy technologies which would explicitly limit the amount of information advertisers could learn about users past behavior in a quantifiable way.<sup>1</sup> If advertisers are only able to see differentially private signals about user behavior, for example, then we can precisely quantify the amount of information that the advertiser’s signal contains about the actions of the user. As we increase the level of privacy, we would naively expect to see several effects: first, the amount of information that the advertiser learns about the user should decrease. Second, the utility of the advertiser should decrease, since she is now less able to precisely target her advertisements. Finally, if the user really was experiencing disutility from the way that the advertiser had been targeting her ads, the user’s utility should increase.

These expectations are not necessarily well grounded, however, for the following reason: in strategic settings, as the information content of the signal that the advertiser receives changes, he will change the way he uses the information he receives to target ads. Similarly, given the way that user behavior is used in ad targeting, a sophisticated user may change her browsing behavior. Hence, it is not enough to statically consider the effect of adding privacy technologies to a system, but instead we must consider the effect in equilibrium. Each privacy level defines a different strategic interaction which results in different equilibrium behavior, and a-priori, it is non-obvious the effect that the privacy technology will have on the equilibrium outcome.

In this paper, we consider a very simple two-stage model of advertising that may be targeted based on past user behavior. The model has three rational agents: a consumer, a seller, and an advertiser. The consumer has a value for the good being sold by the seller, and also a type which determines which of several ads the advertiser benefits most from showing the consumer. The consumer’s value and type are drawn from a joint distribution, and the two are correlated — hence, information about the consumer’s value for the good being sold is relevant to the advertiser when determining what ad to show her.

The game proceeds in two stages. In the first stage, the seller determines a price to set for the good he is selling — then the consumer determines whether or not she wishes to buy the good. In the second stage, the advertiser receives some signal about the purchase decision of the consumer in the first round. The signal may be noisy; the correlation of the signal with the consumer’s purchase decision reflects the level of privacy imposed on the environment, and is quantified via differential privacy. As a function of the signal, the advertiser performs a Bayesian update to compute his posterior belief about the consumer’s type, and then decides which ad to show the consumer. The consumer has a preference over which ad she is shown — for example, one might be for a credit card with a lower interest rate, or for a job with a higher salary. Hence, the consumer does not necessarily act myopically in the first round when deciding whether to purchase the seller’s good or not, and instead takes into account the effect that her purchase decision will have on the second round.

We characterize the equilibria of this model as a function of the level of differential privacy provided by the signal the advertiser receives. We show that in this model, several counter-intuitive phenomena can arise. For example, the following things may occur in equilibrium as we increase the privacy level (i.e. decrease the correlation between the user’s purchase decision and the signal received by the advertiser):

1. The signal received by the advertiser can actually contain *more* information about the agent’s type, as measured by mutual information (Figure 3). Similarly, the difference in the advertiser’s posterior belief about the true type of the agent after seeing the purchase/ did not purchase noisy bits can increase (Figure 2). (Interestingly, this difference does not necessarily peak at the same privacy level as the mutual information between the consumer’s type and the advertiser’s signal).

---

<sup>1</sup>An alternative approach to limiting the information that advertisers can collect is to explicitly try to limit how they use that information to avoid unfairness. See [Dwork et al. \[2012\]](#) for work in this direction.

2. Consumer utility can decrease, and advertiser utility can increase (Figure 4, 5).
3. More generally, these quantities can behave in complicated ways as a function of the privacy parameter  $\epsilon$ : they are not necessarily monotone in  $\epsilon$ , and can even be discontinuous, and equilibrium multiplicity can vary with  $\epsilon$  (Figure 2, 6, 7).

Our work also gives a precise way to derive the functional form of *value of privacy* for players (at least in our simple model), in contrast to a large prior literature that has debated the right way to impose exogenously a functional form on player privacy cost functions Ghosh and Roth [2013], Xiao [2013], Chen et al. [2013], Nissim et al. [2014]. In contrast to these assumed privacy cost functions, which increase as privacy guarantees are weakened, we show that players can actually sometimes have a negative marginal cost (i.e. a positive marginal utility) for *weakening* the privacy guarantees of a mechanism.

In summary, we show that even in extremely simple models, privacy exhibits much richer behavior in equilibrium compared to its static counterpart, and that decisions about privacy regulation need to take this into account. Our results serve as a call-to-arms — that especially when making policy decisions about privacy technologies, *equilibrium effects* need to be analyzed, rather than just static effects, because it is otherwise possible that the introduction of a new privacy technology or regulation can have exactly the opposite effect as was intended.

## 1.1 RELATED LITERATURE

Differential privacy is a quantitative privacy measure, first introduced by Dwork et al. [2006], that formally quantifies the effect that the behavior of a single individual can have on a signal computed from his behavior. There is a vast literature on algorithms satisfying differential privacy, and we refer readers to Dwork and Roth [2014] for a textbook introduction. Differential privacy has been studied in the context of game theory and mechanism design since McSherry and Talwar [2007], who showed that it could be used as a *tool* in mechanism design. Subsequently, a moderately sized literature has emerged studying how differential privacy can be used to design mechanisms that incentivize truthful behavior even in settings in which agents have costs for privacy loss — see e.g. Xiao [2013], Ghosh and Roth [2013], Chen et al. [2013], Nissim et al. [2012, 2014] for a representative but not exhaustive sample. Blum et al. [2015] study a sequential coordination game inspired by financial markets, and show that when players play un-dominated strategies, the game can have substantially higher social welfare when players are given differentially private signals of each other’s actions, as compared to when they are given no signal at all. Here, like in other work, however, privacy is viewed as a constraint on the game (i.e. it is added because of its own merits), and does not increase welfare compared to the full information setting. Ghosh and Ligett [2013] study a simple model of data procurement in which user costs for privacy are a function of the number of other users participating — and study the ability of a mechanism to procure data in equilibrium. In Ghosh and Ligett [2013], agents have explicitly encoded values for their privacy loss. In contrast, in our model, agents do not care about privacy except insofar as it affects the payoff-relevant outcomes of the game they are playing — differential privacy in our setting is instead a parameter defining the game we analyze.

A small literature in economics and marketing has looked to understand the effect of privacy in repeated sales settings. The earliest paper is by Taylor [2004] who studies a setting where buyers purchase from firm 1 in period 1 and firm 2 in period 2. The author shows that counter-intuitively, strategic consumers may prefer that their purchase decision be made public, while strategic sellers may prefer to commit to keep purchase decisions private. Calzolari and Pavan [2006] derive similar results in a general contracting setting.

More recently, Conitzer et al. [2012] consider a setting where a buyer purchases twice from the same firm, and the firm cannot commit to the future price in period 1, and may condition on the consumer’s purchase decision. They consider the effect of allowing the buyer to purchase privacy, i.e. “hide” his first period purchase decision from the seller, and show that in equilibrium, having this ability may make buyers worse off (and the firm better off) than in its absence.

We build on these papers by here modeling privacy as a continuous choice variable in the spirit of differential privacy, rather than the discrete choice (purchase decision revealed or not) considered in previous papers. This allows us to analyze the quantitative effect of privacy on welfare and profit as a continuous quantity, rather than a binary effect, and in particular lets us show for the first time that *increasing* privacy

protections (i.e. decreasing the correlation between the advertiser’s signal and the buyer’s action) can actually increase the information contained in the signal about the buyer’s type.

## 2 MODEL AND PRELIMINARIES

We study a two period game with a single consumer and two firms. The first firm has a single good to sell to the consumer and wishes to maximize its expected profit. The second firm is an advertiser who wishes to show a targeted ad to the consumer. We will also refer to the first firm as the *seller*, and the second firm as the *advertiser*.

1. In period 1, the consumer has a privately known value  $v \in [0, 1]$  for the good, drawn from a distribution with CDF  $F$  and density  $f$ . The seller posts a take-it-or-leave-it price  $p$  for the good, and the consumer makes a purchase decision. We assume that the seller’s price is not observed by the advertiser.
2. In period 2, the consumer can have one of two types,  $t_1$  and  $t_2$ , where the probability of having each type depends on her value  $v$  from period 1. Specifically,  $\Pr(t_1) = g(v)$  and  $\Pr(t_2) = 1 - g(v)$ , for a known function  $g : [0, 1] \rightarrow [0, 1]$ . The advertiser may show the buyer one of two ads,  $A$  and  $B$ . He gets payoff  $s_{1A}$  and  $s_{2A}$  respectively from showing ad  $A$  to a consumer of type  $t_1$  and  $t_2$ , and payoffs  $s_{1B}$  and  $s_{2B}$  from showing ad  $B$  to a consumer of type  $t_1$  and  $t_2$  respectively. The consumer gets additional utility  $\delta$  from being shown ad  $A$  over  $B$ .<sup>2</sup>

ASSUMPTION 1. The following two assumptions are made regarding the distribution of buyer’s value and type:

1. The distribution of the buyer’s value satisfies the non-decreasing hazard rate assumption, i.e.  $\frac{f(v)}{1-F(v)}$  is non-decreasing in  $v$ .
2. The probability that a buyer of value  $v$  is of type  $t_1$ ,  $g(v)$ , is non-decreasing in  $v$ .
3. Buyers prefer ad  $A$ , i.e.  $\delta > 0$ .

Regarding the payoffs to the advertiser, we assume that  $s_{1A} > s_{1B}$  and  $s_{2B} > s_{2A}$ .

