Adverse Selection in Online "Trust" Certifications

Benjamin Edelman
Harvard Business School
1 Soldiers Field Rd.
Boston, MA 02163
bedelman@hbs.edu

ABSTRACT

Widely-used online "trust" authorities issue certifications without substantial verification of recipients' actual trustworthiness. This lax approach gives rise to adverse selection: The sites that seek and obtain trust certifications are actually less trustworthy than others. Using a new dataset on web site safety, I demonstrate that sites certified by the best-known authority, TRUSTe, are more than twice as likely to be untrustworthy as uncertified sites. This difference remains statistically and economically significant when restricted to "complex" commercial sites. In contrast, competing certification system BBBOnline imposes somewhat stricter requirements and appears to provide a certification of positive, albeit limited, value.

Categories and Subject Descriptors

K.5.2 [**Legal Aspects of Computing**]: Government Issues – *regulation.*

General Terms

Economics, Legal Aspects, Security

Keywords

Adverse selection, certification, reputation, trust, Internet

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 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 "lemons" [1], others have worried about similar problems in markets with hidden types – like bad drivers wanting more car insurance than good drivers [6], and healthy people disproportionately buying annuities [8].

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 research manages to address these problems. For example, [6] focuses on novice drivers, who have less private information about their own type (since they have not yet started to drive), letting economists 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 might affect other drivers.

This paper applies an adverse selection model to a new market: web sites and their associated "trust"-type certifications. With a new data source, I analyze characteristics generally unobservable both to consumers and to trust 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. These 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 low-quality sites disproportionately seek and receive certification, substantially reducing 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. Exploring the broader Internet offers the potential for a better match, but with important risks: Untrustworthy sites might send users spam (if users register or otherwise provide 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. A safe-looking site 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 their integrity. But as usual in adverse selection, bad sites pretend they're good.

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. For example, [16]

ICEC '09, August 12-15, 2009, Taipei, Taiwan

presents a model of enforcement of online privacy breaches, finding mandatory government standards appropriate only for the most serious harms.

At the other extreme, users might be left entirely on their own. In complete *caveat emptor*, no regulator, computer maker, or IT department helps cure a user's problems. In some respects, *caveat emptor* is a reasonable description of the current state of affairs. (IT departments cannot protect users from getting ripped off, and even computer experts often feel powerless to stop spam.) But unaccountability carries substantial costs – leading users to take excessive precautions, and 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 of fraud and abuse.

Finally, there exists a middle approach between the extremes of government regulation and *caveat emptor*: 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 good firms receive positive evaluations, and where consumers use only sites with positive ratings. [16] suggests that rating organizations are appropriate for a broad class of online interactions.

1.2. Trust Authorities

Most prominent among non-governmental rating organizations are so-called "trust" certification authorities. These organizations set out specific criteria for membership, often focusing on privacy or on online safety more generally. The organizations reward their members by offering seals to be placed on recipients' web sites, typically on registration forms and checkout pages. To date, the best-known trust authorities are TRUSTe and BBBonline.

In principle, trust 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, critics question the effectiveness of certain trust authorities. [13] offers a stinging critique: Trust authorities have granted multiple certifications to firms under investigation by the FTC for privacy policy violations; trust authorities have declined to pursue complaints against major companies whose privacy breaches were found to be "inadvertent"; and in one case a trust authority even failed to abide by *its own* privacy policy. [15] raises similar concerns: In a 2004 investigation after user complaints, TRUSTe gave Gratis Internet a clean bill of health. Yet subsequent New York Attorney General litigation uncovered Gratis' exceptionally far-reaching privacy policy violations – selling 7.2 million users' names, email addresses, street addresses, and phone numbers, despite a privacy policy exactly to the contrary.

As a threshold matter, trust authorities' substantive standards often seem to duplicate existing duties or practices. Consider the obligations in TRUSTe's Program Requirements [19]. The first listed rule, requiring an email unsubscribe function, duplicates Sec.5.(a)(4)(A) of the federal CAN-SPAM Act. Similarly, credit card network rules exactly overlap with TRUSTe's requirement of SSL encryption (or similar technology) to protect sensitive credit card numbers. [5] reports that TRUSTe initially lacked any substantive requirements whatsoever (requiring only the *presence* of a privacy policy). Low standards match the predictions of [14], 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, [18] reports only two certifications revoked in TRUSTe's ten-year history. TRUSTe's small staff has little apparent ability to detect infractions. Instead, TRUSTe's posted procedures emphasize user complaints and sites' self-certifications. When violations have been uncovered, the proof has come from outside complaints, not from TRUSTe itself.

