Does Anyone Read the Fine Print? A Test of the Informed Minority Hypothesis Using Clickstream Data

Yannis Bakos,[*] Florencia Marotta-Wurgler,[**] and David R. Trossen[***]

February 16, 2009

ABSTRACT

The cornerstone of the law and economics approach to standard form contracts is the "informed minority" hypothesis: In competitive markets, a minority of term-conscious buyers is enough to discipline sellers from offering unfavorable terms. Although the informed minority argument is widely invoked to limit intervention in consumer transactions, there has been little direct investigation of its validity. This paper follows the Internet browsing behavior of 50,373 households with respect to 93 online software companies to study the extent to which potential buyers access the relevant standard form contract, the end user license agreement. Under a range of definitions of potential buyers, we find that between 99% and 99.9% of potential buyers choose not to access the available license agreement. This casts doubt on the validity of the informed minority argument in a market where it has been invoked by both theorists and courts. The very small probability that a license agreement is accessed does increase, to an extent, when the license is featured more prominently on the seller's website.

[*] Stern School of Business, New York University.
[**] New York University School of Law.
[***] Boalt Law School, University of California at Berkeley.
This research was supported with a grant from the NET Institute (www.NETinst.org). The authors are grateful to Ronald Mann, Jeffrey Wurgler, and the participants in the 2008 Law and Entrepreneurship Conference, the Workshop of Information Systems Economics (WISE), and Dennis E. Hermreck, Daniel Priest, Robert Taylor, Michael Tonkinson and Joel Willcher for excellent research assistance.

1. Introduction

The enforceability of standard form contracts has long been debated by legal academics and policy makers. One of the most prominent questions has been whether the content or disclosure of standard form contracts in consumer transactions should be regulated. The rise of Internet commerce has only heightened the importance of this inquiry, as consumers navigate through myriad "browsewrap" and "clickwrap" contracts. Advocates of regulation argue that form contracts are generally unfair and used by sellers to impose self-serving terms on consumers. Those favoring freedom of contract contend that regulation is unnecessary because sellers will attempt to attract potential buyers by offering standard form contracts that reflect buyers' preferences. At the heart of this dispute are buyers' attitudes towards fine print and whether they participate in market mechanisms to ensure economically efficient terms.

It is safe to assume that the majority of consumers typically does not read fine print. Standard form contracts are too long, hard to understand, or seemingly unimportant to take the time to read them and give meaningful assent. The feared result is the well-known market failure due to imperfect information: when most buyers do not factor contract terms into their purchase decisions, sellers will lack incentives to provide more than minimal protection. Concern for consumer welfare has resulted in numerous academic articles as well as laws and initiatives to regulate these markets. For example, in addition to existing contract law doctrines to protect buyers from abusive terms, such as unconscionability and unfair surprise, several state consumer laws prohibit the use of forum selection clauses and disclaimers of implied warranties in consumer contracts.[2] On the federal front, laws such as the Truth in Lending Act and the Magnuson-Moss Warranty Act seek to decrease reading and search costs by requiring standardized disclosure of mandated terms.[3] More recently, there has been heated debate on whether online contracts such as "Terms of Use," privacy policies, and software license agreements should be enforceable or subject to mandatory disclosure rules or contain mandatory provisions.[4]

---

[2] See, e.g., notes 33, 34, infra.

[3] See 15 U.S.C. §§2301-2312 (1976); 15 U.S.C. 1601.

[4] See, e.g., Mark Lemley, Terms of Use, Minn. L. Rev. 91 (2006) (arguing that browsewraps should be enforced only in cases involving sophisticated commercial parties that are repeat players); HUGH COLLINS,

Defenders of freedom of contract have generally rejected intervention by relying on reputational constraints and on the "informed minority" argument. In a seminal article, Alan Schwartz and Louis Wilde argue that sellers won't necessarily offer one-sided terms even when the majority of buyers don't read standard form contracts.[5] In their model, non-reading buyers benefit from an "informed minority" whose willingness to pay for the product is sufficiently sensitive to the quality of the standard terms. When all buyers have the same taste for quality and sellers are unable to discriminate between reading and non-reading buyers, sellers will offer the terms preferred by all buyers. Scholars have used this argument extensively to resist regulation, arguing that blanket measures may introduce their own inefficiencies.[6] For example, imposing a uniform standard would prevent sellers from using contracts to signal their quality or cater to heterogeneous buyer tastes. These scholars advocate rules only to facilitate search by those consumers aspiring to join the informed minority.

Although the informed minority argument is pervasive throughout contract scholarship, law, and policy initiatives advocating a free market for standard form terms, there has been practically no systematic empirical analysis of its validity. This paper addresses this void by examining the extent to which buyers actually do read online standard form contracts and what factors influence the probability of readership.

Specifically, we examine the extent to which potential buyers of software read End User License Agreements (EULAs). EULAs are contracts that govern the use of software products. For a sample of software companies who offer their products online, we use potential buyers' clickstream information to study the extent to which they access the EULA. Our clickstream data set tracks the detailed browsing behavior of thousands of Internet users over a period of time. For each user in the panel, we observe the exact sequence of web pages (urls) accessed in a particular visit to a site and the time spent on each page. The data set also includes detailed demographic characteristics about the head of household, such as age, gender, income, and geographical

---

REGULATING CONTRACTS 279–86 (1999) (arguing that mandatory terms would reduce the contractual one-sidedness caused by market imperfections).

[5] Alan Schwartz & Louis Wilde, Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis, 127 U. Pa. L. Rev. 630-82 (1979).

[6] See, e.g., Douglas Baird, The Boilerplate Puzzle, 104 Mich. L. Rev. 933, 933 (2006); Clayon Gillette, Pre-Approved Contracts for Internet Commerce, 42 Hous. L. Rev. 975 (2005).

location. We track the visits of 49,956 households to 93 software companies in the month of January 2007. For each of these companies, we collect all URLs that correspond to EULAs. We then identify non-casual browsers and study their shopping behavior in each of the 93 companies to measure the fraction that voluntarily accesses the contract page.

Our main finding is that regardless of how strictly we define a "shopper" (or "potential buyer"), only about one out of one thousand shoppers accesses a product's EULA. Such a small number of contract readers casts doubt on the existence of an informed minority that aggressively polices against one-sided terms, at least in the context of software sold online. Still, we test whether a fraction of 0.1% of informed shoppers is sufficiently large to induce sellers to offer "good" terms in the software market. We estimate the marginal cost of providing one pro-buyer term, Maintenance and Support, and find that sellers would find it more cost-effective to lose all informed buyers (assuming each would decline to buy if the given term isn't offered) than to offer this one term, a result that would likely persist for a fraction of informed buyers 1-2 orders of magnitude higher than 0.1%. An informed minority of 0.1% is also orders of magnitude smaller than the required informed minority size in the theoretical examples of Schwartz and Wilde.

We then focus on those factors that are related with the probability that a EULA will be accessed. This probability, although again very small, increases as contracts are made more prominently available on the sellers' websites. This finding is robust to several definitions of shoppers and "shopping visits" and does support the general intuition that lower search costs do increase the probability that consumers will become informed. That said, even contracts that are prominently featured on seller's websites (only one "click" away) are accessed by very few shoppers. Apparently, search costs are still too high for the creation of an informed minority, or else shoppers rationally or irrationally ignore contract terms in their purchase decisions. Regardless of the interpretation, this finding presents interesting questions for the regulation of disclosure. We also find that shoppers are more likely to access the EULAs of smaller companies or companies that offer somewhat suspicious products such as freeware. Shoppers (the few that care) might be rationally deciding to ignore the EULAs of larger, more established companies, relying instead on company reputation or familiarity.

The paper is organized as follows. Section 2 offers academic and legal background on the informed minority hypothesis. Section 3 explains our methodology in the context of a simple model. Section 4 presents data. Section 5 discusses the results and Section 6 concludes.

2. The Informed Minority Hypothesis: Academic and Legal Background

In a typical standard form contract scenario, a buyer purchases a good or service and is presented with a pre-printed form contract with terms pertaining to dispute resolution, remedies for product failure, and warranties, among others, with little opportunity to negotiate over the terms. Despite the benefits associated with standard form use, such as reduction in drafting and negotiation costs, academics and policy makers have expressed concerns about their fairness.

The concern that standard form contracts will be biased towards drafters stems from the common belief that because buyers do not read or understand the contract terms, sellers will impose unfair and one-sided terms. Many academics have shared this view for decades. For instance, David Slawson has argued that standard form contracts are not democratic because buyers will most certainly not read their terms and thus cannot be understood to have agreed to them.[7] He concludes that "[f]or the very reason that these terms are imposed rather than agreed upon, they are almost universally unfair." Similarly, Todd Rakoff has long held that standard form consumer contracts are likely one-sided and should have a presumption of invalidity. He posits that since it is unlikely that consumers will read, understand, and shop for products with favorable terms, sellers will necessarily take advantage of consumer ignorance and include one-sided terms in their standard form agreements.[8] Failure to read has been explained as a rational response, given the low probability that the contingencies specified in the contract will

[7] W. David Slawson, Standard Form Contracts and the Democratic Control of Lawmaking Power, 84 Harv. L. Rev 529, 530 (1975).

[8] Todd D. Rakoff, Contracts of Adhesion: An Essay of Reconstruction, 96 Harv. L. Rev. 1173-1284 (1983), (stating that "the ideal adherent who would read, understand, and compare several forms is unheard of in the legal literature, and, I warrant in life as well.") at 1226.

materialize and consumers' inability to alter the terms.[9] More recently, scholars have provided behavioral accounts explaining the ways in which consumers' biases might prevent them from reading or even understanding terms, making them susceptible to seller manipulation.[10]

2.1. Theory

This view is not universal, however. Steven Salop and Joseph Stiglitz have shown that the existence of uninformed consumers need not prevent a competitive outcome. Specifically, the authors explore the conditions under which a market with consumers heterogeneous in their willingness and ability to become informed about product prices might reach a perfectly competitive price equilibrium. Even when uninformed consumers exist, a market might yield such an equilibrium if enough informed consumers do shop for the competitive price: "[T]here is an informational externality at work between efficient and inefficient information-gathers. Those agents who become informed give an external economy to the uninformed; the weight of their search keeps prices lower. In fact, if there are enough informed agents, the market price will settle down to the perfectly competitive price."[11]

Alan Schwartz and Louis Wilde extend this argument to a legal situation where consumers vary in their ability to become informed about standard form contract terms. The authors show that if a sufficient number of buyers are informed about the price and contract terms of a given product, sellers who cannot discriminate between buyer types will offer the product with efficient terms at a competitive price to all buyers. This is because when markets are sufficiently competitive, the cost to the seller of losing a critical mass of informed consumers outweighs the benefits of offering self-serving terms to those uninformed infra-marginal

---

[9] See, e.g., Avery Katz, Your Terms or Mine? The Duty to Read Fine Print in Contracts, 21 RAND J. of Econ. 518-33 (1990); Melvin A. Eisenberg, The Limits of Cognition and the Limits of Contract, 47 Stan. L. Rev. 211, 240-41 (1995); Gillette, supra, note 6.

[10] See, e.g., Xavier Gabaix & David Laibson, Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets, 121 (2) Q. J. of Econ. 505-40 (2006); Oren Bar-Gill, Seduction by Plastic, 98 Nw. U. L. Rev. 1373 (2004); Russell Korobkin, Bounded Rationality, Standard Form Contracts, and Unconscionability, 70 U. Chi. L.Rev. 1203, (2003); Ronald J. Mann, "Contracting" for Credit, 104 Mich. L. Rev. 899 (2006).

[11] The authors also mention that "[o]n the other hand, by shopping at high-priced stores, the uninformed inflict an external diseconomy on the informed; these informed must gather costly information to obtain the lower price." Steven Salop & Joseph Stiglitz, Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion, 44 Rev. Econ. Stud., 493 (1977), at 494.

consumers.[12] The authors conclude that imperfect information alone is not sufficient to warrant market intervention, a conclusion that would become the cornerstone of the law and economics view of standard form contracts. Subsequent theoretical inquiry has focused on whether markets are sufficiently competitive and whether there is in fact an informed minority of buyers that will police suspicious terms.