The payoff assumption corresponds to the case where type  $t_1$  is the “high type” and ad  $A$  is the “better” ad. For example, the ad could be for a credit card. The advertiser can offer either a card with a low interest rate (ad  $A$ ) or high interest rate (ad  $B$ ), and the consumer’s purchase history may reveal his creditworthiness. The former distributional assumption is standard in mechanism design. The latter assumption amounts to saying that high-value buyers are more likely to be the ones the advertiser wants to target with the “good” ad.

The advertiser neither observes the consumer’s type, nor directly observes his purchase decision. Following the consumer’s decision, the advertiser will learn (partial) information about the consumer’s action in the first period. We write  $b$  to denote the bit encoding the consumer’s decision in the first period — that is,  $b = 1$  if the consumer purchased the good in period 1, and  $b = 0$  otherwise. The advertiser does not learn  $b$  exactly, but rather a noisy version  $\hat{b}$  that has been flipped with probability  $1 - q$ , for  $q \in [1/2, 1]$ . The advertiser observes the noisy bit  $\hat{b}$  and then performs a Bayesian update on his beliefs about the consumer’s type, and displays the ad that maximizes his (posterior) expected payoff. The consumer knows that her period 1 purchase decision will affect the ad she sees in period 2. She seeks to maximize her total utility over both periods, and thus is not myopic.

The parameter  $q$  measures the correlation of the reported bit with the actual purchase decision of the consumer, which can also be quantified via differential privacy:

DEFINITION 1 (Dwork et al. [2006]). A signal  $\hat{b} : \{0, 1\} \rightarrow \{0, 1\}$  satisfies  $\epsilon$ -differential privacy if for every  $b \in \{0, 1\}$  and for every  $b' \in \{0, 1\}$ :

$$\Pr[\hat{b}(b) = b'] \leq \exp(\epsilon) \Pr[\hat{b}(1 - b) = b']$$

<sup>2</sup>Since the customer’s preferences in period 2 are independent of her type, it does not matter whether she knows her period 2 type from the beginning or learns it at the start of period 2.

Differential privacy quantifies in a strong sense what an observer can learn about the input bit from the output — see [Kasiviswanathan and Smith \[2014\]](#) for an analysis of the semantics of this guarantee — and among other things provides an upper bound on the mutual information between the input and output bit.

In our setting, it is easy to translate the parameter  $q$  into an  $\epsilon$ -differential privacy guarantee for the Period 1 purchase decision as follows:

$$\frac{q}{1-q} = e^\epsilon \iff q = \frac{e^\epsilon}{1+e^\epsilon}.$$

In the two extremal cases, full privacy ( $q = 1/2$ ) corresponds to  $\epsilon$ -differential privacy with  $\epsilon = 0$ , and no privacy ( $q = 1$ ) corresponds to  $\epsilon$ -differential privacy with  $\epsilon = \infty$ . By varying  $q$  from  $1/2$  to  $1$ , we will be able to measure changes to the equilibrium outcomes for all possible privacy levels.

### 3 EQUILIBRIUM ANALYSIS

To begin, we observe that any equilibrium must be such that the consumer follows a cutoff strategy in Period 1: there exists a marginal consumer with value  $v^*$ , such that any consumer with value  $v \in [0, v^*)$  does not buy the good, and any consumer with value  $v \in [v^*, 1]$  does. We formally verify this in the Appendix (Proposition 7).

#### 3.1 PERIOD 2

The advertiser sees  $\hat{b}$ , the noisy purchase decision bit, and performs a Bayesian update on its prior over types. This allows us to define the advertiser’s posterior given an observed  $\hat{b} = j$ , for  $j \in \{0, 1\}$ . Recall that  $\hat{b} = b$  with probability  $q$ , and  $\hat{b} = 1 - b$  with probability  $1 - q$ .

Plugging in the probabilities for each  $j \in \{0, 1\}$ , we see that the advertiser’s posterior when he understands the consumer is following a threshold strategy with cutoff  $v^*$  is as follows:

$$\begin{aligned} r(\hat{b} = 1, v^*, q) &= \frac{(1-q) \int_0^{v^*} g(v)f(v)dv + q \int_{v^*}^1 g(v)f(v)dv}{(1-q)F(v^*) + q(1-F(v^*))}, \\ r(\hat{b} = 0, v^*, q) &= \frac{q \int_0^{v^*} g(v)f(v)dv + (1-q) \int_{v^*}^1 g(v)f(v)dv}{qF(v^*) + (1-q)(1-F(v^*))}. \end{aligned}$$

Here  $r(\hat{b}, v^*, q)$  is the advertiser’s posterior belief that the consumer is of type  $t_1$  after seeing the noisy bit  $\hat{b}$ , given that  $v^*$  is the marginal consumer in Period 1 and given the noise level  $q$ .

The advertiser wishes to maximize his expected payoff, so his Bayesian optimal decision rule will be to show ad  $A$  if and only if,

$$s_{1A}r(\hat{b}, v^*, q) + s_{2A}(1 - r(\hat{b}, v^*, q)) > s_{1B}r(\hat{b}, v^*, q) + s_{2B}(1 - r(\hat{b}, v^*, q)).$$

That is, he will show ad  $A$  if it maximizes his expected payoff. Rearranging this in terms of his posterior, he will show ad  $A$  if and only if,

$$r(\hat{b}, v^*, q) > \frac{s_{2B} - s_{2A}}{s_{1A} - s_{2A} - s_{1B} + s_{2B}} := \eta.$$

We define this fraction to be  $\eta$  for shorthand. Notice that  $\eta$  does not depend on any game parameters other than the advertiser’s payoff for each outcome,  $s_{1A}$ ,  $s_{2A}$ ,  $s_{1B}$ , and  $s_{2B}$ . Further, by our assumptions on the ranking of these four (Assumption 1) we have ensured that  $\eta \in [0, 1]$ .

The following lemma will be useful throughout. It says that fixing a cutoff strategy  $v^*$ , the seller’s posterior on seeing a noisy “purchased” bit is increasing as the amount of noise decreases (i.e.  $q$  increases), and similarly the seller’s posterior on seeing a noisy “did not purchase” bit is decreasing.

LEMMA 1. *Fixing  $v^*$ ,  $r(1, v^*, q)$  is increasing in  $q$  and  $r(0, v^*, q)$  is decreasing in  $q$ .*

PROOF. Define  $\alpha_1 = \frac{\int_0^{v^*} g(v)f(v)dv}{F(v^*)}$ ,  $\alpha_2 = \frac{\int_{v^*}^1 g(v)f(v)dv}{1-F(v^*)}$ . Note that since  $g(\cdot)$  is non-decreasing,  $\alpha_1 \leq \alpha_2$ . Therefore,  $r(1, v^*, q)$  can be written as:

$$r(1, v^*, q) = \frac{(1-q)F(v^*)\alpha_1 + q(1-F(v^*))\alpha_2}{(1-q)F(v^*) + q(1-F(v^*))}.$$

This is the convex combination of  $\alpha_1$  and  $\alpha_2$  with weights  $\frac{(1-q)F(v^*)}{(1-q)F(v^*) + q(1-F(v^*))}$  and  $\frac{q(1-F(v^*))}{(1-q)F(v^*) + q(1-F(v^*))}$  respectively. Next, note that for  $q \in [1/2, 1]$  the weight on  $\alpha_2$  is increasing in  $q$ , and the weight on  $\alpha_1$  correspondingly decreasing. To see this, differentiate the weight with respect to  $q$  and observe that it is always positive. Therefore  $r(1, v^*, q)$  is increasing in  $q$ .

Finally, note that  $r(0, v^*, q) = r(1, v^*, 1 - q)$ , so the latter claim follows.  $\square$

The following proposition is an important property of the advertiser's posterior, i.e. that in any equilibrium, seeing a noisy "purchased" bit always results in a higher assessment of type  $t_1$  than a noisy non-purchased bit.

LEMMA 2. *For any period 1 cutoff value  $v^*$  and any noise level  $q$ , the advertiser's posterior probability of the consumer having type  $t_1$  given noisy bit  $\hat{b} = 1$  is higher than his posterior belief of type  $t_1$  given noisy bit  $\hat{b} = 0$ . Formally, for all  $v^*, q$ , it holds that  $r(1, v^*, q) \geq r(0, v^*, q)$ .*

PROOF. The proposition follows from Lemma 1, the fact that  $r(0, v^*, q) = r(1, v^*, 1 - q)$ , and  $q \geq 0.5$ .  $\square$

In light of this, there are only three different strategies that the advertiser could use in equilibrium:

1. Show ad  $A$  to a consumer with noisy bit  $\hat{b} = 1$  and ad  $B$  to consumer with noisy bit  $\hat{b} = 0$ . This is characterized by the following inequalities:

$$r(1, v^*, q) > \eta \quad \text{and} \quad r(0, v^*, q) < \eta \tag{1}$$

2. Always shows ad  $A$ , regardless of the observed noisy bit  $\hat{b}$ . This is optimal for the advertiser when the parameters are such that:

$$r(1, v^*, q) > \eta \quad \text{and} \quad r(0, v^*, q) > \eta \tag{2}$$

3. Always shows ad  $B$ , regardless of the observed noisy bit. This is optimal for the advertiser when the parameters are such that:

$$r(1, v^*, q) < \eta \quad \text{and} \quad r(0, v^*, q) < \eta \tag{3}$$

In the latter two cases, consumers will behave myopically in the first round because their purchase decision doesn't affect their payoff in the next round. The seller can then maximize period 1 profits by posting the monopoly price for the distribution  $F$ . Thus cases 2 and 3 can only occur when the posterior induced by the monopoly price satisfies (2) or (3).

