Why We Can't Be Bothered to Read Privacy Policies
Models of Privacy Economics as a Lemons Market

Tony Vila, Rachel Greenstadt, David Molnar
Harvard University
{*avila@fas,greenie@eecs,dmolnar@hcs*}*.harvard.edu*

May 25, 2003

Abstract

Consumers want to interact with web sites, but they also want to keep control of their private information. Asymmetric information about whether web sites will sell private information or not leads to a lemons market for privacy. We discuss privacy policies as signals in a lemons market and ways in which current realizations of privacy policies may fail to be effective signals. As a result of these shortcomings, we consider a "lemons market with testing," where consumers have a cost of determining whether a site meets their privacy requirement. Our model explains empirical data concerning privacy policies and privacy seals. We end by discussing cyclic instability in the number of web sites that sell consumer information.

1 Introduction

People generally equate e-commerce with violations of their personal property. They are concerned that buying online will result in unwanted spam email, their personal information being sold to marketing organizations and possibly even their identity or credit card information stolen. Recent survey data indicated that 92% of consumers are concerned about the misuse of their personal information online,[8] and privacy concerns are the number one reason why individuals choose to stay off the Internet[9]. Others simply decide that loss of privacy is an inevitable consequence of doing business these days. If we believe that people value privacy, why is there not an efficient market for it? This is the question that this paper seeks to address.

So, what is the state of privacy on the internet, and how has it evolved? In order to answer this question, we need a definition of what it means to protect privacy. For the purposes of this paper, we will define this as following the fair information practices principles as delineated by the FTC[6]. These principles are:

- Notice - Web sites provide consumers with clear and conspicuous notice of their information practices. This would include what information they collect, how they collect it, whether they provide the other properties, whether they disclose this information to other entities, and whether other entities are collecting information through the site.
- Choice - Web sites offer consumers choices as to how their personal information is used beyond the use for which the information was provided.
- Access - Web sites offer consumers reasonable access to the information a site has collected about them, and an opportunity to delete it, or correct inaccuracies.
- Security - Web sites take reasonable steps to protect the security of the information collected from consumers.

In 2000, the FTC found that that only 20% of randomly sampled web sites partially implemented all four fair information practices. The percentage was higher (42%) for the most popular web sites. There had been hope that privacy on the internet could be improved through seal programs[15] or P3P[12], but by 2000 these programs had not seen wide adoption[6, 7].

Recently, however, things are looking up for privacy protection. A similar study to the FTC report was done in 2002 by the Progress and Freedom Foundation[10]. The survey found that web sites are collecting less information, notice is more prevalent, prominent and complete. Choice also increased, with the percentage of the most popular sites offering consumers a choice about sharing information with third parties jumped from 77% to 93%. With the introduction of P3P enabled browsers, P3P adoption was growing (5% in the random domain and 25% in the most popular domain). On the other hand, sites displaying seals were still a very small proportion of the sites (12%).

We attempt to explain these trends and understand where privacy protection and violations may go in future. In section 3, we present a simplified model of privacy as a lemons market with signaling. In section 4, we com-

plicate the model by adding a cost to the consumer to search for a signal. We conclude with discussion of the model and future directions.

2 Related Work

Varian defines "privacy rights" as "the right not to be annoyed" and focuses on assignment of property rights in privacy as a means to establish a market[16]. Our work shows that this market may not be efficient in the presence of asymmetric information.

Acquisti has a general discussion of economic incentives for and against privacy-enhancing technologies, such as anonymizing web proxies; his paper also describes work on how information sharing between vendors in the presence of a strategic consumer leads naturally to a privacy-protecting regime[2]. One example of such sharing is the work on privacy policies by Calzolari and Pavan, in which buyers are allowed to choose between a contract whose terms are public(shared with all vendors) or private(shared only with a single vendor); they show that in this case buyers choose the appropriate contract to maximize their privacy[5]. Our work, in contrast, focuses on the information available to the consumer about the vendor.