Echoing the beliefs of many academics and courts, Schwartz and Wilde state that "[g]enerally, there are a significant number of informed consumers in any given market prior to legal intervention,"[13] although the evidence cited to support this statement is inconclusive.[14] The main issue for them is whether these consumers are able to access adequate information at a reasonable cost. The cheaper it is to do so, the larger (and thus the more effective) this informed minority will be. The determination of the exact proportion of informed consumers necessary to yield a competitive equilibrium is a complicated exercise.[15] Nevertheless, many scholars have

[12] Alan Schwartz and Louis Wilde, supra, note 5. In their model, the equilibrium result changes with the number of shoppers. A competitive equilibrium will result if the number of informed consumers is "substantial." If the number is too small to support a competitive equilibrium, but it is nonetheless significant, the market will reach an equilibrium with a cluster of prices. When there is only a small minority of comparison shoppers, the price distribution will converge towards a monopoly price (or terms). Schwartz and Wilde at 562.

[13] Schwartz and Wilde, id, at 636.

[14] Although the authors cite evidence supporting the view that a significant portion of consumers visit more than one store to engage in price comparison, the evidence cited to support the view that shoppers engage in term comparison is inconclusive. The authors cite a study that examines the effect of the Truth in Lending Act on consumers' knowledge of the interest rates charged by their lenders. The author finds that prior to the passing of the Act, about 14% of the families in the sample estimated accurately the interest rates they paid on their loans and 33% of families did not know their true rates of interest on their loans. After the passing of the Act, the percentage of families with accurate estimates increased to 21% (a statistically significant difference), and the percentage of families with inaccurate estimates of their interest costs decreased to 25%. Lewis Mandell, Consumer Perception of Incurred Interest Rates: An Empirical Test of the Efficacy of the Truth in Lending Act: An Empirical Test of the Truth-in-Lending Law, 26 J. of Finance 1143, 1153 (1971). While the study is able to show that the Truth-in-Lending Act was modestly effective in helping consumers understand the true rate of interest on their existing loans, it says nothing as to whether consumers were informed about the menu of prices or contract terms available to them when they were shopping among creditors. Furthermore, of all contract terms, price is likely to be the most salient one.

[15] Schwartz and Wilde provide a numerical example where the presence of one third of informed consumers would generate a competitive equilibrium with respect to price. Changing the relative costs, however, would require a different percentage of informed consumers. The authors expect that there will be fewer consumers informed about terms in a given market because shopping for terms is costlier than shopping for price, but contend that if one third of those price savvy shoppers are also term savvy, a competitive equilibrium could result. For a critique of this illustration, see. R. Ted Cruz and Jeffrey Hinck, Not My Brother's Keeper: The Inability of an Informed Minority to Correct for Imperfect Information, 47 Hastings L. J. 636 (1995). See also, Oren Bar-Gill and Elizabeth Warren,

repeatedly relied on the informed minority argument to support freedom of contract in mass market transactions.[16]

Courts have also relied on the informed minority argument (or at least a belief that enough buyers are sensitive to contract terms) in deciding whether certain standard form contract terms provisions should be enforceable. For example, in *ProCD vs. Zeidenberg*, Judge Easterbrook enforced a resale restriction in a shrinkwrap license by noting that "[t]erms of use are no less a part of 'the product' than are the size of the database and the speed with which the software compiles listings. Competition among vendors, not judicial revision of a package's contents, is how consumers are protected in a market economy."[17]

Other scholars have relied on the plausibility of the informed minority argument to resist regulation, but express some reservations. Clayton Gillette argues that standard form contracts should be enforceable as long as the interests of those uninformed buyers are represented by the informed minority.[18] Robert Hillman and Jeff Rachlinski explore the role of the informed minority argument in standard form contracting online.[19] They conclude that although the low cost of becoming informed on the Internet is likely to increase number of informed consumers, the free-rider problem introduced by those uninformed consumers is likely to result in an under-

---

Making Credit Safer, U. of Pa. L. Rev. 157 (2008). But see, Alan Schwartz, How Much Irrationality Does the Market Permit? 37 J. L. Stud. (2008).

[16] For example, in proposing an "investment theory" of product warranties, George Priest relies on the informed minority argument to address the concern that most buyers will not factor warranty terms in their purchase decisions. George Priest, A Theory of the Consumer Product Warranty, 90(6) Yale L. J (1981) at 1347. Douglas Baird also relies on the existence of the informed minority to defend freedom of contract. He notes that "[the] typical buyer cannot rely on her own expertise or her ability to dicker with her seller. When the market works effectively, however, she benefits from the presence of other, more sophisticated buyers." Douglas Baird, supra note 6. See also Duncan Kennedy, Distributive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power, 41 Md. L. Rev. 563, 614-15 (1982). For examples of references to the informed minority argument outside standard form contracts, see Richard A. Epstein, The Unintended Revolution of Products Liability Law, 10 Cardozo L. Rev. 2193, 2204 (1989); Alan Schwartz, The Case Against Strict Liability, 60 Fordham L. Rev. 819, 826 (1992).

[17] ProCD v. Zeidenberg, 86 F. 3d 1447 (7th Cir., 1996).

[18] Clayton Gillette, Rolling Contracts as an Agency Problem, Wis. L. Rev.679, at 690 (2004). Gillette notes that "[t]he question is whether such a group of informed buyers exists and whether it shares contract terms preferences with those non-reading buyers…Indeed, I will suggest that the most difficult issue in finding surrogates for nonreading buyers is that one set of buyers may have very different preferences from another set." At 690.

[19] Robert Hillman and Jeff Rachlinski, Standard Form Contracting in the Electronic Age, NYU L. Rev., (2002).

production of knowledgeable buyers.[20] Others have pointed out that many consumers are unlikely to behave rationally in regards to boilerplate, thus making the "informed" minority less effective. Russell Korobkin notes that even when choosing to become informed, boundedly rational buyers are unlikely to consider *all* the contract terms of product attributes in making a purchase decision.[21] Consequently, sellers will offer one-sided *non-salient* terms.

Finally, a group of scholars believes that the informed minority flat out does not exist or does not work. Melvin Eisenberg argues that that a competitive equilibrium is unlikely to result in the presence of imperfect information: "Typically [a competitive equilibrium] will not occur, because most form takers will find it irrational to engage in search and deliberation on any given form."[22] Recently, Omri Ben-Shahar has advocated abandoning recent disclosure proposals that seek to increase the "opportunity to read" standard form contracts.[23] He asserts that because nobody reads fine print, regardless of reduced reading costs in environments such as the Internet, rules that focus on increasing contract disclosure are useless, if not dangerous. Lastly, Victor Goldberg questions the existence of the informed minority as well as whether sellers will indeed find it more profitable to cater to the readers than to take advantage of the non-readers.[24]

As an aside, it is important to note that law and economics scholars have also relied on other arguments to support the conclusion that markets might behave competitively when consumers are imperfectly informed. When sellers are constrained by reputation, sellers will find it in their best interest to offer terms preferred by buyers to protect their reputational investment.

---

[20] The authors also note that "[r]ational calculation alone cannot explain consumers' nearly universal failure to read standard-forms. In some circumstances, the market should produce a sufficient number of consumers who attend to the unlikely contingencies covered by the standard form such that businesses feel disciplined." Id.

[21] Korobkin, supra note 10. For other behavioral accounts of why the existence of informed consumers might not result in competitive outcomes, see Laibson and Gabaix, supra note 10, and Bar-Gill, supra note 10.

[22] Melvin Eisenberg, The Limits of Cognition and the Limits of Contract, 47 Stan. L. Rev. 211, 243 (1995), 243-44.

[23] Omri Ben-Shahar, The Myth of the "Opportunity to Read" in Contract Law, (mimeo, 2008). See also Robert Hillman, On-Line Boilerplate: Would Mandatory Website Disclosure of E-Standard Terms Backfire?, 104 Mich. L. Rev (2006).

[24] Victor P. Goldberg, The "Battle of the Forms": Fairness, Efficiency, and the Best-Shot Rule, 76 OR. L. REV. 155, 165 (1997) ("Others will presume that the random buyer they run into will not have read the form and that, by stacking the deck, the seller can perhaps gain more from the nonreaders that it loses to the readers.") at 165. See also Shmuel Becher, Asymmetric Information in Consumer Contracts: The Challenge is Yet to be Met, Am. Bus. L. J. Vol. 45 (2008).

Sellers might also offer one-sided terms to all consumers, only to later relax them to accommodate reasonable buyer complaints.[25] Similarly, in the case of experience goods or repeat purchases, buyers who do not read terms might ultimately become familiar with the contents of the sellers' boilerplate.[26]

2.2. Prior Evidence

Despite the theoretical importance of the informed minority, there has been little empirical investigation of its existence. The extant evidence includes a couple of studies that survey individuals' attitudes towards fine print in various contexts. In a study of reading practices of online standard form contracts, Robert Hillman surveys 92 contracts students at Cornell Law School and finds that only four percent of those who purchased products online claim to read standard form contracts "as a general matter."[27] Almost sixty percent of respondents, however, reveal that they would be prompted to read the contract depending on the type of vendor, the price of the product purchased, and the term. In another study, Shmuel Becher and Esther Unger-Aviram survey 147 students in law and other areas and ask about the likelihood that respondents will read standard terms in different scenarios (presented as vignettes in the survey). They find that 60% of respondents claim they skim or read parts of a standard form contract before entering a transaction. They also find that a significant number of respondents would be inclined to read contracts after having purchased the product. Although these surveys provide important information about consumer propensities to read standard form contracts, they are based on self-reported behavior or hypothetical commercial scenarios and the survey subjects are highly unrepresentative. The evidence is suggestive but cannot be used to confidently assess the existence of the informed minority.

Three other studies provide indirect evidence regarding the existence of the informed minority. In a study of 62 warranties of an array of consumer durable goods, George Priest

---

[25] See Gillette, supra note 6; Lucian A. Bebchuk & Richard A. Posner, One-Sided Contracts in Competitive Consumer Markets, 104 Mich. L. Rev. 827 (2006).

[26] Priest, supra note 16.

[27] Robert Hillman, Online Consumer Standard Form Contracting Practices: A Survey and Discussion of Legal Implications, in IS CONSUMER PROTECTION AN ANACHRONISM IN THE INFORMATION ECONOMY? (Jane K. Winn, Ed. 2006).

argues that warranties are not biased towards sellers, but rather reflect by the relative ability of buyers and sellers to prevent and insure against loss. Florencia Marotta-Wurgler analyzes the terms of 647 online EULAs and shows that almost all of them are more restrictive than the relevant default rules, but less restrictive than the legal minimum.[28] In a study of contracting practices by online retailers, Ronald Mann and Travis Siebeneicher find that few sellers offer excessively one-sided terms (and many present their contracts in ways that would be deemed unenforceable).[29] Of course, the findings in these studies could be explained by sellers being reputationally constrained and not due to the existence of the informed minority per se.

2.3. Current Legal Landscape

Given the risk of market failure, the law governing standard form contracts affords courts great flexibility in deciding whether to enforce consumer agreements involving standard terms. Courts that believe that sufficient market pressure (either from an informed minority or because sellers are constrained by reputation) will ensure competitive market outcomes routinely enforce boilerplate terms in consumer disputes.[30] For example, in *ProCD v. Zeidenberg*, mentioned above, Judge Easterbrook enforced a commercial-use restriction clause in a shrinkwrapped software license agreement, thus rejecting the buyer's argument that he couldn't have agreed to "hidden" terms. He explained that the burden is on the buyer to protect his own interests by stating that "ours is not a case in which a consumer opens a package to find an insert saying 'you owe us an extra $10,000' and the seller files suit to collect. Any buyer finding such a demand can prevent formation of the contract by returning the package, as can any consumer who concludes that the terms of the license make the software worth less than the purchase price. Nothing in the UCC requires a seller to maximize the buyer's net gains."

Other courts, however, do not believe that buyers are as careful shoppers as Judge Easterbrook portrays them to be. In fact, many rely on the doctrine of unconscionability to

[28] Florencia Marotta-Wurgler, What's in a Standard Form Contract? An Empirical Analysis of Software License Agreement, 4 J. Emp. L. Stud. 677 (2007); Florencia Marotta-Wurgler, Competition and the Quality of Standard Form Contracts: The Case of Software License Agreements 5(3) J. Emp. L. Stud. 447 (2008).

[29] Ronald Mann & Travis Siebeneicher, Just One Click: The Reality of Internet Contracting, 108 (4) Colum. L. Rev. (2008).