We call the equilibrium when the advertiser follows the first strategy a *discriminatory advertising equilibrium*. The latter two are referred to as *uniform advertising equilibria A* and *B* respectively

Define the myopic monopoly price as  $p_M$ , i.e.  $p_M$  solves:

$$p_M - \frac{1 - F(p_M)}{f(p_M)} = 0. \tag{4}$$

The following proposition discusses existence and properties of uniform advertising equilibria.

PROPOSITION 1. *Fixing other parameters of the game:*

1. For  $q = \frac{1}{2}$  there is either a uniform advertising equilibrium  $A$  or  $B$ , but never both.
2. In the former case: uniform advertising equilibria  $A$  exist for all  $q \in [\frac{1}{2}, \bar{q}_2]$ , where  $\bar{q}_2$  is the largest solution to  $r(0, p_M, q) = \eta$  in  $[\frac{1}{2}, 1]$ , if any. Further, there are no uniform advertising equilibria  $B$  for any  $q$ .

3. In the latter, conversely, uniform advertising equilibria B exist for all  $q \in [\frac{1}{2}, \bar{q}_3]$ , where  $\bar{q}_3$  is the largest solution to  $r(1, p_M, q) = \eta$  in  $[\frac{1}{2}, 1]$ , if any. Further there are no uniform advertising equilibria A for any  $q$ .

PROOF. At  $q = 1/2$ , the signal  $\hat{b}$  is complete noise, and the advertiser's posterior on types will be exactly his prior. This means that the advertiser must show the same ad to all consumers, which corresponds exactly to a uniform advertising equilibrium A or B, depending on whether his prior probability of type  $t_1$  is larger or smaller than  $\eta$ . The corner case where the prior exactly equals  $\eta$  is ignored.

Note that  $r(1, p_M, \frac{1}{2}) = r(0, p_M, \frac{1}{2})$ , and both are either  $> \eta$  or  $< \eta$ . Further, by Lemma 1,  $r(1, p_M, q)$  is increasing in  $q$ , while the  $r(0, p_M, q)$  is decreasing in  $q$ .

Therefore the system of equations (2) or (3) can only be satisfied on some interval  $[\frac{1}{2}, \bar{q}_j]$  if at all for  $v^* = p_M$ .  $\square$

To collect everything we have shown so far. There are 3 kinds of possible equilibria in this game:

1. *Discriminatory Equilibrium*: Advertiser shows ad A on seeing a noisy purchase bit and ad B and seeing a noisy non-purchase bit. The cutoffs followed by the consumer,  $v^*$  is such that (1) is satisfied.
2. *Uniform Advertising Equilibrium A*: Advertiser always shows ad A, regardless of bit. In this case buyer purchases myopically, and seller charges the myopic monopoly price  $p_M$ . Further,  $r(\cdot)$  evaluated at  $v^*$  equaling the myopic monopoly price  $p_M$  satisfies (2). This equilibrium exists for all  $q$  on some interval  $[\frac{1}{2}, \bar{q}_2]$ , if at all.
3. *Uniform Advertising Equilibrium B*: Advertiser always shows ad B, regardless of bit. In this case buyer purchases myopically, and seller charges the myopic monopoly price  $p_M$ . Further,  $r(\cdot)$  evaluated at  $v^*$  equaling the myopic monopoly price  $p_M$  satisfies (3). This equilibrium exists for all  $q$  on some interval  $[\frac{1}{2}, \bar{q}_3]$ , if at all.

By observation, the two types of uniform advertising equilibria cannot coexist in the same game.

In a uniform advertising equilibrium, the period 1 behavior is straightforward. We now finish the analysis by characterizing period 1 behavior under discriminatory advertising equilibria.

### 3.2 PERIOD 1 BEHAVIOR IN DISCRIMINATORY ADVERTISING EQUILIBRIA

In this kind of equilibrium the consumer is aware that her period 1 purchasing decisions will affect the ad she sees in period 2. She will buy in period 1 if and only if her surplus from purchasing at price  $p$  plus her continuation payoff from having purchased (i.e. expected utility from the ad that will be shown) is greater than the continuation payoff from not having purchased. Formally, a consumer with value  $v$  will purchase in period 1 if the following holds:

$$(v - p) + q\delta \geq (1 - q)\delta. \quad (5)$$

For the marginal consumer with value  $v^*$ , this inequality must hold with equality.

$$v^* = p + (1 - 2q)\delta.$$

Define  $p_1(q)$  to be the seller's optimal price charged at noise level  $q$ . Further define  $v^*(q)$  as the implied cutoff type at noise level  $q$ , i.e.  $v^*(q) = p_1(q) + (1 - 2q)\delta$ . Note that both  $p_1(q)$  and  $v^*(q)$  are continuous functions of  $q$ .

LEMMA 3. *Assuming a discriminatory advertising equilibrium exists for a neighborhood of  $q \in [\frac{1}{2}, 1]$ , the optimal price  $p_1(q)$  is increasing in  $q$  while the cutoff type  $v^*(q)$  is decreasing.*

PROOF. From equation (5) above, if the seller charges a price of  $p$  in a discriminatory advertising equilibrium, then the buyer purchases if her value exceeds  $p + (1 - 2q)\delta$ . Therefore the seller chooses  $p$  to maximize his net profit,

$$p(1 - F(p + (1 - 2q)\delta)).$$

Differentiating with respect to  $p$ , the optimal price ( $p_1(q)$ ) in a discriminatory advertising equilibrium solves:

$$p_1(q) - I(p_1(q) + (1 - 2q)\delta) = 0, \quad (6)$$

where  $I(v) = \frac{1-F(v)}{f(v)}$ . Applying the implicit function theorem, we see that

$$p_1'(q) - I'(p_1 + (1 - 2q)\delta)(p_1'(q) - 2\delta) = 0.$$

Therefore, since  $I'$  is negative, it follows that  $p_1'(\cdot)$  must be positive. Further, we have that  $p_1'(q) - 2\delta$  is negative, i.e. the cutoff value who buys at the optimal price charged is decreasing in  $q$ .  $\square$

The following proposition says that whenever a discriminatory advertising equilibrium exists at a given  $q$ , the price is higher and more customers buy than in either uniform advertising equilibrium.

LEMMA 4. *For any  $q \in [1/2, 1]$ , the period 1 price in a discriminatory advertising equilibrium (if it exists), is higher than the monopoly price (i.e. the price charged in a uniform advertising equilibrium), which is higher than the purchase cutoff employed by a consumer in a discriminatory advertising equilibrium. Formally,*

$$v^*(q) \leq p_M \leq p_1(q).$$

PROOF. Note that at  $q = \frac{1}{2}$ ,  $p_1(q) = v^*(q) = p_M$ . The result now follows since  $p_1'(\cdot)$  is positive, while  $v^*(\cdot)$  is negative.  $\square$

Finally, to resolve existence, which follows easily from the definitions.

OBSERVATION 1. A discriminatory advertising equilibrium exists at all  $q$  such that  $r(1, v^*(q), q) \geq \eta$  and  $r(0, v^*(q), q) \leq \eta$ .

Next, note that  $v^*(q)$  is a continuous function of  $q$ . Therefore, equilibria of type 1 may exist for possibly multiple disjoint intervals in  $(\frac{1}{2}, 1]$ .

## 4 ILLUSTRATIONS VIA AN EXAMPLE

In this section we highlight some of the counter-intuitive effects that result from changing the level of differential privacy constraining the advertiser's signal, by means of an explicit family of simple examples. The phenomena we highlight are quite general, and are generally not brittle to the choice of specific parameters in the game. For simplicity of exposition, we highlight each of these phenomena in the simplest example in which they arise. For the remainder of this section, we take the distribution of buyer values,  $F$ , to be the uniform distribution on  $[0, 1]$ . We also set the buyer's probability of having type  $t_1$  to be exactly his value — i.e. we take  $g(v) = v$  for all  $v \in [0, 1]$ . Finally, we set the additional utility that a buyer gets from being shown ad A to be  $\delta = 1$ . The value of parameter  $\eta$  (along with its defining parameters  $s_{1A}$ ,  $s_{1B}$ ,  $s_{2A}$ , and  $s_{2B}$ ) will vary by example, and will be specified when relevant.

The static monopoly price in this game is  $p_M = 1/2$ , and the price and cutoff value in a discriminatory equilibrium (if it exists at a given  $q$ ) are

$$p_1(q) = \frac{1}{2} + (2q - 1)\frac{\delta}{2} \quad \text{and} \quad v^*(q) = \frac{1}{2} - (2q - 1)\frac{\delta}{2}.$$

Below we plot these values as a function of  $q$ . Note that in this particular example, the discriminatory price and cutoff value are linear in  $q$  (because values are distributed uniformly), although this need not be the case in general.