TRUSTe's "Watchdog Reports" [20] also indicate a lack of focus on enforcement. TRUSTe's postings reveal that 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.[1]

Finally, trust authorities are paid by the same companies they certify; in the language of [11], trust authorities are "captured." With this revenue model, authorities have little short-run incentive to seek higher standards: Any such pressure would discourage renewals and future applications – reducing revenues.

Even the creators of trust authorities report disappointment in their development. TRUSTe co-founder Esther Dyson called TRUSTe "a little too corporate," and said TRUSTe lacks the "moral courage" to criticize violations [5]. Similarly, the Electronic Frontier Foundation, another TRUSTe co-founder, told the FTC that "it is time to move away from a strict self-regulation approach" [7].

Table 1 reports selected untrustworthy sites certified by TRUSTe, along with a general statement of the sites' respective practices. As of January 2006, TRUSTe listed all these sites among its certified members.

Facing allegations of low substantive standards, lax enforcement, and ethical compromise, it is unclear what direct benefits site certifications offer to consumers. But at least some consumers seem to regard certification as a significant positive signal. For example, in recruiting web sites to get certified, TRUSTe offers an endorsement from certificate recipient Realty Tracker, which says TRUSTe "convey[ed] trust" and "built confidence" with site visitors, yielding "an increase in registrations." See TRUSTE's Realty Tracker Case Study [17].

Moreover, firms are well-equipped to evaluate claimed benefits to certification: Firms can randomly include or omit a seal, thereby measuring whether a seal increases registrations and sales. Indeed, year after year, hundreds of firms seek and renew TRUSTe certification – suggesting that firms find certification valuable. Furthermore, in the related context of comparison shopping sites, [4] empirically confirms the benefits of certification: Merchants with seals can charge a price premium without losing customers.

Even well-known web sites tout their safety certifications. For example, the Microsoft's Online Privacy Policy index features the

[1] For example, TRUSTe failed to update its Watchdog Reports list [20] between June 2004 and spring 2006, an omission corrected only after circulation of a draft of this article. Even in 2009, Watchdog Reports suffer broken links, missing reports, and contradictory document titles.

Table 1: Selected Untrustworthy Sites Certified by TRUSTe

Domain	Description
Direct-revenue.com	Makes advertising software known to become installed without consent. Tracks what web sites users visit, and shows pop-up ads. Blocks many attempts at removal, and automatically reinstalls itself, including by disguising itself as a printer driver. Deletes competing advertising software from users' PCs. Faced litigation by the FTC and the New York Attorney General, plus multiple consumer class actions.
Funwebproducts.com	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. Shows seven sponsored links above the first organic result – overwhelming users with ads.
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 observed becoming installed without user consent. Monitors what web sites users visit, and sends this information to Webhancer's servers.

TRUSTe name and logo adjacent to the page's title and Microsoft logo. eBay presents its TRUSTe certification on its main registration page (a necessary step for all new users joining eBay).

Whatever the actual merits of certification authorities as arbiters of trust, some government authorities seem to regard these organizations as an appropriate step forward. For example, the FTC's 1999 "Self-Regulation and Privacy Online" [9] endorsed private-sector trust authorities as an alternative to comprehensive regulation of online privacy and safety.

The FTC's 1999 recommendation [9] specifically cites two well-known certification systems: TRUSTe's Web Privacy Seal and BBBOnLine's Privacy Seal. My subsequent analysis focuses on these authorities due to their prevalence, their relatively large member lists, and the public availability of their member lists.

2. THEORY OF ADVERSE SELECTION IN TRUST AUTHORITIES

Suppose certain trust authorities issue certifications of trustworthiness without rigorous assessment of recipients' true trustworthiness. Certifications seek to signal consumers that the certified firms are in fact highly likely to be trustworthy. But if untrustworthy firms can easily get certified, the signal drops in value: Seeing a certified firm, a consumer would rightly worry that the firm is not truly trustworthy.

