Adverse Selection in Online "Trust" Certifications

Benjamin Edelman
Harvard University
bedelman@fas.harvard.edu

October 15, 2006
working draft

Abstract

Widely-used online "trust" authorities issue certifications without substantial verification of the actual trustworthiness of recipients. Their lax approach gives rise to adverse selection: The sites that seek and obtain trust certifications are actually significantly less trustworthy than those that forego certification. I demonstrate this adverse selection empirically via a new dataset on web site characteristics and safety. I find that TRUSTe-certified sites are more than twice as likely to be untrustworthy as uncertified sites, a difference which remains statistically and economically significant when restricted to "complex" commercial sites. I also present analogous results of adverse selection in search engine advertising – finding ads at leading search engines to be more than twice as likely to be untrustworthy as corresponding organic search results for the same search terms.

Keywords: Adverse selection, certification, reputation, trust, Internet, search engines.

I thank seminar participants at Harvard University's Department of Economics, Business School, and Department of Computer Science, and at the 2006 Workshop on the Economics of Information Security (University of Cambridge). I am grateful to Robert Akerlof, Ross Anderson, Peter Coles, Chris Dixon, Andrei Hagiu, Ariel Pakes, David Parkes, Al Roth, Stuart Schechter, and five anonymous reviewers for helpful comments and suggestions.

1 Introduction

When agents have hidden types, contract theory warns of bad results and potentially even market unraveling. Since Akerlof's defining "lemons" (1970), others have worried about similar problems in other markets – such as bad drivers wanting more car insurance than good drivers (Chiappori and Salanie 2000), and healthy people disproportionately buying annuities (Finkelstein et al, 2004).

In general, it is difficult to empirically assess the significance of adverse selection problems. For example, used car markets are made more complicated by idiosyncratic details – unobservable car characteristics, local markets, and casual sellers. Some work manages to address these problems. For example, Chiappori and Salanie focus on novice drivers, who have less private information about their own type (since they have not yet started to drive), such that economists can observe most relevant characteristics. But these special cases bring problems of their own. Researchers may be less interested in the absence of adverse selection among novice drivers' insurance purchases, and more interested in the adverse selection that (perhaps) actually does affect most other drivers.

This paper applies an adverse selection model to a new market – Internet web sites and their associated "trust"-type certifications. With a new data source, I analyze characteristics generally unobservable both to consumers and to certification authorities. Unmasking sites' otherwise-hidden types provides an unusual opportunity to measure the magnitude of adverse selection occurring in this market.

Beyond adverse selection, trust certifications are also of interest in their own right. Such certifications have played an important role in the policy debate as to regulation of online privacy and safety, and typical Internet users see such certifications remarkably

frequently. Yet adverse selection significantly taints trust certifications: My analysis indicates that adverse selection substantially reduces overall certification quality. In particular, I find that sites certified by the best-known authority, TRUSTe, are more than twice as likely to be untrustworthy as uncertified sites.

1.1. The Basic Web Site Safety Problem

Consumers seeking online services face a serious problem in deciding what sites to use. Consumers could stick with "known-good" big names, but such a narrow focus would reduce match quality, denying users the rich diversity of Internet content. But venturing into the unknown Internet carries important risks: Untrustworthy sites might send users spam (if users register or otherwise provide their email addresses), infect users' computers with viruses or other harmful code (if users install the programs that sites offer), or simply fail to deliver the promised merchandise (if users make purchases). Ex ante, users have no easy way to know which sites to trust. Even if a site looks safe, it could turn out to be a wolf in sheep's clothing.

These online interactions reflect a two-sided market – with sites actively making decisions about how to present themselves. Good sites want to demonstrate that they're good. But as in the usual moral hazard framework, bad sites pretend they're good too.

Facing numerous untrustworthy or even malicious sites, some analysts call for government regulation. In principle, a government agency might examine web sites in search of spam, scams, and harmful programs. To some extent, the FTC and state attorneys general perform such investigations – though their efforts address only a small portion of bad actors. As a practical matter, government intervention seems inapt. See

Tang et al. (2005), presenting a model of enforcement of online privacy breaches, finding mandatory government standards appropriate only for serious harms.

At the other extreme, users might be left entirely on their own. In complete laissez faire, no regulator, computer maker, or even IT department helps cure a user's problems. In some respects, laissez faire is a reasonable description of the current state of affairs. (IT departments can't protect users from getting ripped off, and many a computer expert feels powerless to stop spam.) But unaccountability carries substantial costs – leading users to take excessive precautions, preventing the formation of otherwise-profitable relationships. Users would buy more products, join more sites, and download more programs were it not for their well-founded fears in a laissez faire regime.

Finally, there exists a middle approach between the extremes of government regulation and laissez faire: A non-governmental rating organization. Such an organization would identify specific bad practices, then evaluate sites' behaviors. If evaluations were accurate and low-cost, such ratings might support an equilibrium where all good firms receive positive evaluations, and where all consumers use only those sites with positive ratings. Tang et al. describe this approach as appropriate for a broad class of online behaviors. But there are reasons to doubt its effectiveness. Extending sites to take a continuum of types, a single binary certification may not convey adequate information about all possible site behaviors. (Lizzeri 1999) Insufficient precision is particularly likely if consumers are heterogeneous in preferences, especially in their assessments of objectionable behaviors. Finally, it is hard to identify what specific behaviors are "bad," particularly when firms have strong economic incentives to blur the boundaries. So in practice, a rating authority may be less expedient than Tang et al. hope.

1.2. Certification Authorities

Most prominent among non-governmental rating organizations are so-called "trust" certification authorities. These rating organizations set out specific criteria for membership, often focusing on privacy or on safety more generally. The organizations reward their members by offering seals to be placed on recipients' web sites, often at the point where site operators want to reassure users of sites' legitimacy and trustworthiness (e.g. at registration forms and checkout pages). Particularly well-known certification authorities are TRUSTe and BBBonline.

In principle, certification authorities might set and enforce substantive and procedural provisions sufficiently rigorous that certified members are highly likely to satisfy reasonable consumers' expectations of safety. But in practice, there is reasonable basis to doubt the effectiveness of certain certification authorities. LaRose and Rifon (2002) offer a stinging critique: Certification authorities have granted multiple certifications to firms under investigation by the FTC for privacy policy violations; certification authorities have failed to pursue complaints against major companies whose privacy breaches were found to be "inadvertent"; and in one case a certification authority even failed to abide by its own privacy policy. Singel (2006) also questions the effectiveness of TRUSTe's effort: In a 2004 investigation after user complaints, TRUSTe gave Gratis Internet a clean bill of health. Yet a subsequent New York Attorney General statement and settlement indicated that Gratis had committed exceptionally far-reaching privacy policy violations – selling 7.2 million users' names, email addresses, street addresses, and phone numbers, despite a privacy policy that prohibited such a sale.

As a threshold matter, certification authorities' substantive standards often seem substantially duplicative with existing duties or practices. Consider the requirements summarized in TRUSTe's Program Requirements. For example, the first listed rule, allowing an email unsubscribe function, duplicates Sec.5.(a)(4)(A) of the federal CAN-SPAM Act. Similarly, the first security rule, using SSL encryption or similar technology to protect sensitive information like credit card numbers, is already widespread due to credit card network rules. See also Boutin (2002), reporting TRUSTe initially lacking any *substantive* requirements whatsoever (i.e. requiring only the *presence* of a privacy policy).[1] Such low standards match the predictions of Lizzeri (1999), finding that, under general conditions, a certification intermediary prefers only to reveal whether quality exceeds some minimal standard.