[30] Enforcement will occur absent fraud, duress, or misrepresentation.

invalidate one-sided contract terms that appear to be the result of seller abuse.[31] A term will be struck down under this doctrine if it is found to be procedurally and substantively unconscionable. The procedural aspect of the inquiry focuses on whether the buyer is deprived of an opportunity to meaningfully assent to the terms because the contract is hard to find, in miniscule print, hard to understand, or because the buyer lacks meaningful choice. The substantive aspect asks whether a particular term is so one-sided that it would "shock the conscience" of an informed buyer. Although this doctrine gives courts little guidance on how to effectively distinguish mutually beneficial clauses from exploitative ones, courts that believe that failure to read might result in seller abuse are eager to strike down terms.[32]

In addition, various state legislatures who consider that failure to read leaves consumers particularly vulnerable have enacted various mandatory terms in consumer contracts. For example, the states of Idaho[33], North Carolina, and Montana[34] have statutes prohibiting the enforcement of forum selection clauses in consumer transactions.[35] California has enacted the California Arbitration Act that mandates a waiver of arbitration fees for low income consumers and requires arbitration organizations to make all consumer arbitration decisions publicly available.[36] On the federal front, there are several laws that aim to increase the size of the

[31] UCC 2-302. Comment 1 states that "The principle is one of the prevention of oppression and unfair surprise…and not of disturbance of allocation of risks because of superior bargaining power."

[32] For example, the California Court of Appeals recently struck down as unconscionable an arbitration clause with a class action waiver that visibly included in the fine print of a cell phone service agreement. After noting that contracts of adhesion should be carefully examined, the court stated that "[t]he possibility of overreaching is even greater in ordinary consumer transactions involving relatively inexpensive goods or services because consumers have little incentive to carefully scrutinize the contract terms or to research whether there are adequate alternatives with different terms, and companies have every business incentive to craft the terms carefully and to their advantage. The unconscionability doctrine ensures that companies are not permitted to exploit this dynamic by imposing overly one-sided and onerous terms." Gatton v. T-Mobile USA, Inc., 152 Cal. App. 4th 571, 585 (Cal. App. 1st Dist. 2007).

[33] Idaho Code § 29-110.

[34] Mont. Code36§ 18-1-403.

[35] Several courts also refuse to enforce forum selection clauses. See, e.g., Fidelity & Deposit Co. v. Gainesville Iron Works, Inc., 125 Ga. App. 829 (Ga. Ct. App. 1972). For a more detailed review of prohibitions on dispute resolution clauses, see Florencia Marotta-Wurgler, "Unfair" Dispute Resolution Clauses: Much Ado About Nothing? in (BOILERPLATE: FOUNDATIONS OF MARKET CONTRACTS, Omri Ben-Shahar, ed.) (Cambridge University Press, 2007)

[36]California Code of Civil Procedure Sections 1284.3, 1281.96.

informed minority by decreasing reading and comparison shopping costs. The most famous example of such law is the Magnuson-Moss Warranty Act, a disclosure law enacted in 1975 to regulate the form and content of consumer product warranties.[37] The act requires that sellers who choose to provide warranties draft them in clear language and present them to consumers in a standardized fashion.[38] It also seeks to reduce shopping costs by requiring that warranties be available for inspection prior to purchase.[39] Yet despite these protective measures, whether buyers read terms or not, the law is that buyers are under a duty to read standard form contracts and are thus deemed to have given "blanket assent" to reasonable terms.[40]

More recently, there have been numerous proposals to protect consumers in mass market transactions, especially those that take place over the Internet, by seeking to facilitate contract access to the informed minority. Some of the most prominent and contentious proposals involve software and other information goods. For example, the American Law Institute (ALI) has proposed new Principles of Software Contracts to harmonize and increase certainty of the laws governing software transactions online. One of its main goals is to "promote reading and the opportunity to read terms" as a way of alleviating market failures.[41] For example, the *Principles*' rules require software vendors to post the terms of their license agreements in their corporate website, thus effectively eliminating "pay now, terms later" contracts (where buyers cannot access the contract until after they had paid for the product). The proposed law would also deem all browsewrap contracts (i.e., contracts included in hyperlinks on sellers' websites that don't have to be expressly agreed to, such as websites' "Terms of Use") unenforceable.

---

[37] 15 U.S.C. §§2301-2312 (1976).

[38] 15 U.S.C. §2302(a).

[39] 15 U.S.C. §2302(b)(1)(a). Another prominent example is the Truth in Lending Act (TILA), a disclosure law that seeks to protect consumers in credit transactions by requiring a clear and standardized disclosure of the essential terms and costs associated with the deal. 15 U.S.C. 1601 note.

[40] KARL LLEWELLYN, THE COMMON LAW TRADITION: DECIDING APPEALS 370-371(196).

[41] ALI Principles at 96. The reporter explains that "because case reports and the websites of watchdog groups already evidence vendors' use of unsavory terms, [the] Principles assume that market pressure is insufficient in software retail markets to assure the production of reasonable terms, both in presentation and substantive content…The preferred strategy of [the] Principles is to draft rules that promote reading terms before committing to a transfer, which, in turn, should decrease the instances of market failure." ALI Principles at 94-96. For a detailed account of these proposals and an analysis rejecting the "opportunity to read" approach, see Omri Ben-Shahar, supra note 23.

Another effort to create a uniform and cohesive body of law for computer information including software has been the Uniform Computer Information Transactions Act (UCITA), a model act drafted by the ALI and NCCUSL. The act, which has been met with strong opposition, has so far been enacted only in Maryland and Virginia.[42] One of the reasons UCITA has been rejected is that, in contrast to the ALI's *Principles*, it allows enforcement of "pay now, terms later" contracts as long as buyers are made to assent to the terms and are able to return the software after having had an opportunity to review the EULA. While the drafters of UCITA advocate this form of contracting because it decreases contracting costs, critics worry that that sellers will take advantage of the higher reading costs presented by post-purchase availability of terms and impose one-sided terms on unwitting buyers. Ultimately, whether the approach endorsed by the drafters of the *Principles* or the drafters of UCITA is the most appropriate depends on whether there is market failure and whether increased disclosure would ameliorate it relative to the costs.

3. Methodology

To understand the preferable legal treatment of consumer standard form contracts or the most desirable proposal for reform, it seems critical to document whether a minority of informed buyers capable of disciplining the market exists. Our approach is to study the browsing and shopping behavior of online consumers. We track the detailed browsing and shopping behavior of Internet users to the corporate websites of 93software companies. We examine the extent to which apparent shoppers (by various definitions) seek to become informed about the EULAs that govern the featured software.

Online software markets provide a particularly good environment to study the existence of the informed minority. First, while non-price information such as the associated contractual rights and restrictions is important for all types of products, it is a particularly significant consideration for information goods such as software as terms form an integral part of the way the product is used. Second, there is a wide variation in the way EULAs are presented to consumers online. Some vendors require buyers to explicitly agree to terms by using "clickwrap"

[42] See Va. Code Ann. 59.1-501.1 to 509.2 (Michie. 2001); Md. Code Ann. Comm. Law 22-101 to 22-816 (Supp. 2002).

contracts, while others post their terms in their websites (some more prominently than others) so it is up to buyers to seek out the terms and become informed. This variation allows us to address the important question whether buyers are more likely to read contracts that are less hidden. Third, shopping for competing goods and the terms that governs them is cheap and easy online relative to most commercial settings. To the extent the informed minority exists, here is where we are likely to find it. Finally, several recent debates on legal reform in standard form contracts focus on electronic contracts in general and software contracts in particular. A study of the existence of the informed minority in online software markets places us at the center of these debates.

3.1. A Stylized Model

To explain our empirical approach to documenting the size of the informed minority, it is helpful to start with a simple model. Consider a setting with a single seller, homogeneous buyers and two types of possible standard form contracts, offering what we will call "good" and "bad" terms that are respectively more and less favorable to the buyers. Buyers value the rights and restrictions incorporated in the standard form contract (e.g., warranty terms, the ability to transfer the product, etc); good terms are valued more than bad terms, resulting in corresponding product valuations of either V_g or V_b, with $V_g > V_b$. We will assume that $V_g > 0$, but it is possible that $V_b < 0$. Good terms are more expensive for the seller to provide than bad terms, resulting in corresponding product costs of c_g and c_b, with $c_g > c_b \geq 0$. The price p charged by the seller is exogenously determined (perhaps as a result of the cost of the product or service and competitive conditions).

The seller moves first by deciding to offer good or bad terms, and each buyer decides whether or not to become informed about the terms of the standard form contract. Buyers incur a positive cost c_r if they choose to become informed about the terms, reflecting the cost of finding and reading the standard form contract. Informed buyers purchase with probability b_1 if the terms are good and b_3 if the terms are bad ($b_1 > b_3$), and uninformed buyers purchase with probability b_2 ($b_1 > b_2 > b_3$). A seller providing bad terms will also incur a penalty cost $F \geq 0$ if discovered by a buyer that becomes informed and does not purchase. This can be thought as the cost of damage to the seller's reputation, perhaps because the buyer may inform other buyers,

write negative reviews, etc. This setting is similar to the empirical framework to follow and the corresponding extensive form game is depicted in Figure 1.

There are two types of equilibria in this setting:

- A pure strategy equilibrium always exists, with $g = r = 0$, corresponding to the case where the seller offers "bad" terms and buyers do not become informed.
- A mixed strategy equilibrium where the seller will offer "good" terms with probability $g = \frac{c_r(b_1 - b_3)}{p - V_b}$ and buyers will read with probability $r = \frac{c_g - c_b}{(b_1 - b_3)(p - c_b + F)}$ will exist when the values of the parameters in our setting allow *g* and *r* to satisfy the constraints $0 < g < 1$ and $0 < r < 1$.

Under certain conditions only the pure equilibrium exists, for instance when the reading cost c_r or the cost of providing good terms c_g are sufficiently high. A pure equilibrium is inefficient if $V_g - V_b > c_g - c_b$, as it corresponds to an inefficient provision of terms because the buyers value good terms above the seller's cost to provide them.[43]

3.2. An Empirical Framework

Next we outline a framework that we use to help empirically measure the existence and size of the informed minority. It is similar in spirit to the model described above. Visitors to the websites of the companies in our sample, described below, can be classified into potential buyers or users visiting for other reasons, such as looking for online instruction manuals for a product they already own, to search user forums for troubleshooting information, or for entertainment—e.g., to watch the "Mac vs PC" commercials. We denote by *s* the fraction of potential buyers ("shoppers"); non-shoppers make up the remaining fraction $1 - s$. We denote by e_1 the fraction

[43] We also analyzed a variation of the above setting with buyers that are heterogeneous in their cost of becoming informed. Without loss of generality, we assumed two types of buyers, with corresponding costs c_H and c_L ($c_H > c_L > 0$) if they choose to become informed about contract terms, possibly reflecting differences in education or experience or ability to interpret standard form contracts. In this case a separating equilibrium can exist where only buyers with low reading cost become informed about contract terms with positive probability, and thus create an externality on the high cost buyers. While in this case the size of the informed minority may depend on the exogenous distribution of buyer types, the essence of our analysis of the ability of the informed minority to discipline the sellers remains applicable.

of shoppers and by e_2 the fraction of non-shoppers that read the online EULAs. Finally, we denote by b_1 the fraction that purchase the product ("buyers") among shoppers that read the EULA and by b_2 the fraction of buyers among shoppers that do not read the EULA. This process is depicted in Figure 2. In this setting, the fraction e_1 of shoppers that read the online EULAs represents the informed minority.[44]

In what follows we empirically estimate the number of visitors in our sample for each of the six categories shown in Figure 2. We estimate the number of buyers, shoppers and non-shoppers, readers and nonreaders. We use access to a EULA page as a proxy for reading (generous in the sense that some EULA accesses are by accident or inconsequential to the buying decision), and we use initiating the checkout process as a proxy for buying, and we use other contextual information to identify non-shoppers.

We can break down readers into se_1b_1 readers that buy and $se_1(1-b_1)$ readers that do not buy. In addition, there are $s(1-e_1)b_2$ buyers that are not readers, and $s(1-e_1)(1-b_2)$ shoppers that neither read nor buy. Few non-shoppers would be expected to read EULAs and thus we would expect $(1-s)e_2$ to be small. Finally, the (large) majority of non-shoppers that do not read EULAs is $(1-s)(1-e_2)$. In the end, our estimates of the fraction e_1 of shoppers that make up the informed minority is $\frac{se_1b_1 + se_1(1-b_1)}{se_1b_1 + se_1(1-b_1) + (1-e_1)b_2 + (1-e_1)(1-b_2)}$. We now turn to our sample.