3 Privacy as a Lemons Market

The "lemons market" was introduced by Akerlof as an example of asymmetric information [1]. In the original example, buyers and sellers trade in a market with two types of cars: "good" cars, worth a high amount, and "lemons" worth relatively little. Buyers are unable to distinguish a good car from a lemon before buying, and therefore will offer less than the full price of a "good" car to offset the chance of buying a lemon. As a result, no owner of a good car elects to sell; the market is flooded with lemons.

To begin with, we can think of a consumer choosing among web sites that may respect her privacy ("Respecting" sites) or may not ("Defecting") with no way to determine beforehand which is which. Then privacy in web sites looks like the lemons market. As a result, we would expect all web sites to not respect privacy.

In the context of web sites, we can make this more formal as follows. Suppose web sites fall into two categories: Respecting(R) sites that do not sell private information and Defecting(D) sites that do sell such information. A customer may choose to buy or not buy with a site. If the customer buys from a Respecting site, it gains B. If it buys from a Defecting site, it

obtains $B - V$, where V is the cost to the customer of a privacy violation. The resulting payoff matrix is

$$\begin{pmatrix} & Respects & Defects \\ Buys & B & B-V \\ Doesn't & 0 & 0 \end{pmatrix}$$

4 Privacy Signals

The lemons market for privacy motivates the introduction of privacy *signals*. For instance, a web site may adopt a strict privacy policy to demonstrate its commitment to keeping customer information private. Alternatively, web sites may acquire reputations concerning their handling of customer information. In general, a signal is a means by which privacy-respecting sites can differentiate themselves from their non-respecting competitors. We will describe formally how these signals work, discuss some candidate signals in the web site privacy market, and then show how shortcomings in these candidate signals motivate our move away from signals to a market with testing.

Signalling is well studied in the context of a lemons market; if a signal is low cost for "good" players and high cost for "lemon" players, then consumers can reliably use the signal to separate good players from lemons[14]. Assuming that such a signal exists, we can show a separation in the web site privacy market.

Web sites now fall fall into four classes: Respecting who do not signal, Respecting who do signal, Defecting who do not signal, and Defecting sites who do signal. Consumers now fall into three classes: Buy from a site, Don't buy from a site, Only buy from a site that presents the signal that they follow fair information practices.
B = the benefit the consumer gets from a transaction
V = the cost for the consumer of having their privacy violated
P = the benefit the firm gets from the transaction
S_R, S_D = the cost to the respectful or defecting firm to send the signal guaranteeing privacy
I = the benefit the firm gets from selling the consumer's personal information.

The payoff matrix is then

$$\begin{pmatrix} & Respects-NS & Respects-S & Defects-NS & Defects-S \\ Buys & B,P & B,P-S_R & B-V,P+I & B-V,P+I-S_D \\ Doesn't & 0,0 & 0,-S_R & 0,0 & 0,-S_D \\ BuysSignal & 0,0 & B,P-S_R & 0,0 & B-V,P+I-S_D \end{pmatrix}$$

Now the rational choice of each player depends on the relationship between S_R, P, and $S_D - I$. A Defecting site will send the signal only if its cost for doing so is less than the benefit it gains from the transaction and from selling the consumer's private information. Formally, if $P < S_D - I$, a Defecting site will not send the signal. A Respecting site, in turn, will send the signal if its cost of doing so is less than the benefit it gains from the transaction, or $P > S_R$. Therefore, if $S_R < P < S_D - I$, then all and only the respecting web sites will send the signal, and the rational consumer will only make transactions with these signalling web sites. This is the desired separation of the market. The separation requires that the signal be high cost for Defecting sites and low cost for Respecting sites, i.e. $S_D - S_R$ must be at least I.

Do signals with high cost for Defecting sites and low cost for Respecting sites exist in the real world? Privacy policies are the most obvious candidates for such a signal. The recent P3P standard provides a way for sites to mechanically codify privacy policies [12]. User interfaces such as the AT&T P3P Privacy Bird give customers easy ways to tell whether a site's P3P policy matches their individual preferences [3]. Implementing a P3P policy costs a significant amount of time and effort, demonstrating a commitment on the part of the web site to privacy.