To provide a positive signal, a certification must increase a rational consumer's assessed probability that a site is trustworthy. Suppose a rational consumer has a prior belief $P(t)$ that a given site is trustworthy. The consumer then receives a signal ("*s*") of trustworthiness ("*t*"). The consumer updates according to Bayes Rule:

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

Expanding the denominator using the Law of Total Probability:

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

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:

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

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

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

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

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

Equation 6 offers an intuitive result: For a certification to increase a consumer's assessment of the probability that a certified site is safe, the certification must be given to trustworthy sites more often than it is given to untrustworthy sites.

2.1. Testing for Adverse Selection at Trust Authorities

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.) Then an informative certification requires:

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

Adverse selection 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 continues the approach in the adverse selection literature. Consider [8], finding that annuitants live longer than non-annuitants – a negative relationship between claimed type (annuity purchase) and outcome (lifetime). [6] uses 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). [10] extends these correlations with the equilibrium assumption that average price in a given market must reflect average quality in that market. [10] 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. [21] 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, [21]'s measure matches (7).

Others studying online trust authorities have also worried of adverse selection. For example, [13] finds that privacy policies at certified sites allow more invasive data collection than policies at uncertified sites. But where [13] hand-scores 200 sites, I use automation to evaluate hundreds of thousands of sites, and I consider axes of trustworthiness other than privacy policy loopholes. In addition to investigating the quality of certified sites, [12] specifically considers certifiers lowering their standards to attract more sites. But [12] studies only 34 well-known sites certified as of 2001 – limiting the generality of their findings. In contrast, my analysis includes more recent data and far more sites.

2.2. 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 or 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 predict that worthless trust authorities will collapse. But as an empirical matter, trust authorities have existed for some time and show no sign of disappearing. Although inconsistent with a world of fully-informed consumers, the persistence of trust authorities makes sense under reasonable assumptions. For example, suppose some users are slow learners – drawing inference about certification based on the quality of sites certified *in prior periods*. Then an initial batch of high-quality certified sites would effectively subsidize future certifications.[2] Alternatively, if some users are naïve (mistakenly trusting certifications that are actually worthless), certification would be profitable if naïve users are sufficiently widespread relative to the cost of certification. In extensions (available on request), I have sketched a model of these effects.

The slow-learner model offers an empirical prediction: The average quality of certified sites should decrease over time. Suppose a trust authority happened to start with members that truly are trustworthy, producing a favorable initial reputation with users. (Consider the alternative: If a certifier began by certifying untrustworthy sites, it would have little hope of building a positive reputation with users.) In the face of slow learning, that favorable reputation would take some time to dissipate. In the interim, untrustworthy firms can profit from certification. The resulting hypothesis:

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

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 consumers knew sites' actual trustworthiness, there would be no hidden types and no opportunity for adverse selection. But new data collection systems allow analysis of sites' actual behaviors 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 to visit web sites and attempt to measure their safety. SiteAdvisor's automation uncovers site characteristics that are otherwise difficult for users to discern. For example, one SiteAdvisor system 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 system 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 systems check for excessive pop-ups, security exploits, scams, links to other bad sites, and more.

SiteAdvisor's measurements are imperfectly correlated with trust authorities' stated rules. For example, a site could send its registrants hundreds of emails per week, yet still receive a TRUSTe certification. Nonetheless, SiteAdvisor's approach is highly correlated with the behaviors *users* deem objectionable. Without understanding the subtleties of trust authorities rules, users seem to regard certifications as general indicators of good business practices. Any site failing SiteAdvisor's tests is a site likely to present substantial concern to typical users. I therefore consider SiteAdvisor data a good proxy for sites' true trustworthiness – for the outcomes users actually care about, even when those outcomes differ from trust authorities' official requirements.

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 the Internet Archive (archive.org).

Table 2 presents SiteAdvisor's policies and compares these policies with the requirements of TRUSTe and BBBOnLine.

Equations 6 and 7 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 analyses include specifications with various controls, including a crude measure of site commerciality (.COM versus .ORG versus other

[2] This chronology seems to match the history of TRUSTe, which was founded by a set of trustworthy companies to serve their regulatory goals. In particular, those companies preferred private-sector certification as an alternative to threatened FTC regulation of online privacy practices. Only later did TRUSTe begin to certify sites with more controversial practices.