Figure 1 shows that as  $q$  increases, the discriminatory equilibrium price increases and cutoff value (value of marginal consumer that purchases in period 1) decreases. Relative to a uniform advertising equilibrium, more consumers purchase the good in period 1, and at a higher price in a discriminatory equilibrium. Observe further that at  $q = 1$  (i.e. no privacy), all consumers purchase in period 1, regardless of their value. This is because the value in period 2 of hiding information relevant to their type exceeds the loss that they take by buying at a loss in period 1. Here, when the consumers are offered no privacy protections, they in effect change their behavior to guarantee their own privacy.

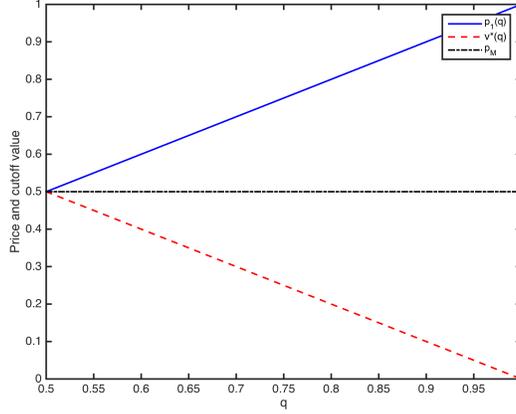


FIGURE 1: A plot of the monopoly price  $p_M$  and the equilibrium discriminatory price  $p_1$  and cutoff value  $v^*$  as a function of  $q$ . In a uniform equilibrium, the cutoff value is equal to the monopoly price because consumers behave myopically.

The existence and types of equilibria in this game depend on the advertiser’s posterior, given that the prices and cutoff values above will arise in period 1. The advertiser’s posterior beliefs about the consumer’s type for each realization of  $\hat{b}$  are given below for both the discriminatory and uniform advertising equilibria.

$$\begin{aligned}
 r(1, v^*(q), q) &= \frac{-2q^3 + 5q^2 - 3q + 1}{4(q^2 - q + \frac{1}{2})} \\
 r(0, v^*(q), q) &= \frac{-2q^3 + 5q^2 - 3q}{4(q^2 - q)} = \frac{3 - 2q}{4} \\
 r(1, p_M, q) &= \frac{1 + 2q}{4} \\
 r(0, p_M, q) &= \frac{3 - 2q}{4}
 \end{aligned}$$

To illustrate the existence of various equilibria, we plot the advertiser’s possible posterior beliefs below as a function of  $q$ .

Note the following counter-intuitive fact: the advertiser’s posterior, having seen a noisy “purchased” bit, i.e.,  $\hat{b} = 1$ , in a discriminatory equilibrium is non-monotone in the noise level  $q$ . Statically, if we were to increase the privacy level (i.e. decrease  $q$ ), we should always expect the advertiser’s posterior to be *less informative* about the consumer’s type, but as we see here, in equilibrium, increasing the privacy level can sometimes make the advertiser’s posterior more accurate. This can occur because of the two competing implications of adding less noise (i.e. increasing  $q$ ): on the one hand, the observed bit  $\hat{b}$  is less noisy, and is thus a more accurate indicator of the consumer’s purchase decision in period 1. On the other hand, as  $q$  increases, a larger fraction of consumers buy in period 1; the pool of consumers who do purchase the item is “watered down” by low-valued consumers who are unlikely to have type  $t_1$ . This can be viewed as a larger fraction of consumers modifying their behavior to guarantee their own privacy, as the privacy protections inherent in the market are weakened. The dominating effect on the posterior depends on  $q$  and the game parameters.

Similar non-monotonicities can never be observed when the advertiser sees  $\hat{b} = 0$  in a discriminatory equilibrium because these two implications are no longer at odds. As we reduce the amount of noise added, the noisy bit  $\hat{b}$  is still more likely to be accurate. In addition, the maximum value  $v^*(q)$  of a consumer who did not purchase is decreasing in  $q$ , ensuring that only the lowest valued (and thus the least likely to have type  $t_1$ ) consumers do not purchase in period 1. These two effects conspire to ensure that  $r(0, v^*(q), q)$  is monotonically decreasing in  $q$ .

Figure 2 can be also be used to illustrate the existence of equilibria in this family of games. Specifying a value of  $\eta$  for the game determines the type of equilibria (if any) that exist at each  $q$ , according to Conditions

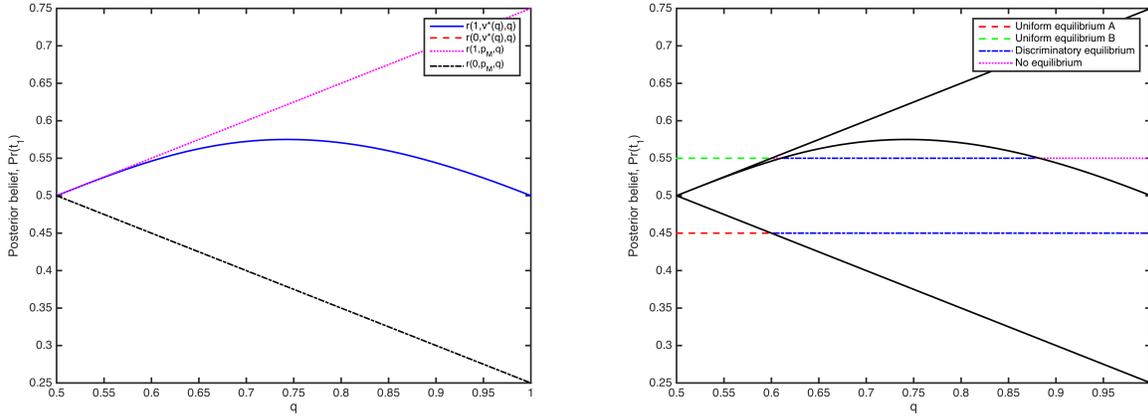


FIGURE 2: On the left is a plot of the advertiser’s possible posteriors in both discriminatory and uniform equilibria, given the corresponding prices and cutoff values of Period 1 and an observation of  $\hat{b}$ . In this example,  $r(0, v^*(q), q) = r(0, p_M, q)$  for all  $q$ . On the right, we show for  $\eta = 0.45$  and  $\eta = 0.55$  how these posteriors correspond to different equilibrium types as  $q$  varies.

(1), (2), and (3). This can be easily visualized using Figure 2.

Equilibria need not exist for all ranges of  $q$ : it is possible for none of Conditions (1), (2), or (3) to be satisfied for a given  $\eta$  and  $q$ . However, in all games, there is a uniform equilibrium at  $q = 1/2$ , where consumers behave myopically in period 1 and then no information is shared with the advertiser. Equilibria also need not be unique for a given  $q$ ; a discussion of equilibrium multiplicity is deferred to Section 5.

Next we show that in settings for which a discriminatory equilibrium exists for a range of  $q$ , the mutual information between the noisy bit  $\hat{b}$  and the consumer’s type can be non-monotone in  $q$ . In particular, as we increase the privacy protections of the market (i.e. decrease  $q$ ), we can sometimes end up *increasing* the amount of information about the consumer’s type present in the advertiser’s signal! This is for similar reasons to those that lead to non-monotonicity of the advertiser’s posterior belief — as the market’s privacy protections decrease, consumers change their behavior in order to guarantee their own privacy. Figure 3 plots the mutual information between  $\hat{b}$  and the consumer’s type as a function of  $q$  in a discriminatory equilibrium.<sup>3</sup> Note that although both the advertiser’s posterior and the mutual information between the consumer’s purchase decision and the signal exhibit similar non-monotonicities in  $q$ , they do not peak at the same value of  $q$ !

These phenomena together suggest that in the range of  $q$  in which the mutual information is decreasing, the advertiser might actually prefer that the market include stronger privacy guarantees. Indeed, Figure 4 plots the advertiser’s utility in a discriminatory equilibrium as a function of  $q$ , where  $s_{1A} = s_{2B} = 1$  and  $s_{1B} = s_{2A} = 0$ . This setting of parameters gives  $\eta = 1/2$ , where a discriminatory equilibrium exists for all  $q$ .

As predicted, the advertiser’s utility is non-monotone in  $q$ . However, the advertiser’s utility is not maximized at the same value of  $q$  that maximizes the mutual information. Thus, the advertiser’s interests are not necessarily aligned with the goal of learning as much information about the consumer as possible — and are certainly not incompatible with privacy technologies being introduced into the market. Indeed, the ideal level of  $q$  for the advertiser is strictly on the interior of the feasible set  $[1/2, 1]$ .

We would also like to understand how consumer surplus and profit vary with  $q$ . These are confounded by the fact that for a fixed level of  $\eta$ , an equilibrium type may or may not exist for a given  $q$ . To simplify the analysis to not account for this existence problem, consider the following thought experiment: For each equilibrium type, and any  $q$ , pick  $\eta$  such that the appropriate one of Conditions (1), (2), or (3) is satisfied. Fixing the advertiser’s equilibrium behavior,  $\eta$  only affects the advertiser’s payoff, not the seller’s or buyer’s. Figure 5 plots the consumer’s surplus and the seller’s profit as a function of  $q$  for each equilibrium type, under this artificial thought experiment.

<sup>3</sup>As illustrated by Figure 2, there are games for which a discriminatory equilibria exist for the relevant range of  $q$ , e.g., when  $\eta = 1/2$ , a discriminatory equilibrium exists for all  $q \in [1/2, 1]$ .