At the same time, relying on privacy policies alone is problematic. What prevents a site from publicizing a strict policy but then reneging on the policy and selling information anyway? Put another way, where does the cost differential between Respecting and Defecting sites come from for privacy policies? One answer may lie in the legal and public relations exposure to a Defecting site that collects information despite the presence of a privacy policy. For example, Real Networks suffered public criticism when its software was found to gather and report information[11]. Unfortunately, this sort of discovery happens rarely, and may not provide enough of an incentive against violating the policy.

Reputations offer a potential alternative. There is empirical evidence that reputations can work in electronic commerce to differentiate sellers.

Resnick, Zeckhauser, Swanson, and Lockwood show that reputation on eBay does lead consumers to pay an average of 7.6% higher prices to sellers with high reputation over others with low reputations[13]. Yamagichi and Matsuda show experimentally that reputation can alleviate a lemons market [17]. Unfortunately, unlike eBay, no centralized, often-updated repository of web site privacy reputations exists. We cannot depend on a customer having knowledge of the web site's previous actions, or even other customer's reports of the web site's actions.

These issues cause us to take a different approach. Instead of focusing on the difference in cost to the web sites of sending a signal, we suggest focusing on the cost to the consumer of *testing* whether the web site is privacy-respecting. The resulting privacy market with testing is the focus of the next section.

5 Testing in the Lemons Market

Even if a web site has a privacy policy, the policy may not function as a signal because it is not read. People value their time and effort, and are unlikely to spend time finding or reading information that is boring or confusing to them, when there are many other more entertaining options.

In general, the opportunity cost of collecting signals is higher online than in the physical world, and as a consumer participates in many of these transactions, they are even less likely to want to repeat such monotonous and frustrating actions. Relative to the small scale of most interned transactions, such as buying a book or simply visiting a web site, comprehending a privacy policy is much less acceptable than when buying a house or getting significant medical treatment. Simply put: in the time you could read and check Amazon.com's privacy policy, you could drive to Barnes and Noble and buy the book.

So to add an aspect to the traditional "signal" model of the Lemons Market, we include the factor T, the cost for a consumer to check if a firm is sending the aforementioned signal. This is analogous to hiring your own mechanic to check if a car is a lemon or not before buying it. For every signal a firm would create to represent its attitude towards your personal information, there are costs associated with it such as:

- Read the rather long privacy policy
- Check consumer responses and e-trust web sites
- Install the P3P bird program

Of course these are extremely heterogeneous actions and costs that we are looking at. Different firms would require different amounts of effort in checking on these signals, and each type of checking would take a different effort. Reading a click-through contract is not the same thing as researching in Consumer Reports. And different sectors will definitely have a different T associated for each. But when trying to find the value of a firm in a specific sector, the uncertainty is only how much you'll have to research them, or how convoluted their particular policy is to read or enforce. The consumer does not know the real value of this effort cost beforehand. Therefore, it is often unknown what exact effort will be spent before you have already found the signal, ie., how much the firm conforms to fair information practices. For the purposes of simplicity, we assume that consumers a priori assume the cost of T, based on their experience and conventional wisdom; they perceive T as the average of all firms, T_a.

This becomes especially tricky when different firms manipulate their particular testing cost based on whether they want consumers to check for signals. Unfortunately this information is not of much help to the consumer, since if they knew the testing cost beforehand, that would a way of effortlessly gaining information about the signal and relative intentions of the company. So what happens when this cost T enters the traditional signaling payoff matrix?
Allow for the variables representing:
B = the benefit the consumer gets from a transaction
T = the cost to test for the consumer
V = the cost for the consumer of having their privacy violated
P = the benefit the firm gets from the transaction
S = the cost to the firm to send the signal guaranteeing privacy
I = the benefit the firm gets from selling the consumer's personal information.