4. Data

Our data are derived from a "clickstream" data set for January 2007, representing a sample of over 90,000 U.S. households. These data were made available to us by our data

[44] It is possible that for some shoppers accessing the EULA will not affect their probability of buying the product. For instance, some shoppers may not know what a EULA is, or discover after accessing the EULA that they are not capable to comprehend its language, or may access it accidentally or out of curiosity. To the extent that such accesses of the EULA do not make a shopper part of the "informed minority," e_1 is likely to overestimate the informed minority fraction of shoppers; the corresponding adjustments are likely to be relatively small, however. Furthermore, we explore the role of these concerns by studying the role of the time spent on the EULA page. On the other hand, because we don't consider other ways in which shoppers might become informed about the terms (e.g., via word of mouth, or repeat purchases), there is the possibility that e_1 might underestimate the size of the informed minority. We address this explicitly in Section 5.

provider, a major online research company, which recruited a representative panel of U.S. households that agreed to install in their computers a data collection plug-in that recorded the url address of each webpage visited on these computers. The resulting data set contains the exact sequence of web pages visited by each member of the panel, including the amount of time spent on each page. This is a very large data set, containing the browsing behavior of 92,411 users for January 2007.[45] Monthly churn is approximately 15%, i.e. each month about 15% of users in the sample drop out, in which case they are not included in that month's data.

To avoid self-selection biases, the panel of households was selected to be demographically and geographically balanced and representative of the entire population of U.S. households with Internet access.[46] The information captured for each web page visited by a panelist in the raw data is coded with a user identifier that anonymously but uniquely identifies each panelist and a session identifier that delimits each panelist's web browsing into separate "sessions." Additional information captured includes the url of each page visited, the time that webpage was accessed, the time spent on that page, whether this was a secure (i.e., encrypted) connection, the web server delivering the web page, and a unique identifier for the company or division owning that web server. The recorded page views comprise the bulk of this massive data set; additional files included non-personally identifiable demographic information about the panelists, and a corporate hierarchy identifying the parents, if any, for the divisions or companies owning the web servers that appear in the data (e.g., office.microsoft.com and mail.hotmail.com would be identified as companies or divisions having the same corporate parent, Microsoft).

4.1. Sample Construction

In view of the size of this data set, we restrict our analysis to companies in one type of market and one type of contract. Specifically, we study user visits to software companies that sell or distribute their products through their corporate websites and make their EULAs available on their site for users to peruse. We relied on our data provider's detailed classification of different

[45] Information was captured for 6,355,922 user sessions in January 2007, with 461,027,284 corresponding web page views.

[46] The company has vast experience in measuring Internet and media audiences. The panel is considered one of the largest representative media research samples currently around. The company used random digit dialing to recruit participants. It updated demographic information regularly, implemented various procedures to keep the panel updated, and ensured that tracking is unobtrusive to prevent any distortions in behavior.

markets to select visits to only those classified as "software companies." We subsequently identified for inclusion in our data set two types of software companies that make their products available for online purchase or downloading: retailers, and freeware providers. Retailers license their software for a price through their corporate website. Freeware providers offer their software for free to anyone wishing to download it. Examples include browser toolbars, plug-ins, and browsers.[47] We are interested in observing users' propensities to become informed about the terms of these two types of software. For the purpose of a sufficiently homogenous sample of sellers, we exclude subcategories such as vendors not making their products available for online purchase or downloading, peer-to-peer software providers, and web hosting companies. This process resulted in 1,199 software companies for further analysis.

A preliminary data analysis at this stage shows that well under 1% of website visitors access a EULA page (definitions of visits follow below); we thus decided to exclude from our analysis companies with fewer than 50 unique visitors that viewed at least two pages during their visit. Our interest is to analyze users with genuine intent to purchase, or "shoppers," and users that view a single page are unlikely to have such intent. We identified 197 companies that satisfied these conditions, and for each one of these companies we obtained the web page addresses (URLs) of all the EULAs made available on the company's website. To do this, we visited each company' company website and used a combination of manual browsing, Google search within the company's website and, if available, search of the website provided by the company. In addition, we searched all page views in the clickstream data corresponding to these companies to identify possible EULA pages (e.g., pages whose web address contained "eula" or "legal" or "terms"), which we subsequently investigated manually. We were able to identify and collect all EULAs made available online for 93 companies, our final sample of companies for which we examine the browsing behavior of visitors. During this process, we also excluded companies that did not make their EULAs available online. The final sample includes 81 retail companies, and 12 freeware companies.

[47] We classify a company as "retail" if it offers the core or a non-trivial amount of software for sale, even if it also offers software for free. For example, Adobe offers several free plug-ins, such as Shockwave, or a PDF reader. However, we classify it as "retail."

4.2. Company and product characteristics

It is possible that consumers will feel less need to scrutinize the terms in EULAs from companies that are large or that have been in existence for a long time, as they may assume that such companies are more trustworthy and fair. To test this possibility, we obtained information about each company's annual revenue, year of incorporation, and whether it was publicly traded. We collected this information from Hoovers.com, Yahoo Finance, or direct communications with the companies themselves.

Panel A in Table 1 reports summary statistics for the company characteristics for each of the three types of companies analyzed. For retail companies, average revenue is $1.52 billion with a standard deviation of $6.85 billion, a number obviously driven by a few large firms. Median revenue for this category is $6 million. The average age of these companies, measured as 2008 minus the year of incorporation, is 14.7 years old (the median is 13). Eight percent of the companies in this category are publicly traded. In contrast, the median age of freeware firms is five-and-a-half years (the average is eight). Eight percent of these companies are publicly traded. Given the range of company sizes, markets, and age, restricting the sample to companies that make their EULAs available online did not create any significant biases.

We collect several product characteristics. For each company, we recorded one "flagship" product per company. Many small and medium companies market one main product, so we select that product. For larger companies, we selected the product accounting for the largest fraction of sales, and when this information was not available, we selected the product most prominently featured on the company's website. Because users might more inclined to become informed about the EULA terms of higher priced products, we obtained the price for the flagship or representative product as well as median price of all products for each firm. We recorded whether the product is a single or multi-use license, because multi-seat licenses are likely to have higher prices; and whether the product is offered to developers. We note whether the company offers a trial version of selected product and of the majority of its products because it might affect users' propensity to read terms.[48] We also note whether the product is directed to

[48] Trial versions are generally offered with limited functionalities over a limited time period. In constructing the sample for Marotta-Wurgler (2007), the author found that the majority of trial licenses are noticeably different from the product licenses (e.g., the trial license reads "Trial License" and is generally shorter), such that a user would not consider them being substitutes.

business users or to members of the general public. Finally, we classify each product into one of 150 software product categories, such as antivirus and word processing, as classified by Amazon.com, one of the largest Internet retailers of software.

Panel B in Table 1 reports summary statistics for the product characteristics for each of the two types of companies. The average product price for retail companies is $395 and the median price is $58.40. For each company we compute the median price of all software products listed on the website, and the average of those median prices is $352; the median is $49. About 70% of retail products and 100% of freeware is targeted to members of the general public (or very small businesses) as opposed to larger businesses. Finally, the majority of companies (84%) offer a trial version of their featured product—or the product in our sample. Seventy-eight percent of them offer trial versions for most of their offerings.

4.3. Contract characteristics

We want to measure whether users choose to become informed about terms as well as possible. We thus attempt to collect *all* the EULAs that can be accessed in a company's website. As noted above, many companies only sell one product so they post just the EULA that governs the use of that product. Other companies sell many products, yet all are governed by a single EULA posted on the site. Others post different EULAs for different products. Finally, some firms post the EULAs for all their current and past versions of all their products. The EULA collection process obtained 240 EULAs for the 81 retail companies and 34 EULAs for 12 freeware companies.

An important policy question is whether in the absence of an informed minority to promote efficient standard terms, the appropriate remedy is regulation of disclosure rather than regulation of terms. Increased disclosure, it is argued, would promote the emergence of an informed minority, thus dispensing with the need for further regulation. The underlying premise is that improved accessibility of the contract would result in improved readership. To examine this claim, we measured the accessibility of each of the selected products' EULAs in each of the companies' websites. Specifically, we used the methodology in Marotta-Wurgler (2009) to measure EULA accessibility, defined as their location relative to the most direct clickstream path of purchase (or, in the case of free software, download) within the company's website, starting

from the home page.[49] Given that all companies in our sample make their software available online, the most direct clickstream path to purchase is easy to determine.

The resulting "distance" measure equals the minimum number of clicks necessary to access a product EULA from the product purchase path. For instance a value of 0 would indicate that the EULA was presented as part of the product purchase process, typically presented to the buyer and then requiring that the buyer take an action agreeing to the terms, such as clicking "I Agree." A buyer doesn't need to make any additional clicks to access this EULA, as she is required to assent to it. A value of 1 would indicate that the EULA could be accessed by clicking on a link that appeared during the purchase process (i.e., it was one click away). Certain companies require buyers to agree to the EULA terms, but these terms are not presented on the screen during the purchase process; instead the buyer has to click on a link and read the terms. In these cases the accessibility of the terms is lower than when the value is 0, as an additional click is required to access the terms. On the other hand, the buyer is made aware of the existence of the EULA and is required to agree the terms, and thus accessibility is higher than when the EULA is simply one click away. In view of the above considerations, such intermediate cases are assigned a value of 0.5. We follow Marotta-Wurgler (2008) and label those EULAs with distance scores of 0 and 0.5 as "Forced" because users are forced to acknowledge them in the checkout process.

The distribution of the location of EULAs in our sample is shown in Table 2. For the two types of companies, there is wide variation in the accessibility of terms. For retail companies, about 30% of sellers "force" buyers to acknowledge and agree to terms. Over 40% of sellers locate their license one click away from the product purchase path. The other sellers in the sample make their EULAs harder to access by placing them up to six clicks away from the most obvious path to purchase. The mean location score for all EULAs of retail companies is 1.28. For freeware companies, a smaller fraction of sellers, 16.67%, require buyers to agree to the EULA. Nearly 60% of sellers in this category place their EULAs one click away from the most obvious path to download or purchase. Around 25% of these sellers place their EULAs two or more clicks away. The mean distance is also 1.28 for freeware products.

[49] Florencia Marotta-Wurgler, Are "Pay Now, Terms Later" Contracts Worse for Buyers? Evidence from Software License Agreements, J. Legal Stud. (forthcoming, 2009).

4.4. Definition of a user visit

We aim to estimate the fraction of "shoppers," or visitors with some potential to purchase during this visit, that becomes informed about EULA terms. Our data provider reports *all* the Internet browsing activity of its users. Given the nature of our data, identifying such shopping-oriented visits is challenging, as a large fraction of visitors may be browsing without intent to purchase. They may be looking for online instruction manuals for a product they already own, for example, or searching user forums for product support information. This is more of an issue with larger diversified websites that provide significant non-product oriented content such as Microsoft, Adobe or Symantec. Smaller companies tend to have sparser websites focused on supporting the purchase process. As noted earlier, our data provider tracks the particular web servers that host the URLs accessed.

To restrict our analysis to visitors with intent or potential to purchase, we exclude visits made exclusively to servers that are not related to the retail aspect. For example, Microsoft and Adobe, and Symantec locate their user forums and patches and rebate pages on separate servers (e.g., www.adobeforums.com, symantecrebates.com). This helps us weed out many visitors without potential intent to purchase. Thus, we only focus on those servers that are more directly related to shopping or purchasing activities. We then adopt several approaches to more precisely identify shopping-oriented visits.

We define a user visit as all page views (url accesses) from a company's website within a single user "session." One way of identifying shoppers is by examining the intensity of a company visit. A user with intent to purchase is likely to view several pages in the retail side of the company's website. For our analysis we included visits with at least two page views, as a minimal definition, as visits with just one page view were not likely to belong to users with intent to purchase.[50] That is, our broadest and most generous definition of a shopping visit includes all visits by users who accessed at least two pages in a company's website.[51] The

[50] The literature that studies Internet shopping and browsing behavior generally excludes visits to companies with only one page view, as they have been found to be mostly mistakes or the result of redirects. See Randolph E. Bucklin & Catarina Sismeiro, A Model of Web Site Browsing Behavior Estimated on Clickstream Data, J. of Marketing Res. Vol. XL, 249-267 (2003).

[51]This definition has been used extensively in the literature to define a valid company visit. See, e.g., Wendy W. Moe & Peter S. Fader, Dynamic Conversion Behavior at e-Commerce Sites, Mgmt. Sci 50 (2004); Catledge, Lara D. and James E. Pitkow, Characterizing Browsing Behaviors on the World Wide Web, Computer Networks and ISDN Sys, 27 (6) (1995), 1065–73.

second, more restrictive definition includes all visits by users who accessed at least five pages in a given company. This is likely to exclude casual browsers somewhat more effectively.[52]

A problem with this way of identifying shopping-oriented visits is that a count of urls doesn't precisely identify the purpose of the users' browsing behavior, as it does not reveal whether the visitor is shopping for the perfect security suite for her home computer, is looking for product documentation in the "maintenance and support" section for a product already purchased, or is simply browsing with no interest in the company's products. It would thus be possible to overestimate the number of shopping-oriented visits, which would cause the informed minority to appear to be smaller than it actually is.