Tellingly, strikingly few certificates have been revoked. For example, TRUSTe's Fact Sheet (2006) reports only two certifications revoked in TRUSTe's ten-year history. Of course there are multiple explanations for an absence of revocations. Suppose certificate recipients think a certification authority will detect infractions with high probability and will impose harsh punishments. Then that authority will attract only "good" sites, and the authority might never actually detect any wrongdoing, nor ever actually have to punish any recipient. But this tough-dog theory stands contrary to observed facts: For example, TRUSTe has only a small staff, with little obvious ability to detect violations of its rules. TRUSTe's posted procedures reveal substantial focus on sites' self-certifications and on user complaints. Rule violations at TRUSTe member sites have repeatedly been uncovered by independent third parties, not by TRUSTe itself.

[1] According to some industry sources, TRUSTe claims that most applicants must modify their sites or practices to meet TRUSTe's requirements. But TRUSTe does not report these changes, even in general terms. It is therefore difficult to assess what benefits, if any, these changes provide to users.

TRUSTe's "Watchdog Reports" page also indicates a lack of focus on enforcement. According to TRUSTe's posted data, users continue to submit hundreds of complaints each month. But of the 3,416 complaints received since January 2003, TRUSTe concluded that *not a single one* required any change to any member's operations, privacy statement, or privacy practices, nor did any complaint require any revocation or on-site audit. Other aspects of TRUSTe's watchdog system also indicate a lack of diligence.[2]

Finally, as Greenstadt and Smith (2005) point out, certification authorities are "captured" – paid by the same companies they certify. Certification authorities have little incentive to antagonize their customers: Any such pressure would harm the authority's profits by discouraging renewals and future applications.

Even the creators of certification authorities seem unhappy with their development. Boutin (2002) quotes TRUSTe co-founder Esther Dyson conceding that TRUSTe is "a little too corporate" and that TRUSTe lacks the "moral courage" to criticize violations. Similarly, TRUSTe co-founder Electronic Frontier Foundation admitted in a 1999 letter to the FTC that "legislation is needed" to protect users' privacy and that "it is time to move away from a strict self-regulation approach."

Facing allegations of low substantive standards, lax enforcement, and ethical compromise, it is unclear what direct benefits site certifications offer to consumers. Furthermore, consumers are unlikely to place substantial value on a promise to assist with dispute resolution: TRUSTe's lenient disposition of watchdog reports indicates that consumers' complaints rarely cause changes in members' business practices.

[2] TRUSTe failed to update its Watchdog Reports list, https://www.truste.org/consumers/watchdog_reports.php, between June 2004 and spring 2006, an omission corrected only after circulation of this article. Furthermore, the posted reports show a striking inattention to detail. For example, the first seven months' reports all bear the title "Watchdog Report for October 2000" (correct only for the first of those reports), and the next 37 reports all say "Watchdog Report for May 2001" (a title also inaccurate for all but the first).

Despite certification authorities' limited substantive role in assuring good online practices, at least some consumers seem to regard a certification authority's certification as a significant positive signal. For example, in promoting its service to potential applicants, TRUSTe touts its benefits to certificate recipient Realty Tracker, which says TRUSTe "convey[ed] trust" and "built confidence" with Realty Tracker's visitors, yielding "an increase in registrations." See Realty Tracker Case Study. See also LaRose and Rifon, characterizing certification authorities' seals as "persuasion attempt[s]."

Firms are well-equipped to evaluate claims of benefits to certification: Firms could randomize their inclusion of TRUSTe or similar seals, thereby determining whether seals actually increase registrations and sales. In the related context of comparison shopping sites, Baye and Morgan (2003) empirically confirm the benefits of certification seals: Merchants with seals can charge a price premium over uncertified merchants, without losing customers.

Whatever the actual merits of certification authorities as arbiters of trust, some government authorities seem to regard certification authorities as an appropriate step forward. See the FTC's 1999 "Self-Regulation and Privacy Online," finding private-sector certification authorities preferable to direct FTC regulation.

The FTC specifically cites two well-known certification systems: TRUSTe's Web Privacy Seal and BBBOnLine's Privacy Seal Program. These certification authorities are the focus of my subsequent analysis, due to their prevalence, their relatively large member lists (compared to other certification authorities), and their decisions to post their member lists (providing data necessary for my analysis). I largely focus on TRUSTe, the first online trust certification authority, the largest, and (it seems) still the best-known.

1.3. Search Engines as Arbiters of Trust

Though less explicitly focused on trust, search engines also play a prominent role in influencing users' decisions as to what sites and services are safe to use. High placement at a search engine conveys a kind of endorsement – that a given site is (or is believed to be) among the best resources for a given search term. (Gaudeul 2004)

Empirical work finds that users value high search engine rankings. Consumers believe highest-ranked sites are most likely to serve their interests (Marable 2003), and top-ranked sites have the highest click-through rates. (Joachims 2005) Because users tend not to understand the difference between paid search engine advertising and ordinary "organic" listings (Consumer Reports WebWatch 2002), Marable's result likely applies to all search results, not just organic results.

Search engines present two distinct potential notions of certification. First, a site might be said to be "certified" (or at least endorsed by the search engine) if it appears high in organic search results, rather than further down or (perhaps) not at all. Second, a site might be said to be certified if it appears as a sponsored result, e.g. a search engine advertisement. Subsequent analysis will test each of these hypotheses separately.

The remainder of this paper largely groups search engines together with certification authorities. I use the term "certification authority" to refer specifically to certificate-granting organizations such as TRUSTe, while I use the broader term "trust authority" to include search engines also.

2 Theory of Adverse Selection in Trust Authorities

Suppose that, as described above, certain trust authorities issue certifications of trustworthiness without rigorous assessment of recipients' true trustworthiness. This

market structure creates significant reason to worry of adverse selection among certification recipients. Certifications of trustworthiness seek to signal consumers that the certified firms are in fact highly likely to be trustworthy. But if untrustworthy firms can obtain certifications just as easily as trustworthy firms, then consumers have little reason to conclude that a certified firm is trustworthy: Rational consumers would rightly worry that certified firms obtained certifications despite actually being untrustworthy.

To provide consumers with the intended positive signal, a trust authority's certification must increase a rational consumer's assessed probability that a given site is trustworthy. Suppose a rational consumer has some prior belief $P(t)$ that a given site is trustworthy, before receiving a signal (denoted "s"**)** of trustworthiness ("t"). Such a consumer should update his probability according to the usual Bayes Rule formula:

$$P(t|s) = \frac{P(s|t)\,P(t)}{P(s)} \tag{1}$$

Expanding the denominator using the Law of Total Probability:

$$P(t|s) = \frac{P(s|t)\,P(t)}{P(s|t)\,P(t) + P(s|\bar{}t)\,P(\bar{}t)} \tag{2}$$

For consumer's assessment of site trustworthiness to increase as a result of a site's certification, it must be the case that $P(t|s) > P(t)$, which implies:

$$\frac{P(s|t)}{P(s|t)\,P(t) + P(s|\bar{}t)\,P(\bar{}t)} > 1 \tag{3}$$

Rearranging further, using the fact that $P(t) = 1 - P(\bar{}t)$:

$$P(s|t) > P(s|t)\,P(t) + P(s|\bar{}t)\,P(\bar{}t) \tag{4}$$

$$P(s|t)\,P(\bar{}t) > P(s|\bar{}t)\,P(\bar{}t) \tag{5}$$

$$P(s|t) > P(s|\bar{}t) \tag{6}$$

Equation 6 offers an intuitive result: For a certification to cause a consumer to conclude a certified site is more safe than the consumer thought ex ante, the certification must be given to trustworthy sites more often than it is given to untrustworthy sites.

Equation 6 yields an empirical strategy for testing site certifications: Compare the certification rates of trustworthy sites with the certification rates of untrustworthy sites. Alternatively, further rearranging confirms that it is equivalent to compare the trustworthiness rates of certified sites, relative to the trustworthiness rates of uncertified sites. (See Appendix for proof.) For a valid certification that increases consumers' ex post assessment of site trustworthiness, certified sites must be more likely to be trustworthy than are uncertified sites. Formally, a valid certification requires

$$P(t|s) > P(t|\bar{s}) \tag{7}$$

The preceding adverse selection model offers a clear empirical prediction: That the inequality in (7) should fail. In particular, if adverse selection substantially affects these certifications, then certified sites should be *less* safe than uncertified sites.