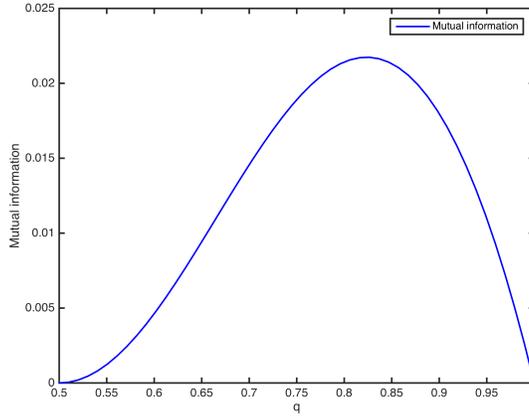


FIGURE 3: A plot of the mutual information between  $\hat{b}$  and the consumer's type in a discriminatory equilibrium.

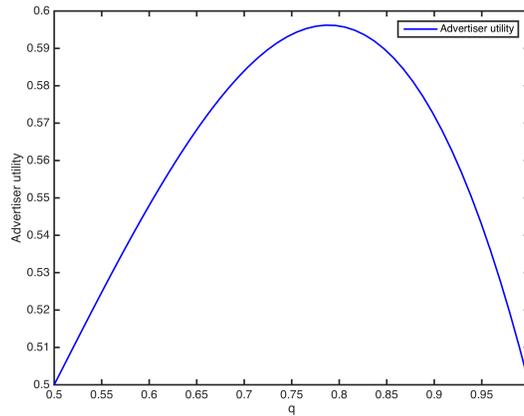


FIGURE 4: A plot of the advertiser's utility in a discriminatory equilibrium, where  $s_{1A} = s_{2B} = 1$  and  $s_{1B} = s_{2A} = 0$ . These parameters correspond to  $\eta = 1/2$ , which ensures that a discriminatory equilibrium exists for all  $q \in [1/2, 1]$ .

In a discriminatory equilibrium, the consumer surplus is convex in  $q$ ; and although at certain points on the curve, consumers have positive marginal utility for increased privacy (i.e. a smaller value of  $q$ ), the consumer's globally optimal value is  $q = 1$ . This means that given the equilibrium effects, consumers fare the best with no privacy and would prefer to have their purchase decision revealed exactly. The seller also prefers  $q = 1$ , because as  $q$  increases, more consumers purchase the good at a higher price in equilibrium. Unsurprisingly, consumer surplus and revenue are constant across  $q$  in uniform advertising equilibria.

In addition to changes for a fixed equilibrium type, varying  $q$  can also change the type of equilibria that exist in the game. Figure 2 showed that small changes in  $q$  can cause new equilibria to spring into existence or disappear. Thus the consumer surplus, profit, and advertiser's utility can jump discontinuously in  $q$ . We illustrate this phenomenon with two examples: the first example illustrates changes between a uniform advertising equilibrium B, a discriminatory equilibrium, and no equilibrium. The second shows a change from a uniform advertising equilibrium A to a discriminatory equilibrium.

For the first example, set  $s_{1A} = .5$ ,  $s_{2B} = .6$ , and  $s_{1B} = s_{2A} = .05$ , which implies  $\eta = .55$ . As illustrated in Figure 2, as we increase  $q$  from  $1/2$  to  $1$ , there is first a uniform advertising equilibrium B, then equilibria briefly cease to exist, then discriminatory equilibria exist, and finally no equilibria exist for large  $q$ . Each change of equilibrium existence results in discrete jumps in the consumer surplus, profit, and advertiser

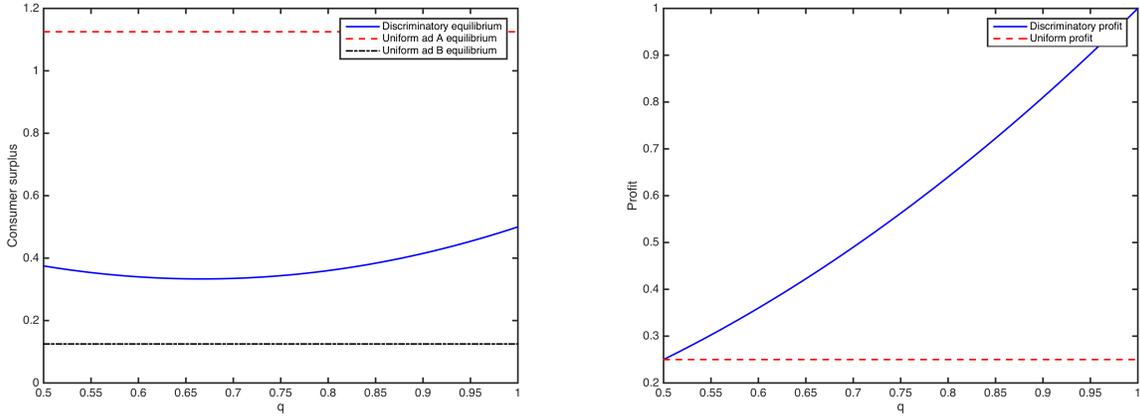


FIGURE 5: On the left is a plot of consumer surplus in all three equilibrium types. On the right is a plot of the seller’s profit in both discriminatory and uniform equilibria. The seller’s profit in a uniform equilibrium doesn’t depend on the ad shown in period 2.

utility. Figure 6 illustrates below.

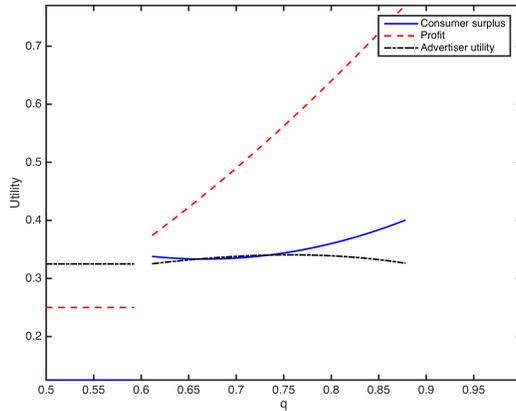


FIGURE 6: A plot of consumer surplus, profit, and advertiser utility in equilibrium when  $s_{1A} = .5$ ,  $s_{2B} = .6$ , and  $s_{1B} = s_{2A} = .05$ . Discontinuities correspond to changes of the equilibrium type that exists, or an absence of equilibria altogether.

For the second example, set  $s_{1A} = .6$ ,  $s_{2B} = .5$ , and  $s_{1B} = s_{2A} = .05$ , which implies  $\eta = .45$ . In this game, as  $q$  increases from  $1/2$  to  $1$ , the equilibrium type changes discretely from a uniform advertising equilibrium A to a discriminatory equilibrium, also illustrated in Figure 2. This change also causes discontinuities in the consumer surplus, profit, and advertiser utility. Note here that a tiny decrease in the level of privacy (i.e. increase in  $q$ ) can cause a precipitous drop in welfare. Figure 7 illustrates this effect.

## 5 MULTIPLICITY OF EQUILIBRIA AND OTHER RESULTS

In this section we build on the example of the previous section to provide some formal results about equilibrium multiplicity, welfare, etc. as a function of the promised level of privacy offered to the consumer. Proofs are deferred to the Appendix.

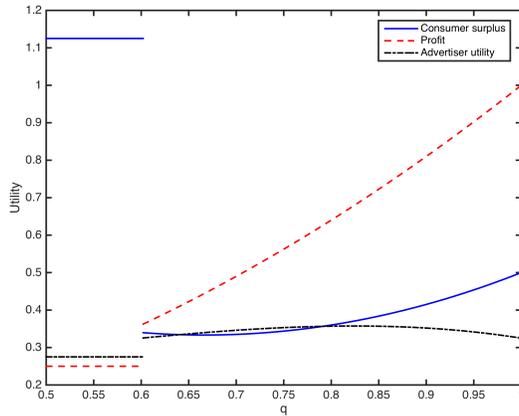


FIGURE 7: A plot of consumer surplus, profit, and advertiser utility in equilibrium when  $s_{1A} = .6$ ,  $s_{2B} = .5$ , and  $s_{1B} = s_{2A} = .05$ . Discontinuities correspond to changes of the equilibrium type that exists.

## 5.1 MULTIPLICITY

PROPOSITION 2. *A discriminatory equilibrium and a uniform advertising equilibrium A can coexist in the same game for the same level of  $q$ .*

PROPOSITION 3. *If a discriminatory equilibrium and a uniform advertising equilibrium A coexist at a given noise level  $q$ , then the buyer prefers the uniform equilibrium to the discriminatory equilibrium, regardless of her value  $v$ . On the other hand, the seller prefers the discriminatory equilibrium.*

Proposition 3 is intuitive, so we omit a formal proof. To see the first claim, observe that a buyer faces both a lower price in period 1 and sees a better ad in period 2, so she is always better off in the uniform advertising equilibrium, regardless of her value. To see the latter, observe that the discriminatory equilibrium allows the seller to sell to more consumers ( $v^*(q) < p_M$ ) at a higher price ( $p_1(q) > p_M$ ).

TAKEAWAY Small changes in  $q$  can make uniform advertising equilibria cease to exist, and can therefore have a discrete impact on welfare and revenue, as they cause the equilibrium to shift discontinuously from uniform to discriminatory. At these boundaries, the buyer may (strictly, discontinuously) prefer slightly less privacy while the seller may (strictly, discontinuously) prefer more privacy!