At the other extreme, a user that has selected a product and initiated a checkout or payment process has demonstrated shopping intent. Thus, we use the initiation of the checkout process as the strictest criterion to identify visits with intent to purchase. We identify such events by identifying and subsequently recognizing for the 93 companies in our sample the web page addresses that would be utilized during the checkout and payment process.[53] While knowing that a user started a checkout or payment process provides no guarantee that the transaction was completed, it indicates an extremely high likelihood that a transaction was indeed contemplated. In this sense, this definition is overly restrictive, as it excludes shopping visits that do not result in the initiation of a checkout process.[54]

A possible concern with this definition is that it would exclude shoppers who access the EULA and decide not to purchase, which would bias the estimated size of the informed minority if the conversion rates for readers and non-readers differ. On the other hand, it would be desirable to define shopping sessions as the ones where a checkout process was initiated if we

[52] Several studies have shown that visits with higher page views are more likely to result in purchases. Bucklin & Sismeiro, supra note 50; Wendy Moe, Hugh Chipman, Ed George, & George McCulloch, A Bayesian Tree Model of Online Purchasing Behavior Using In-Store Navigational Clickstream Data (working paper, 2002). Moe et al identify different browsing behaviors by shoppers to a particular website and find that shoppers view an average of 4.01 pages in a given session.

[53] Some companies have integrated handling of part or all of the checkout and payment process within their own website, while other companies outsource parts of the checkout and/or payment processes. For instance, when a visitor proceeds to checkout, they may be redirected to a company like Digitalriver.com that will process the transaction. This mode of checkout is common. Sometimes, in addition to the checkout company, a payment company is involved in completing the transaction such as PayPal.

[54] Given the low conversion rates in electronic commerce, such visits are likely to represent the majority of shopping visits.

assumed that "true" shoppers will purchase from *some* merchant (while they may visit many), and that among these shoppers the ones that constitute the informed minority (i.e., the ones that access EULAs) are equally likely to do so in any of the merchants they visit, then the behavior of the visitors that initiate a checkout session is representative or the shoppers as a whole, and their likelihood to access EULAs provides an appropriate estimate for the size of the informed minority. These assumptions are not unreasonable, given that our sample consists of mostly all major software vendors and several other smaller ones in several markets, and that browsing behavior is followed for an extensive period of time. We address this further in Section 5.2.

To summarize, the three measures described above establish the shopping intent of a session with increasing strictness. The first definition is likely to overestimate the number of visits with intent to consider purchase, while the third definition likely underestimates such visits. The actual number of shopping sessions will lie somewhere in between. As our definitions of a shopping visit become stricter, estimates of the informed minority become more conservative, and the actual number is likely to lie somewhere in between as well.

4.5. Defining shopping visits: Single sessions versus monthly aggregates

A last issue with defining a shopping visit is ascertaining the length of time of a particular visit. Our data provider and the industry in general define user sessions as periods of web browsing activity separated by at least 30 minutes of inactivity. Under this definition, a user can have multiple visits to a given company in a day, a week, or a month. We adopt this definition because it is consistent with related literature.[55] We refer to all page views from a unique company's website within a single user session as a "company visit" by that user. For example, consider an uninterrupted session where a particular user first visits Symantec (a company in our sample), then visits BananaRepublic.com, and then visits McAfee (another company in our sample). This session yields two unique "company visits": one to Symantec and one to McAfee. If, after visiting McAfee, the user goes back to Symantec within the same uninterrupted session, we aggregate that second visit to Symantec with first visit to that company. This aggregation facilitates analysis and does not cause any problems. If this user is deciding whether to buy from

[55] See Moe & Fader, supra note 51.

McAfee or Symantec, aggregation still allows us to see whether the user accessed EULAs when deciding which product to purchase.

Despite its popularity, the uninterrupted session measure might be too narrow. It is conceivable that a user's shopping activity on a given company spans several days or even weeks. Research on Internet shopping behavior reveals that because visiting and "traveling" to a store on the web is cheap and easy, users are more inclined to visit the company several times over an extended period before finally deciding whether or what to purchase.[56] Thus, users accrue information about a product over time and across several visits. If this is the case, then the uninterrupted session measure will overestimate the number of visits with intent to purchase. To correct for this issue, we adopt the methodology of Johnson, Moe, Fader, Bellman, and Lohse (2004).[57] They conclude that repeated visits to a company within a month are likely to correspond to the same shopping cycle. We thus aggregate visits to a unique company in a given month and present these aggregated sessions as an alternative measure of a company visit with intent to purchase. The problem here is that, for some users, we might be likely to compound multiple shopping visits into a single one, thus undercounting the number of shopping visits (and, conversely, overcounting sessions with EULA visits). On the other hand, this measure may allow us to include repeated short visits to a given company that would be excluded under the alternative measures of visit.[58] This definition of a visit might be more robust. As noted earlier, we expect that actual shopping visits will lie somewhere in the middle. We take comfort in the fact that, as it turns out, the results for the various different definitions of visits are broadly similar.

---

[56] Id.

[57] Eric J. Johnson, Wendy W. Moe, Peter S. Fader, Steven Bellman, Gerald L. Lohse, On the Depth and Dynamics of Online Search Behavior, 50 Mgmt. Sci. 3, 299-308 (2004) (finding that less than 1% of all month-long sessions in their sample contained more than one purchasing transaction in a given company).

[58] One problem with our monthly shopping definition is that it assumes that all shopping sessions begin at the beginning of the month. This is likely not the case. As long as shopping sessions are likely to start on *any* day, truncating the sample at the end of a given month shouldn't cause a problem. We examined the number of sessions for each particular day (for visits in the month of January, 2007) and found no significant differences between them.

4.6. Demographic and geographic data

We seek to study characteristics of shoppers that make them more or less likely to become informed about standard terms, so we also include personal information about the households (or heads of households) in the particular sample. Specifically, we have the age and sex of the head of the household, household income, household size, and the presence of children in the household. Because these data are aggregated at the household (and not individual) level, their relationship to eula access might be noisier. Still, important variables like income and location are generally common to all household members. Table 3 reports the summary statistics for the visitors to the companies in our samples that fit each of our definitions of visits.

Panel A includes visitors who accessed a minimum of two pages in at least one of the companies in the sample during a single uninterrupted session as defined by our data provider. This is the most inclusive definition of a visit. The sample is comprised of 50,373 unique visitors. The average age of the users in this group is 46.3. The youngest users in the sample are 18 years old and the oldest reports to be 99. Average income for heads of households is $60,531, with a standard deviation of $39,675. Median income ($37,500) might describe the sample better. Approximately half of the sample is comprised by males. The average number of members in a household is 2.8, and the median number of household members is 3. There are children in about 41% of the households in the sample.

Panel B reports summary statistics for the sample of visitors who accessed a minimum of five page visits in at least one company. Given that this is a more restrictive measure of a visit with intent to purchase, the sample drops to 35,028 unique users. The characteristics of these users are very similar to the most expansive measure of a visit.

Finally, Panel C in Table 3 reports summary statistics for users that satisfy our strictest definition of intent to purchase. These users must have selected a product for purchase and began the checkout process. The sample now drops to 3,352 unique users. This is to be expected, given that online conversion rates are known to be less than two percent. Again, the users in this sub-sample are mostly similar to those in the broader samples. The only difference with users in this group is that average and median income are higher ($62,421 and $75,000, respectively).

5. Results

In this section, we identify those company/shopping visits where the user accessed a EULA. We do this by matching the urls corresponding to all the EULAs we collected to the clickstream of urls accessed by users during their company visits. We compute descriptive statistics of company visits and EULA accesses under alternative definitions of a visit with intent to purchase. Finally, we present regressions to study the determinants of the (as it turns out, low) probability that a EULA will be accessed.

5.1. Company Visits Characteristics and EULA Accesses

We seek to study which fraction of buyers *seek* to become informed about EULA terms in deciding whether to purchase software. As noted before, some sellers make their contracts available to buyers by merely posting them somewhere on their site and others require buyers to agree to them during the purchasing process. Because companies with EULAs with a location score of 0 or 0.5 present their contracts to buyers as part of the checkout process, whether or not buyers intend to become informed about terms, we exclude them from our analysis.[59] We thus restrict our analysis to those companies where, as is desirable, buyers must unambiguously search for and access the EULA to become informed. This leaves us with 66 companies.

Tables 4 and 5 summarize the characteristics of company visits, measured either as uninterrupted sessions (Table 4) or visits by unique users, aggregating all the monthly sessions by individual users (Table 5). In each case, the data are presented for the three alternative definitions of a company visit. We also separate visits according to the type of company visited, with the note that only retailer visits include secure checkout page views (freeware is free, so there is no secure "checkout" process). In addition to the number of company visits under each definition, the left half of Tables 4 and 5 show the mean and median number of pages viewed during such visits, and their mean and median duration of the visits in seconds. These give us a rough sense of what the "sessions" to a particular company look like. In the right half of the

[59] This restriction might be over-inclusive, since some visitors may wish to become informed about the EULA before deciding whether to purchase. But since all visitors who begin a checkout process "access" the EULA, we have no way of identifying those buyers that select a product to see the EULA from those who merely want to purchase the product and are taken to the EULA page as a result of this.

Tables, we describe the subset of these visits that included a EULA access, the mean and median number of pages viewed before the first EULA access, and the mean and median length of time spent viewing EULAs in visits where a EULA was accessed. These last two measures give us an indication of shoppers' level of care or intent in accessing EULA pages.[60]

Looking at uninterrupted session/visits (Table 4), with the least strict definition of a visit (2 or more pages accessed at the company visited), there are 120,545 such visits to software retailers and 28,007 to freeware providers. This definition includes repeat visitors. For retail companies, an average visit consisted of 12.4 page views and a 311 second (5.2 minutes) stay. But these numbers are driven by extreme values. The median number of pages visited in any given company was 5 and the median time spent was 105 seconds (1.75 minutes). The number of page views and duration of these visits have skewed distributions. EULAs were accessed in only 55 of the 120.768 visits to software retailers (or 0.05% of all such visits) and in 40 visits to freeware companies (or 0.14%).[61] Users visited an average of 12.2 pages in a given company before they accessed the EULA. The median number of pages accessed before a EULA visit was 7. An important consideration is whether shoppers who access the EULA actually read it. For users in this group, the average time on the EULA page was 47.7 seconds and the median time was 29 seconds. The average number of words of EULAs for retail products in the sample (unreported) is 2,277 with a standard deviation of 1,148 words. The median length is 1,702 words. Given the time spent in the EULA, it is unlikely that users would be able to read their terms in their entirety.[62] Visits to freeware providers have less page views (the median is 4 pages) and are of shorter duration (median time spent is 45 seconds). This is expected, as freeware sites tend to be sparser. EULAs are accessed 0.14% of the times. The median time spent on EULAs is also 29 seconds, and the median length of these EULAs is 1,754.

When a visit is defined to require 5 or more pages accessed at the company visited, there are 67,769 uninterrupted session/visits to software retailers and 13,520 to freeware companies.

[60] All EULA urls in our sample include *only* the text of the EULA, so this should be a fairly reliable measure for time spent reading the EULA.

[61] We define a EULA visit as spending at least one second in the EULA url. This is a very broad definition of a EULA visit, since contracts cannot be read in such a short time.

[62] The average reading rate of American adults is 250 to 300 words per minute. Bailey, R.W & Bailey, L.M, Reading speeds using RSVP, User Interface Update (1999).

The median number of pages viewed in a given visit to a retailer is now 10 pages and the median length of the visit is 185 seconds (or 3.1 minutes). Similarly skewed distribution of visit page views and duration persist, and EULAs were accessed slightly more frequently, in 50 visits to software retailers (or 0.07%) and 30 visits to freeware companies (or 0.22%). The median number of pages visited before accessing a EULA was 8 pages for retail companies and 4 pages for freeware providers. At least for retail companies, this suggests that there might be some "shopping" among products within the site before the EULA is accessed.

Finally, limiting our consideration to visits to software retailers that included initiation of a secure checkout session, the number of visits falls to 5,509, with similar median page views per visit, but about twice as long mean and median durations. This is expected, as purchasers are likely to take a longer time checking out and processing the transaction. In this restricted sample, there are 5 voluntary accesses of a EULA in the course of purchase, constituting 0.11% of all visits. The median number of pages accessed is 20 for users in this group, suggesting even more intense shopping within the site. The median time spent in the EULA also doubles for users in this group Interestingly, out of *all* sessions with EULA visits, 4% (if we use the 2 page visit definition) or 8% (if we use the 5 page visit definition) of them resulted in purchases. This number is much higher than the average 2% conversion rate in Internet purchases.