HYPOTHESIS 1: Certified sites are less safe than uncertified sites.

Analyzing correlations between trustworthiness and certification is analogous to the approach in the existing adverse selection literature. Consider Finkelstein (2004), finding that annuitants live longer than non-annuitants – a negative relationship between claimed type (annuity purchase) and outcome (lifetime). Chiappori and Salanie (2000) use a similar method to demonstrate the absence of adverse selection in car insurance for novice drivers in France – finding no correlation between the conditional distributions of claimed type (insurance purchase) and outcome (insurance claims). Genesove (1993) extends these correlations with the equilibrium assumption that average price in a given market must reflect average quality in that market. He then regresses auction bids on

variables including a type-determining variable (there, whether a given used car was sold by a dealer who exclusively sells used cars), interpreting a significant coefficient as evidence of adverse selection at used car dealers. Villeneuve (2003) offers a specific measurement of "intensity of adverse selection," calculated as the quotient between the prevalence of some action (e.g. buying insurance) in a subsample, versus the action's prevalence in the full population. Rearranging terms, Villeneuve's measure matches (7).

Others studying online certification authorities have also worried of adverse selection. See LaRose and Rifon, finding that privacy policies at certified sites allow more invasive data collection than policies at uncertified sites. But where LaRose and Rifon hand-score 200 sites, I consider *hundreds of thousands* of sites using automated methods, and I consider axes of trustworthiness other than privacy policy loopholes. In addition to investigating the quality of certified sites, Jamal et al (2003) specifically consider certifiers lowering their standards to attract more sites. But Jamal et al study only 34 well-known sites certified as of 2001 – restrictions limiting the generality of their finding that certifications tend to deliver what they promise. In contrast, I include current data and more sites – letting me analyze all certified sites, not just top sites.

2.1. Trust Authorities in Equilibrium

Critics might reasonably doubt whether uninformative certifications can exist in equilibrium. Suppose, as hypothesized above, that trust authorities suffer adverse selection – such that certified sites are actually less deserving of trust, on average, then uncertified sites. Alternatively, suppose trust authorities award certifications randomly, uncorrelated with sites' actual trustworthiness. In equilibrium, users should learn that so-called "trust" certifications are actually uninformative. Then users should discount –

ignore! – those certifications. But if consumers ignore the certifications, sites have no incentive to become certified. Then certification schemes should disappear altogether.

It is reassuring to see a prediction that worthless trust authorities will self-destruct. But in practice, we observe that trust authorities *do* exist, have existed for some time, and show no sign of disappearing. A natural null hypothesis is exogenous market forces. For example, the large companies that founded TRUSTe are likely to continue to support it so long as it serves their regulatory goals. So even if TRUSTe would otherwise face extinction, core members may keep it afloat. Similarly, search engines' ordered listings unavoidably issue implicit certifications, i.e. high rankings, and this design decision seem unlikely to change substantially in the short run. So trust authorities may continue to exist for an extended period, even if basic economic theory suggests that they should not.

Although I credit this null hypothesis, the following two sections sketch algebraic models of two alternatives. First, I present the effect of slow-learning users. Second, I consider the continuing influx of novice users and their likely beliefs about certifications.

2.2. Model: A Profit-Maximizing Certifier with Slow-Learning Users

Suppose a trust authority happened to start with "good" with members that truly are trustworthy. This initial good period creates reputation among consumers who, during the initial period, observe that the trust authority's members actually are trustworthy. That reputation might take some time to dissipate, i.e. in the face of slow learning by otherwise-sophisticated consumers. In the interim, untrustworthy firms can use consumers' delayed learning to gain consumers' trust. The resulting hypothesis:

> *HYPOTHESIS 2: Certification authorities do not suffer adverse selection in initial periods, but they suffer adverse selection that worsens over time.*

There is good reason to think a certification authority might start with trustworthy members. When a new certification authority begins operation, its certificate of trustworthiness has no clear value. An online "certificate" is just an image placed on recipients' sites – easily replicated by any other self-styled trust authority or rogue third party. Such a certification offers uncertain initial value to initial recipients. Good-type recipients may want the certification for some intrinsic reason. But bad-type recipients have no reason to seek certification: Since consumers initially do not know what the certification (purportedly) means, they will not defer to it. Furthermore, consider the alternative: If a certifier began by certifying untrustworthy firms, it would have little hope of building positive reputation with users – an extra reason to start with good members.

There is also good reason to think consumers will be slow to learn what certification means. Certification issuers may be less than forthright, lacking incentive to explain their true practices. Users cannot easily link bad experiences back to specific sites and certifications, especially when harm is delayed and when causation is unclear.

For concreteness, I offer a formal model of certification decisions when learning is slow. I largely follow Lizzeri (1999), but I extend his approach to multiple time periods.

Consider a profit-maximizing certification authority that may certify any or all of a set of firms, indexed by i. Firms have qualities q_i drawn from a uniform distribution, $q_i \sim \text{unif}(0,1)$, which the certification authority observes perfectly. The certifier reaps a fixed profit from each certification issued. For example, if the certifier certifies all firms with quality above $\bar{q}$, the certifier's profit is proportional to (and without loss of generality, equal to) $\pi = 1 - \bar{q}$. The certification authority discounts the future at rate δ.

Consumers defer to certifications that they believe to be credible. In particular, consumers seek a guaranteed minimum quality $\mathbf{E}_c[\min(q|\text{cert})] \geq \bar{q} > 0$, where $\mathbf{E}_c$ denotes consumers' beliefs. If consumers learn that certified sites do not meet the $\bar{q}$ level of quality, consumers will stop placing any weight on the certifications.

Consumers assess certified sites using a slow-learning procedure. In the first period, consumers accurately assess certified sites. Each period thereafter, consumers update their assessment of minimum certification quality using a moving average, placing weight ρ on their prior assessment of certification quality, and 1-ρ on the true minimum quality of certified sites. That is

$$E_c[\min(q^t|\text{cert})] = \rho E_c[\min(q^{t-1}|\text{cert})] + (1-\rho)\min(q^t|\text{cert}) \quad (8)$$

Following Baye and Morgan (2003), certification strictly increases a firm's profits – if consumers consider the certification credible. But if consumers find a certification non-credible, the certification is worthless.

Finally, the certification authority is constrained in its choice of issuing rules: It can set one cutoff Q_1 in the first period and a different cutoff Q_2 in the second period, but that second cutoff must then be retained permanently. This is a strong assumption, but it yields stark results to capture the model's intuition. At the conclusion of this section, I consider the implications of relaxing this requirement.

Suppose the certification authority exogenously sets a cutoff rule Q_1 for the first period. Then the certification authority will certify all sites with $q_i \geq Q_1$, yielding profit to the certifier of $\pi_1 = 1-Q_1$. Consumers form correct beliefs of $\min(q^1|\text{cert}) = Q_1$.

Now consider the certification authority's choice of Q_2. The certification authority maximizes future profits, discounted by the certification authority's discount rate:

$$\pi_{\text{infinite}}(Q_2) = (1\text{-}Q_1) + \sum_{t=1}^{\infty}\delta^t(1\text{-}Q_2)\mathbf{1}\{\mathrm{E_c}[\min(q^t|\text{cert})|Q_2] \geq \bar{q}\} \tag{9}$$

Correctly maximizing this summation requires anticipating consumers' learning.