Table 2: Comparison of Selected TRUSTe, BBB Privacy, and SiteAdvisor Policies

Characteristic	TRUSTe Web Site Privacy Seal	BBBOnLine Privacy	SiteAdvisor
Software Downloads	No restriction	No restriction	Rates a site unfavorably if the site offers programs that are, or that bundle, "adware" or "spyware."
Email	No restriction	No restriction	Rates a site unfavorably if the site sends a large number of messages or does not honor requests to unsubscribe.
Web Links	No restriction	No restriction	Rates a site unfavorably if the site links to other sites rated unfavorably.
BBB Membership	No requirement	Required, with a satisfactory record of handling complaints	No requirement
Privacy policy	Compulsory. Site must self-certify its practices. Must disclose information collection and use.	Compulsory. Three dozen rules about privacy policy provisions and site practices.	No requirement
Dispute resolution with consumers	Site must accept consumer complaints and participate in TRUSTe "Watchdog" process.	Site must participate in the BBBOnline Dispute Resolution Process.	n/a
Application or certification fee	Yes, up to $7,999 per year	Yes, up to $7,000 per year	No

extensions) as well as popularity (as measured by a large US ISP).[3] [4] I analyze approximately half a million sites – the web's top sites according to the ISP that provided popularity data.

4. RESULTS AND DISCUSSION

I begin by testing Hypothesis 1 using the method in Equation 7. Comparing the trustworthiness of certified and uncertified sites (within the top web sites reported by my ISP data source), I obtain the results in Tables 3 and 4 for TRUSTe and BBBOnLine (privacy seal program), respectively.

Computing conditional probabilities from Table 3 yields the pie charts in Figure 1. 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 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 an initial confirmation of the adverse selection result posited in Section 2.

The TRUSTe adverse selection result in Table 3 holds in a regression framework that controls for additional variables. In Table 5, column 1 gives 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 significantly negative. That is, on the margin, TRUSTe certification remains associated with a reduction in the probability that a given site is actually trustworthy.

(Throughout all regressions, *** denotes P-values less than 0.001, ** denotes P-values less than 0.01, and * denotes P-values less than 0.05.)

In Table 6, Column 1, I test the suggestion that TRUSTe's negative association with trustworthiness is spurious. Some might

[3] Popularity data comes in rank form, so larger values imply lesser traffic.
[4] By agreement with the ISP, I cannot identify it by name.

Table 3: Trustworthiness by TRUSTe Certification Status

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

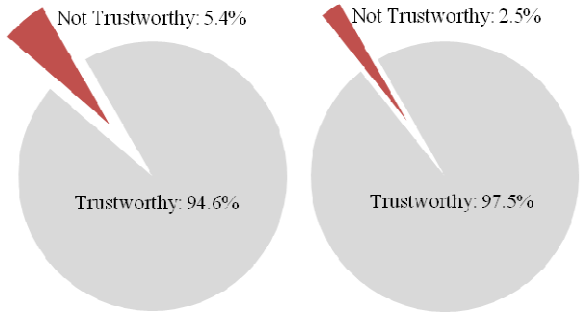

Figure 1: Comparing TRUSTe-Certified and Uncertified Sites

Table 4: Trustworthiness by BBB Privacy Certification Status

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

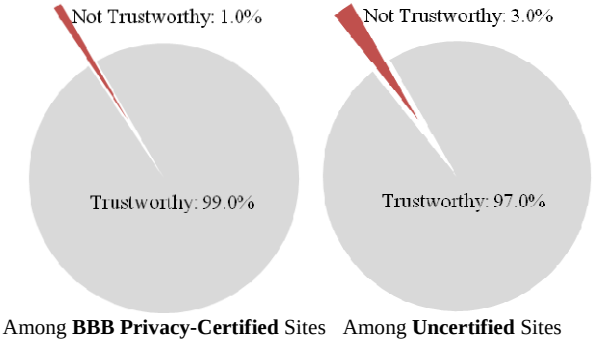

Figure 2: Comparing BBB-Certified and Uncertified Sites

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, 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. (See Figure 2.) This result holds in a regression framework (Table 7) including when controlling for site complexity (Table 6, Column 2). Industry sources attribute BBB's success to BBB's detailed evaluation of applicants. For example, BBBOnLine requires that an applicant be a member of a local BBB chapter (which adds its own requirements), whereas TRUSTe tends to rely primarily on applicants' self-assessments. Though BBB's approach offers important benefits, BBB apparently faces substantial difficulties including 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 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 over time, certification comes to include less trustworthy sites. Using historical TRUSTe membership data preserved by Archive.org, Table 8 and Figure 3 confirm that hypothesis as to TRUSTe. Note that I do not observe sites' prior practices. Instead, I use current trustworthiness as a proxy for historic behavior – effectively assuming that trustworthy sites stay trustworthy, and vice versa.