$$\begin{pmatrix} & Respects & Defects \\ Tests & B-T, P-S & -T, 0 \\ Doesn't & B, P-S & B-V, P+I \end{pmatrix}$$

Let us label p as the portion of firms that respect privacy and send the signal, with $(1-p)$ as the portion of firms that sell information and dont send the signal appropriately. We can then find the relative utility of deciding to test a site instead of buying without being aware.

$$
\begin{aligned}
(Tests) &= p(B-T)+(1-p)(-T) \\
&= pB-T \\
U(Doesn't) &= p(B)+(1-p)(B-V) \\
&= B-V+pV \\
U(Tests)-U(Doesn't) &= pB-T-(B-V+pV) \\
U(Tests)-U(Doesn't) &= pB-T-B+V-pV \\
U(Tests)-U(Doesn't) &= -(1-p)(B-V)-T
\end{aligned}
$$

When p approaches 1, and all firms respect privacy, then the consumer has great incentive to not test web sites, since the only difference is that he is paying T. As long as the system is making firms behave, then free riders emerge who dont want to take the cost of testing. When p approaches 0, and all firms sell personal information, than the consumer has great incentive to test web sites. Since we assume the cost of someone's privacy being violated to be higher than the benefit of the transaction and the cost of testing (which is a significant assumption of privacy economics, but a necessary one), then $-(B-V)-T$ is quite positive. We can do similar calculations with labeling q as the portion of the consumers who test, and $(1-q)$ as the portion of consumers who dont bother to test privacy policies. Checking the relative benefits reveals such:

$$
\begin{aligned}
U(Respects) &= q(P-S)+(1-q)(P-S) \\
&= P-S \\
U(Defects) &= q(0)+(1-q)(P+I) \\
&= P+I-qP-qI \\
U(Respects)-U(Defects) &= P-S-(P+I-qP-qI) \\
U(Respects)-U(Defects) &= -S-I+qP+qI \\
U(Respects)-U(Defects) &= -(S+I)+qP+qI
\end{aligned}
$$

When q approaches 1, and all consumers test web sites, then the firm has significant incentive to create a respectful privacy market, by P-S. When p approaches 0, and no consumers test web sites, then the firm is very likely to not respect privacy or send the signal, and stands to gain S+I.

This suggests a dramatic instability in the privacy market:

1. When all firms respect privacy (as many consumers would likely believe from real world experience), no consumers will test for signals.

2. When no consumers test for signals, all firms will sell personal information.

3. When most firms sell personal information, all consumers will start to test a web site's privacy policy (which is about where the market is now).

4. When all consumers test for signals, all firms will establish commitments to respect privacy (which is where the data suggests the market is headed).

5. Return to step 1 ad infinitum.

These broad trends will continue to revolve around various values of p and q as time moves on. Any attempt to reach a perfect market where all firms respect personal information, and consumers knowingly pay the premium for that, will dissolve, largely because of free rider problems, and firms leaping to take advantage of pauses in security. What will eventually emerge (but not directly) are stable middle ground values for p and q: p* and q*.

$$\begin{aligned} U(Tests) - U(Doesn't) &= 0 \\ -(1-p)(B-V) - T &= 0 \\ p^* &= \frac{B-V-T}{B-V} \\ U(Respects) - U(Defects) &= 0 \\ -(S+I) + q^*P + q^*I &= 0 \\ q^* &= \frac{S+I}{P+I} \end{aligned}$$

If testing and non-testing consumers reach the right balance, then firms get equal benefits from deciding to respect or not to respect privacy. Similarly, if a certain mix of firms signal and don't signal, then consumer's loss to testing equals how much they lose from having their information violated on average. This is the only Nash equilibrium in the payoff matrix, although it is not the most stable situation, since individual consumers or firms will

only be indirectly and slowly affected by any change they make strategy, and not immediately drawn back to the equilibrium.

This conclusion definitely reflects the available data on the privacy market.