Consider consumers' future beliefs of certification quality, as a function of Q_2:

$$\begin{aligned}
\mathrm{E_c}[\min(q^1|\text{certified})] &= Q_1 \\
\mathrm{E_c}[\min(q^2|\text{certified})] &= Q_1\rho + Q_2(1\text{-}\rho) \\
\mathrm{E_c}[\min(q^3|\text{certified})] &= Q_1\rho^2 + \rho(1\text{-}\rho)Q_2 + (1\text{-}\rho)Q_2 \\
\mathrm{E_c}[\min(q^4|\text{certified})] &= Q_1\rho^3 + (1\text{-}\rho)(\rho^2 + \rho + 1)Q_2 \\
\mathrm{E_c}[\min(q^t|\text{certified})] &= Q_1\rho^{t\text{-}1} + (1\text{-}\rho)\left(\sum_{i=1}^{t-2}\rho^i\right)Q_2 \\
&= Q_1\rho^{t\text{-}1} + (1\text{-}\rho)\left(\frac{1}{1\text{-}\rho} - \frac{\rho^{t-1}}{1\text{-}\rho}\right)Q_2 \\
&= Q_1\rho^{t\text{-}1} + (1\text{-}\rho^{t-1})Q_2
\end{aligned} \tag{10}$$

The certification remains credible so long as $\mathbf{E}_c[\min(q^t|\text{certified})] \geq \bar{q}$. Solving for t:

$$\begin{aligned}
&Q_1\rho^{t-1} + (1\text{-}\rho^{t-1})Q_2 \geq \bar{q} \\
&\rho^{t-1} \geq \frac{\bar{q} - Q_2}{Q_1 - Q_2} \\
&t \leq \frac{\ln(\bar{q} - Q_2) - \ln(Q_1 - Q_2)}{\ln(\rho)} + 1 = t^*
\end{aligned} \tag{11}$$

The certification authority therefore receives future discount profits of:

$$\pi_{\text{infinite}}(Q_2) = (1\text{-}Q_1) + \sum_{t=1}^{\infty}\delta^t(1\text{-}Q_2)\mathbf{1}\left\{t < \frac{\ln(\bar{q} - Q_2) - \ln(Q_1 - Q_2)}{\ln(\rho)} + 1\right\} \tag{12}$$

Computing the summations, using the standard result that $\sum_{t=0}^{\infty} ar^t = a/(1\text{-}r)$, and separating piecewise according to the two possible ranges for Q_2:

$$\pi_{\text{infinite}}(Q_2) = \begin{cases} (1\text{-}Q_1) + (1\text{-}Q_2)\delta/(1\text{-}\delta) & Q_2 \geq \bar{q} \\ (1\text{-}Q_1) + (1\text{-}Q_2)(1\text{-}\delta^{[\ln(\bar{q} - Q_2) - \ln(Q_1 - Q_2)]/\ln(\rho) + 1})\delta/(1\text{-}\delta) & Q_2 < \bar{q} \end{cases} \tag{13}$$

The top result indicates that if the certifier sets a $Q_2 \geq \bar{q}$, the certification will remain credible forever, and the certifier will receive an infinite stream of certification payments. But if the certifier sets $Q_2 < \bar{q}$, as in the bottom result, the certification only remains credible through $t=t^*$ given by (11), truncating the certifier's payments accordingly.

Taking first-order conditions of (13) yields a lengthy but explicit optimal Q_2, omitted here for brevity. More useful are two comparative statics on Q_2. All else equal, the authority sets a low Q_2 if δ is small, e.g. the certification authority highly discounts the future. Similarly, if ρ is sufficiently large, consumers remember the past so strongly that the certification authority can profitably certify low-quality sites and reap higher profits from doing so, without consumers quickly noticing.

Suppose the certification authority sets a $Q_2<\bar{q}$. From periods 2 through t^*, the true minimum quality of certified sites will be below consumers' desired $\bar{q}$, that is $\min(q^t|\text{certified})<\bar{q}$, Nonetheless, for any $Q_2>0$, certified sites will remain at least somewhat safer than uncertified sites, namely $\min(q^t|\text{certified})>\min(q^t)$ and $E[q^t|\text{certified}]>E[q^t]$. For certified sites to be *worse* than uncertified sites, at least some good sites must elect not to get certified – a result developed in the next section.

These results rely in part on the requirement that Q_2 stay fixed over time, not adjusting from period to period. If the certification authority were able to adjust Q_2, it would raise Q_2 just before the t^* from (11), thereby maximizing profits while avoiding losing credibility. Nonetheless, there would still exist a time period, e.g. from periods 2 through t^*, during which consumers are mistaken about firm trustworthiness.

2.3. Model: Novice Users and Site Heterogeneity

The prior section suggests that certification authorities might persist for an extended period due to users' slow learning. But then *all* firms should get certified – contrary to empirical observation that many sites don't. This section offers a second model of certification decisions, providing intuition on why only some sites get certified.

Suppose firms receives a continuous stream of naïve new consumers, who mistakenly believe all certified firms are trustworthy. If novices are sufficiently numerous, firms might find it profitable to continue to present trust certificates, even though experienced users know the certifications are worthless. If certifications cost sufficiently little, they may offer a measurable benefit with no substantial downside.

Suppose proportion p_i of consumers at a given firm i are naïve, while the remaining $1\text{-}p_i$ are sophisticated. Sophisticated consumers make purchases from that firm with probability s_i, and sophisticated consumers ignore sites' certifications. But if a firm is certified, naïve consumers purchase with a higher probability r_i+s_i. (For naïve consumers, r_i gives the marginal increase in purchase probability when a firm obtains certification.) A certificate costs $\boldsymbol{c}$ (per year). A given firm receives n_i visitors per year and earns profit π_i from each sale. A rational firm obtains certification if the certification increases profits, i.e.:

$$\text{[profits from certification]} > \text{[profits if not certified]} \tag{14}$$

Specifically:

$$\begin{aligned}&\text{[profits from naïve consumers if certified]}\\ &\quad + \text{[profits from sophisticated consumers if certified]}\\ &\quad - \text{[costs of certification]}\\ &\qquad\qquad > \text{[profits from all consumers if not certified]}\end{aligned} \tag{15}$$

Substituting:

$$[n_i\pi_i][p_i(r_i+s_i) + (1\text{-}p_i)s_i] - c \; > n_i\pi_i s_i \tag{16}$$

Rearranging and canceling, a firm obtains a certification if:

$$n_i\pi_i p_i r_i > c \tag{17}$$

In general, the prior section assumed n_i, π_i, r_i, and s_i were all strictly positive, and that c was relatively small. These are good assumptions for typical commercial web

sites, with substantial naïve users (p nonnegligible) who are confused about the value of certifications (s nonnegligible), and with certification cost low relative to site size (c small). The model predicts that all such sites should seek and obtain certifications.

But consider a firmwith more sophisticated users. Their sophistication could enter via a small p_i – a firm where very few users are naïve, i.e. because they understand the true meaning of certification. Alternatively, consumer sophistication could enter via a small s_i – no increase in purchases due to certification, i.e. because consumers already know about the site and would have made purchases even without a certification. For such firms, the left side of Equation 17 may be smaller than the right, making certification unprofitable.

This approach also models other typical site characteristics. Consider variation in firms' ex-ante reputation. For example, a certification probably cannot boost eBay's already-good reputation – so eBay's s_i is small. In contrast, obscure firms have big s_i's because certifications are more likely to increase their perceived trustworthiness. The model also anticipates heterogeneity in firms' compliance costs, following Tang et al. Firms with prohibited or borderline practices face higher effective costs of certification, i.e. big c_i's. But some trustworthy firms might have high effective c_i's too, because their complex operations or high-paid staff (e.g. attorneys) make it particularly costly to confirm compliance. Firms on both extremes might therefore forego certification.

This is a static model; it does not develop and cannot predict equilibrium outcomes. In particular, in this model consumers do not update their beliefs according to firms' behavior, nor do firms change their behavior to suit changes in consumers' decision-making processes (since consumers' decisions do not change). I offer this model to help

understand *observed* outcomes – that some sites get certified and others do not. I omit a more general model because this model is sufficient to motivate subsequent results and because I consider this market far from equilibrium. (Many new users are arriving, and many new sites are appearing, with widespread hidden information about site types.)