Aggregating all monthly sessions of an individual user into a monthly visit (shown in Table 5) leads to similar results. In most cases the total number of visits is reduced as multiple visits by individual users are combined. (The average number of sessions per user is (unreported) 3.2.) Importantly, the numbers for the broadest definition of a visit, at least 2 page views, indicate that this category is capturing a non trivial number of casual browsers with little intent to shop.[63] An exception is software retailer visits that included initiation of a secure checkout session. Visits with secure checkout increased, albeit moderately, because combining visits for certain users on a monthly basis resulted in a qualifying monthly visit replacing two or more non-qualifying uninterrupted session visits.

[63] As a rough reference, Nielsen reports that, an audience of 164,890 users in July 2008 visited an average of 22.5 pages in each domain visited and spent an average of 53 seconds in each page. See http://www.nielsen-online.com/pr/pr_080812.pdf.

For the same reason the number of visits with EULA accesses in some cases increased while in other cases decreased. The overall results of Table 5, however, are consistent with Table 4, indicating that our results are robust to the precise definition of company visits.

5.2. Interpreting the results: Estimating the size of the informed minority

Coming back to the empirical framework of Figure 2, visitors to the websites of the companies in our sample can be classified into potential buyers or users visiting for other reasons, such as looking for online instruction manuals for a product they already own, to search user forums for troubleshooting information, or for entertainment—e.g., to watch the "Mac vs. PC" commercials. We measure the total number of page views during each visit, as well as whether a EULA was accessed and whether a secure checkout session was initiated. This data, reported for individual sessions in Table 4 and for monthly visitors in Table 5, allows us to estimate the number of readers, buyers and shoppers by using access to a EULA page as a proxy for reading, initiating the checkout process as a proxy for buying, visits with 5 or more page views as a proxy for identifying shoppers and visits between 2 and 5 page views as a proxy for identifying non-shoppers. Based on the data in Table 5, we estimate the number of monthly visitors in our sample for each of the six categories shown in Figure 2:

The se_1b_1 readers that buy and $se_1(1-b_1)$ readers that do not buy are 6 and 43respectively. There are $s(1-e_1)b_2$ or 3,528 buyers that are not readers, and $s(1-e_1)(1-b_2)$ or 37,120 shoppers that neither read nor buy. Few non-shoppers would be expected to read EULAs and so it is not surprising that $(1-s)e_2$ is small; in our sample it equals 4(out of 22,575 visits). Finally, the large majority of non-shoppers do not read EULAs; this number is $(1-s)(1-e_2)$ or 22,571 based on the above proxies. We thus arrive at an estimate that the fraction e_1 of shoppers that are in the informed minority of $\frac{se_1b_1 + se_1(1-b_1)}{se_1b_1 + se_1(1-b_1) + (1-e_1)b_2 + (1-e_1)(1-b_2)} = \frac{49}{40{,}697} = 0.12\%$.

It is possible that considering all visitors with 5 or more page views as shoppers will overestimate the number of shoppers. An alternative estimate could be obtained by assuming that among actual shoppers, the "conversion ratio" to initiate a checkout session among non-readers is the same as that for readers at 6/49 = 12.2% (which is higher than purchase conversion ratios of 2-5% cited in the marketing literature, but reasonable if only about 1 in 3 checkout sessions

we capture result in actual purchases). In that case the informed minority fraction for all shoppers would be the same as the fraction for buyers, i.e. 6/3534 = 0.17%.[64,65]

The bottom line is that the fraction of visitors that access EULAs is very small, on the order of 0.1%. While a number of alternative estimates can be calculated, these estimates point to that fraction being well under 1%. Can such a small informed minority protect buyers and discipline sellers into providing efficient contract terms, thus preventing a market failure? The literature offers few meaningful suggestions as to how large the informed minority needs to be, and these are typically provided in the context of illustrative examples. Schwartz and Wilde offer an example where the informed minority needs to be 20% to 30% to be effective. Our estimates here are imperfect, but they are two orders of magnitude smaller.[66]

Theoretically, the size of informed minority required to induce sellers to provide good terms depends on the tradeoff between the gross profit from selling an additional unit (determined from the marginal cost of the product and the cost of providing better contract terms). In our model, a mixed equilibrium in the case where all consumers face the same cost of becoming informed is characterized by a fraction $r = \frac{c_g - c_b}{(b_1 - b_3)(p - c_b + F)}$ becoming informed about contract terms. This fraction becomes smaller as the incremental cost of providing good

[64] Most models of the informed minority predict that the conversion ratio for non-readers would be the same or higher as the conversion ratio for readers, as the latter will be less likely to purchase the product if they are not satisfied with the terms of the EULA. An upper bound on the size of the informed minority can be obtained if we assume that $b_2 = 100\%$, i.e., that 100% of non-readers proceed to purchase the product. In that case the informed minority would be 49/(3534+43) = 1.37% of the total number of shoppers.

[65] As mentioned in section 4.4, if we assume that "real" shoppers will purchase from *some* merchant (while they may visit many), and that among these shoppers the ones that constitute the informed minority are equally likely to access a EULA in any of the merchants they visit, then the behavior of visitors in our sample that initiate a checkout session is representative of shoppers as a whole, and their likelihood to access EULAs (0.17%) provides an estimate for the size of the informed minority among these most determined shoppers.

[66] The estimates presented above are based on monthly visits as reported in Table 5. This is conservative in the sense that using visits defined as individual sessions would result in lower estimates for the size of the small minority. Specifically, using session data from Table 4 would result in 5 readers that buy, 45 readers that do not buy, 5,504 buyers that are not readers, and 62,215 shoppers that neither read nor buy, and 52,776 non-shoppers that include only 5 readers. The fraction e_1 of shoppers in the informed minority would be 50/67769 = 0.074%. The fraction of readers that initiate checkout sessions would be 5/50 = 10%, and assuming the same conversion ratio for non-readers would give 5/5509 = 0.09% as the informed minority. A conversion ratio of 100% for non-readers would give an upper bound for the informed minority of 50/(5509+45) = 0.91%

terms decreases, as the probability that shoppers who become informed about the terms will drop out if they encounter bad terms increases, as the marginal cost of the product decreases, and as the damaging consequences for the seller of being "discovered" to offer bad terms increase. Of course, this general theoretical conclusion is rather unhelpful. Given certain values for these unknown parameters, *any* fraction of shoppers being term conscious could support an informed minority equilibrium. So, ultimately, one must look at the fraction 0.1% with a modicum of subjective judgment, even though it is within rounding error of zero. To us, it seems a rather tiny fraction, given the imperfectly competitive nature of software markets. Still, we explore the possibility of an informed minority equilibrium in this market.

While, as mentioned above, given appropriately chosen values of the parameters in our setting *any* fraction of shoppers that choose to become informed could support an informed minority equilibrium, the market for software maintenance and support ("M&S") can be used to derive an estimate for the likely range of one of these parameters, the marginal cost of "good" terms. Specifically, M&S is a key term in software EULAs[67] and thus the marginal cost of M&S provides a floor for the marginal cost of offering good EULA terms.[68]

To estimate the cost of M&S terms, we obtained product price and annual M&S price for 520 software products from the 42 software companies in the sample of Marotta-Wurgler (2007) that provided M&S separately on a periodic basis (i.e., did not charge per incident). The average annual M&S price to product price ratio for these software products had a mean of 0.26, median of 0.20 and standard deviation of 0.22. Since there was high inter-company correlation, we focused on company means. Figure 3 shows the distribution for the 40 companies remaining after dropping two outliers with too high ratios, which has a mean of 0.29, median of 0.24 and standard deviation of 0.16.

[67] Marotta-Wurgler (2007) identifies and measures "23 important and common terms that allocate rights and risks between buyers and sellers of software" and M&S is one of these 23 terms.

[68] The analysis of 647 EULAs in Marotta-Wurgler (2007) gives M&S an average value of 0.68 on a scale from 0 to +1 where 0 indicates the default terms in the absence of any EULA provisions (no M&S) and +1 indicates free M&S for 31 days or more. About three quarters of the sample companies commit to a "free" (i.e., included in the base price of the software) M&S period in their EULAs with these periods ranging from 60 days to 2 years, with a mean of 292 days and a mode and median of 1 year. Thus M&S provisions in EULAs are significantly more favorable to consumers than default, and likely to constitute an important fraction of the cost of offering pro-consumer EULA terms.

Thus a year of M&S for software products[69] is on average priced at 25-30% of the product price; since M&S costs are primarily variable (labor) costs, if the market for M&S was perfectly competitive, this would provide an indication of the marginal cost of M&S and thus a floor on the marginal cost of pro-consumer EULA terms. There are several reasons why 25-30% of product price may be too high an estimate: consumers may be more likely to purchase M&S from the seller of the software, and thus software companies may price as a two part tariff, with a lower price for the upfront purchase (the software product) and a higher price for the subsequent purchase (M&S); consumers that purchase M&S are likely to have higher M&S costs due to adverse selection and/or moral hazard; software companies may have substantial market power in providing M&S due to barriers to entry for competitors that are not as familiar with their product or consumers' propensity to purchase M&S from the seller of the original software. All of the above factors would result in a M&S-to-product price ratio that is higher than the cost of providing M&S. On the other hand, M&S is only one of 23 key EULA terms, which include several other types of warranties and permissions to copy or distribute the software that can impose opportunity costs. Furthermore, M&S pricing is similar in enterprise software markets, where significant competition exists from third-party M&S providers and purchase of M&S contracts is almost universal.

On balance, it seems appropriate to assume that the cost of M&S is at least 20% of the product price, or 0.2*p* in the notation of section 3.1. The fraction of consumers that become informed in that setting is given by $r = \frac{c_g - c_b}{(b_1 - b_3)(p - c_b + F)}$ and dividing numerator and denominator by *p* we get $r = \frac{(c_g - c_b)/p}{(b_1 - b_3)(1 - c_b/p + F/p)}$. If we assume that $(c_g - c_b)/p \geq 0.2$ and since $(b_1 - b_3) \leq 1$ and $1 - c_b/p + F/p \leq 1 + F/p$, we get $r \geq \frac{0.2}{1 + F/p}$, or $F \geq (0.2/r - 1)p$. What this means is that when $r \approx 0.001$ (or 0.1%) as we saw in our data, in order for an informed minority equilibrium to be supported, one would need *F*, or the *expected* cost from an informed consumer discovering a EULA with bad terms, to be at least 199*p*, or $19,900 for a product

[69] As mentioned above, one year was the most common as well as the median duration of free M&S for the companies that provided such a period of free M&S.

costing $100.[70] We believe this to be impossible, thus we conclude that our data is not consistent with an informed minority equilibrium.

5.3. Determinants of EULA visits

Although few potential shoppers in our sample actually click on the EULA, it is interesting to examine what characteristics of the company, product, user, and website that separate those that do from those that don't. In order to assess the relationships between our sample variables and EULA accesses we estimated the following logistic regression:

$$\begin{aligned} EULA\_Access = {} & \alpha + \beta_1 Freeware\_dummy + \beta_2 Median\_Price + \beta_3 EULA\_Location \\ & + \beta_4 Log\_Revenue + \beta_5 Public\_Company + \beta_6 Non\_EULA\_Pages\_Viewed \\ & + \beta_7 Gender + \beta_8 Log\_Income + \beta_9 Log\_Age + \varepsilon \end{aligned}$$

The dependent variable EULA_Access is whether a EULA was accessed during a particular company visit. We included a dummy for whether this is a freeware company. Users may be concerned about hidden costs when downloading a freeware and thus may be more inclined to read the corresponding EULAs. Similarly, consumers may be more concerned about the terms associated with expensive products, as these costs involve a higher cost to the consumer. To test for this possibility, we included in the regression the natural log of median price for products sold online by that company.

We would expect that making the EULA harder to find would significantly reduce the probability consumers will access it. To test this, we included the location of the EULA (defined in Table 2 as the number of clicks it takes to access the EULA) as an independent variable. Consumers may be less likely to access EULAs from companies they trust; for instance, this may be the case for larger companies, companies with a long-time industry presence, or public companies that might be subject to higher regulatory and shareholder scrutiny. To test for this we included independent variables for the natural log of a company's revenues (as a proxy for size) and status as a public company. Extensive visits with a large number of page views may have

[70] In this example, this can be conceptualized as a seller losing 199 sales as a result of someone discovering a bad term.

increased likelihood of including a EULA access, as, for instance, they may be more likely to represent serious shoppers that are likely to access the EULA as part of their due diligence on their prospective purchase. We include the number of pages accessed (other than EULAs).