Constant fluctuations in the benefits of reading privacy policies or not, or towards respecting privacy or not, would describe the internet market much better. Future goals of every consumer being savvy and protecting their personal information would appear to be at least somewhat impractical. If not already reached, eventually some people will act responsibly, while others will not, in perpetuity.

6 Conclusions and Future Directions

Our models explain previous trends in the web site privacy market. In the Introduction, we saw that despite the fact that more web sites follow the Fair Information Practices today than in 2000, the number of web sites with privacy seals has not increased proportionally. Through our analysis of privacy seals as a signal, we showed that this non-adoption can be explained because a privacy seal does not have a lower cost for privacy-respecting sites than for privacy-defecting sites.

Our models also give insight into the structure and future of privacy for web sites. Recall that our model for testing yielded a single equilibrium point, namely

$$(p^*, q^*) \quad = \quad (\frac{B-V-T}{B-V} \ , \frac{S+I}{P+I} \ )$$

We now show that the market does not move directly to that equilibrium point. The continuing progression of privacy policies and consumer protection software, and fluctuating statistics regarding privacy protection, and the conclusions of our model both agree that instead the market oscillates around the equilibrium point. We suggest three reasons why.

1) Time to reach equilibrium is large. The simplest model is where every player is shortsighted, has perfect information, and can costlessly change their strategy each turn. We found a cycling through the four possible absolutes, of consumers testing or not testing, and companies signaling and

not signaling. In this state, no one will ever reach the equilibrium point. Instead, assume that actors take a certain amount of time to find out information (like the proportion of their opposites who are following certain strategies) and to time. Each turn, only a certain portion of each actors will switch their strategy if switching benefits them. Even more, assume that this speed is directly proportional to the utility difference between one strategy and another, ie., the greater the benefit their is to switch to the other strategy, the more people will switch each turn. A system of two parametric equations could now be used to calculate how many people will test or how many firms will signal at any given turn. In this sub-model, the proportion of consumers and companies will frequently meet their part of the equilibrium point, but will "overshoot" the point, because their opposite number (the companies or consumers) are not in the correct proportion similarly. These should spiral around, until they eventually come to a stand still with both portions at the equilibrium point.

This is an interesting approach, because the velocity at which people change strategy is not actually equivalent to how fast the market reaches the equilibrium point. Too slow a speed, would mean no change at all from the present situation. Too fast a speed, guarantees more over shoots, the extreme of which would be eternal cycling between the four absolutes. Velocity, and changes in it, would only have second order effects. So consumer protection companies (or business consultants) who endeavor to spread information regarding and abilities to switch strategies, may either be irrelevant, or even preventing the market from reaching the equilibrium point, the most efficient position that the market can reach.

2) The point is not stable. Ideal equilibria are defined by reinforcing factors. A ball at rest in a valley is stable because if it starts to go to either side, gravity will pull it back down. This equilibrium point does not have immediate reinforcements. If there is a sudden shock that changes the portion of consumers who test, no consumers are directly affected and encouraged to restore the ideal portion. Instead, companies will have some incentive to deviate, and from that consumers will have incentive to deviate again, slowly oscillating once more around the equilibrium point.

3) The equilibrium point may change. The internet environment is such that benefits from purchases, benefits for consumer information, and signaling and testing costs may change. But even if they are all relatively static, In particular, T_a can be viewed as an endogenous variable that is dependent

on respectful firms trying to lower the cost of testing (they present T_1), and non-signaling firms who want to make testing a hassle to consumers by writing purposefully obtuse policies (they present T_2). Since we have

$$T_a = pT_1 + (1-p)T_2$$

the actual equilibrium point for p is affected by, and changes with, the different firms out there. As more non-respectful firms enter the market, they raise the cost of T_a, changing the incentives for consumers, so that they are less likely to bother testing. This aspect reduces even further the effectiveness of possible reductions in T (like the P3P bird), since defecting firms can foil that by not adhering to those improvements, and in fact making them more difficult. A "one-armed bandit" approach to finding a firms true testing cost could also greatly change the landscape. By gradually estimating the cost and possible signal that a firm is sending, some interesting dynamics might affect the utility predictions[4].