2.4. Adverse Selection in Search Engine Results

The empirical economics literature confirms a worry of adverse selection in search engine advertising. Animesh et al. (2005) examine relationships between product type, quality (e.g. trustworthiness), and advertising strategy. Following Darby and Karny (1973), Animesh et al. separate search terms according to product type – distinguishing between search goods (with characteristics identifiable through pre-purchase inspection), experience goods (characteristics revealed only through consumption), and credence goods (where even ex-post observation does not reveal all characteristics). For experience and credence goods, Animesh et al. find that lower quality firms bid higher, but they find a positive relationship between quality and bids for search goods. Animesh et al. therefore find an adverse selection effect in search engine advertising for experience and credence goods, though not for search goods. Animesh's intuition: Users recognize and patronize trustworthy firms selling search goods. But when users want experience and credence goods, users can't distinguish between trustworthy and untrustworthy firms. Untrustworthy firms make higher profits (e.g. by selling low-quality goods at full price), so untrustworthy firms can afford to bid higher for search engine ads in these categories.

Animesh et al. consider the *intensive* margin of search engine advertising – *how much* a site bids for search ads. Animesh et al. therefore effectively test the hypothesis of higher-ranked pay-per-click sites being safer than lower-ranked sites. But adverse

selection can also present itself at the *extensive* margin – whether sites advertise through search advertising *at all*. This extensive analysis is the focus of my subsequent analysis.

In contrast to search engine ads, where bids largely determine placement, organic results are intended to select for *high-quality* sites. As described in Google's much-cited PageRank specification (Brin and Page 1998), modern search engines generally evaluate sites in part based on their inbound links (links from other sites). "Bad" sites find it harder to obtain inbound links: Others don't want to link to sites they consider untrustworthy. So link-based rating systems may make search engines' organic listings more trustworthy and less subject to adverse selection or manipulation. In particular:

HYPOTHESIS 3: Organic results are safer than sponsored results.

HYPOTHESIS 4: Higher-quality search engines have safer organic results.

Here, quality refers most naturally to use of PageRank-type ratings, but more loosely to user appraisal of search engine relevance. Industry sources indicate that Google and Yahoo increasingly have comparable organic search quality, with Microsoft and Ask somewhat behind, all in that order. See e.g. Webmasterbrain (2006).

Not all analysts believe search engine advertising faces adverse selection. Overture founder Bill Gross reportedly commented that "the best way to clean up search results [is] to use money as a filter." (Hansell 2001) Gross effectively asks what distinguishes high-quality sites from low-quality sites. In Gross's view, the difference is that only high-quality sites can afford to advertise. Hypothesis 3 suggests an alternative: That low-quality sites are equally (or better) able to advertise, but that high-quality sites can more easily obtain favorable organic placement via links from other high-quality sites. I distinguish between these theories in my test of Hypothesis 3.

Section 1.3 offers an additional notion of search engines issuing certifications: That a search engine implicitly endorses a site by granting it a high ranking in organic listings. This notion of certification offers a corresponding question for site safety: high-ranked sites could be more or less safe than lower-ranked sites. But here, there is little reason to worry of adverse selection: A robust mechanism, namely PageRank and other link analysis, guards high rankings. The resulting hypothesis:

HYPOTHESIS 5: High-ranked organic sites are no less safe than other sites.

3 Empirical Strategy

The preceding hypotheses call for analysis of "true" trustworthiness of a large number of sites. In general this data is difficult to obtain. If such data were readily available to consumers, there would be no hidden type problem be no opportunity for adverse selection. Furthermore, in general market participants have more information than researchers working after-the-fact. But the peculiarities of online trust make it possible to examine, measure, and analyze sites' trustworthiness, even though consumers and trust authorities largely lack this information.

To determine sites' "true" trustworthiness, I use data from SiteAdvisor. (Disclosure: SiteAdvisor is a for-profit firm, and I serve on its advisory board.) To protect consumers from unsafe web sites, SiteAdvisor runs automated systems ("robots") to visit web sites and attempt to measure their safety. SiteAdvisor's robots uncover site characteristics that are otherwise difficult for users to discern. For example, one SiteAdvisor robot provides a different single-use email address to each web form it finds. SiteAdvisor measures how many messages are subsequently sent to that address – identifying sites and forms that yield junk mail. Another SiteAdvisor robot downloads all

programs it finds, installs each program on a separate virtual computer, then scans for spyware – assessing the possibility of infection at each site. Other robots check for excessive pop-ups, security exploits, scams, links to other bad sites, and more.

SiteAdvisor's measurements are imperfectly correlated with trust authorities' rules. For example, a site could send hundreds of emails per week to its registrants, yet still receive a TRUSTe certification and still qualify to advertise at major search engines. Nonetheless, SiteAdvisor's approach is highly correlated with the behaviors *users* actually find objectionable. Users are unlikely to understand the subtleties of trust certifications; rightly or wrongly, users seem to regard such certifications as general seals of approval and of good business practices. Any site failing SiteAdvisor's tests is a site likely to be of substantial concern to typical users. I therefore consider SiteAdvisor data a good proxy for sites' true trustworthiness – for the outcomes users actually care about.

Separately, I need data on trust authorities' member lists. I obtain member lists from the current web sites of TRUSTe and BBBOnLine, and I obtain yearly historic TRUSTe member lists from date-stamped data at archive.org.

For assessment of search engines' implicit endorsements of trustworthiness, I use a crawler to extract search engine results and ads as of January 2006. I extract data for 1,397 popular keywords, including all Google Zeitgeist 2005 keywords (popular search terms) as well as corresponding lists from other search engines. I extract data from the top five search engines: Google, Yahoo, AOL, MSN, and Ask. For each search term, I extract the top 50 organic results and up to the first 50 ads (if available).

Despite the apparent simplicity of Equations 6 and 7, they hide considerable complexity. These equations might be taken to call for conditioning on other site

characteristics – for example, comparing certified sites with other commercial sites rather than with a full cross-section of sites. My empirical strategy includes specifications with various controls, including a crude measure of site commerciality (.COM versus .ORG versus other extensions) as well as popularity (as measured by an ISP).

Throughout, I analyze approximately half a million sites – generally the top sites according to the ISP that provided me with popularity data. In many specifications, I add information about site popularity, again as measured via this ISP. My "traffic" data comes in rank form, so larger values counterintuitively imply *smaller* amounts of traffic.

4 Results and Discussion

4.1. Certification Authorities

I begin by testing Hypothesis 1 using the method in Equation 7. Comparing the trustworthiness of certified and uncertified sites, I obtain the results in Tables 1 and 2 for TRUSTe and BBBOnLine (privacy seal program), respectively. Notice that TRUSTe-certified sites are *less* likely to actually be trustworthy: Only 94.6% of TRUSTe-certified sites are actually trustworthy (according to SiteAdvisor's data), whereas 97.5% of all tested sites (and 97.5% of non-TRUSTe sites) are trustworthy. That is, TRUSTe-certified sites are more than twice as likely to be untrustworthy as uncertified sites. This analysis gives a basic initial confirmation of the adverse selection result posited in Section 2.

The TRUSTe adverse selection result in Table 1 holds in a regression framework that controls for additional variables. Table 3 Column 1 gives a probit estimation of the relationship between TRUSTe certification and true site trustworthiness. Column 2 adds site traffic – addressing the worry that popular sites are exogenously both safer and more likely to be certified. Column 3 adds a notion of site type – dummies for .COM sites and

for .ORG's. In each specification, the TRUSTe certification coefficient remains statistically significantly negative. That is, on the margin, TRUSTe certification remains associated with a reduction in the probability that a given site is actually trustworthy.