Finally, certain demographic characteristics of a visitor may be significantly associated with the probability to access a EULA. High-income visitors may be more educated and thus better able to understand the language (and importance) of EULAs, and thus more likely to access them. Or, high-income visitors have a higher opportunity cost, and be less likely to spend time reading a EULA. We include the natural log of income, as well as age and a dummy for gender (of the head of household) taking a value of 1 for men. Standard errors are clustered by visitor.

We estimate the above model for the six alternative definitions of a company visit that we introduced, namely for uninterrupted as well as monthly aggregated user sessions, and for each of the above requiring viewing at least two page views at the company visited, at least five page views, and initiation of a secure checkout process. We had to drop EULA_Location as an independent variable when visits were defined to include initiating a secure checkout process, as in that case all EULA accesses had the same EULA_Location score, making it impossible to estimate β_4 by logit regression. The regression results are in Table 6.

The coefficient for the freeware dummy β_1 is positive and significant for the monthly visits, offering some support to the hypothesis that consumers are more likely to access EULAs for free products. This would be consistent with consumers assuming there will be a "catch" in products offered for free or perhaps mistrusting that type of offers in general, or with consumers' focus shifting to terms since there is no price information to be analyzed.

The coefficient for the median product price is positive and significant in the analysis for both uninterrupted and monthly sessions when visits are defined as containing at least two or five page views. This may be because consumers collect more information, including EULA terms, when purchasing expensive products, since the consequences of a poor decision are likely to be higher. Curiously, however, this relationship is reversed for those visits where a checkout process is initiated. It is possible that buyers of expensive products were committed to the product already, or lacking alternatives, or else lacking useful alternatives, and thus relatively uninterested in the standard terms.

The coefficient for EULA location is negative and highly significant. This shows a strong correlation between a EULA's accessibility and probability it will be accessed. Terms made available on seller websites but hidden away will not be discovered and accessed by consumers. This relationship between accessibility and probability of access is shown in Figure 4. Apparently, online, more than one click away is too far.

The coefficient for company revenue is negative as expected, and strongly significant. This perhaps suggests that consumers view reading the EULA as less necessary for known, trusted companies. The level of significance is lower for sessions with secure checkout initiated, possibly because establishing such a session already reflects a certain degree of trust. The coefficient for public status is also highly significant in all specifications. Controlling for size, users are more likely to visit the EULAs of public companies.

The coefficient for the number of non-EULA page views during a visit is positive and significant when visits are defined as containing at least two page views, with or without a secure checkout session, but not significant when visits are defined as containing at least 5 page views (for the uninterrupted sessions).

In terms of demographic characteristics, the coefficient for gender has a negative sign throughout but it is not statistically significant. Women appear to be more likely to access EULAs than men. The effect of income, is positive and only barely significant, indicating that consumers with higher income might be more sophisticated and thus more prone accessing the EULA.

In unreported regressions we examine whether the option to download the trial version of a product affects user's propensity to access the EULA. As noted in Table 1, a large fraction of sellers offer trial versions. A plausible hypothesis is that users that become familiar with the trial version of a product might be less inclined to read the EULA of the retail version, perhaps due increased familiarity and comfort with the product. There is virtually no relationship between the presence of trial versions and users' likelihood of accessing EULAs for the broadest definitions of visits, but when visits are defined as beginning the checkout process, the coefficient on this variable is negative and highly significant, thus offering some support to the "familiarity" hypothesis.

Finally, we study whether shoppers are less likely to read the EULAs of products that are more likely to be purchased repeatedly. Users that become familiar with a product that is

continuously updated, like Microsoft Office, might not feel the need to read the EULA to assess its quality. Other products, such as test preparation software, are less likely to be purchased repeatedly. We create a dummy variable that equals one if the company markets products that are in our judgment likely to be repeat purchases. However, we find no relationship between the nature of the use of the software and users' propensity to access EULAs.

These regressions show some determinants of EULA access, but the message of the previous section is that the most important term in the regression is the constant term: EULAs are rarely read by anyone.

6. Discussion and implications

Consumer access to the terms of standard form contracts has been at the center of a legal and policy debate, and a major question has been whether disclosure of terms in standard form contracts that govern consumer transactions should be regulated. A related debate has focused on the enforceability of terms and possible need to regulate disclosure for software in general, and software purchased online in particular.

These debates center around the validity of the informed minority hypothesis, the view that comparison shoppers for standard terms help sustain efficient equilibria in the provision of those terms. In this paper we take advantage of a unique data set to investigate the extent to which consumers actually do access the terms of certain standard form online contracts. That is, we attempt to measure directly the "informed minority." We believe that we have measured this aspect of consumer behavior with reasonable precision.

Our results inform these debates in three ways. First, we cast doubt on the very existence of the informed minority: very few consumers become informed about contract terms standard form online contracts, at least in our context. Specifically, the fraction of shoppers that accesses EULAs, the license agreements that govern the use of software, is on the order of 0.1%. It is difficult to benchmark this number against a theoretical norm, but we note it is a number that is two full orders of magnitude smaller than examples offered in the literature for the required size of the informed minority. Even under generous assumptions, or alternative definitions of "shopper" and "visit," it is hard to envision that number growing even to 1%. The informed minority (if it exists in our context) may or may not be informed, but is definitely a minority. We

also show that it is doubtful that its size is large enough to generate a competitive equilibrium in this market.

Another conclusion is that disclosure matters. The probability of accessing the standard form contract remains low but increases sharply with accessibility. In our context, putting information more than one click away dramatically reduces the probability this information will be accessed by consumers. One might interpret this finding as supporting regulation of disclosure of contract terms, perhaps mandating direct accessibility of these terms. But, when evaluated in the context of how low the fraction of informed consumers is to begin with, the effectiveness of such regulation might be questionable. Even a large percentage increase will leave the fraction of informed consumer very small.

Finally, consumers seem to be guided somewhat by the reputation and trustworthiness of the seller, as well as other income and price considerations in deciding whether to become informed about EULA terms. They are more likely to access these terms when faced with a smaller (presumably less reputable or familiar) seller, or circumstances they may consider suspicious, such as offers for freeware, as seems intuitive.

Figure 1. A Stylized Model.

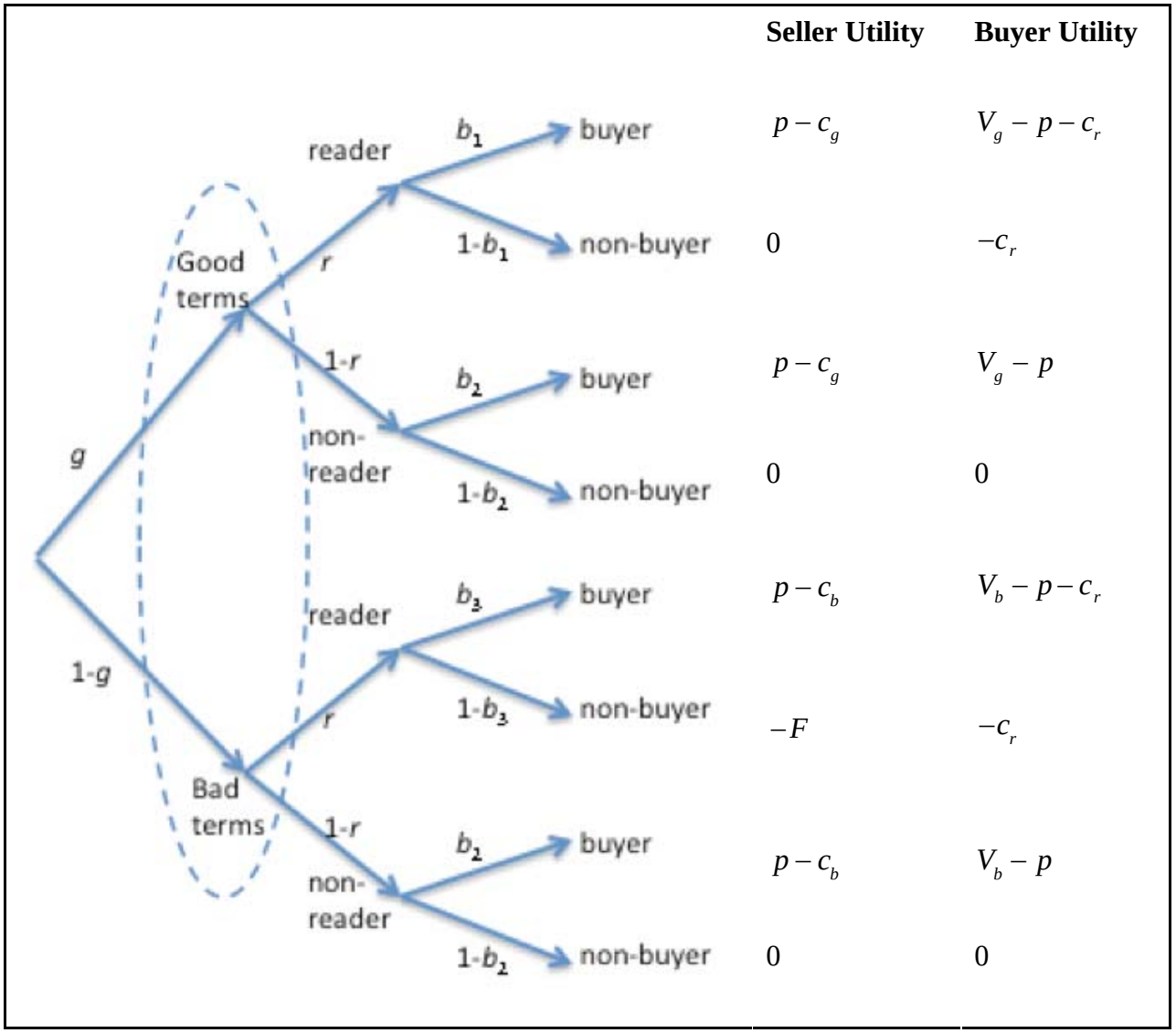


Figure 2. Empirical Framework

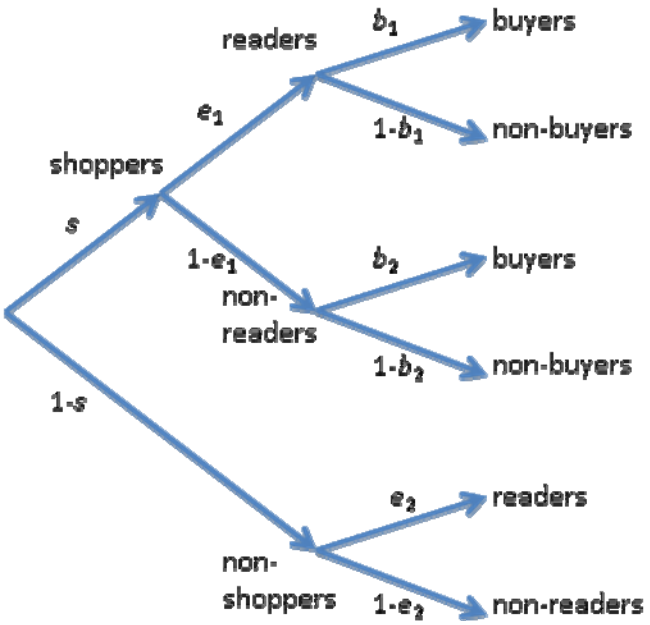
Visitor types	Fraction of visitors
	
	se_1b_1
	$se_1(1-b_1)$
	$s(1-e_1)b_2$
	$s(1-e_1)(1-b_2)$
	$(1-s)e_2$
	$(1-s)(1-e_2)$