PROPOSITION 4. *A discriminatory equilibrium and a uniform advertising equilibrium B can coexist in the same game for the same level of  $q$ .*

PROPOSITION 5. *In any game where both a discriminatory equilibrium and a uniform advertising equilibrium B exist for the same value of  $q$ , average consumer welfare is always higher under the discriminatory equilibrium, but individual consumers may have different preferences for these two equilibria.*

TAKEAWAY In a game where a discriminatory equilibrium and uniform advertising equilibrium B coexist, buyers prefer the discriminatory equilibrium. Since the uniform advertising equilibrium B exists for an interval of “low”  $q$ , a buyer therefore may prefer less privacy!

## 5.2 WELFARE COMPARATIVE STATICS AND PREFERENCES OVER LEVELS OF PRIVACY

PROPOSITION 6. *In settings where a discriminatory equilibrium exists for a range of  $q$ , the equilibrium (consumer) welfare can be increasing in  $q$ .*

**TAKEAWAY** Note that the example in the previous section already demonstrates that this can be true on some interval of  $q$ , see for example Figure 5. The proof in the appendix demonstrates this formally. Note that this implies the buyer’s preferences over different levels of privacy may be complex, and her welfare is not necessarily the monotonically increasing in her privacy as is often assumed.

**OBSERVATION 2.** The seller prefers the largest  $q$  consistent with discriminatory equilibrium — he gets a higher price and more demand. If a discriminatory equilibrium ceases to exist at some interior  $q$ , the seller will at that point prefer that the market provide *more* privacy.

**TAKEAWAY** The seller prefers discriminatory equilibria over uniform advertising equilibria, and the least privacy that is consistent with a discriminatory equilibrium if given the choice over privacy levels.

**OBSERVATION 3.** For any level of noise  $q$ , the advertiser always prefers a discriminatory equilibrium to a uniform equilibrium if both exist.

To see this, note that in any uniform equilibrium, the advertiser’s net utility is the same as his ex-ante utility from showing that ad (since he ignores information from period 1). Since he chooses to act on the information he gets in a discriminatory equilibrium, his net utility must exceed the ex-ante utility of showing the same ad.

**TAKEAWAY** The advertiser will always prefer levels of  $q$  consistent with discriminatory equilibria. Among these, however, he may prefer strictly interior levels of  $q$ , for example, as demonstrated by Figure 4 previously.

## 6 CONCLUDING REMARKS

A rich body of work on differential privacy has developed in the computer science literature over the past decade. Broadly caricatured, this literature provides algorithms for accurate data analyses (of various sorts), subject to guaranteeing  $\epsilon$ -differential privacy to individual entries in the dataset. Typically,  $\epsilon$  can be set to be any value (implicitly to be chosen by the entity that controls access to the dataset), and mediates a trade-off between the strength of the privacy guarantee offered to the data owners and the accuracy of the analysis promised to the user of the data. This trade-off is typically viewed in simple terms: higher values of  $\epsilon$  (i.e. less privacy) are better for the analyst because they allow for higher accuracy, the reasoning goes, and worse for the privacy-desiring individual. Of course, if the dataset is already gathered, this reasoning is correct.

In this paper, we add some caveats to this folk wisdom in a simple stylized model. The dataset here consists of the individuals’ purchase decision of a good. The data analyst in our model is an advertiser. Individuals do not care about the privacy of their purchase decision for its own sake, but rather, care about privacy only insofar as it affects how ads are targeted at them. A crucial point is that in our model, at the time of choosing the privacy policy, these purchase decisions have not yet been made. As a result, the price of the good, the purchase decision of the individual, and the advertising policy of the advertiser all depend on the announced privacy policy. Evaluations of the privacy preferences of seller, advertiser, and buyers must take into account everyone’s equilibrium incentives. As we demonstrated, these can be the opposite of the simple static trade-offs we are used to, and reasoning about them correctly can be complex.

As the literature expands from (the already hard) questions of privately analyzing existing datasets, to thinking of setting privacy policies that influence future user behavior and the datasets that result from this behavior, the equilibrium approach we espouse here will be important. We hope this paper serves as a call-to-arms to reasoning about privacy policy in such settings, while also highlighting the difficulties.

## REFERENCES

- Avrim Blum, Jamie Morgenstern, Ankit Sharma, and Adam Smith. Privacy-preserving public information for sequential games. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 173–180. ACM, 2015.
- Giacomo Calzolari and Alessandro Pavan. On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, 130(1):168–204, 2006.

- Yiling Chen, Stephen Chong, Ian A Kash, Tal Moran, and Salil Vadhan. Truthful mechanisms for agents that value privacy. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 215–232. ACM, 2013.
- Vincent Conitzer, Curtis R Taylor, and Liad Wagman. Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science*, 31(2):277–292, 2012.
- Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 1(1):92–112, 2015.
- Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2014.
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC '06*, pages 265–284, 2006.
- Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226. ACM, 2012.
- Arpita Ghosh and Katrina Ligett. Privacy and coordination: computing on databases with endogenous participation. In *ACM Conference on Electronic Commerce*, pages 543–560, 2013.
- Arpita Ghosh and Aaron Roth. Selling privacy at auction. *Games and Economic Behavior*, 2013. ISSN 0899-8256. doi: <http://dx.doi.org/10.1016/j.geb.2013.06.013>. URL <http://www.sciencedirect.com/science/article/pii/S0899825613000961>. Preliminary Version appeared un the Proceedings of the Twelfth ACM Conference on Electronic Commerce (EC 2011).
- Shiva P Kasiviswanathan and Adam Smith. On the ‘semantics’ of differential privacy: A bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1):1, 2014.
- Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.
- Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 774–789. ACM, 2012.
- Kobbi Nissim, Salil Vadhan, and David Xiao. Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 411–422. ACM, 2014.
- Laura Ryan. Feds investigate ‘discrimination by algorithm’, September 2014. URL <http://www.nationaljournal.com/tech/feds-investigate-discrimination-by-algorithm-20140915>. [Online; Retrieved 8/5/2015].
- Curtis R Taylor. Consumer privacy and the market for customer information. *RAND Journal of Economics*, pages 631–650, 2004.
- David Xiao. Is privacy compatible with truthfulness? In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 67–86. ACM, 2013.

## A APPENDIX

We now show that the only equilibria of this game have a cutoff strategy in period 1.

PROPOSITION 7. *All equilibria have the property that in period 1, there exists a threshold value  $v^*$  such that the consumer buys if and only if  $v > v^*$ .*

PROOF. Assume not. Then there exists  $v, v'$  such that  $v' < v$ , and in equilibrium consumers with value  $v'$  buy in period 1, while consumers with value  $v$  do not. We consider the three possible equilibrium types, and show that each one leads to a contradiction.

If this is a discriminatory equilibrium, then consumers with  $\hat{b} = 1$  are shown ad  $A$ , and consumers with  $\hat{b} = 0$  are shown ad  $B$ . Then it must be the case that the consumer's utility satisfies:

$$u(v', \text{buy}) \geq u(v', \text{not buy}) \quad \text{and} \quad u(v, \text{not buy}) \geq u(v, \text{buy})$$

This implies:

$$\begin{aligned} & u(v', \text{buy}) + u(v, \text{not buy}) \geq u(v', \text{not buy}) + u(v, \text{buy}) \\ \iff & [v' - p + q\delta] + [(1 - q)\delta] \geq [(1 - q)\delta] + [v - p + q\delta] \\ \iff & v' \geq v \end{aligned}$$

This is a contradiction because  $v > v'$ .

If this is a uniform advertising equilibrium A, then all consumers are shown ad  $A$  and receive utility  $\delta$  in period 2. It must be the case that:

$$\begin{aligned} & u(v', \text{buy}) \geq u(v', \text{not buy}) \\ \iff & v' - p + \delta \geq \delta \\ \iff & v' \geq p \end{aligned}$$

Also,

$$\begin{aligned} & u(v, \text{not buy}) \geq u(v, \text{buy}) \\ \iff & \delta \geq v - p + v_A \\ \iff & p \geq v \end{aligned}$$

These two facts above imply that  $v' \geq v$ , which is a contradiction since we have assumed that  $v > v'$ .

If this is a uniform advertising equilibrium B, then all consumers are shown ad  $B$  and receive zero utility in period 2.

$$u(v', \text{buy}) \geq u(v', \text{not buy}) \iff v' \geq p \quad \text{and} \quad u(v, \text{not buy}) \geq u(v, \text{buy}); \iff p \geq v$$

Again, these facts imply  $v' \geq v$ , which is a contraction. □

PROPOSITION 2. *A discriminatory equilibrium and a uniform advertising equilibrium A can coexist in the same game for the same level of  $q$ .*

PROOF. Suppose the distribution  $F$  of buyers' values is uniform on  $[0, 1]$ , and first fix  $q = 1$  (i.e. no noise).

At these values, from equation (6), we know that the price charged in a discriminatory equilibrium, if it exists, is

$$p^* = \frac{1 + \delta}{2},$$

while the cutoff of types who buy is

$$v^* = \frac{1 - \delta}{2}$$

Suppose the distribution of period 2 types is such that  $g(v)$  is the step function that is 0 below  $v^*$  and 1 above  $v^*$ . In this case, the discriminatory equilibrium exists at  $q = 1$  because we have not added any noise, and consumers who do not buy are certainly type  $t_2$  and should be shown ad  $B$ , while consumers who purchase are almost certainly type  $t_1$  and should be shown ad 1.