In Table 5, I test the suggestion that TRUSTe's negative association with trustworthiness is spurious. Some might worry: TRUSTe's members tend to operate complex web sites, and complex sites can fail SiteAdvisor's automated testing in more ways than simple (static, non-interactive) sites. So perhaps the untrustworthiness of TRUSTe's members reflects only that complex sites both 1) get certified by TRUSTe, and 2) fail automated trustworthiness tests. I reject this hypothesis by restricting analysis to domains that offer downloads and/or email signup forms. Restricting my analysis to this subset of domains, I find that TRUSTe certification remains significantly negative.

Notably, Tables 2 and 4 indicate that BBBOnLine's privacy seal does not suffer significant adverse selection. Unlike TRUSTe's certified sites, BBB-certified sites are slightly more likely to be trustworthy than a random cross-section of sites. Industry sources attribute BBB's success to BBB's detailed evaluation of applicants, including requiring membership in a local BBB chapter (with associated additional requirements), whereas TRUSTe tends to rely primarily on applicants' self-assessments. Though BBB's approach offers important benefits, BBB apparently faces substantial difficulties: A backlog of applicants and a slow application approval process (in part caused by the additional required evaluations). BBB's web site reports only 631 certificates have been issued to date, and it is unclear whether BBB could scale its process to evaluate orders of magnitude more sites. Section 5 expands on the policy ramifications of these differences.

Hypothesis 2 conjectured that certification authorities' membership becomes less trustworthy over time. Table 6 and Figure 1 confirm that hypothesis. Note that I do not observe sites' prior practices. I use current trustworthiness as a proxy for historic behavior – effectively assuming that trustworthy sites stay trustworthy, and vice versa.

4.2. Search Engines and Search Engine Advertising

Similar adverse selection plagues search engines advertising, as set out in Hypothesis 3. Table 7 demonstrates that untrustworthy sites are overrepresented among ads at all five tested search engines, relative to their presence in organic listings for the same terms: Rows 5 through 8 (percent of sponsored results that are untrustworthy) are all larger than rows 1 through 4 (untrustworthiness of organic results). An ANOVA test confirms that these differences are highly significant ($P<0.001$).

Table 7 shows a striking result for organic listings at Yahoo – *zero* untrustworthy sites in the top organic position at Yahoo (for the 1,397 keywords tested), compared to 2%+ at every other search engine. Conscious of Yahoo's history of manual organization (e.g. via the Yahoo Directory), some industry sources suggest that Yahoo manually selects top organic sites for top keywords. Such a procedure could produce exceptionally safe top organic results for top keywords, albeit at a cost in Yahoo staff time.

The relative untrustworthiness of search engine ads raises questions as to why search engines' organic listings are safer. Hypotheses 4 and 5 both speak to this question, positing that organic listing algorithms (e.g. PageRank) successfully block untrustworthy sites from high organic positions. Hypothesis 4 suggests that search engines with inferior organic listing algorithms (e.g. Ask) have less trustworthy organic results. Table 7 confirms this claim – finding Ask's organic results less safe than other

search engines (statistically significant in an ANOVA test). This confirmation of Hypothesis 4 suggests that other search engines' safer organic results result from their better organic listing algorithms. Hypothesis 5 approaches the same question from a different perspective, looking for relationships between site safety and organic ranking. Table 8 reports that higher-ranked organic sites are safer, although this result is driven by subsequent pages of results (Column 4), not placement within the first page (Column 2).

Affirmation of Hypotheses 4 and 5 tends to reinforce claims that rigorous organic listing algorithms increase organic listing safety. Adverse selection is not a foregone conclusion. To the contrary, it can be prevented by appropriate market institutions and even automated systems. Edelman and Rosenbaum (2006) offer additional suggestions.

5 Policy Implications

The markets at issue – explicit "trust" certifications and search engine advertising – are new and developing, subject to ongoing adjustment and redesign. In established industries, norms and fundamental product characteristics substantially limit how markets develop. But in these developing markets, institutional reforms could improve outcomes.

Policy responses could encourage trust authorities to consider the externalities of their actions. If a trust authority says a site is trustworthy when it is not, the trust authority currently can largely ignore the consequences of its error. (Indeed, in the short run trust authorities benefit from such errors: Certification authorities receive fees for each certification issued, and search engines get paid for showing advertisements. But no fees result from refusing certifications or ads.) However, suitable sanctions could change trust authorities' calculations. An appropriate sanction would force trust authorities to consider the harm consumers suffer from reliance on an erroneous endorsement.

Regulators have clear mechanisms to alter certification authorities' behavior. For example, law or regulation could require that certification authorities devote more resources to finding improper uses of their certifications. Granting certifications without proper investigation could expose a certification entity to liability (roughly on a negligence theory). Alternatively, policy could attempt to encourage enforcement. If certification authorities were required to publish consumer complaints, the resulting transparency would help assure appropriate responses, and more consumers would participate. Meritorious complaints could also be rewarded directly, i.e. via bounties.

Search engines present similar opportunities for policy intervention. A 2002 FTC rule requires search engines to label their sponsored links. But further FTC rules could extend this duty, as could litigation under existing statutes. For example, search engines could be required to exercise reasonable diligence in selecting and approving advertising buyers. Duties might track a basic common law notion of negligence – with increased care required when selling a large amount of advertising, when doing business with an otherwise-unknown company, and when selling ads in categories known to include many untrustworthy advertisers. Search engines might satisfy these duties, at least in part, by offering improved procedures for users to complain about particular ads.

Although I focus on a single market, my results fall within a broader context of regulators responding to incentives. A certificate issuer is paid a fee when it issues a certification, but it gets nothing for rejecting an application. Similarly, examiners at the US Patent Office get credit for each patent application approved – but far less credit for each application rejected. (Ravicher 2005) The notorious 1850 Fugitive Slave Act also set skewed compensation – granting a judge twice as large a fee for sending a defendant

to slavery as for finding him free. Facing these incentives, it's little wonder that the PTO issues so many patents, or that Fugitive Slave Act courts enslaved so many victims.

For those favor who prefer self-regulation over direct government intervention, BBBOnLine's Privacy seal offers a possible way forward, boasting members more trustworthy than average sites. BBB's tradition of self-regulation seems to help – creating institutional protection against lax review, and blunting short-run incentives to issue unearned certifications. BBB also benefits from its regional network of evaluators, whose proximity to applicants lets them better assess trustworthiness. Yet BBB's small member list and apparent delays make it an unlikely solution to the full problem of online safety. BBB's separate Reliability seal features more members – 27,000+ sites – but with correspondingly less scrutiny performed on each member. BBB's Reliability members turn out to be less trustworthy than its Privacy members – further suggesting that BBB's Privacy approach may not scale to certify dramatically more sites.

Meanwhile, my analysis offers practical lessons for regulators, users, and certification authorities. Regulators should hesitate to assume self-regulatory bodies will assess would-be members correctly; self-regulators' incentives diverge substantially from a reasonable social utility function. Users should also be wary of supposed certifications; sophisticated users ought to question why sites proclaim their certification and what those certifications really mean. Finally, certification authorities might rightly reconsider their practices – realizing that their credibility is on the line and that, in the long run, users will come to disbelieve certifications that are granted too easily.

6 Tables and Figures

*Throughout, *** denotes P-values less than 0.001, while ** denotes P-values less than 0.01 and * denotes P-values less than 0.05.*

6.1. Conditional Probability Analysis of Trust Certifications

These tables reflect analysis of top sites, as reported by a major US ISP based on its customers' web usage.

	TRUSTe-certified	Not certified
Trustworthy	874	515,309
Not Trustworthy	50	13,148

Table 1: Trustworthiness by TRUSTe Certification Status

Associated conditional probabilities:

P(trustworthy|certified) = 94.6% P(untrustworthy|certified) = 5.4%
P(trustworthy|uncertified) = 97.5% P(untrustworthy|uncertified) = 2.5%

	BBB-certified	Not certified
Trustworthy	284	515,898
Not Trustworthy	3	13,196

Table 2: Trustworthiness by BBB Privacy Certification Status

Associated conditional probabilities:

P(trustworthy|certified) = 99.0% P(untrustworthy|certified) = 1.0%
P(trustworthy|uncertified) = 97.0% P(untrustworthy|uncertified) = 3.0%

6.2. Regression Analysis of Trust Certifications

These tables reflect analysis of top sites, as reported by a major US ISP based on its customers' web usage.