These reasons suggest new directions if one wants to achieve an efficient and reliable marketplace. Simply making consumers more aware of the cost of privacy violation, or trying to decrease the cost of testing (via programs like the P3P bird) cannot make an absolutely efficient market. They can, in the long term, reduce the ratios of testing consumers and respectful firms, but only that. Even grouping all firms under one trusted intermediary has drawbacks, because as soon as all consumers trust that intermediary and no longer test it, it has every incentive to abuse its resources, this time with complete market power, making even an exact mixed-strategy equilibrium unlikely. Traditional incremental and individual-agent based approaches all have troubles when if the goal is to protect the entire market with regards to privacy.

Instead, our model suggests that one needs to provide firms with direct incentives to respect personal information. Permanent and enforced laws against certain uses of such information, or absolute reductions of T to 0 (such as by the government taking on the testing itself) are the only methods at the moment that can raise p and q to 1 in a stable solution. Future research could focus on whether these conclusions are preserved even after augmenting our model to be more robust.

7 Acknowledgments

We thank Professor Ed Glaeser and Professor Richard Zeckhauser for advice and suggestions on directions of research. We also thank Stuart Schecter for helpful discussions.

References

[1] George A. Ackerhof. The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, pages 488–500, August 1970.

[2] Alessandro Acquisti. Protecting privacy with economics: Economic incentives for preventive technologies in ubiquitous computing environments. In *Workshop on Socially-informed Design of Privacy-enhancing Solutions,4th International Conference on Ubiquitous Computing (UBICOMP 02)*, September 2002. `http://guir.berkeley.edu/privacyworkshop2002/papers/acquisti-ubicomp-09%-19-02.pdf`.

[3] AT&T. P3p privacy bird. `http://www.privacybird.com`.

[4] A.N. Burnetas and M. N. Katehakis. Asymptotic bayes analysis for the finite horizon one armed bandit problem. *Probability in the Engineering and Informational Sciences*, 2002. to appear.

[5] Giacomo Calzolari and Alessandro Pavan. Optimal design of privacy policies. `http://faculty.nwu.edu/faculty/pava/ODPP.pdf`.

[6] Federal Trade Commission. Privacy online: Fair information practices in the electronic marketplace. `http://www.ftc.gov/reports/privacy2000/privacy2000.pdf`, 2000.

[7] EPIC. Pretty poor privacy. `http://www.epic.org`, 1999.

[8] Center for Democracy and Technology. Surveys main page. `http://www.cdt.org/privacy/guide/introduction/surveyinfo.html`, 2002.

[9] Heather Green, Mike France, Marcia Stepanek, and Amy Borrus. Our four-point plan. *Business Week Online*, March 2000.

[10] William F. Adkinson Jr, Jeffrey A. Eisenach, and Thomas M. Lenard. Privacy online: A report on the information practices and policies of commercial web sites. http://www.pff.org/publications/privacyonlinefinalael.pdf, 2002.

[11] Brian McWilliams. Real networks hit with privacy lawsuit. http://www.internetnews.com/bus-news/article.php/8161_235141.

[12] Dierdre Mulligan, Ann Cavoukian, Ari Schwartz, and Michael Gurski. P3p and privacy: An update for the privacy community. http://www.cdt.org/privacy/pet/p3pprivacy/shtml, 2000.

[13] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on ebay: A controlled experiment. http://www.si.umich.edu/~presnick/papers/postcards/, 2002.

[14] Michael Spence. Job market signalling. *The Quarterly Journal of Economics*, pages 355–74, 1973.

[15] TrustE. Truste statistics. http://www.truste.org/bus/pub_bottom.html, 2003.

[16] Hal Varian. Economic aspects of personal privacy. http://www.sims.berkeley.edu/~hal/Papers/privacy/, 1996.

[17] Toshio Yamagichi and Masafumi Matsuda. Improving the lemons market with a reputation system. Technical report, University of Hokkaido, 2002.