The myopic monopoly price under  $F$  is  $p_M = 1/2$ . Note that if customers were then to purchase myopically, i.e. whenever their value exceeds the price, then the posterior probability of type  $t_1$  given that the customer did not purchase is exactly  $\delta$ . For  $\eta < \delta$ , we then have that the seller charging the myopic monopoly price, the customers purchasing myopically, and the advertiser showing everyone ad  $A$  constitutes a uniform advertising equilibrium A.

In the same game, we can show that both types of equilibria co-exist for a continuous range of  $q < 1$  as well. Recall from Proposition 2 that it is sufficient for a uniform advertising equilibrium A to exist if  $r(0, p_M, q) > \eta$ . Given myopic behavior on the part of the consumers, and the seller setting the monopoly price in period 1, the advertiser's posterior having seen noisy bit  $\hat{b} = 0$  is as follows:

$$\begin{aligned} r(0, p_M, q) &= \frac{q \int_0^{1/2} \mathbb{1}_{v > \frac{1-\delta}{2}} dv + (1-q) \int_{1/2}^1 dv}{qF(1/2) + (1-q)(1-F(1/2))} \\ &= \frac{q\frac{\delta}{2} + (1-q)\frac{1}{2}}{q\frac{1}{2} + (1-q)\frac{1}{2}} \\ &= \frac{q\frac{\delta}{2} + (1-q)\frac{1}{2}}{\frac{1}{2}} \\ &= \delta q + (1-q) \end{aligned}$$

A uniform advertising equilibrium A will exist whenever  $r(0, p_M, q) > \eta$ . That is, for any  $q \in [1/2, 1]$  satisfying

$$\delta q + (1-q) > \eta \iff q > \frac{\eta - 1}{\delta - 1}$$

Since we have already assumed that  $\eta < \delta$ , then  $1 > \frac{\eta - 1}{\delta - 1}$ , so there is a non-empty range of  $q$  satisfying this condition.

We now verify that there is a continuous range of  $q$  for which there also exists a discriminatory equilibrium of this game. By equation (6), the discriminatory equilibrium price  $p_1(q)$  must satisfy  $p_1(q) - I(p_1(q) + (1 - 2q)\delta) = 0$ , where  $I(v) = \frac{1-F(v)}{f(v)}$ . Plugging in the distribution  $F$  as  $U[0, 1]$ ,

$$\begin{aligned} p_1(q) - (1 - p_1(q) - (1 - 2q)\delta) &= 0 \\ \iff p_1(q) &= \frac{1}{2} - (1 - 2q)\frac{\delta}{2} \end{aligned}$$

The Period 1 cutoff value is then

$$v^*(q) = p_1(q) + (1 - 2q)\delta = \frac{1}{2} + (1 - 2q)\frac{\delta}{2}$$

There exists a discriminatory equilibrium at  $q \in [1/2, 1]$  if both  $r(1, v^*(q), q) > \eta$  and  $r(0, v^*(q), q) < \eta$ . By construction,  $r(1, v^*(q), q) > \eta$  is satisfied for all  $q \in [1/2, 1]$ . We now compute the advertiser's posterior

$r(0, v^*(q), q)$ .

$$\begin{aligned}
r(0, v^*(q), q) &= \frac{q \int_0^{v^*(q)} \mathbb{1}_{v > \frac{1-\delta}{2}} dv + (1-q) \int_{v^*(q)}^1 dv}{qF(v^*(q)) + (1-q)(1-F(v^*(q)))} \\
&= \frac{q(v^*(q) - \frac{1-\delta}{2}) + (1-q)(1-v^*(q))}{qv^*(q) + (1-q)(1-v^*(q))} \\
&= 1 - \frac{q^{\frac{1-\delta}{2}}}{qv^*(q) + (1-q)(1-v^*(q))} \\
&= 1 - \frac{\frac{1}{2}q(1-\delta)}{(1-q) - (1-2q)[\frac{1}{2} + (1-2q)\frac{\delta}{2}]} \\
&= 1 - \frac{\frac{1}{2}q(1-\delta)}{(1-q) - \frac{1}{2}(1-2q) - \frac{1}{2}(1-2q)^2\delta} \\
&= 1 - \frac{\frac{1}{2}q(1-\delta)}{\frac{1}{2}[1 - (1-2q)^2\delta]} \\
&= 1 - \frac{q(1-\delta)}{1 - (1-2q)^2\delta}
\end{aligned}$$

For there to be a discriminatory equilibrium at  $q$ , it must be the case that  $r(0, v^*(q), q) = 1 - \frac{q(1-\delta)}{1-(1-2q)^2\delta} < \eta$ . Although there is not a nice closed form description of the  $q \in [1/2, 1]$  satisfying this condition, we note that this expression is differentiable (and thus continuous), everywhere except when

$$1 = (1-2q)^2\delta \iff q = \frac{1 \pm \sqrt{\frac{1}{\delta}}}{2}$$

If we restrict  $\delta < 1$ , then this  $q$  will fall outside of our range of interest, and this expression  $r(0, v^*(q), q)$  is differentiable on  $[1/2, 1]$ .

We now take the derivative of  $r(0, v^*(q), q)$  with respect to  $q$ , and see that it is negative, so the function is monotone decreasing.

$$\frac{\partial r(0, v^*(q), q)}{\partial q} = - \frac{(1-\delta)[1 - (1-2q)^2\delta] - [-2\delta(1-2q)(-2)q(1-\delta)]}{[1 - (1-2q)^2\delta]^2}$$

We are only interested in the sign of this expression, and the denominator is clearly positive, so we will proceed only with the numerator (without the negative sign in front of it).

$$\begin{aligned}
(1-\delta)[1 - (1-2q)^2\delta - 4\delta(1-2q)q] &= (1-\delta)[1 - \delta(1-2q)(1+2q)] \\
&= (1-\delta)[1 - \delta(1-4q^2)] \\
&= (1-\delta)^2 + (1-\delta)4\delta q^2
\end{aligned}$$

Since  $0 < \delta < 1$ , this expression is positive, so plugging it back into the original expression (with a negative in front) means that  $\frac{\partial r(0, v^*(q), q)}{\partial q}$  is negative, so  $r(0, v^*(q), q)$  is a monotone decreasing function. This means that there exists a continuous range of  $q$  for which the condition  $r(0, v^*(q), q) < \eta$  is satisfied. Since it is satisfied at  $q = 1$  and  $r(0, v^*(q), q)$  is decreasing in  $q$ , then the condition must be satisfied for all  $q \in [1/2, 1]$ .

To summarize, when  $0 < \eta < \delta < 1$ , there exists a non-trivial range of  $q$ , namely  $q \in [\frac{\eta-1}{\delta-1}, 1]$ , for which there exists both a discriminatory equilibrium and uniform advertising equilibrium A in this game.  $\square$

**PROPOSITION 4.** *A discriminatory equilibrium and a uniform advertising equilibrium B can coexist in the same game for the same level of  $q$ .*

PROOF. Consider the same example as in Proposition 2, where  $v \sim U[0, 1]$  and  $g(v)$  is 1 if  $v > \frac{1-\delta}{2}$  and 0 otherwise.

There will exist a uniform advertising equilibrium B in this game at  $q$  if and only if  $r(1, p_M, q) < \eta$ , where  $p_M = \frac{1}{2}$ . The advertiser's posterior in this setting after seeing  $\hat{b} = 1$  is

$$\begin{aligned} r(1, p_M, q) &= \frac{(1-q) \int_0^{1/2} \mathbb{1}_{v > \frac{1-\delta}{2}} dv + q \int_{1/2}^1 dv}{qF(1/2) + (1-q)(1-F(1/2))} \\ &= \frac{(1-q)\frac{\delta}{2} + q\frac{1}{2}}{q\frac{1}{2} + (1-q)\frac{1}{2}} \\ &= \frac{(1-q)\frac{\delta}{2} + q\frac{1}{2}}{\frac{1}{2}} \\ &= (1-q)\delta + q \end{aligned}$$

A uniform advertising equilibrium B exists for all  $q$  satisfying

$$(1-q)\delta + q < \eta \iff q < \frac{\eta - \delta}{1 - \delta}$$

Restricting  $\eta > \frac{1+\delta}{2}$  ensures that there is a non-empty interval of  $q \in [1/2, 1]$  satisfying this condition.