Φ(Site Trustworthiness)	(1)	(2)	(3)
Constant	1.96***	1.89***	1.96***
	(0.003)	(0.005)	(0.011)
TRUSTe Certification	-0.356***	-0.302***	-0.276***
	(0.068)	(0.080)	(0.068)
Site Traffic Rank		1.30×10^{-7}***	1.30×10^{-7}***
		(6.24×10^{-9})	(6.24×10^{-9})
Site Type Dummies			Yes

Table 3: Probit of Site Trustworthiness on TRUSTe Certification and Site Characteristics

Φ(Site Trustworthiness)	(1)	(2)	(3)
Constant	1.96***	1.89***	1.96***
	(0.004)	(0.005)	(0.011)
BBB Privacy Certification	0.349	0.395	0.416
	(0.217)	(0.217)	(0.217)
Site Traffic Rank		1.32×10^{-7}***	1.31×10^{-7}***
		(6.25×10^{-9})	(6.25×10^{-9})
Site Type Dummies			Yes

Table 4: Probit of Site Trustworthiness on BBB Site Certification and Site Characteristics

Φ(Site Trustworthiness)	(1)	(2)
Constant	1.67***	1.67***
	(0.002)	(0.002)
TRUSTe Certification	-0.187*	
	(0.074)	
BBB Privacy Certification		-0.439
		(0.236)
Site Traffic Rank	9.40×10^{-8}***	9.52×10^{-8}***
	(1.00×10^{-8})	(1.00×10^{-8})
Site Type Dummies	Yes	Yes

Table 5: Probit of Site Trustworthiness on Site Certification and Site Characteristics, Among Complex Sites (with web forms and/or software downloads)

6.3. Historical Analysis of Trust Certifications

Date	Num. TRUSTe-Certified Sites	% Untrustworthy
January 1998	28	0.00%
July 1998	61	0.00%
January 1999	319	2.19%
May 1999	430	1.63%
January 2000	1467	1.77%
August 2000	1527	1.70%
January 2001	1550	2.26%
January 2002	1532	2.61%
January 2003	1208	2.24%
April 2004	1225	2.29%
July 2004	1331	2.70%
November 2004	1172	2.99%
February 2005	1263	2.93%
April 2005	1269	3.07%
January 2006	1554	3.41%

Table 6: Historical Analysis of Trustworthiness of TRUSTe-Certified Sites

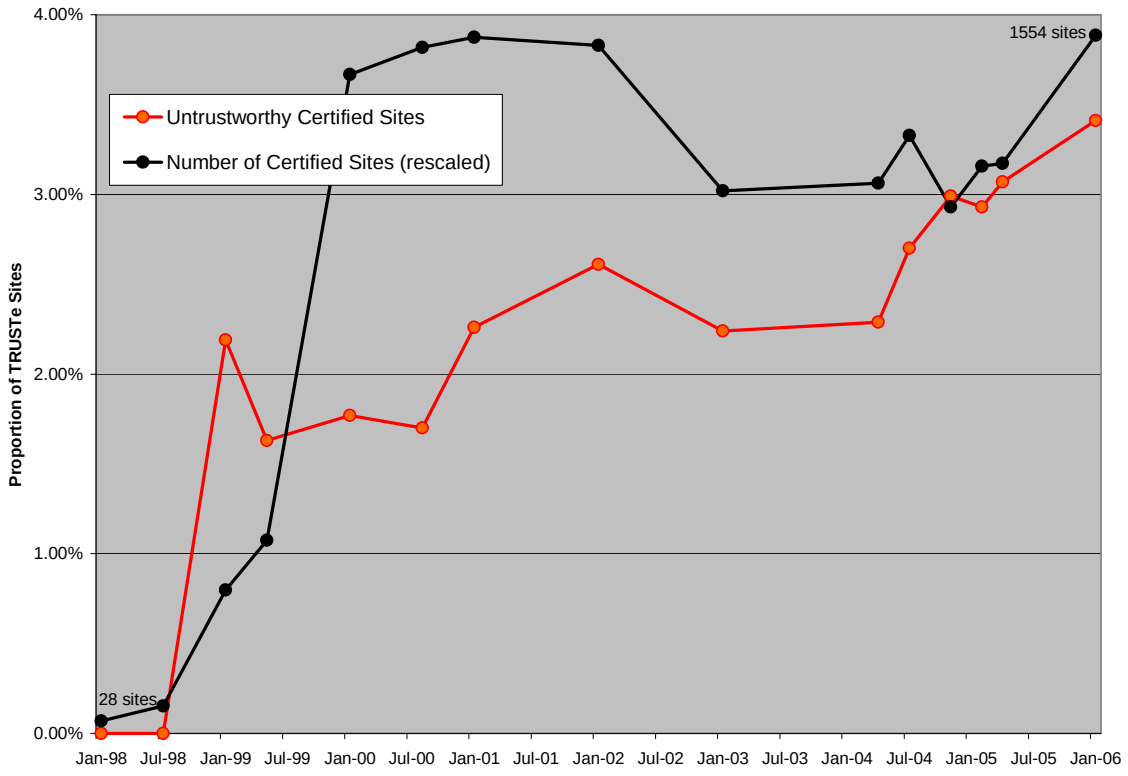

Figure 1: Historical Analysis of Trustworthiness of TRUSTe-Certified Sites

6.4. Untrustworthy Sites Certified by TRUSTe

The table below reports selected untrustworthy sites certified by TRUSTe, along with a general statement of the sites' respective practices. As of January 2006, TRUSTe listed all of these sites among its certified members.

Domain	Description
Direct-revenue.com	Direct Revenue makes advertising software known to become installed without consent. Tracks what web sites users visit, and shows pop-up ads. Historically, blocks many attempts at removal, automatically reinstalls itself, and deletes certain other programs from users' PCs. Faces litigation by the New York Attorney General, plus multiple consumer class actions.
Funwebproducts.com	This site, among other Ask.com toolbar distributors, installs a toolbar into users' web browsers when users agree to install smileys, screensavers, cursors, or other trinkets. Moves a user's Address Bar to the right side of the browser, such that typing an address into the standard top-left box performs a search rather than a direct navigation.
Maxmoolah.com	Offers users "free" gifts if they complete numerous sequential partner offers. Privacy policy allows sharing of user' email addresses and other information with third parties. In testing, providing an email address to Maxmoolah.com yielded a total of 485 distinct e-mails per week, from a wide variety of senders.
Webhancer.com	Makes online tracking software, sometimes installed without consent. Monitors what web sites users visit, and sends this information to Webhancer's servers.

6.5. Search Engines and Search Engine Advertising

These tables reflect analysis of a sample of 1,397 popular keywords obtained from industry sources.

	% Untrustworthy				
Which Result	Google	Yahoo	MSN	AOL	Ask
Top 1 Organic	2.73%	0.00%	2.03%	2.75%	3.23%
Top 3 Organic	2.93%	0.35%	2.24%	2.73%	3.24%
Top 10 Organic	2.74%	1.47%	2.56%	2.56%	2.94%
Top 50 Organic	3.04%	1.55%	2.46%	2.79%	3.12%
Top 1 Sponsored	4.44%	6.35%	6.17%	6.87%	7.99%
Top 3 Sponsored	5.33%	5.72%	6.16%	6.87%	7.99%
Top 10 Sponsored	5.89%	5.14%	6.37%	6.35%	8.31%
Top 50 Sponsored	5.93%	5.40%	6.01%	7.20%	8.20%

Table 7: Site Trustworthiness by Search Engine Placement Time and Position

Φ(Site Trustworthiness)	(1)	(2)	(3)	(4)
Constant	1.800***	1.844***	1.935***	1.888***
	(0.0164)	(0.025)	(0.010)	(0.013)
Organic Ranking	0.0153***	0.0039	-0.0059***	-0.0047***
Position	(0.0022)	(0.0025)	(0.0005)	(0.0005)
Search Engine Dummies		Yes		Yes
Result Restriction	Top 10	Top 10	All	All

Table 8: Probit of Site Trustworthiness on Organic Search Engine Ranking

7 Appendix: Reversibility of Conditionals in Bayes Rule Analysis, when Outcome and Signal are Both Binary

The body of the paper claims that, in the case in which both s and t are binary, $P(s|t)<P(s|\bar{t})$ if and only if $P(t|s)<P(t|\bar{s})$. This section provides the proof.