Figure 3. Annual maintenance to product price ratio, 40 companies

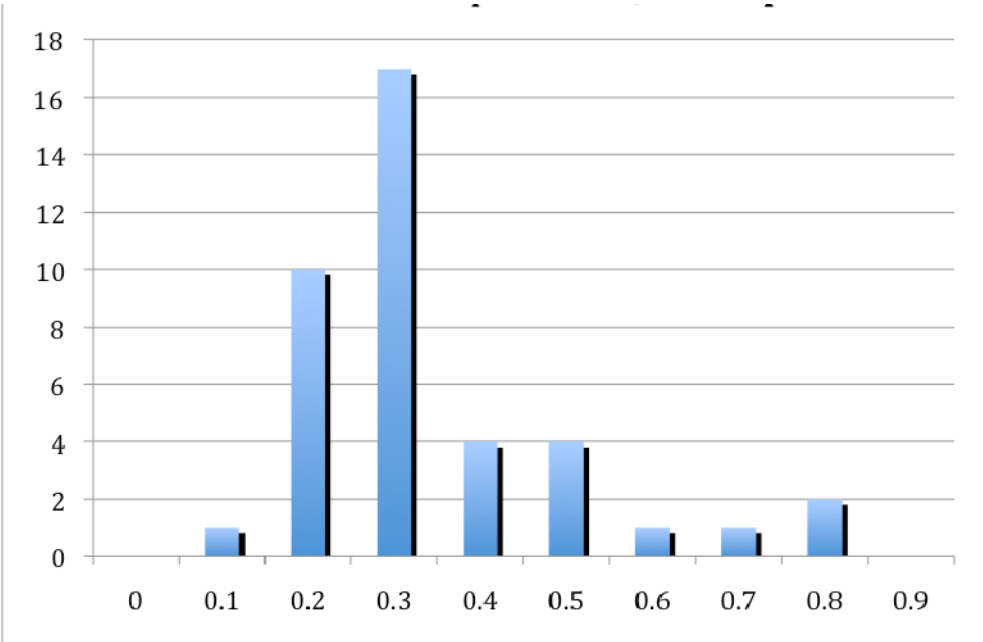
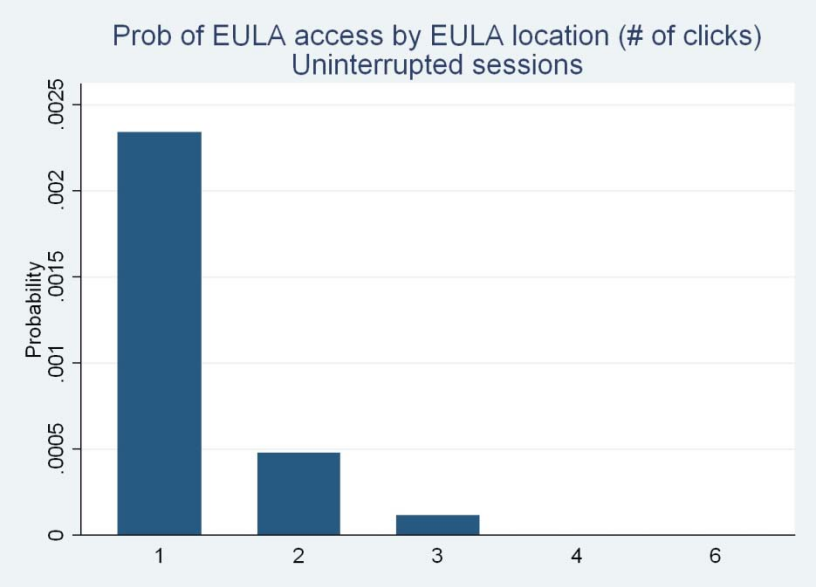


Figure 4. Probability of EULA access by location



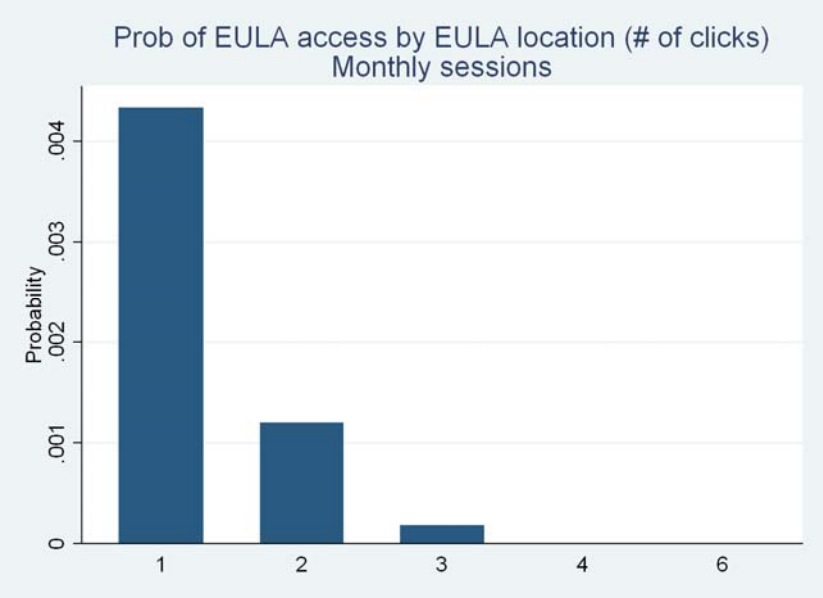


Table 1. Company and Product Characteristics

	N	Mean (s.d)	Minimum	Median	Maximum
Panel A. Company Characteristics					
Retail					
Revenue (Millions $)	81	1,520 (6,850)	0.1	6	51,100
Age (years)	81	14.7 (10)	2	13	55
Public Company	81	0.27 (0.45)	0	0	1
Freeware					
Revenue (Millions $)	12	1,160 (4,010)	0.1	0.1	13,900
Age (years)	12	8.1 (6.7)	2	5.5	26
Public Company	12	0.08 (0.29)	0	0	1
Panel B. Product Characteristics					
Retail					
Consumer Product	81	0.68 (0.47)	0	1	1
Price ($)	81	394.8 (1,025.4)	9.97	58.40	5,295.00
Median Price ($)	81	351.9 (1,014.88)	1	49	5000.00
Trial (featured product)	81	0.84 (0.37)	0	1	1
Trial (most products)	81	0.78 (0.42)	0	1	1
Freeware					
Consumer Product	12	0.83 (0.39)	0	1	1

Table 2. EULAs According to Accessibility

	Retail	Freeware	All Companies
Distance in Clicks from Buyer's Clickstream	**N (%)**	**N (%)**	**N (%)**
0	3 (3.70)	-	3 (3.23)
0.5	22 (27.16)	2 (16.67)	24 (25.81)
1	34 (41.98)	7 (58.33)	41 (44.09)
2	12 (14.81)	2 (16.67)	14 (15.05)
3	7 (8.64)	1 (8.33)	8 (8.60)
4	2 (2.47)	-	2 (2.15)
5	-	-	-
6	1 (1.23)		1 (1.08)
Total	81 (100.0)	12 (100.0)	93 (100.0)
Mean distance	1.28	1.25	1.28

Table 3. User Characteristics

	N	Mean (s.d)	Minimum	Median	Maximum
Panel A. Users accessing at least 2 pages in at least one sample company					
Age (years)	50,373	46.25 (13.78)	18	46	99
Gender (1= Male)	50,373	0.50 (0.50)	0	0	1
Income ($)	50,373	60,531 (39,675)	12,500	37,500	150,000+
Household Size	50,373	2.78 (1.27)	1	3	5+
Presence of Children (1= Yes)	50,373	0.41 (0.49)	0	0	1
Panel B. Users accessing at least 5 pages in at least one sample company					
Age (years)	35,028	46.34 (13.70)	18	46	99
Gender (1= Male)	35,028	0.50 (0.50)	0	1	1
Income ($)	35,028	60,624 (39,737)	12,500	37,500	150,000+
Household Size	35,028	2.79 (1.27)	1	3	5+
Presence of Children (1= Yes)	35,028	0.41 (0.49)	0	0	1
Panel C. Users initiating checkout in at least one sample company					
Age (years)	3,352	47.29 (13.96)	18	46	99
Gender (1= Male)	3,352	0.51 (0.50)	0	1	1
Income ($)	3,352	62,421 (40,981)	12,500	75,000	150,000+
Household Size	3,352	2.78 (1.25)	1	3	5+
Presence of Children (1= Yes)	3,352	0.41 (0.49)	0	0	1

Table 4. Company and EULA visits. Visits measured as uninterrupted sessions

	N of company visits	Mean N of pg. clicks per company visit (s.d.)	Median N of pg. clicks per company visit	Mean length of company visit in seconds (s.d.)	Median length of company visit in seconds	N of EULA visits (% of company visits)	Mean N of pg. viewed before EULA access (s.d)	Median N of pg. viewed before EULA access	Mean length of EULA access in seconds (s.d.)	Median length of EULA access in seconds
Panel A. At Least 2 Pages Accessed During Visit										
Retail	120,545	12.4 (26.9)	5	310.9 (713.6)	105	55 (0.05%)	12.15 (16.6)	7	47.7 (45.4)	29
Freeware	28,007	13.6 (36.9)	4	163.51 (618.8)	43	40 (0.14%)	7.45 (15.4)	3	99.6 (237.8)	29
Panel B. At Least 5 Pages Accessed During Visit										
Retail	67,769	19.9 (34)	10	439 (903.1)	185	50 (0.07%)	13.2 (17)	8	46.9 (47.3)	27
Freeware	13,520	25.2 (50.6)	12	239.2 (856)	67	30 (0.22%)	9.5 (17.4)	4	60.6 (104.5)	20.5
Panel C. At Least 1 Secure Checkout Page Accessed During Visit										
Retail	5,509	13.1 (30.7)	5	586.5 (1,939)	218	5 (0.11%)	20.4 (16.8)	20	84 (60.5)	60

Table 5. Company and EULA visits. Visits measures as monthly aggregates of uninterrupted sessions

Company	N of company visits	Mean N of pg. clicks per company visit (s.d.)	Median N of pg. clicks per company visit	Mean length of company visit in seconds (s.d.)	Median length of company visit in seconds	N of EULA visits (% of company visits)	Mean N of pg. viewed before EULA access (s.d)	Median N of pg. viewed before EULA access	Mean length of EULA access in seconds (s.d.)	Median length of EULA access in seconds
Panel A. At Least 2 Pages Accessed During Visit										
Retail	63,272	23.67 (79.2)	7	592.4 (2,880)	161	53 (0.08%)	21.5 (43.04)	9	51.9 (47.5)	30
Freeware	11,010	35.8 (292.8)	4	415.9 (2,811)	75	42 (0.38%)	10.9 (22.1)	3.5	102.5 (234.9)	29
Panel B. At Least 5 Pages Accessed During Visit										
Retail	40,697	35.3 (96.9)	14	837 (2,562)	292	49 (0.12%)	23.1 (44.4)	10	50.16 (47.82)	29
Freeware	5,370	70.6 (416.5)	11	741.5 (3,993)	148	34 (0.63%)	13.2 (24.1)	4	104.9 (251.2)	25
Panel C. At Least 1 Secure Checkout Page Accessed During Visit										
Retail	3,534	34 (79.1)	10	1,419 (4,681.5)	450	6 (0.22%)	24.2 (23.3)	16	99 (60.1)	94

Table 6. Determinants of EULA visits. Logit Regressions

Dependent Variable: EULA access						
	Uninterrupted sessions			Monthly aggregate sessions		
	At Least 2 Pages	At Least 5 Pages	At Least 1 Secure CP	At Least 2 Pages	At Least 5 Pages	At Least 1 Secure CP
	(1)	(2)	(3)	(4)	(5)	(6)
Freeware dummy	1.61 (1.05)	0.99 (1.20)	-	2.89*** (0.88)	2.61*** (0.91)	-
Ln Med. Price	0.38** (0.15)	0.33* (0.17)	-1.38*** (0.11)	0.44*** (0.13)	0.41*** (0.15)	-1.34*** (0.11)
Location	-0.67*** (0.20)	-.82*** (0.27)	-	-0.66*** (0.15)	-0.69*** (0.17)	-
Ln Revenue	-0.38*** (0.07)	-0.38*** (0.09)	-0.13* (0.08)	-0.42*** (0.06)	-0.42*** (0.07)	-0.14* (0.07)
Public	2.17*** (0.59)	2.23*** (0.70)	5.84*** (0.51)	2.22*** (0.52)	2.40*** (0.60)	0.04*** (0.43)
N Pgs. Viewed	0.06*** (0.01)	0.03 (0.02)	0.04*** (0.01)	0.06*** (0.01)	0.03*** (0.01)	0.04*** (0.01)
Gender	-0.33 (0.23)	-0.22 (0.25)	-0.06 (0.19)	-0.30 (0.21)	-0.30 (0.01)	-0.10 (0.19)
Ln Income	0.15 (0.16)	0.21 (0.18)	0.29** (0.14)	0.19 (0.15)	0.23 (0.16)	0.20 (0.13)
Ln Age	0.50 (0.37)	0.29 (0.40)	0.70** (0.30)	0.43 (0.35)	0.31 (0.37)	0.43 (0.31)
N	148,552	81,289	5.509	74,282	46,067	3,536
R^2	0.11	0.11	0.63	0.14	0.14	0.64

Note: * = denotes significance at 0.1 level, ** at 0.05 and *** at 0.01