A discriminatory equilibrium exists for any  $q$  such that  $r(1, v^*(q), q) > \eta$  and  $r(0, v^*(q), q) < \eta$ . Recall from the proof of Proposition 2 that the equilibrium price is  $p_1(q) = \frac{1}{2} - (1-2q)\frac{\delta}{2}$  and the equilibrium cutoff value is  $v^*(q) = \frac{1}{2} + (1-2q)\frac{\delta}{2}$ . Also recall that  $r(0, v^*(q), q) < \eta$  is equivalent to the condition  $1 - \frac{q(1-\delta)}{1-(1-2q)^2\delta} < \eta$ . We now expand  $r(1, v^*(q), q)$ .

$$\begin{aligned} r(1, v^*(q), q) &= \frac{(1-q) \int_0^{v^*(q)} \mathbb{1}_{v > \frac{1-\delta}{2}} dv + q \int_{v^*(q)}^1 dv}{qF(v^*(q)) + (1-q)(1-F(v^*(q)))} \\ &= \frac{(1-q)(v^*(q) - \frac{1-\delta}{2}) + q(1-v^*(q))}{(1-q)v^*(q) + q(1-v^*(q))} \\ &= 1 - \frac{(1-q)\frac{1-\delta}{2}}{(1-q)v^*(q) + q(1-v^*(q))} \\ &= 1 - \frac{\frac{1}{2}(1-q)(1-\delta)}{q + (1-2q)[\frac{1}{2} + (1-2q)\frac{\delta}{2}]} \\ &= 1 - \frac{\frac{1}{2}(1-q)(1-\delta)}{q + \frac{1}{2}(1-2q) + \frac{1}{2}(1-2q)^2\delta} \\ &= 1 - \frac{\frac{1}{2}(1-q)(1-\delta)}{\frac{1}{2}[1 + (1-2q)^2\delta]} \\ &= 1 - \frac{(1-q)(1-\delta)}{1 + (1-2q)^2\delta} \end{aligned}$$

To finish specifying the game parameters, set  $\delta = .9$  and  $\eta = .984$ . With these parameters, both discriminatory equilibria and uniform advertising equilibria B exist for a continuous range of  $q$ . We will verify this for one  $q$  value in this range, namely  $q = .8$ . First, observe that a uniform advertising equilibrium B exists because  $r(1, p_M, .8) = .98 < .984 = \eta$ . Next, observe that  $r(0, v^*(.8), .8) \approx .88166 < .984 = \eta$  and  $r(1, v^*(.8), .8) \approx .98484 > .984 = \eta$ . Thus a discriminatory equilibrium exists as well.  $\square$

PROPOSITION 5. *In any game where both a discriminatory equilibrium and a uniform advertising equilibrium B exist for the same value of  $q$ , average consumer welfare is always higher under the discriminatory equilibrium, but individual consumers may have different preferences for these two equilibria.*

PROOF. In any uniform advertising equilibrium B, the myopic monopoly price is charged, and by assumption, all consumers are shown the inferior ad. Therefore the ex-ante consumer welfare is  $\int_{p_M}^1 (v - p_M)f(v)dv$ .

At any  $q$  where a discriminatory equilibrium exists, the ex-ante consumer welfare is:

$$\begin{aligned} & \int_{v^*(q)}^1 (v - p_1(q))f(v)dv + (q(1 - F(v^*(q))) + (1 - q)F(v^*(q)))\delta \\ &= \int_{v^*(q)}^1 (v - p_1(q) - (1 - 2q)\delta)f(v)dv + (1 - q)\delta. \end{aligned}$$

Note that the latter term is positive. Further, recall from Proposition 4,  $v^*(q) = p_1(q) + (1 - 2q)\delta$ , and that  $v^*(q) \leq p_M$ . Therefore  $\int_{v^*(q)}^1 (v - p_1(q) - (1 - 2q)\delta) \geq \int_{p_M}^1 (v - p_M)f(v)dv$ .

We will now see that if both a discriminatory equilibrium and a uniform advertising equilibrium B exist, then consumers with values  $v \in [0, p_M]$  will prefer the discriminatory equilibrium, while the preferences of consumers with values  $v \in [p_M, 1]$  depend on the game parameters  $\delta$  and  $\eta$ , as well as the noise level  $q$ .

For consumers with values  $v \in [0, v^*(q)]$ , they prefer the discriminatory equilibrium because they still don't purchase the good in period 1, but they have a chance at the better ad in period 2. For consumers with values  $v \in [v^*(q), p_M]$ , under the discriminatory equilibrium, they receive utility  $(v - p_1(q)) + \delta q$  for buying the good at price  $p_1(q)$  and being shown the better ad with probability  $q$ . Under the uniform advertising equilibrium B, they receive utility 0 for not buying in period 1, and then being shown ad B with probability 1. For all consumers with value in this range, the discriminatory equilibrium is preferred because

$$v - p_1(q) + q\delta > (1 - q)\delta > 0,$$

where the first inequality is because the consumer maximized her utility by buying in the discriminatory equilibrium, and the second inequality is because  $\delta > 0$  and  $q < 1$ .

For consumers with values  $v \in [p_M, 1]$ , under the discriminatory equilibrium, they receive utility  $v - p_1(q) + \delta q$  for buying the good at price  $p_1(q)$  and being shown the better ad with probability  $q$ . Under the uniform advertising equilibrium B, they receive utility  $v - p_M$  for buying at price  $p_M$  in period 1, and then being shown ad B with probability 1. These consumers will prefer the discriminatory equilibrium if and only if

$$v - p_1(q) + \delta q > v - p_M \iff \delta q > p_1(q) - p_M$$

Intuitively, the term  $\delta q$  captures the consumer's bonus in period 2 from the possibility of being shown the better ad. The term  $p_1(q) - p_M$  is the additional amount the consumer must pay in period 1 to get the good. Then the consumer will prefer discrimination whenever the increase in utility from seeing the better ad outweighs the increased price she must pay in period 1.

In the example used in Propositions 2 and 4, the discriminatory equilibrium is preferred:

$$\begin{aligned} \delta q - (p_1(q) - p_M) &= \delta q - \left( \frac{1}{2} + (2q - 1)\frac{\delta}{2} - \frac{1}{2} \right) \\ &= \delta q - (2q - 1)\frac{\delta}{2} \\ &= \delta \left( q - \frac{1}{2} \right) > 0 \end{aligned}$$

Since  $\delta > 0$  and  $q > 1/2$ , this term is always positive, so discrimination is always preferred in this game. However, this is not necessarily the case in all games. We'll next see an example of a game where the players with values  $v \in [p_M, 1]$  prefer the uniform advertising equilibrium B.

Consider now the value distribution that is inverse exponential, with parameter  $\lambda$ , truncated to lie in the domain  $[0, 1]$ . That is,

$$f(v) = \frac{\lambda e^{\lambda v}}{e^\lambda - 1} \text{ and } F(v) = \frac{e^{\lambda v} - 1}{e^\lambda - 1}$$

The inverse hazard rate of this function is

$$I(v) = \frac{1 - F(v)}{f(v)} = \frac{1 - \frac{e^{\lambda v} - 1}{e^\lambda - 1}}{\frac{\lambda e^{\lambda v}}{e^\lambda - 1}} = \frac{e^\lambda - e^{\lambda v}}{\lambda e^{\lambda v}} = \frac{1}{\lambda}(e^{\lambda(1-v)} - 1)$$

Then the uniform advertising equilibrium  $B$  price  $p_M$  solves

$$p_M - \frac{1}{\lambda}(e^{\lambda(1-p_M)} - 1) = 0$$

and the discriminatory equilibrium price  $p_1(q)$  solves

$$p_1(q) - \frac{1}{\lambda}(e^{\lambda(1-v^*(q))} - 1) = 0,$$

where  $v^*(q) = p_1(q) + (1 - 2q)\delta$ .

Choosing  $\delta = .5$ ,  $q = 1$ , and  $\lambda = 1$ , we see that  $p_1(q) = .56324$  and  $p_M = .557146$ . Thus

$$\delta q = .055 < .06094 = p_1(q) - p_M,$$

so these consumers prefer the uniform advertising equilibrium B. □

**PROPOSITION 6.** *In settings where a discriminatory equilibrium exists for a range of  $q$ , the equilibrium (consumer) welfare can be increasing in  $q$ .*

**PROOF.** To simplify calculations, suppose again that valuations are distributed on the entire positive real line. Next consider ex-ante equilibrium welfare of the buyer as a function of  $q$  in the discriminatory equilibrium:

$$\int_{v^*(q)}^{\infty} (v - p_1(q))f(v)dv + q\delta(1 - F(v^*(q))) + (1 - q)\delta F(v^*(q)).$$

Differentiating with respect to  $q$  and collecting terms we have:

$$-f(v^*(q))v^{*'}(q)(v^*(q) - p_1(q) - (1 - 2q)\delta) + \delta(1 - 2F(v^*(q))) - p_1'(q)(1 - F(v^*(q))).$$

Recall that by the definitions of  $v^*(q)$  we have

$$v^*(q) - p_1(q) - (1 - 2q)\delta = 0$$

Therefore, the derivative of ex-ante welfare in a discriminatory equilibrium w.r.t.  $q$  equals

$$\delta(1 - 2F(v^*(q))) - p_1'(q)(1 - F(v^*(q))).$$

Recall from the proof of Lemma 3 we have

$$\begin{aligned} p_1'(q) - I'(p_1 + (1 - 2q)\delta)(p_1'(q) - 2\delta) &= 0, \\ \implies p_1'(q) &= 2q \left( \frac{I'(p_1 + (1 - 2q)\delta)}{I'(p_1 + (1 - 2q)\delta) - 1} \right). \end{aligned}$$

Substituting in, the derivative of welfare w.r.t.  $q$  equals

$$\delta(1 - 2F(v^*(q))) - 2q \left( \frac{I'(p_1 + (1 - 2q)\delta)}{I'(p_1 + (1 - 2q)\delta) - 1} \right) (1 - F(v^*(q))).$$

Note that for  $I'$  small (e.g. close to 0 like an exponential distribution that has constant hazard rate), and  $F(v^*) < \frac{1}{2}$ , this is positive.

To specify one such game, the example given in Section 4 has consumer welfare that is increasing in  $q$ , for  $q$  greater than roughly 0.67. (See Figure 5.) Selecting, e.g.,  $\eta = \frac{1}{2}$  in that game will ensure that discriminatory equilibria exist for all  $q$  in that range. □