For s and t binary, there are four possible combinations of values of s and t. Let the values within the table below denote the respective probabilities, with a+b+c+d=1.

	s	$\bar{s}$
t	a	b
$\bar{t}$	c	d

The definition of conditional probability yields the following four identities:

$$P(s|t)=\frac{a}{a+b} \qquad P(s|\bar{t})=\frac{c}{c+d} \qquad P(t|s)=\frac{a}{a+c} \qquad P(t|\bar{s})=\frac{b}{b+d} \qquad (18), (19), (20), (21)$$

Suppose $P(s|t)<P(s|\bar{t})$. Substituting, then cross-multiplying and expanding:

$$\frac{a}{a+b} < \frac{c}{c+d} \qquad (22)$$

$$a(c+d) < c(a+b) \qquad (23)$$

$$ac + ad < ac + bc \qquad (24)$$

Subtracting ac from each side:

$$ad < bc \qquad (25)$$

Adding ab to each side:

$$ab + ad < ab + bc \qquad (26)$$

$$a(b+d) < b(a+c) \qquad (27)$$

$$\frac{a}{a+c} < \frac{b}{b+d} \qquad (28)$$

Substituting, using identities (20) and (21):

$$P(t|s) < P(t|\bar{s}) \qquad (29)$$

So $P(s|t)<P(s|\bar{t})$ ➔ $P(t|s) < P(t|\bar{s})$. But all steps in the analysis are reversible, which proves the converse and completes the proof.

8 References

Akerlof, George. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(4), pp. 488-500.

Animesh, Animesh, Vandana Ramachandran, and Siva Viswanathan. 2005. "Quality Uncertainty and Adverse Selection in Sponsored Search Markets." .NET Institute Working Papers: No. 05-27.

Baye, Michael, and John Morgan. 2003. "Red Queen Pricing Effects in E-Retail Markets."*SSRN Working Paper*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=655448 (accessed July 20, 2006).

Boutin, Paul. 2002. "Just How Trusty is Truste?" *Wired*. April 9, 2002. http://www.wired.com/news/exec/0,51624-0.html (accessed July 20, 2006).

Brin, Sergey, and Lawrence Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems*, 30(1-7), pp. 107-117.

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108, pp. 56-78.

Consumer Reports Web Watch. "A Matter of Trust: What Users Want From Web Sites." 2002. http://www.consumerwebwatch.org/pdfs/a-matter-of-trust.pdf (accessed July 20, 2006).

Darby, Michael and Edi Karny. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics*, 16(1), pp. 67-88.

Edelman, Benjamin, and Hannah Rosenbaum. 2006. "The Safety of Internet Search Engines." http://www.siteadvisor.com/studies/search_safety_may2006.html (accessed July 20, 2006).

Electronic Frontier Foundation. 1999. Letter to the FTC. Original at http://www.eff.org/pub/Privacy/Email_Internet_Web/19991020_req_to_prtc_com3.html, quoted in relevant part at http://yro.slashdot.org/article.pl?sid=99/11/05/1021214 (accessed July 20, 2006).

Finkelstein, Amy, and James Poterba. 2004. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market." *Journal of Political Economy*, 112(1), pp. 183-208.

FTC. 2002. "Re: Complaint Requesting Investigation of Various Internet Search Engine Companies for Paid Placement and Paid Inclusion Programs." http://www.keytlaw.com/FTC/Rules/paidplacement.htm (accessed July 20, 2006).

FTC. 1999. "Self-Regulation and Privacy Online." FTC. July 1999. http://www.ftc.gov/os/1999/07/privacy99.pdf (accessed July 20, 2006).

Gaudeul, Alexandre. 2004. "Internet Intermediaries' Editorial Content Quality." *WUSTL Industrial Organization, Economics Working Paper Archive*.

Genesove, David. 1993. "Adverse Selection in the Wholesale Used Car Market." *Journal of Political Economy*. 101(4), pp. 644-665.

Greenstadt, Rachel and Michael Smith. 2005. "Protecting Personal Information: Obstacles and Directions" in *Proceedings of the Fourth Annual Workshop on Economics and Information Security*. Cambridge, Massachusetts.

Hansell, Saul. 2001. "Paid Placement is Catching On in Web Searches." *New York Times*. June 4, 2001.

Jamal, Karim, Michael Maier, and Shyam Sunder. 2003. "Privacy in E-Commerce: Developing of Reporting Standards, Disclosure, and Assurance Services in an Unregulated Market." *Journal of Accounting Research*. 41(2), pp. 285-309.

Joachims, Thorsten, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. "Accurately Interpreting Clickthrough Data as Implicit Feedback," in *Proceedings of the Conference on Research and Development in Information Retrieval*.

LaRose, Robert, and Nora Rifon. 2002. "Your Privacy Is Assured – Of Being Invaded: Web Sites with and without Privacy Seals." http://www.msu.edu/~larose/es2003post.htm (accessed July 20, 2006).

Lizzeri, Alessandro. 1999. "Information Revelation and Certification Intermediaries." *The RAND Journal of Economics.* 30(2), pp. 214-231.

Marable, Leslie. 2003. "Consumer Reaction to Learning the Truth About How Search Engines Work." http://www.consumerwebwatch.org/pdfs/false-oracles.pdf (accessed July 20, 2006).

Ravicher, Daniel. 2005. "Statement Before the Subcommittee on Courts, the Internet, and Intellectual Property." http://www.pubpat.org/Ravicher%20Statement%20on%20Patent%20Act%20of%202005.pdf (accessed July 20, 2006).

Singel, Ryan. 2006. "'Free IPod' Takes Privacy Toll." *Wired*. March 16, 2006. http://www.wired.com/news/technology/0,70420-0.html (accessed July 20, 2006).

Tang, Zhulei, Yu Hu, and Michael Smith. 2005. "Protecting Online Privacy: Self-Regulation, Mandatory Standards, or Caveat Emptor" in *Proceedings of the Fourth Annual Workshop on Economics and Information Security*. Cambridge, Massachusetts.

"TRUSTe Case Study - Realty Tracker." TRUSTe. http://www.truste.org/pdf/Realty_Tracker_Case_Study.pdf (accessed July 20, 2006).

"TRUSTe Fact Sheet." http://www.truste.org/about/fact_sheet.php (accessed July 20, 2006).

"TRUSTe Program Requirements." http://www.truste.org/requirements.php (accessed July 20, 2006).

"TRUSTe Watchdog Reports." https://www.truste.org/consumers/watchdog_reports.php (accessed July 20, 2006).

"US Patent and Trademark Office FY 2005 Fee Schedule." http://www.uspto.gov/web/offices/ac/qs/ope/fee2004dec08.htm (accessed July 20, 2006).

Villeneuve, Bertrand. 2003. “Mandatory Pensions and the Intensity of Adverse Selection in Life Insurance Markets.” *The Journal of Risk and Insurance*. 70(3), pp. 527-548.

Webmasterbrain Search Engine Experiment. 2006. http://www.webmasterbrain.com/seo-tools/seo-experiments/the-search-engine-experiment/test-results/ (accessed July 20, 2006).