While I focus on online trust authorities certifying web site practices, other certifications seek to verify different aspects of behavior. For example, [2] and [3] examine the Certification Authorities (CAs) that issue electronic signatures for use in public key infrastructure – evaluating whether CAs issue certifications not properly justified under governing policy, and assessing the incentives that influence CAs' operations. My finding of disproportionate unwarranted certifications tracks [3]'s prognosis of substantial quality uncertainty in public key certificates.

5. POLICY IMPLICATIONS

The framework of [1] offers suggestions to address problems of information asymmetry, but these responses appear to be inapt or unsuccessful in this context. To [1]'s suggestion of *guarantees* comes the problem that, at least as currently structured, online trust authorities are in no position to offer a meaningful guarantee of certified sites' practices. Indeed, TRUSTe's Terms and Conditions specifically prohibit a user from relying on TRUSTe's certifications, and BBBOnLine's Terms of Use disclaim liability for listings. Furthermore, while a guarantee would certainly benefit users, a heightened level of verification would present certification authorities with higher costs in certification, substantial liability in case of breach by a certified site, or both. So certification authorities are unlikely to offer guarantees voluntarily.

[1] next suggests the use of *brand names* to remedy information asymmetries. To some extent "TRUSTe" and "BBB" present useful brand names that consumers can recognize and, in due course, credit or discount as appropriate. But at least in the context of TRUSTe, the value of the brand – through historical placement on well-known trusted sites like Microsoft and eBay – is then diluted by less trustworthy sites that later received and promoted TRUSTe certification. [1] presupposes that a brand name will elect to protect and preserve its reputation, but TRUSTe's certifications indicate otherwise.

Table 6: Probit of Site Trustworthiness on TRUSTe Certification and Site Characteristics

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

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 7: Probit of Site Trustworthiness on BBB Privacy 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

Finally, [1] notes the possibility of *licensing*. Certainly government oversight of online trust authorities could rein in certifications too easily granted. Conceivably some middle ground could preserve a portion of the decentralization, flexibility, and cost-saving benefits of self-regulation while adding additional control through government oversight. But to those who founded online trust authorities in the spirit of self-regulation, detailed government oversight is likely to be viewed as highly undesirable.

Seeing an apparent failure by at least some well-known trust authorities, the FTC might reasonably revisit its 1999 decision to favor certification-based self-regulation in lieu of substantive FTC oversight. But if regulators sought to retain the basic approach of self-regulatory certifications, they have ample tools to improve certification outcomes.

For one, trust authorities might appropriately face liability for their most egregious misclassifications. Within the framework of [1], this approach essentially comprises a compulsory regulation-mandated guarantee – but if market forces do not inspire trust authorities to provide such guarantees, regulation could assist. At present, 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: Certifying a site yields a fee, while no fee results from denying a certification.) But if a trust authority falls short of a reasonable standard of care, it might properly face liability on a negligence theory. (Analogous liability has been sought, with mixed results, as to erroneous certifications by rating agencies and auditing firms in the financial sector.)

In a narrower change, regulators could require trust authorities to publish consumers' complaints about certified sites, or regulators could receive and tabulate such complaints. (Analogously, the Department of Transportation tracks and summarizes consumer complaints about airlines.) The resulting transparency would help assure that trust authorities appropriately investigate problems brought to their attention.

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. Indeed, after the completion of a draft of this article, BBB closed the Privacy program to new applicants – seemingly a response to the limited number of sites that had chosen to participate in that program. BBB's separate Reliability seal features far more members (some fifty thousand), but with correspondingly less scrutiny on each member. In separate analysis, I found that BBB's Reliability members are somewhat less trustworthy than its Privacy members – providing further reason to doubt whether the BBB's Privacy approach can scale to certify dramatically more sites.

My analysis offers practical lessons for regulators, users, and trust authorities. Regulators should not assume self-regulatory bodies will assess would-be members correctly, for self-regulation incentives diverge substantially from social optima. Users should also be wary of supposed certifications – questioning what certifications really mean and why sites boast of certification. Finally, trust authorities might rightly reconsider their practices – realizing that, in the long run, users will come to disbelieve certifications that are granted too easily.

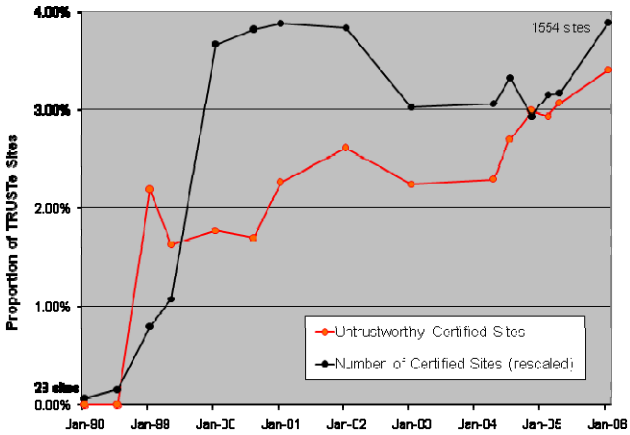

Figure 3: Historical Analysis of Trustworthiness of TRUSTe-Certified Sites

Table 8: Historical Trustworthiness of TRUSTe-Certified Sites

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%

6. REFERENCES

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

[2] Backhouse, James, et al. 2005. "A Question of Trust – An Economic Perspective on Quality Standards in the Certification Services Market." Communications of the ACM.

[3] Backhouse, James, et al. 2005. "Spotting Lemons in the PKI Market: Engendering Trust by Signaling Quality." *Electronic Commerce and the Digital Economy*, ed. Michael Shaw. M.E. Sharpe, Inc., Advances in Management Information Systems Series Editor, Vladimir Zwass.

[4] Baye, Michael, and John Morgan. 2003. "Red Queen Pricing Effects in E-Retail Markets." SSRN Working Paper 655448 (accessed May 8, 2009).

[5] Boutin, Paul. 2002. "Just How Trusty is Truste?" *Wired*. April 9, 2002. http://www.wired.com/news/exec/0,51624-0.html (accessed May 8, 2009).

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

[7] Electronic Frontier Foundation. 1999. Letter to the FTC. Ori ginal 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 May 8, 2009).

[8] 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.

[9] FTC. 1999. "Self-Regulation and Privacy Online." FTC. July 1999. http://www.ftc.gov/os/1999/07/privacy99.pdf (accessed May 8, 2009).

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

[11] 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.

[12] 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.

[13] 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 May 8, 2009).

[14] Microsoft. "Online Privacy Notice Highlights." http://privacy.microsoft.com/ (accessed May 8, 2009).

[15] Singel, Ryan. 2006. "'Free IPod' Takes Privacy Toll." *Wired*. March 16, 2006.

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

[17] "TRUSTe Case Study - Realty Tracker." TRUSTe. http://www.truste.org/pdf/Realty_Tracker_Case_Study.pdf (accessed May 8, 2009).

[18] "TRUSTe Fact Sheet." http://www.truste.org/about/fact_sheet.php (accessed May 8, 2009).

[19] "TRUSTe Program Requirements." http://www.truste.org/requirements.php (accessed May 8, 2009).

[20] "TRUSTe Watchdog Reports." https://www.truste.org/consumers/watchdog_reports.php (accessed May 8, 2009).

[21] 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.

7. APPENDIX: REVERSIBILITY OF CONDITIONALS IN BAYES RULE ANALYSIS, WHEN SIGNAL AND OUTCOME 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	$-s$
t	a	b
$-t$	c	d

The definition of conditional probability yields the following identities:

$$P(s|t) = \frac{a}{a+b} \qquad P(s|\text{-}t) = \frac{c}{c+d} \qquad (8), (9)$$

$$P(t|s) = \frac{a}{a+c} \qquad P(t|\text{-}s) = \frac{b}{b+d} \qquad (10), (11)$$

Suppose $P(s|t)<P(s|\bar{t})$. Substituting from (8) and (9), then cross-multiplying and expanding:

$$\frac{a}{a+b} < \frac{c}{c+d} \qquad (12)$$

$$ac + ad < ac + bc \qquad (13)$$

Subtracting ac from each side, adding ab to each side, and regrouping:

$$ab + ad < ab + bc \qquad (14)$$

$$\frac{a}{a+c} < \frac{b}{b+d} \qquad (15)$$

Substituting, using (10) and (11):

$$P(t|s) > P(t|\bar{s}) \qquad (16)$$

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